Introdução aos Processos Estocásticos

para estudantes de Matemática

Artur O. Lopes e Sílvia R. C. Lopes

Dedicado a Daniel, Luciana e Luna

Sumário

Prefácio

O presente texto apresenta uma introdução à Teoria dos Processos Estocásticos para alunos que se encontram em algum programa de graduação (ou, no começo do mestrado) em Matemática. O texto foi escrito de tal forma que (praticamente) não requer que o leitor tenha feito um curso de Probabilidade para entender o seu conteúdo. Observamos que este livro foi escrito visando apenas alunos do Bacharelado e Mestrado em Matemática.

O Capítulo 1 foi elaborado com a intenção de dar ao leitor uma ideia inicial intuitiva do que é um Processo Estocástico. Ele apresenta alguns exemplos elementares e destaca qual o ponto de vista correto pelo qual se deve encarar a teoria, e, ainda, quais são algumas das perguntas básicas em que se está interessado. Diferentemente dos outros ele tem um caráter informal.

O livro está estruturado de tal jeito que é possível seguir duas rotas distintas na sua leitura. Para o leitor que deseja um contato preliminar do assunto, sem o conhecimento mais profundo de Teoria da Medida, sugerimos a leitura na ordem das seções apresentadas no índice do livro.

Para aquele que prefere uma abordagem dos Processos Estocásticos - já de início - de uma maneira mais bem formalizada do ponto de vista matemático, sugerimos que leia primeiro o Capítulo 1, e a seguir, se dirija ao Capítulo 5. Após esta leitura então volte ao Capítulo 2, e a seguir, o Capítulo 3 e 4.

No segundo capítulo, tratamos de Cadeias de Markov a tempo discreto. O pré-requisito para esta seção é apenas o conhecimento de resultados básicos sobre matrizes e um pouco de Álgebra Linear. No terceiro apresentamos alguns resultados sobre limites de variáveis aleatórias, entre eles a Lei dos Grandes Números e o Teorema Central do Limite. Numa primeira leitura, sem maiores prejuízos para o que segue, se poderia pular a demonstração dos resultados mais sofisticados

No quarto capítulo analisamos as Cadeias de Markov a tempo contínuo. Neste capítulo se usa alguns resultados de equações diferenciais e sistemas lineares de equações diferenciais. No apêndice 4.5 apresentamos uma breve introdução às equações diferenciais.

Muitos exemplos são discutidos ao longo da apresentação. Alguns exercícios são propostos ao longo do texto e recomendamos o leitor tentar resolvê-los no momento em que aparecem. Finalmente, ao fim de cada capítulo, apresentamos uma lista extensa de exercícios cobrindo o material desenvolvido.

Por fim, no Capítulo 5, Seção 5.1, apresentamos uma apanhado geral de certos resultados da Teoria da Medida. A partir daí, introduzimos as definições e demonstramos alguns dos resultados básicos da Teoria dos Processos Estocásticos. Apresentamos alguns resultados básicos sobre processos independentes e ergódicos.

Ao fim do texto decrevemos brevemente alguns tópicos que são de grande importância: martingales, movimento Browniano e difusões. Um tratamento matemático mais completo destes assuntos foge ao escopo deste livro. Quando se deseja demonstrar resultados sobre "quase todos os caminhos amostrais" se necessita de Teoria da Medida, companhia inseparável dos Processos Estocásticos.

Esclarecemos ao leitor que todo o material do livro é apresentado de forma matematicamente precisa. As demonstrações são desenvolvidas com correção. No entanto, algumas questões mais complexas da teoria requerem uma maior maturidade analítica e uma formalização matemática mais sofisticada. Isto foi deixado para o fim. Com o objetivo de atingir um público mais amplo, optamos por seguir a presente sequência. Tratamos aqui com muitos detalhes o caso de sistemas com espaço de estados discreto (mas não exclusivamente).

O presente texto sofreu influência na sua redação daqueles livros que no nosso entender introduzem a Teoria dos Processos Estocásticos da maneira mais elegante possível:

P. Billingsley, Probability and Measure, Wiley, 1995

G. Grimmett and D. Stirzaker, Probability and Random Processes, Oxford Press, 1994

D. W. Stroock, An introduction to Markov Processes, Springer Verlag, 2005

S. Ethier and T. Kurtz, Markov Processes, John Wiley, 1986

Karlin and Taylor, A First Course in Stochastic Processes, Academic Press, 1975

Karlin and Taylor, A Second Course in Stochastic Processes, Academic Press, 1975

J. Norris, Markov Chains, Cambridge Press, 1997

R. Durrett, Probability: Theory and Examples, Cambridge Press

J. Kuntz, Markov Chains revisited, arXiv (2020)

Finalmente, gostaríamos de agradecer a vários colegas e estudantes que leram e nos auxiliaram a corrigir várias imperfeições em versões preliminares do presente texto: Jairo Mengue, Adriano Tort, Carlos Felipe Lardizabal, Rafael R. Souza, Cleonis Figueira, Adriana N. de Oliveira, Marcelo Disconzi, Everton Artuso, Gustavo Muller, Matheus Stapenhorst, Rafael Pereira, Thomas J. Spier, Maicon Karling, Jader Brasil, Josué Knorst, Luisa Borsato, Gustavo Pessil, Guilherme Feltes, Hermes Ferreira, Luiz Felipe Pereira, Thomas Jacq e Filipe Jung.

Porto Alegre, 20 de janeiro de 2020

Artur O. Lopes e Sílvia R. C. Lopes

1

Introdução

Vamos apresentar inicialmente algumas definições preliminares, exibir alguns exemplos de natureza simples e discutir também algumas das ideias básicas da Teoria da Probabilidade e dos Processos Estocásticos. Certas partes desta seção têm um caráter levemente informal. É a única do livro com tais características.

O objetivo aqui é descrever certos aspectos essenciais do ponto de vista probabilístico ou estatístico (de se analisar um problema matemático).

Considere fixado um conjunto K. Uma probabilidade P em K é uma lei que associa a cada subconjunto A de K um valor real não negativo $P(A)$ menor ou igual a um. O valor $P(K)$ se assume ser igual a 1. Algumas propriedades mais (Definição 1.2) serão requeridas para P.

Fixado o conjunto K, denotamos por $\mathfrak{p}(K) = \{A : A \subset K\}$, o conjunto das partes de K.

Infelizmente, em muitos casos interessantes não se pode definir P (com as tais propriedades que desejamos) em todos os subconjuntos A de K, ou seja sobre a classe de conjuntos $\mathfrak{p}(K)$.

Sendo assim, é necessário introduzir a seguinte definição:

Definição 1.1. *Uma família $\mathcal{A}$ de subconjuntos de um conjunto K é chamado de σ-álgebra sobre K no caso em que:*

a) o conjunto K pertence a $\mathcal{A}$,

b) se A pertence a $\mathcal{A}$, então o complemento $K - A$ também pertence a $\mathcal{A}$,

c) se $(A_n)_{n\in\mathbf{N}}$ é uma coleção enumerável de conjuntos em $\mathcal{A}$, então a união $\cup_{n\in\mathbf{N}} A_n$ também está em $\mathcal{A}$.

Segue da definição que $\emptyset$ está em $\mathcal{A}$.

Note que $\mathcal{A} \subset \mathfrak{p}(K)$. Note também que se $(A_n)_{n\in\mathbf{N}}$ é uma coleção enumerável de conjuntos em $\mathcal{A}$, então a interseção $\cap_{n\in\mathbf{N}} A_n = K - \cup_{n\in\mathbf{N}}(K - A_n)$ está em $\mathcal{A}$.

Os conjuntos A em $\mathcal{A}$ são chamados de *conjuntos $\mathcal{A}$-mensuráveis*, ou, simplesmente *conjuntos mensuráveis*, se estiver claro qual a σ-álgebra $\mathcal{A}$ a qual estamos nos referindo.

Quando K for um conjunto finito ou enumerável a única σ-álgebra que iremos considerar sobre K será $\mathfrak{p}(K)$.

Observamos que quando K é finito, então $\mathfrak{p}(K)$ é finito, mas quando K é enumerável infinito, então $\mathfrak{p}(K)$ não é enumerável.

Definição 1.2. *Uma probabilidade P sobre uma σ-algebra $\mathcal{A}$ de subconjuntos de K é uma lei que associa a cada A em $\mathcal{A}$ um número real $P(A) \geq 0$ tal que*

a) $P(K) = 1$,

b) se E_n , $n \in \mathbf{N}$, é uma coleção enumerável de subconjuntos de $\mathcal{A}$ tal que $E_n \cap E_m = \emptyset$, para $m \neq n$, então $P(\cup_{n\in\mathbf{N}} E_n) = \sum_{n\in\mathbf{N}} P(E_n)$.

Como $K \cup \emptyset = K$, então $P(\emptyset) = 0$.

Chamamos o par $(K, \mathcal{A})$, como definido acima, de espaço mensurável.

Uma vez fixada a probabilidade P sobre a σ-álgebra $\mathcal{A}$ chamamos a tripla $(K, \mathcal{A}, P)$ de espaço de probabilidade.

Um exemplo simples aparece quando jogamos um dado: existem seis faces e intuitivamente sabemos que cada uma tem probabilidade $1/6$ de sair. Podemos modelar tal problema com $K = \{1, 2, 3, 4, 5, 6\}$, e $P(\{i\}) = 1/6$, onde $i \in K$.

A σ- algebra $\mathcal{A}$ é o conjunto das partes $\mathfrak{p}(K)$. A probabilidade de sair 1 ou 2 é 2/6. Ou seja

$$P(\{1,2\}) \,=\, P(\{1\}) + P(\{2\}) \,=\, 1/6 \,+\, 1/6 \,=\, 2/6.$$

Este é um problema não determinístico. Não podemos afirmar qual face irá sair ao lançar o dado, podemos apenas falar na probabilidade de sair uma certa face.

A ideia que o conceito geral de probabilidade traduz é que um conjunto $A \in \mathcal{A}$ tal que $P(A) = 0$ é de natureza desprezível (probabilisticamente falando), ou seja, $w \in A$ não ocorre em termos probabilísticos (ou, estatísticos). Por sua vez, um conjunto A tal que $P(A) = 1$ traduz o fato que $w \in A$ ocorre com certeza em termos probabilísticos. Estamos tratando aqui de eventos aleatórios, sendo assim só se pode falar da probabilidade de algo ocorrer. O conjunto A descreve um conjunto de elementos w com certas propriedades. Quanto maior for o valor $P(A)$, maior a probabilidade de $\omega \in A$ ocorrer.

No exemplo do dado considerado antes, apenas o conjunto vazio tem probabilidade zero. Neste caso, não aparece de maneira muito transparente o que estamos tentando destacar acima. O problema é que tomamos um exemplo muito simples. Outros mais complexos aparecerão em breve.

Quando K for finito ou enumerável, ou seja, da forma $K = \{k_1, k_2, ..., k_n\}$ ou $K = \{k_1, k_2, ..., k_n, ...\}$, a probabilidade P sobre $\mathcal{A} = \mathfrak{p}(K)$ fica especificada apenas pelos valores $p_i = P(\{k_i\})$. Note que assumimos que $\sum_{j=1}^{\infty} P(\{k_i\}) = 1$.

É usual denotar o espaço K onde P atua por Ω e denominá-lo de espaço amostral. Um elemento w em Ω será denominado de uma amostra de Ω ou também de um evento.

Definição 1.3. *Considere K equipado com uma σ-álgebra $\mathcal{A}$ e V outro conjunto equipado com uma σ-álgebra $\mathcal{G}$. Uma função $\phi : K \to V$ é chamada de mensurável se $\phi^{-1}(A) \in \mathcal{A}$ para todo $A \in \mathcal{G}$.*

Consideraremos inicialmente apenas funções mensuráveis $\phi : K \to V$, em que V é finito ou enumerável, $V = \{v_1, v_2, ..., v_i, ...\}$, e contido em $\mathbb{R}$ (ou ainda em $\mathbb{R}^m$ para algum m natural). O conjunto K não precisa necessariamente estar dentro de um $\mathbb{R}^k$.

Uma função mensurável $\phi : (K, \mathcal{F}) \to (\mathbb{R}, \mathcal{R})$ é também denominada de Variável Aleatória.

Quando $\mathcal{A}$ for $\mathfrak{p}(K)$, o conjunto das partes de K, então qualquer $\phi : K \to V$ é mensurável (para qualquer σ-álgebra que consideremos sobre V).

Sobre $\mathbb{R}$ será considerado muitas vezes uma sigma-algebra denominada de sigma-algebra de Borel e denotada algumas vezes por $\mathcal{R}$ e que será propriamente definida mais tarde (ver seção 5).

Esta sigma-algebra $\mathcal{R}$ contém todos os intervalos (e assim uniões e interceções enumeráveis de intervalos). Um ponto qualquer x_0 é tal que $\{x_0\} = \cap_{n\in\mathbb{N}}(x_0 - 1/n, x_0 + 1/n)$, e assim o conjunto $\{x_0\}$ está também em $\mathcal{R}$. Na parte inicial do livro não iremos necessitar de tal conceito.

Em qualquer caso é usual a notação $X : (\Omega, \mathcal{A}) \to (V, \mathcal{G})$, ou seja se utiliza letras maiúsculas para descrever a função $\phi = X$ com as propriedades acima descritas.

É comum denotar os elementos de Ω por w e os elementos onde X toma valores por x, logo x denota um elemento em V (a letra minúscula x correspondendo a maíuscula X). Assim se considerarmos uma $Y : (\Omega, \mathcal{A}) \to (V, \mathcal{G})$ os elementos de V serão denominados de y. Vamos apresentar agora dois exemplos.

Um exemplo de um jogo que ilustra os conceitos acima: considere um dado com seis faces e que será jogado uma vez. O jogo é o seguinte: se sair a face 1 ou 2 ganhamos 1 real, caso contrário ganhamos 2 reais. Considere $X : \{1, 2, 3, 4, 5, 6\} = \Omega \to \{1, 2\}$, tal que

$$X(1) = 1,\ X(2) = 1\ , X(3) = 2\ , X(4) = 2\ , X(5) = 2,\ X(6) = 2.$$

A função X descreve o que vamos ganhar quando se joga o dado em função da face que sai. Neste caso é natural concluir que

$$P(\{w : X(w) = 1\}) = 1/3 \quad \text{e} \quad P(\{w : X(w) = 2\}) = 2/3.$$

Seguindo a notação descrita acima, a σ-álgebra a ser considerada em Ω é $\mathcal{A} = \mathfrak{p}(\{1, 2, 3, 4, 5, 6\})$. Ainda $V = \{1, 2\}$ e $\mathcal{G} = \mathfrak{p}(\{1, 2\})$.

É fácil ver que soma, produtos e compostas de funções mensuráveis determinam novas funções mensuráveis.

Como outro exemplo de probabilidade, considere uma cidade que possui população de $N = 10.000$ habitantes e que cada habitante utiliza um e apenas um de dois provedores de internet. Suponhamos que num dado dia 7.000 habitantes utilizam o provedor 1 e 3.000 usam o provedor 2.

Neste caso é natural tomar Ω como o conjunto dos 10.000 habitantes, $\mathcal{A}$ como a classe das partes $\mathfrak{p}(\Omega)$ e ainda considerar a probabilidade P em $\mathfrak{p}(\Omega)$ tal que

$$P(A) = \frac{\text{número de pessoas no conjunto } A \subset \Omega}{10.000}.$$

Seja $X : \Omega \to V = \{1, 2\}$, que associa cada habitante a seu provedor no dia em questão. Como Ω e V são finitos (logo $\mathcal{A} = \mathfrak{p}(\Omega)$) então X é mensurável.

Aqui terminam os exemplos.

Seja $\phi : (K, \mathcal{A}, P) \to (V, \mathcal{G})$ mensurável, então fica naturalmente definido uma probabilidade $\tilde{P}$ sobre $(V, \mathcal{G})$, através de $\tilde{P}(B) = P(\phi^{-1}(B))$, para cada $B \in \mathcal{G}$. É fácil ver que $\tilde{P}$ é uma probabilidade sobre $(V, \mathcal{G})$.

Por exemplo, para uma coleção de $E_n \in \mathcal{G}$ disjuntos,

$$\tilde{P}(\cup_n E_n) = P(\phi^{-1}(\cup_n E_n)) = P(\cup_n \phi^{-1}(E_n)) = \sum_n P(\phi^{-1}(E_n)) = \sum_n \tilde{P}(E_n).$$

Observação: algumas vezes denotaremos tal $\tilde{P}$ por P_ϕ, para enfatizar que foi obtida de P através de ϕ.

Dado $X : (\Omega, \mathcal{A}, P) \to (V, \mathcal{G})$, diremos que $\tilde{P} = P_X$ é a probabilidade (em V) induzida pela função mensurável (variável aleatória) X e pela probabilidade P (em Ω).

Voltando ao nosso problema original, tal probabilidade P_X é denominada de *distribuição da variável aleatória* X. Quando X toma valores reais, ou seja $V = \mathbb{R}$, a probabilidade P_X estará definida para subconjuntos B da reta real.

Em breve veremos que a distribuição P_X de X é na verdade mais importante que a própria probabilidade P.

No exemplo que consideramos acima com dois provedores de internet no qual definimos $X : \Omega \to V = \{1, 2\}$, obtemos, a partir da probabilidade P inicialmente considerada, uma nova probabilidade $\tilde{P} = P_X$ definida acima como $\tilde{P}(\{1\}) = 0.7$ e $\tilde{P}(\{2\}) = 0.3$. Usando a propriedade aditiva da probabilidade, então é claro que disto segue que $\tilde{P}(\{1, 2\}) = 1$. Ainda $\tilde{P}(\emptyset) = 0$. Deste modo ficaram explícitos os valores de $\tilde{P} = P_X$ sobre $\mathcal{P}(\{1, 2\})$. Este é um dos exemplos mais simples de probabilidade que conseguimos imaginar. Podemos então dizer que utilizar o provedor 1 tem probabilidade 0.7 e utilizar o provedor 2 tem probabilidade 0.3, ou seja $P_X(\{1\}) = 0.7$ e $P_X(\{2\}) = 0.3$. É preferível a notação $P(X = 1) = 0.7$ e $P(X = 2) = 0.3$.

Acima $P(X = 1)$ significa $P(\{w : X(w) = 1\})$, etc...

Fixado X, para simplificar a notação, muitas vezes não se diferencia a probabilidade P (inicialmente considerada) da distribuição P_X, omitindo assim expressões com $\tilde{P}$ ou P_X (que age sobre subconjuntos de V) e usando apenas a letra P (que age sobre subconjuntos de Ω). Logo, quando esta claro de qual X falamos, não deve ser motivo de confusão falar de $P(B)$ para um conjunto $B \subset V$. Neste caso, $P(B) = P(\{\omega \in \Omega : X(\omega) \in B\})$.

Um exemplo ilustrativo é o seguinte: seja $\Omega = [0, 1] = \{x \in \mathbb{R} : 0 \leq x \leq 1\}$ e $X : [0, 1] \to \mathbb{R}$. As σ-álgebras "naturais"a serem consideradas neste caso serão descritas com detalhe na seção 5.

Suponha que X seja uma função mensurável. Vamos ver no capítulo 5 que se X é contínua então ela é mensurável. Uma probabilidade P natural a ser considerada aqui é aquela que a um subconjunto da forma $[a,b]$ (um intervalo contido em $[0,1]$) dá o valor $P([a,b]) = b-a$. Observamos que os intervalos serão conjuntos mensuráveis na σ-álgebra natural. No entanto, esta sigma-algebra não será $\mathfrak{p}([0,1])$ (conforme capítulo 5).

Desejamos introduzir o conceito de integral de uma função mensurável com respeito a uma probabilidade.

Se $X = I_{[a,b]}$ é natural dizer que $\int X(x)dP(x) = \int XdP = \int I_{[a,b]}dP = b-a$. Por exemplo, se $X = I_{[0.3,0.7]}$, então $\int XdP = 0.4$.

Ainda se $X = 5\, I_{[0.3,0.7]} + 8\, I_{[0.9,1.0]}$ então

$$\int X(x)\, dP(x) \;=\; 5\, P[0.3, 0.7] + 8\, P[0.9, 1.0] \;=\; 5\ 0.4 + 8\ 0.1 = 2.8.$$

Isto porque qualquer conceito razoável de integral deveria ser linear na função a ser integrada.

Suponha que $x_0 = 0 < x_1 < x_2 < ... < x_{n-1} < x_n$. Considere os intervalos $c_n = [x_k, x_{k+1}]$, $k = 0, ..., n-1$. Seja agora $X = \sum_{k=0}^{n-1} \alpha_n I_{c_n}$, onde $\alpha_k \in \mathbb{R}$, $k = 0, ..., n-1$. Parece também natural que

$$\int X\, dP = \sum_{k=0}^{n-1} \alpha_k P([x_k, x_{k+1}]) = \sum_{k=0}^{n-1} \alpha_k\, (x_{k+1} - x_k).$$

Vamos pensar por exemplo que os peixes num grande lago tem comprimento entre 0 e 1 metro e que a probabilidade do peixe ter comprimento com valor no intervalo $[a,b]$ é $b-a$ (a unidade de medida é metro). Estes peixes são pescados e vendidos no mercado. Após medições de alguns peixes colhidos no lago e o faturamento da venda dos mesmos ao longo dos meses se chegou a conclusão que

1) os peixes com comprimento no intervalo $[0, 1/4]$ são vendidos a 10 reais,

2) os peixes com comprimento no intervalo $(1/4, 3/4]$ são vendidos a 15 reais,

e,

3) os peixes com comprimento no intervalo $(3/4,1]$ são vendidos a 25 reais.

Vamos supor que os peixes são pescados de forma que com a probabilidade $b-a$ eles tem comprimento entre a e b. Vamos denotar por P tal probabilidade. Com isto queremos dizer, por exemplo, que se forem pescados 400 peixes, então o número N dos que tem comprimento entre $3/4=0.75$ e 1 metro seria tal que

$$\frac{N}{400}=1/4=1-3/4=P([3/4,1]).$$

Este P é o mesmo que vimos acima.

Assim, o valor

$$10\,P([0,1/4])+15\,P(1/4,3/4]+25\,P(3/4,1]=$$

$$10\ \times 0.25+15\ \times 0.5+25\ \times 0.25=2.5+7.5+6.25=16.25$$

parece indicar o valor médio de venda de peixe. Assim, se pescarmos e vendermos 400 peixes receberíamos em reais

$$400\times 16.25=6500.$$

Dizemos que $10\ \times 0.25+15\ \times 0.5+25\ \times 0.25=16.25$ nos dá o valor esperado de venda de peixe.

Note que se $X=10\,I_{[0,1/4]}+15\,I_{[1/4,3/4]}+25\,I_{[3/4,1]}$ então

$$\int XdP=10\,P([0,1/4])+15\,P((1/4,3/4])+25\,P((3/4,1])=16.25.$$

É usual denotar por $\mathbb{E}(X)$ (valor esperado da variável X segundo P) esta integral $\int X\,dP$. No caso seria o valor esperado do preço do peixe vendido.

Para uma função contínua $X:[0,1]\to\mathbb{R}$ a integral $\int_0^1 X(x)dx$ é um conceito bem conhecido do Cálculo Diferencial e Integral. A integral de Riemann de X é obtida através do limite de somas de Riemann da forma

$$\sum_{k=0}^{n-1} X(y_k)\,(x_{k+1}-x_k),$$

onde $y_k \in [x_k, x_{k+1}]$.

Este limite vai acontecer quando n for grande e todos os comprimentos envolvidos $|x_{k+1} - x_k|$, $k = 0, 1, ..., n-1$, forem uniformemente pequenos.

Em resumo os valores $\sum_{k=0}^{n-1} X(y_k)\,(x_{k+1} - x_k)$ são aproximações da integral $\int X(x)dx$.

Assim, dada uma função contínua $X : [0, 1] \to \mathbb{R}$, a integral $\int X(x)dx$ descreve de alguma forma o valor esperado da variável X segundo a P descrita acima. Será então natural denotar $\mathbb{E}(X) = \int X(x)dx$.

Vamos agora voltar ao caso geral. Dada uma função mensurável $X : \Omega \to \mathbb{R}$ desejamos dar sentido ao que seria $\int X\,dP$. Ou seja, dar um sentido a integral de X com relação a P.

Se A é um conjunto mensurável será então natural definir $\int I_A\,dP = P(A)$. Ainda, se $X = \sum_{i=1}^n \alpha_i I_{A_i}$, onde os A_i são mensuráveis disjuntos dois a dois, e os α_k são números reais, então

$$\int X\,dP = \sum_{i=1}^{n} \alpha_i P(A_i).$$

Suponha agora que $U_{i=1}^n A_i = \Omega$.

Para uma função mensurável geral $X : \Omega \to \mathbb{R}$, pode-se tomar uma aproximação da forma acima $\sum_{i=1}^n X(y_i)P(A_i)$ (com n grande e $U_{i=1}^n A_i = \Omega$), onde cada $y_i \in A_i$ e as probabilidades $P(A_i)$ são uniformemente pequenas. No limite se teria o valor $\int X\,dP$. O paralelo com a integral de Riemann descrita acima fica evidente. Esta é uma descrição bem informal do que será rigorosamente apresentado no Capítulo 5.

Neste caso o valor esperado de X segundo P seria $\int XdP = \mathbb{E}(X)$. Na Definição 2.16 abaixo vamos voltar a abordar este conceito em um caso interessante em que $\int XdP$ pode ser calculado de uma maneira simples e natural.

Vamos agora definir o que é um processo estocástico.

Definição 1.4. *Seja* $(\Omega, \mathcal{A}, P)$ *espaço de probabilidade,* $(S, \mathcal{G})$ *um espaço mensurável e ainda uma família de variáveis aleatórias* X_t *indexadas por um parâmetro* $t \in T$*, onde* $T \subset \mathbb{R}$ *(isto é, cada* $X_t : (\Omega, \mathcal{A}, P) \to (S, \mathcal{G})$ *é mensurável). Dizemos que tal* $(X_t)_{t\in T}$ *é um processo estocástico. No presente texto* S **é sempre finito ou enumerável** *e assim* $\mathcal{G} = \mathfrak{p}(S)$*.*

Definição 1.5 (**Espaço de índices ou parâmetros temporais**)**.** *O conjunto* $T \neq \emptyset$ *contido em* $\mathbb{R}$ *é denominado espaço de parâmetros temporais, ou, índices do processo. O conjunto* T *possui uma ordem e vamos pensar que para cada* $t \in T$ *a variável* X_t *descreve o que acontece com o processo no tempo* t*.*

Dois casos importantes são:

Parâmetro Temporal Discreto - $T = \mathbb{N}$*, ou* $\mathbb{Z}$*, ou ainda* $\{1, 2, ..., n\}$*.*

Parâmetro temporal Contínuo - $T = [a, b]$*, ou* $T = \{t \in \mathbb{R} : t \geq 0\} = \mathbb{R}^+$ *ou ainda* $T = \mathbb{R}$*.*

Definição 1.6 (**Espaço de Estados**)**.** *É o conjunto* S*, ou seja, o elenco dos possíveis valores de cada variável aleatória* X_t*.*

Quando S é finito, S será **SEMPRE** descrito na forma $\{1, 2, ..., m\}$ ou ainda da forma $\{1, 2, ..., m\}^k$, onde m e k são números naturais.

Por exemplo, se S denota o conjunto de três possíveis canais de TV, escolheremos para cada um deles um número de 1 a 3.

Se S for enumerável infinito S será **SEMPRE** descrito na forma $S = \mathbb{N}$, ou, $S = \mathbb{Z}$ (dependendo do caso).

Quanto ao parâmetro temporal, o único caso que iremos tratar aqui neste capítulo é quando $T = \mathbb{N}$. Alguns exemplos iniciais serão para T finito.

Para todo $w \in \Omega$ fixado, e t também fixo, $X_t(w)$ determina o valor do processo no tempo t avaliado em w e algumas vezes é denotado por w_t. Quando $w \in \Omega$ está fixo e t variável, os valores $X_t(w) = w_t$ descrevem a evolução temporal ao longo do tempo $t \in T$. Usamos a letra grega ω para denotar $\omega = \{w_t\}_{t\in T} \in S^T$ associado a um certo w. Observe que $w \in \Omega$ e $\omega \in S^T$. Por exemplo, se $S = \{1, 2, 3\}$ e $T = \mathbb{N}$, um ω poderia ser a sequência ordenada infinita $(2, 2, 3, 1, 2, 3, 1, 2, 3, 3, ...) = (w_1, w_2, w_3, ..., w_n, ...)$.

Note ainda que usaremos aqui a seguinte notação: se o conjunto A é definido por

$$A = \{w : X_{t_1}(w) = a_1, X_{t_2}(w) = a_2, ..., X_{t_n}(w) = a_n\} =$$

$$\{X_{t_1} = a_1, X_{t_2} = a_2, ..., X_{t_n} = a_n\}$$

e B por

$$B = \{w : X_{s_1}(w) = b_1, X_{s_2}(w) = b_2, ..., X_{s_m}(w) = b_m\},$$

então

$$A \cap B = \{w : X_{t_1}(w) = a_1, X_{t_2}(w) = a_2, ..., X_{t_n}(w) = a_n,$$

$$X_{s_1}(w) = b_1, X_{s_2}(w) = b_2, ..., X_{s_m}(w) = b_m\},$$

ou seja, sem maior preocupação com a ordem dos tempos envolvidos.

Concretamente, se

$$A = \{X_2 = 1, X_6 = 3, X_8 = 4\},$$

e

$$B = \{X_1 = 4, X_2 = 1, X_8 = 3, X_9 = 2\},$$

podemos denotar $A \cap B$, indistintamente como

$$A \cap B = \{X_1 = 4, X_2 = 1, X_6 = 3, X_8 = 3, X_8 = 4, X_9 = 2\},$$

ou como

$$A \cap B = \{X_2 = 1, X_6 = 3, X_8 = 4, X_1 = 4, X_2 = 1, X_8 = 3, X_9 = 2\}.$$

Note que neste caso

$$\{X_1 = 4, X_2 = 1, X_6 = 3, X_8 = 3, X_8 = 4, X_9 = 2\} = \emptyset,$$

porque não pode existir w tal que $X_8(w) = 4$ e $X_8(w) = 3$.

Não confunda este conjunto com

$$\emptyset \neq \{X_1 = 4, X_2 = 1, X_6 = 3, X_8 \in \{3,4\}, X_9 = 2\} =$$
$$\{X_1 = 4, X_2 = 1, X_6 = 3, X_8 = 3, X_9 = 2\} \cup$$
$$\{X_1 = 4, X_2 = 1, X_6 = 3, X_8 = 4, X_9 = 2\}.$$

Note ainda que

$$\{X_1 \in \{2,4\}, X_2 = 1, X_4 \in \{3,4\}\} \neq$$
$$\{X_1 = 4, X_2 = 1, X_4 = 3\} \cup$$
$$\{X_1 = 2, X_2 = 1, X_4 = 4\}.$$

Prosseguindo com o exemplo da cidade com $N = 10.000$ habitantes, vamos supor que a cada mês se faz uma enquete e cada pessoa informa qual internet está usando. Vamos estabelecer que se vai realizar enquetes em três oportunidades seguidas com intervalo de um mês. Fica assim determinado que $T = \{1, 2, 3\}$, $S = \{1, 2\}$ e X_t descreve qual provedor uma determinada pessoa estava utilizando no dia da t-ésima enquete, $t \in \{1, 2, 3\}$.

Para simplificar assumimos que, em cada mês, cada pessoa w utiliza um e apenas um provedor. Neste caso, é natural tomar $\Omega = \{1, 2, ..., N\}$, e escolher uma amostra w representa escolher um habitante da cidade. Neste caso, $X_t(w) = w_t \in S$, $t \in \{1, 2, 3\}$, descreve o provedor utilizado pelo indivíduo (ou amostra) w na t-ésima enquete.

Um elemento ω poderia ser, por exemplo, $\omega = (1, 2, 1) \in S^T$. Este ω corresponde a indivíduos w tais que usaram a internet 1 no mês 1, a internet 2 no mês 2 e a internet 1 no mês 3.

As perguntas que estaremos preliminarmente interessados em analisar envolvem por exemplo: qual o valor de

$$P(X_1 = 2, X_2 = 2, X_3 = 1) = P(X_1 \in \{2\}, X_2 \in \{2\}, X_3 \in \{1\}) =$$

$$P(\{w \in \Omega, \text{ tais que } X_1(w) = 2, X_2(w) = 2, X_3(w) = 1\}\,)?$$

Para efetuar este cálculo, contamos cada indivíduo que na primeira enquete usava o provedor 2, na segunda o mesmo provedor 2, e na terceira trocou para o provedor 1. A seguir dividimos o número obtido por $N = 10.000$.

Os possíveis w_t tomariam valores em

$$S^T = \{1,2\}^3 = \{1,2\} \times \{1,2\} \times \{1,2\}.$$

Note que diferentes pessoas $w \in \Omega$ podem determinar o mesmo valor $\omega = (w_t)_{t\in T} \in \{1,2\}^3$.

Um conceito de fundamental importância em probabilidade é o de probabilidade condicional.

Definição 1.7. *Fixado $(\Omega, \mathcal{A}, P)$, denotamos por*

$$P(A \,|\, B) = \frac{P(A \cap B)}{P(B)},$$

a probabilidade de ocorrer A dado que ocorreu B. Isto só faz sentido, é claro, se $P(B) \neq 0$.

Por exemplo, para saber qual a probabilidade de um estudante do colégio B passar no exame vestibular da universidade A, considera-se o quociente

$$\frac{\text{número de estudantes do colégio } B \text{ que passaram na universidade } A}{\text{número de estudantes do colégio } B}.$$

Definição 1.8. *Fixada uma probabilidade P, dizemos que o evento definido pelo conjunto A é independente do evento definido pelo conjunto B se*

$$P(A \,|\, B) = P(A),$$

ou seja, se

$$P(A \cap B) = P(A)\, P(B).$$

Esta propriedade descreve o fato que para os conjuntos A e B em consideração, a ocorrência de A não influencia, em têrmos estatísticos, a ocorrência ou não de B.

Dadas funções X, Y o sentido de

$$P(X = a \,|\, Y = b) \quad \text{é} \quad P(\{w_1 |\, X(w_1) = a\} \mid \{w_2 \,|\, Y(w_2) = b\}\,).$$

Dadas funções X, Y, Z, V o sentido de

$$P(X = a\,, Y = b|\, Z = c,\, V = d) \quad \text{é} \quad P(\{\, X = a, Y = b\} \mid \{\, Z = c\ , V = d\}\,).$$

Definição 1.9. *Sejam $(X, \mathcal{F}, P)$, onde $X(w) \in S$, e $(Y, \mathcal{F}, P)$, onde $Y(w) \in S_1$, onde S e S_1 finitos. Diremos que X é independente de Y, se para quaisquer elementos $a \in S \subset \mathbb{R}$ e $b \in S_1 \subset \mathbb{R}$, vale que*

$$P(X = a \,|\, Y = b) = P(X = a),$$

ou seja, se

$$P(X = a,\, Y = b) = P(\{X = a\} \cap \{Y = b\}) = P(\{X = a\})\, P(\{Y = b\}).$$

Como exemplo, considere X a variável que descreve a face que sai quando se joga um dado pela primeira vez e Y a variável que descreve a face que sai quando jogamos o dado pela segunda vez.

Sejam $a, b \in \{1, 2, 3, 4, 5, 6\}$ fixados, então

$$P(X = a,\, Y = b) = 1/36 \,=\, (1/6)^2.$$

Isto porque, temos ao todo 6^2 possibilidades de saídas de pares ordenados de faces $(x, y) \in \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$. O par (a, b) corresponde a apenas uma possibilidade. Cada par tem a mesma chance de sair, logo tem mesma probabilidade.

Sendo assim,

$$P(X=a,\, Y=b) = 1/36 \,=\, 1/6\;\, 1/6 \,=\, P(X=a)\, P(Y=b).$$

Logo, a face que sai no primeiro lançamento do dado é independente do que sai no segundo lançamento. É usual colocar o tempo como parâmetro e assim denominar $X = X_1$ e $Y = X_2$. Se fôssemos lançar a moeda uma terceira vez, o resultado seria X_3.

Dadas funções X, Y, Z, V dizemos que X, Y são independentes de Z, V se para quaisquer a, b, c, d vale

$$P(X=a\,, Y=b|\, Z=c,\, V=d) \,=P(X=a\,, Y=b).$$

Da mesma forma, Dadas funções $X_1, X_2, .., X_n$, dizemos que elas independem das funções $Y_1, Y_2, .., Y_k$ se para quaiquer $x_1, x_2, .., x_n,\; y_1, y_2, .., y_k$ vale

$$P(X_1=x_1, X_2=x_2, ..., X_n=x_n|\, Y_1=y_1,\, Y_2=y_2, ..., Y_k=y_k) \,=$$

$$P(X_1=x_1, X_2=x_2, ..., X_n=x_n).$$

Suponha que funções X, Z, V são tais que X, Z independam de V. Então

$$P(X=a\,|\, Z=c,\, V=d) \,= \frac{P(X=a\,, Z=c, V=d)}{P(Z=c, V=d)} =$$

$$\frac{P(X=a\,, Z=c, V=d)}{P(V=d)}\, \frac{P(V=d)}{P(Z=c, V=d)} =$$

$$P(X=a\,, Z=c\,|\, V=d)\, \frac{P(V=d)}{P(Z=c)\, P(V=d)} =$$

$$P(X=a\,, Z=c)\, \frac{1}{P(Z=c)} = P(X=a\,|\, Z=c) \;\; (*)$$

Ou seja, a informação de V pode ser descartada.

Outra propriedade interessante é a seguinte: suponha que X seja independente de Y e Z.

Então, X é independente de $Y+Z$.

De fato, dados a, b então

$$P(X=a \,|\, Y+Z=b) = \frac{P(X=a\,,\, Y+Z=b)}{P(Y+Z=b)} =$$

$$\sum_c \frac{P(X=a\,,\, Y=c,\, Z=b-c)}{P(Y+Z=b)} =$$

$$P(X=a) \sum_c \frac{P(\,Y=c,\, Z=b-c)}{P(Y+Z=b)} = P(X=a) \qquad (**)$$

É fácil ver de forma semelhante que se X é independente de $Y_1, Y_2, .., Y_k$. X é independente de $Y_1+Y_2+...+Y_k$, então

Voltando ao modelo do uso da internet, poderíamos definir uma variável Y tal que $Y(w)=4$ se a renda mensal do indivíduo w é abaixo de $4.000,00$ reais e $Y(w)=5$ caso contrário. Neste caso, $S_1=\{4,5\}$, e se por acaso X é independente de Y, então existe uma clara indicação de que o uso da internet 1 ou 2 é independente da classe de renda do indivíduo w.

Podemos nos perguntar também: qual a probabilidade de uma pessoa utilizar a internet 1 na terceira enquete, dado que utilizou a internet 2 nas duas primeiras? Será que nesta questão específica existe independência?

Para responder tal pergunta devemos calcular

$$P(X_3=1 \,|\, X_1=2, X_2=2) = \frac{P(X_1=2, X_2=2, X_3=1)}{P(X_1=2, X_2=2)}.$$

Vai existir independência (do que acontece no tempo 3 em função do uso anterior no tempo 1 e 2), se, por acaso,

$$\frac{P(X_1=2, X_2=2, X_3=1)}{P(X_1=2, X_2=2)} = P(X_3=1).$$

Não nos parece natural que vá ocorrer independência, pois existe sempre uma certa dose de inércia na índole das pessoas: se um indivíduo usava a internet 2 no mês 2, então o valor da probabilidade que ele vá continuar usando a internet 2 no mês 3 é maior do que o valor da probabilidade que ele passe a usar a internet 1 no mês 3.

Para responder com certeza a pergunta acima seria necessário obter os dados exatos sobre os habitantes da tal cidade nestas três oportunidades e fazer a conta acima.

Voltemos agora ao exemplo do uso da internet. Consideramos finalmente o caso mais interessante em que $T = \mathbb{N}$ e $X_t : \Omega \to \{1, 2\}$ que vai descrever a evolução temporal ilimitada do uso do provedor de cada habitante w da cidade. Uma pergunta natural que podemos nos fazer neste caso é a seguinte: será que existem os limites

$$\lim_{t\to\infty} P(X_t = 1) = \pi_1,$$

e

$$\lim_{t\to\infty} P(X_t = 2) = \pi_2?$$

Outra questão: será que existe o limite

$$\lim_{t\to\infty} P(X_t = 1 \,|\, X_1 = 2)?$$

Um dos objetivos da teoria é atacar questões desta natureza.

Se assume no modelo que a tendência de mudança no mês subsequente depende basicamente do comportamento dos habitantes no mês atual. Em termos matemáticos, isto poderia ser descrito - por exemplo - como

$$P(X_5 = i \,|\, X_4 = j, X_3 = k, X_2 = r, X_1 = s) = P(X_5 = i \,|\, X_4 = j).$$

Esta propriedade descreve o que vai ser formalizado posteriomente como a propriedade Markoviana.

Definição 1.10. *Para cada amostra $w \in \Omega$ fixada, seja a sequência $\omega = (w_t)_{t\in\mathbb{N}} = (X_t(w))_{t\in\mathbb{N}} \in \{\omega : \mathbb{N} \to S\} = S^{\mathbb{N}}$, que será denominada de caminho amostral.*

Na verdade os $\omega = (w_t)_{t\in\mathbb{N}}$ desempenham na teoria um papel mais fundamental do que os w.

Os exemplos do mundo real, no entanto, muitas vezes aparecem de maneira natural no domínio dos $w \in \Omega$. Por exemplo, no caso do exemplo do uso da internet, analisando a utilização de provedor em tres meses, diferentes pessoas poderiam determinar a mesma sequencia $\omega = (1, 2, 2)$. Ou seja, varios $w \in \Omega$ poderiam determinar o mesmo $\omega \in \{1, 2\}^3$.

Alertamos o leitor que, fixado o processo estocástico $(X_t)_{t\in\mathbb{N}}$, é usual não fazer muita distinção entre w e ω e também entre Ω e $S^{\mathbb{N}}$. Ou seja, podemos falar em $w \in \Omega$, ou $\omega \in S^{\mathbb{N}}$. Preferimos analisar as questões sobre o ponto de vista dos $\omega \in S^{\mathbb{N}}$ que são mais fundamentais do ponto de vista do modelo matemático.

Vamos considerar a seguir uma classe importante de processos estocásticos.

Definição 1.11. *Fixados $(\Omega, \mathcal{A}, \mathcal{P})$, dizemos que o processo X_t tomando valores em S (enumerável) e com parâmetro $t \in T = \mathbb{N}$ é independente se para cada n e cada sequência $t_1 < t_2 < ... < t_n$, $t_i \in T = \mathbb{N}$, e para cada sequência de conjuntos $A_1, A_2, ..., A_n$, onde $A_i \subset S$, $i \in \{1, 2, ..., n\}$, vale que*

$$P(X_{t_1} \in A_1,\, X_{t_2} \in A_2, ...,\, X_{t_n} \in A_n) =$$

$$P(X_{t_1} \in A_1)\, P(X_{t_2} \in A_2)...P(X_{t_n} \in A_n).$$

Vamos voltar ao exemplo do jogo com um dado que mencionamos antes. Como vimos, neste jogo $P(X = 1) = 1/3$ e $P(X = 2) = 2/3$. Vamos agora jogar o dado sucessivamente e X_t vai descrever o que ganhamos na jogada $t \in T = \mathbb{N}$ em função da face que saiu. É natural assumir que para cada t fixo, a variável X_t é descrita também por X como acima.

Uma conta fácil (levando em conta o conjunto das possibilidades) mostra que

$$P(X_1 = 1,\, X_2 = 2,\, X_3 = 1) = \frac{2\,.4\,.2}{6^3} = \frac{16}{216} =$$

$$\frac{2}{6}.\frac{4}{6}.\frac{2}{6} = P(X_1 = 1)\, P(X_2 = 2)\, P(X_3 = 1).$$

Procedendo de maneira semelhante é fácil ver que

$$P(X_{t_1} = a_1,\, X_{t_2} = a_2, ...,\, X_{t_n} = a_n) =$$

$$P(X_{t_1} = a_1)\, P(X_{t_2} = a_2) ... P(X_{t_n} = a_n),$$

para qualquer sequência $t_1 < t_2 < ... < t_n$ e $a_i \in \{1, 2\}, i \in \{1, 2, ..., n\}$.

Mais explicitamente,

$$P(X_{t_1} = a_1,\, X_{t_2} = a_2, ...,\, X_{t_n} = a_n) = (\frac{1}{3})^k\, (\frac{2}{3})^{n-k},$$

onde k é o número de valores 1 entre os n valores $a_1, a_2, ..., a_n$.

Procedendo da forma acima, agora para conjuntos $A_i \subset S$, é fácil de se concluir que o processo estocástico associado a jogar o dado sucessivas vezes (e ver se obtemos $X = 1$ ou $X = 2$) é um processo independente.

Outra questão: vamos jogar o dado n vezes e denotar por $X_1, X_2, ..., X_n$ os resultados obtidos sucessivamente em ordem de aparecimento; qual a probabilidade de se obter k vezes $X_i = 1$ (ou seja, sair a face 1 ou 2 do dado), entre os $i \in \{1, 2, 3.., n\}$? Ora existem

$$C_n^k = \frac{n!}{(n-k)!\, k!}$$

possibilidades de isto ocorrer no universo de 2^n ocorrências de $X = 1$ ou $X = 2$, ou seja dos possíveis resultados X_i que se obtem ao jogar o dado n vezes.

Cada uma das ocorrências tem probabilidade $(\frac{1}{3})^k\, (\frac{2}{3})^{n-k}$. Aqui estamos usando a expressão acima que segue da independência do processo.

Logo a probabilidade que buscamos vale

$$\frac{n!}{(n-k)!\,k!}\,(\frac{1}{3})^k\,(\frac{2}{3})^{n-k}.$$

Mais geralmente, se ocorrer no jogo uma probabilidade p de sair $X=1$ e uma probabilidade $1-p$ de ocorrer $X=2$ em uma jogada, a probabilidade de ocorrer um total de k vezes o $X=1$ em n jogadas é igual a

$$\frac{n!}{(n-k)!\,k!}\,p^k\,(1-p)^{n-k}.$$

Estamos supondo é claro que existe independência entre as sucessivas jogadas.

Esta distribuição é denominada de Binomial (n,p) e denotada por $\mathcal{B}(n,p)$. Para cada k temos um valor e a soma destes valores para $k=0,1,2,..,n$ é igual a 1.

Para checar que a soma destes valores é exatamente igual a 1, podemos usar o Binômio de Newton:

$$1=(p+(1-p))^n=\sum_{k=0}^{n}\frac{n!}{(n-k)!\,k!}\,p^k\,(1-p)^{n-k},$$

Definição 1.12. *Fixados $(\Omega,\mathcal{A},\mathcal{P})$, dizemos que o processo X_t tomando valores em S (enumerável) e com parâmetro $t\in T=\mathbb{N}$ é independente e identicamente distribuído, denotado por i. i. d., se ele é independente, e ainda, para todo t, vale*

$$P(X_t\in A)=P(X_0\in A),$$

para todo subconjunto A de S.

Note que, neste caso, se denotamos $P(X_0=s)=p_s=P(X_t=s)$, para todo $t\geq 0$ e $s\in S$, então

$$P(X_0=a_0,X_1=a_1,X_2=a_2,...,X_n=a_n)=p_{a_0}\,p_{a_1}\,p_{a_2}\,...p_{a_n}.$$

Vamos agora dar um exemplo prático do uso da Teoria das Probabilidades. Uma companhia aérea possui um avião com s lugares. Ela sabe que em geral ocorre que um certo número de pessoas compram a passagem mas não aparecem na hora do voo. Então ela vende v lugares para o voo e $v > s$. Através da experiência passada, a companhia sabe que, em termos estatísticos, existe uma probabilidade p de comparecimento. Ou seja, cada indivíduo, entre as v pessoas que compram passagem, comparece ao voo com probabilidade p. Qual o risco de que o número de pessoas que comparecem ao vôo supere o número de assentos s? Ora, primeiro note que se pode supor independência na análise da questão. Qual probabilidade $r(j)$ de aparecerem j dos v passageiros que compraram a passagem? Resposta:

$$r(j) = \frac{v!}{(v-j)!\,j!}\, p^j\,(1-p)^{v-j}.$$

Logo, a probabilidade de que o número de pessoas que comparecem ao voo supere o número de assentos s é

$$\sum_{j=s+1}^{v} r(j).$$

Estamos supondo para simplificar que todos os passageiros a embarcar estão começando seu voo na mesma cidade, que não existe a questão de voo de conexão, etc... Assim, é razoável supor a independência do comportamento dos envolvidos e da sua capacidade de comparecer ao voo em consideração, ou não.

Definição 1.13. *Fixados $(\Omega, \mathcal{A}, P)$, $(V, \mathcal{V}) = (V, \mathfrak{p}(V))$, onde $V \subset \mathbb{R}$ é finito ou enumerável, e uma função mensurável $X : (\Omega, \mathcal{A}, P) \to (V, \mathfrak{p}(V))$. Chamamos de integral de X em relação a P, o valor real*

$$\sum_{x\in V} x\ P(\{w \text{ tal que } X(w) = x\}),$$

e que será denotado por $\int X dP = \int X(w) dP(w)$.

Como dissemos antes é usual chamar $\int X dP$ *de esperança da variável aleatória* X *e denotar este valor por* $E(X)$.

Note que dadas as variáveis aleatórias $X : \Omega \to V \subset \mathbb{R}$ e $Y : \Omega \to V \subset \mathbb{R}$, então $\int (X+Y) dP = \int X dP + \int Y dP$.

Vamos calcular agora $E(X_1)$ no caso do exemplo do processo independente em que jogamos sucessivamente uma moeda viciada (identificação: cara=1 e coroa=2). Assumimos que sair 1 tem probabilidade 1/3 e sair 2 tem probabilidade 2/3. Segue da definição acima que

$$E(X_1) = \int X_1 dP = 1.P(X_1 = 1) + 2.P(X_1 = 2) = 1\,.\,1/3\; + 2\,.\,2/3\; = 5/3.$$

Este valor corresponde ao lucro médio esperado quando se joga a moeda uma vez. Note que neste caso, pela independencia, também vale para qualquer tempo i que $E(X_i) = 5/3$.

Vamos considerar agora um exemplo em que se vai lançar varias vezes um dado como seis faces (que determina um processo independente). Assuma que $X_1 = 1$ se a face obtida no lançamento foi 1 ou 2, e ainda, $X_1 = 2$ caso contrário. Sendo assim, $E(X_1) = 5/3$, de forma semelhante ao exemplo anterior. Neste novo jogo se assume que X_1 descreve a ganho em uma logada. Para saber a riqueza acumulada até a terceira jogada deveríamos considerar a variável aleatória $S_3 = X_1 + X_2 + X_3$.

A variável S_3 está definida sobre

$$S = \{1,2,3,4,5,6\}^3,$$

ou seja,

$$S_3 : \{1,2,3,4,5,6\}^3 \to \mathbb{N}.$$

Para simplificar a notação é usual considerar que embora inicialmente $X_1 : \{1,2,3,4,5,6\} \to \{1,2\}$, podemos usar a mesma expressão X_1 para denotar $X_1 : \{1,2,3,4,5,6\}^3 \to \{1,2\}$, ou seja, $X_1(w_1, w_2, w_3) = X_1(w_1)$.

O mesmo vale analogamente para X_2 e X_3. Este procedimento é completamente geral e será utilizado no texto em outros casos similares sem nenhuma menção a esta pequena sutileza. Logo X_1, X_2, X_3 e S_3 podem ser consideradas todas sobre o mesmo domínio $\{1,2,3,4,5,6\}^3$.

Como a integral é aditiva na funcão obtemos então que

$$E(S_3) = E(X_1) + E(X_2) + E(X_3) = 3 \,.\, 5/3 = 5.$$

S_3 descreve o ganho em três jogadas e é uma variável aleatória tomando valores em $\mathbb{R}$. Ainda, $E(S_4) = E(X_1) + E(X_2) + E(X_3) + E(X_4) = 20/3$.

Se tivermos que calcular o que se ganha na média neste jogo em três jogadas, a resposta será $E(S_3) = 5$. Observamos que este jogo com dados (descrito acima) é em tudo semelhante ao jogo anterior em que consideramos uma moeda viciada.

Podemos considerar outros jogos em que dependendo das faces que aparecem sucessivamente quando uma moeda é lançada quatro vezes seguidas se ganha uma certa quantia, ou então um jogo de roleta, etc...

Da mesma forma como antes, se pode calcular a esperança do ganho auferido em cada um destes diversos jogos. Se tivermos que tomar uma decisão sobre qual jogo seria mais lucrativo jogar, a escolha mais sábia seria participar daquele que tem o maior valor esperado do ganho.

Vamos apresentar mais um exemplo da importância do cálculo do valor esperado. No caso da venda de passagem da companhia aérea discutido previamente, assuma que para cada passageiro que exceda os s assentos disponíveis será necessário pagar uma diária de hotel (até que saia o próximo vôo no dia seguinte) de 200,00 reais. Qual será a despesa média esperada $L(v)$ oriunda do procedimento de vender v passagens com $v > s$ para um avião de s lugares?

A resposta é

$$L(v) = \sum_{j=s+1}^{v} r(j)\,(j-s)\,200,00.$$

Agora devemos calcular, em função do valor do preço de cada passagem, o lucro obtido com a venda de t passagens e comparar com o prejuízo $L(v)$ oriundo do eventual comparecimento de mais de s passageiros com a venda de v passagens. Para simplificar nossa análise não estamos levando em conta a influência no número de assentos necessários no avião no dia seguinte na eventualidade de comparecimento de mais de s passageiros num certo dia. A questão relevante para a companhia aérea seria então encontrar para qual v ocorre o valor máximo de lucro. Não vamos discutir aqui as eventuais questões éticas envolvidas no problema.

Após estes exemplos vamos voltar ao nosso tópico principal, ou seja os Processos Estocásticos.

Voltamos agora ao caso geral de processo estocástico $(X_t)_{t\in\mathbb{N}}$, sobre um espaço de probabilidade $(\Omega, \mathcal{A}, P)$ fixado, onde $X_t : \Omega \to S$ para um certo conjunto finito ou enumerável S. O caso que realmente nos interessa é quando $t \in \mathbb{N}$, ou seja, t ilimitado. Neste caso, podemos, por exemplo, analisar o comportamento limite de um caminho amostral $(w_t)_{t\in\mathbb{N}}$ (com probabilidade 1) obtido do processo. Por exemplo, no caso em que $S = \{1, 2\}$, podemos nos perguntar: fixado $\omega = (w_1, w_2, w_3, ..., w_n, ...)$ será que existe o limite

$$\lim_{n\to\infty} \frac{\#\,\text{vezes que aparece o 1 entre os valores } w_1, w_2, w_3, ..., w_{n-1}}{n}.$$

Como se pode calcular tal valor? Este tipo de resultado ocorre, por exemplo, para processos independentes e é contemplado pela Lei dos Grandes Números, um dos teoremas mais importantes da Probabilidade (ver seção 2.1 para um enunciado preciso).

Observação 1 -Existem $\omega = (w_t)_{t\in\mathbb{N}}$ para os quais tal limite não existe. O que desejamos saber é se existe um conjunto $K \subset \Omega = \{1, 2\}^{\mathbb{N}}$ tal que $P(K) = 1$ e para $\omega \in K$, vale que existe o limite acima. Esta é a essência da visão probabilística (ou estatística) de analisar o problema. Tal conjunto K

existe mas **não** depende apenas de finitas coordenadas (é mais complexo que isto como veremos no último capítulo).

Concretamente, no caso em que se joga sucessivamente uma moeda, e associamos 1 à cara e 2 à coroa, se $X_n \in \{1, 2\}$ determina a face que saiu na jogada n-ésima, sabemos intuitivamente que quando a moeda é jogada um número grande de vezes, na média, na metade delas sai cara. Este fato segue da Lei dos Grandes Números e do fato que o processo X_n, neste caso, é independente (conforme capítulo 4 e 5).

Afirmamos que existe um conjunto $K \subset \Omega = \{1, 2\}^{\mathbb{N}}$, onde vale que para qualquer $\omega = (w_1, w_2, ..., w_n, ...) \in K$

$$\lim_{n\to\infty} \frac{\#\,\text{vezes que aparece o 1 entre os valores } w_1, w_2, ..., w_{n-1}}{n} = \frac{1}{2}.$$

O conjunto K vai estar numa sigma-algebra $\mathcal{F}$ na qual se pode falar da sua probabilidade $P(K)$ (oriunda da informação de que a cada jogada existe probabilidade meio de sair cara ou coroa). Afirmamos que $P(K) = 1$!!! Este resultado será demonstrado mais tarde no capítulo 5.

Note que é **possível** que saia **sempre** (portanto, infinitas vezes) cara, o que corresponderia ao evento $\omega = (1, 1, 1, 1, 1, 1, 1, 1, ...)$. O fato é que este elemento está **fora do conjunto** K.

Sendo assim, do ponto de vista probabilístico **não é possível** que $\omega = (1, 1, 1, 1, 1, 1, 1, 1, ...)$ ocorra.

Um fato de fundamental importância é que o conjunto K dos caminhos amostrais $\omega = (w_1, w_2, w_3, ..., w_n, ...)$ que tem limite $1/2$ não depende de finitas coordenadas. É necessário conhecer cada w_n, para todo $n \in \mathbb{N}$, para decidir se ω está ou não em K.

Afirmar que $P(K) = 1$ traduz em termos matemáticos precisos o que a nossa intuição nos diz. Fica claro, desta forma, o sentido do estudo do ponto de vista probabilístico de se entender os fenômenos aleatórios.

Voltemos ao exemplo em que uma companhia vende v lugares para um avião de s lugares, onde $v > s$. Vamos supor que a cada dia é oferecido um voo seguindo esta política, onde v e s estão fixos. Suponha que X_n descreva o número de passageiros que aparecem no dia n. O processo estocástico X_n toma valores em $S = \{0, 1, 2, 3, ..., v\}$. É razoável supor que X_n é um processo independente. Conforme calculado antes, $L(v)$ é o valor esperado de gasto com hotel oriundo desta política. Uma das consequências da Lei dos Grandes Números é que, neste caso, se calcularmos o valor médio durante, digamos 100 dias, iremos obter aproximadamente o valor $L(v)$. Mais precisamente, seja L_n o gasto no dia n com hotel para passageiros que não conseguem lugar no voo (quando $X_n > s$), então da LGN segue que

$$\lim_{n\to\infty} \frac{L_1 + L_2 + ... + L_n}{n} = L(v).$$

Sendo assim, o gasto estimado em 100 dias seria aproximadamente $100\, L(v)$.

Desejamos esclarecer um ponto importante sobre como são em geral introduzidos os modelos na teoria.

Considere $T = \mathbb{N}$, as variáveis $X_t : \Omega \to S$, $t \in \mathbb{N}$, a probabilidade P, etc... Como sempre S é finito, ou, infinito enumerável. Considere agora para n fixo, uma seqüência também fixada de tempos ordenados, $t_1 < t_2 < ... < t_n \in T$.

Considere uma sequência $A_1, A_2, ..., A_n \subset S$ fixada. Considere

$$P(\{\, w \in \Omega \;\text{ tais que }\; X_{t_1}(w) \in A_1,\, X_{t_2}(w) \in A_2, ...,\, X_{t_n}(w) \in A_n\}) \;=$$

$$P(X_{t_1} \in A_1,\, X_{t_2} \in A_2, ...,\, X_{t_n} \in A_n).$$

Dizemos que $P(\, X_{t_1} \in A_1,\, X_{t_2} \in A_2, ...,\, X_{t_n} \in A_n\,)$ determina a distribuição conjunta das variáveis $X_{t_1}, X_{t_2}, ..., X_{t_n}$ quando exaurimos todas as possibilidades de A_i,

$$A_1, A_2, ..., A_n \subset S.$$

As informações acima descritas e que são fornecidas pelas distintas possibilidades de todos os n, $t_1 \leq t_2 \leq ... \leq t_n$, e $A_1, A_2, ..., A_n \subset S$ são denominadas **distribuições finito-dimensionais**. Ela são obtidas a partir do conhecimento explícito de P, de Ω, das X_t, onde $t \in T = \mathbb{N}$ etc...

As distintas classes de Processos Estocásticos que são analisadas em diversos livros, em geral, não são apresentadas do modo concreto (uso da internet, passageiros que utilizam uma companhia aérea). Se assume em cada modelo (na maioria das vezes) que o processo satisfaz certas propriedades baseadas na distribuições finito-dimensionais. Isto é natural quando se busca encontrar modelos e não se parte de um exemplo concreto. Após obtermos uma série de distintos modelos teóricos é que fará sentido analisar um determinado problema do mundo real. Então, nos perguntaremos: em qual dos diversos modelos teóricos anteriormente estudados melhor se encaixa o fenômeno natural em análise? Quanto maior for a riqueza e diversidade de modelos que tivermos a nossa disposição, com mais precisão será descrito o fenômeno natural e melhor poderão ser as nossas previsões futuras a seu respeito. Como se diz, "se na sua caixa de ferramentas, a única que está disponível é um martelo, todo problema vai lhe parecer com um prego".

Definição 1.14. *Os conjuntos da forma*

$$\{X_{t_1} \in A_1, ..., X_{t_j} \in A_j, ..., X_{t_n} \in A_n\},$$

para $t_1 < t_2 < ... < t_n$, $A_1, A_2, ..., A_n \subset S$, *fixos, são denominados de conjuntos cilíndricos ou cilindros.*

Se $S = \{1, 2, 3, .., d\}$ usamos a notação $\overline{a_0, a_1, ..., a_k}$, onde $a_j \in \{1, 2, ..., d\}$ para denotar o cilindro

$$\overline{a_0, a_1, ..., a_k} = \{w = w_0, w_1, w_2, .., w_n, ... \,|\, w_j = a_j, j = 0, 1, 2, ..., k\}.$$

Por exemplo, suponha $S = \{1, 2, 3\}$, $n = 2$, $t_1 = 0, t_2 = 2$, $A_1 = \{1, 2\}, A_2 = \{3\}$, assim

$$\{ X_0 \in \{1, 2\}, X_2 = 3 \} = \{ X_0 \in \{1, 2\}, X_1 \in S, X_2 = 3\} \}$$

é um cilindro.

De outra forma, o conjunto acima é descrito por

$$\{w = (1, w_1, 3, w_3, w_4, ..., w_n, ...)\} \cup \{w = (2, w_1, 3, w_3, 1, w_4, ..., w_k, ...)\}.$$

Pode ser também ser escrito como a união disjunta

$$\overline{1, 1, 3} \cup \overline{2, 1, 3} \cup \overline{1, 2, 3} \cup \overline{2, 2, 3} \cup \overline{1, 3, 3} \cup \overline{2, 3, 3}.$$

Repetindo, o que afirmamos anteriormente, em geral não se parte do conhecimento explícito da P, de Ω, das $X_t : \Omega \to S$, etc..., mas sim supõe-se que P satisfaça alguma condição em termos das distribuições finito dimensionais (que devem satisfazer certas condições de compatibilidade que serão descritas posteriormente). Ou seja, se parte da informação do valor da probabilidade P sobre os cilindros de todos os tamanhos n. Para simplificar vamos pensar que o parâmetro temporal está sobre $\mathbb{N}$.

Conjuntos $A \subset S^{\mathbb{N}} = \Omega$ com maior complexidade que os cilindros são de grande importância. Estes estarão numa certa σ-álgebra $\mathcal{F} \subset \mathfrak{p}(\Omega)$ e o valor $P(A)$ será obtido via um processo de extensão da informação da probabildade dos cilindros. O resultado que assegura esta extensão se denomina o teorema de Caratheodori-Kolmogorov (a ser formalizado no capítulo 5).

Observamos que P, Ω e $\mathcal{F}$ não são únicos. Ou seja, podemos obter distintos P, Ω etc..., a partir da mesma informação obtida das distribuições finito-dimensionais. Dito isto, a partir de agora, um processo estocástico $\{X_n\}_{n\in\mathbb{N}}$ tomando valores em S será, nada mais nada menos que uma probabilidade P sobre o conjunto $\Omega = S^{\mathbb{N}}$. Ainda, se $\omega = (w_0, w_1, w_2, ..., w_n, ...) \in \Omega = S^{\mathbb{N}}$, então, assumiremos **"SEMPRE"** que $X_n(w) = w_n$. Sobre a questão de quem é exatamente a σ-álgebra $\mathcal{F} \subset \mathfrak{p}(\Omega)$, definida no conjunto $\Omega = S^{\mathbb{N}}$, diremos no momento apenas que ela contém todos os conjuntos da forma

$$\{w \in \Omega \,:\, X_{t_1} = a_1, X_{t_2} = a_2, ..., X_{t_n} = a_n\} \subset S^{\mathbb{N}},$$

onde $n \in \mathbb{N}$, $t_1 < t_2 < ... < t_n$ e $a_1, a_2, ..., a_n \in S$.

Vamos apresentar agora um exemplo interessante.

Uma matriz $\mathcal{P} = (P_{ij})_{i,j=1,2,..,n}$ da forma n por n é dita linha estocástica se os elementos P_{ij} da matriz são não-negativos e a soma de cada linha é igual a 1.

Para simplificar a exposição considere $S = \{1, 2\}$ e $T = \mathbb{N}$.

Neste caso, um exemplo de matriz estocástica seria

$$\mathcal{P} = \begin{pmatrix} 1/3 & 2/3 \\ 4/7 & 3/7 \end{pmatrix}.$$

Um vetor $\pi = (\pi_1, \pi_2)$ é denominado de vetor de probabilidade (inicial) sobre $S = \{1, 2\}$ se $\pi \in \mathbb{R}^2$ é tal que $\pi_1, \pi_2 \geq 0$ e $\pi_1 + \pi_2 = 1$.

Por exemplo $\pi = (2/5, 3/5)$.

Fixado $\mathcal{P}$ e π vamos definir primeiro a probabilidade P sobre $\Omega = \{1, 2\}^{\mathbb{N}}$ para certos tipos de conjunto.

Por definição,

$$P(X_0 = a_0, X_1 = a_1, X_2 = a_2, X_3 = a_3, ..., X_n = a_n) =$$

$$P(\,\{w \in \Omega \,:\, X_0 = a_0, X_1 = a_1, X_2 = a_2, X_3 = a_3, ..., X_n = a_n\}\,) =$$

$$\pi_{a_0}\, \mathcal{P}_{a_0\, a_1}\, \mathcal{P}_{a_1\, a_2}\, \mathcal{P}_{a_2\, a_3}\, ...\, \mathcal{P}_{a_{n-1}\, a_n},$$

onde $n \in \mathbb{N}$, $a_i \in \{1, 2\}$ e $0 = t_0 < 1 < 2 < ... < n$.

Ainda, por definição

$$P(X_0 = s) = \pi_s,$$

para todo s.

Por exemplo, no caso da matriz dois por dois $\mathcal{P}$, e do vetor de probabilidade π, descrito acima, obtemos que

$$P(X_0 = 2, X_1 = 1, X_2 = 1) =$$
$$\pi_2\, \mathcal{P}_{2\,1}\, \mathcal{P}_{1\,1} = \frac{3}{5}\,\frac{4}{7}\,\frac{1}{3} = \frac{12}{105}.$$

Afirmamos que as regras de compatibilidade (a que nos referimos informalmente) estão satisfeitas para a P dada, e assim, pelo Teorema de Caratheodori-Kolmogorov (ver Capítulo 5), esta P pode ser considerada sobre uma certa σ-álgebra $\mathcal{F} \subset \mathfrak{p}(\Omega)$ e existe um processo estocástico X_t compatível com a informação fornecida pela informação inicial dada pelo valor de P nos cilindros. Todos os conjuntos da forma

$$\{w \in \Omega \,:\, X_0 = a_0, X_{t_1} = a_1, X_{t_2} = a_2, X_{t_3} = a_3, ..., X_{t_n} = a_n\},$$

estão em $\mathcal{F}$.

O processo X_t, ou seja, a família de funções mensuráveis (ou, variáveis aleatórias) $X_t : \Omega = \{1,2\}^{\mathbb{N}} \to \{1,2\} = S$, $t \in T = \mathbb{N}$, que obtemos é tal que $X_t(\omega) = w_t$, se

$$\omega = (w_0, w_1, ..., w_t, ...) \in \Omega = S^{\mathbb{N}} = \{1,2\}^{\mathbb{N}},$$

e $t \in T = \mathbb{N}$.

P é uma probabilidade sobre $\Omega = \{1,2\}^{\mathbb{N}} = S^{\mathbb{N}}$. Assim, por exemplo,

$$P(X_0 = 2, X_1 = 1, X_2 = 1) \,=\, P(\{2\} \times \{1\} \times \{1\} \times S^{\mathbb{N}}).$$

Note que agora faz sentido o valor

$$P(X_{2\,n} = 2, \text{ para todo } n \in \mathbb{N}),$$

pois o conjunto

$$\{X_0 = 2, X_2 = 2, X_4 = 2, ..., X_{2\,n} = 2, ...\} =$$

$$\cap_n^\infty \{X_0 = 2, X_2 = 2, X_4 = 2, ..., X_{2\,n} = 2\},$$

vai estar na $\mathcal{F}$ obtida.

Pode-se mostrar (ver Capítulo 5) que, neste caso,

$$P(X_{2\,n} = 2, \text{ para todo } n \in \mathbb{N}) = \lim_{n\to\infty} P(X_0 = 2, X_2 = 2, X_4 = 2, ..., X_{2\,n} = 2).$$

Este processo é nosso primeiro exemplo de Processo de Markov. Em breve vamos falar mais detalhadamente sobre tais processos.

Vamos apresentar agora uma ilustração concreta da maneira implícita de apresentar um processo estocástico. Para isto será necessário descrever, antes de mais nada, uma propriedade muito importante.

Regra de Bayes: Seja $(\Omega, \mathcal{A}, P)$ um espaço de probabilidade. Considere duas variáveis aleatórias, uma X tomando valores em S_1 e outra Y tomando valores em S_2.

Então, para $s_1 \in S_1$ fixo

$$P(X = s_1) = \sum_{s_2 \in S_2} P(X = s_1 | Y = s_2)\ P(Y = s_2).$$

Esta propriedade segue trivialmente de

$$P(X = s_1) = \sum_{s_2 \in S_2} P(X = s_1 \,\text{e}\, Y = s_2) =$$

$$\sum_{s_2 \in S_2} P(X = s_1 | Y = s_2)\ P(Y = s_2),$$

que por sua vez segue da propriedade b) da definição 1.2.

Uma versão um pouco mais geral desta propriedade afirma:

Regra de Bayes: Seja $(\Omega, \mathcal{A}, P)$ um espaço de probabilidade. Considere uma variável aleatória X tomando valores em S e um conjunto $A \in \mathcal{A}$.

Então,

$$P(w \in A) = \sum_{s \in S} P(w \in A | X = s)\ P(X = s).$$

A demonstração deste fato é a mesma do caso anterior.

De maneira heurística podemos dizer que a regra de Bayes desempenha em probabilidade um papel semelhante ao do teorema fundamental no Cálculo. Mais exatamente, uma informação global (uma integral no Cálculo) $P(w \in A)$ é obtida através de uma totalização de informações localizadas (a derivada no Cálculo) $\sum_{s \in S} P(w \in A | X = s)\ P(X = s)$.

A análise das diversas propriedades de um processo estocástico será tão mais complexa quanto mais intensa forem as relações de dependência entre as variáveis. Nesta hierarquia de dificuldade, os mais simples são os processos independentes, depois seguem os Processos de Markov que serão analisados em breve.

Definição 1.15. *Dizemos que o processo X_t tomando valores em S (enumerável) com parâmetro $t \in T = \mathbb{N}$ é estacionário se para cada n e cada sequência $t_1 < t_2 < ... < t_n$, onde $t_i \in T$, $t > 0$ e para cada sequência de conjuntos $A_1, A_2, ..., A_n$, onde $A_i \subset S$ vale que*

$$P(X_{t_1} \in A_1,\, X_{t_2} \in A_2, ...,\, X_{t_n} \in A_n) =$$

$$P(X_{t_1+t} \in A_1,\, X_{t_2+t} \in A_2, ..., X_{t_n+t} \in A_n).$$

Os processos estacionários são aqueles em que um deslocamento uniforme de um valor t, em todos os tempos envolvidos na distribuição conjunta, não altera sua probabilidade de ocorrencia.

Neste caso, por exemplo, para s fixo qualquer em S

$$P(X_1 = s) = P(X_{1+1} = s) = P(X_2 = s).$$

Consideramos acima $t = 1$, $n = 1$, $a_1 = s$ e $t_1 = 1$.

Ainda,

$$P(X_3 = s) = P(X_{2+1} = s) = P(X_2 = s) = P(X_1 = s),$$

e assim por diante...

Estas propriedades podem não ser verdadeiras se o processo X_t não é estacionário. Por exemplo, no caso do jogo da moeda descrito anteriormente em que o capital inicial era c, temos que $P(X_0 = c) = 1$, mas $P(X_1 = c) = 0$. Logo, neste caso, o processo não é estacionário.

Antes de finalizarmos esta introdução vamos falar brevemente sobre alguns exemplos aplicados da Teoria dos Processos Estocásticos.

Um conjunto finito de observações organizadas cronologicamente no tempo é denominada de uma Série Temporal.

Suponhamos que estes dados foram obtidos da observação ao longo do tempo de algum fenômeno do mundo real. Nosso objetivo inicial seria tentar identificar, se possível, de qual processo estocástico foi originado tal série de dados, e com isto poder fazer previsões do que se pode esperar no futuro.

Distintos Processos Estocásticos poderiam modelar um dado problema, alguns com maior ou menor sucesso.

Nem toda a série de dados tem origem aleatória. Ela poderia ser determinada por um fenômeno determinístico. Existem testes estatísticos que permitem determinar se tal acontece ou não.

Num problema do mundo natural em que existe aleatoriedade, dada uma amostra finita $w_1, w_2, ..., w_n$, é natural perguntar: qual Processo Estocástico $(X_t)_{t \in T}$ poderia melhor modelar tal fenômeno?

Estas questões são muitas vezes bastante delicadas e requerem o uso da Estatística Matemática para que se possa determinar qual o melhor modelo. Por exemplo, dado uma série de dados, será que o modelo de Processo Estocástico que o descreve é estacionário? Será que pode ser descrito por uma Cadeia de

Markov? Caso afirmativo, qual seriam as probabilidades de transição?

Em geral é necessário utilizar testes estatísticos para aceitar ou rejeitar alguma hipótese sobre o modelo. Por exemplo, podemos nos perguntar, num certo exemplo específico, é válida a hipótese de que o Processo Estocástico em análise é uma Cadeia de Markov?

Antes de entrarmos no estudo das Cadeias de Markov e na análise de processos mais gerais vamos descrever algumas áreas de aplicabilidade da Teoria dos Processos Estocásticos em diversos campos da Ciência tais como: Mecânica Estatística, Economia, Engenharia, Biociências, etc.

1. Mecânica Estatística

Alguns dos mais importantes exemplos dos Processos Estocásticos foram desenvolvidos em conexão com os estudo de flutuações e ruídos nos sistemas físicos. Esta teoria pode ser considerada como fundamentação matemática para a Física Estatística.

Estados de Gibbs em Mecânica Estatística: Considere agora um reticulado (lattice em inglês) uni-dimensional (um fio) descrito por posições espaciais discretas, ou seja por $\mathbb{Z}$. Em cada sítio (um elemento em $\mathbb{Z}$) consideramos um spin que pode ser $+$ ou $-$.Existem interações entre os sitios, ou seja se na posição 3 existe um $+$ então isto interfere na probabilidade de se ter um $-$ na posição 4 e talvez até de um $+$ na posição 5. Esta interação está determinada por certas Leis Físicas. Podemos associar o número 1 a $+$ e o número 2 a $-$, deste modo podemos considerar de certa forma $S = \{1, 2\}$. Um certo arranjo de spins (em todos os possíveis sitios) seria um elemento ω no espaço de Bernoulli $\{+, -\}^{\mathbb{Z}}$, ou em $\{1, 2\}^{\mathbb{Z}} = S^{\mathbb{Z}}$, se considerarmos a identificação acima. Se fizermos várias observações em distintos momentos obteremos distintos elementos $\omega \in \{+, -\}^{\mathbb{Z}}$. Sendo assim, o problema não é determinístico. O que faz sentido é perguntar: qual a probabilidade de encontrarmos um arranjo com um spin $+$ na posição 2 e um spin $-$ na posição 7? Ou seja, o que faz sentido do ponto de vista físico é determinar uma probabilidade P no espaço $\Omega = \{+,-\}^{\mathbb{Z}}$ que descreva o sistema acima (no qual, de algum modo que não vamos descrever aqui, estão fixadas as interações). Esta probabilidade P é denominada de estado de Gibbs.

No presente caso a idéia de t como (assim chamado) o parâmetro temporal do Processo Estocástico deve ser substituído pelo conceito de $t \in \mathbb{Z}$ como o ponto do reticulado $\mathbb{Z}$.

Uma informação importante seria, por exemplo, a probabilddade de

$$P(X_0 = +, X_1 = +, X_2 = -\,).$$

Ou seja, a probabilidade do cilindro $\overline{++-}$. Ou, de outra forma, a probabilidade de ocorrer spin para cima na posição 1 e 2 do reticulado e para baixo na posição 3.

O problema mais real seria, na verdade, considerar o lattice tridimensional $\mathbb{Z}^3$, e supor que poderíamos ter em cada sítio (um elemento em $\mathbb{Z}^3$) um spin $+$ ou $-$. Neste caso, estaríamos interessados em probabilidades P sobre $\{+,-\}^{\mathbb{Z}^3}$. A descrição acima é um pouco simplista e esquemática, mas dá uma idéia geral da relevância da Teoria dos Processos Estocásticos em Mecânica Estatística.

2. Modelos Estocásticos em Medicina e Biociências

O tamanho e a composição de uma população estão constantemente sofrendo flutuações aleatórias. É a Teoria de Processos Estocásticos que permite descrever de maneira precisa os mecanismos destas flutuações. Alguns exemplos de fenômenos biológicos nos quais os Processos Estocásticos são adequados:

a) extinção de gerações de famílias

b) as mutações de genes e recombinações de genes na teoria da evolução

c) a distribuição espacial de comunidades de plantas e animais

d) a luta pela sobrevivência entre duas populações que interagem

e) a propagação de epidemias

f) funcionamento de redes de neurônios

3. Engenharia, Computação, Comunicação e Controle

Alguns exemplos de problemas envolvendo comunicação e/ou controle em engenharia e onde aparecem Processos Estocásticos:

a) recepção de sinais de rádio na presença de distúrbios naturais e artificiais

b) reprodução de imagens e sons

c) delineamento de sistemas de controle para processos industriais

d) modelos de funcionamento de máquinas ou de recebimento de mensagens telefônicas numa central.

e) transmissão de informação através de canais.

4. Economia e Ciência da Administração

Processos Estocásticos fornecem um preciso método para estudar flutuações dos valores dos bens econômicos e administrar operações de negócios. Portanto, desempenham um papel importante em Economia, Ciência da Administração e Pesquisa Operacional. Áreas de grande aplicação da teoria dos processos estocásticos são:

a) controle de inventários

b) análise de filas de espera

c) carteiras de investimento e aplicação em bolsa de valores

2

Cadeias de Markov à Tempo Discreto

2.1 Processos e Cadeias de Markov à Tempo Discreto

Trataremos nesta seção de processos estocásticos onde $T = \mathbb{N}$ e $S \subset \mathbb{R}$ é finito ou enumerável.

Se S tem d elementos, denotamos

$$S = \{s_1, s_2, \ldots, s_d\}.$$

Para simplificar a notação, denotaremos **SEMPRE** os elementos de S por inteiros, assim, por exemplo, $S = \{1, 2, 3, \ldots, d\}$, se S tem d elementos. Algumas vezes, considera-se também $S = \{a, a+1, \ldots, b\}$, onde $a < b$ são dois números inteiros em $\mathbb{Z}$.

Se S tem infinitos (enumeráveis) elementos, então denotaremos S por $S = \{1, 2, 3, \ldots, n, \ldots\} = \mathbb{N}$, ou por

$$S = \{\ldots, -n, \ldots, -3, -2, -1, 0, 1, 2, 3, \ldots, n, \ldots\} = \mathbb{Z},$$

ou então por $S = \{s_1, s_2, s_3, ..., s_n, \ldots\}$, e assim por diante...

Vamos relembrar algumas definições antes de começar a apresentar os conceitos fundamentais que serão analisados aqui.

Definição 2.1. *Uma matriz $\#S$ por $\#S$*

$$\mathcal{P} = (P_{i,j}),$$

$i \in S, j \in S$ *é dita estocástica (ou linha estocástica) se para cada i fixo vale que $\sum_{j\in S} P_{i,j} = 1$, e ainda $P_{i,j} \geq 0$.*

Esta matriz quadrada pode ter infinitas colunas (e linhas) se S for infinito.

Algumas vezes usaremos a notação $P(i,j)$, ou mesmo, $\mathcal{P}_{i,j}$, em vez de $P_{i,j}$.

Exemplo 2.1. Quando $S = \{1,2,3\}$, podemos considerar por exemplo

$$\mathcal{P} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 2/5 & 2/5 & 1/5 \\ 0 & 1 & 0 \end{pmatrix}$$

que é linha estocástica.

Note que por esta razão

$$\mathcal{P}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 2/5 & 2/5 & 1/5 \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Observe que no entanto

$$\begin{pmatrix} 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 2/5 & 2/5 & 1/5 \\ 0 & 1 & 0 \end{pmatrix} =$$

$$\begin{pmatrix} 1/3+2/5 & 1/3+2/5 & 1/3+1/5 \end{pmatrix} \neq \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}.$$

A expressão acima tem o sentido de produto de matrizes (no caso uma matriz um por três multiplicada por uma matriz três por três, gerando assim, uma matriz um por três).

Seja u a matriz do tipo $\#S$ por 1 (ou, seja um vetor coluna) tal que tem todas as coordenadas iguais a 1.

Afirmamos que se $\mathcal{P}$ é matriz linha estocástica então

$$\mathcal{P}\, u \,=\, u.$$

Note que, reciprocamente, se a matriz $\mathcal{P}$ tem entradas não negativas, então

$$\mathcal{P}\, u \,=\, u$$

implica que ela é estocástica.

O exemplo acima ilustra isto.

Este fato é simples de demostrar no caso geral de matrizes dois por dois. Por exemplo, no caso $S = \{1, 2\}$ temos que

$$\begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} P_{11} + P_{12} \\ P_{21} + P_{22} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Deixamos o caso geral a cargo do leitor.

$\diamondsuit$

Definição 2.2. *Um vetor $\pi = (\pi_s)_{s \in S}$, com entradas $s \in S$, é dito um vetor de probabilidade sobre S se $\pi_s \geq 0$ e $\sum_{s \in S} \pi_s = 1$.*

Exemplo 2.2. Quando $S = \{1, 2, 3\}$, podemos ter por exemplo

$$\pi = (1/7, 4/7, 2/7).$$

Podemos também expressar tal π na forma de matriz 1 por 3

$$\pi \;=\; (1/7 \;\, 4/7 \;\, 2/7).$$

Dada uma matriz estocástica $\mathcal{P}$ três por três

$$\pi \;\, \mathcal{P} = (1/7 \;\, 4/7 \;\, 2/7) \;\, \mathcal{P},$$

tem o sentido de produto de matrizes (no caso uma matriz um por três multiplicada por uma matriz três por três).

$\diamondsuit$

Utilizaremos estas duas formas

$$\pi = (\pi_1, \pi_2, ..., \pi_s),$$

ou,

$$\pi = (\pi_1 \;\, \pi_2 \,...\; \pi_s),$$

indistintamente sem menção a cada expressão específica.

Como dissemos antes, os processos estocásticos de natureza mais simples são os independentes e identicamente distribuídos. A seguir, em ordem de complexidade, vem os Markovianos que descreveremos a seguir.

Primeiro destacamos o fato que um Processo Estocástico é uma probabilidade P sobre (o espaço dos caminhos amostrais) $S^{\mathbb{N}}$.

Relembre que $\overline{x_0, x_1, ..., x_n} = \{w \in S^{\mathbb{N}}$ tal que w tem suas primeiras $n+1$ coordenadas exatamente iguais a $x_0, x_1, ..., x_n\}$. A informação básica do Processo Estocástico são as medidas destes cilindros $\overline{x_0, x_1, ..., x_n}$.

É usual a notação

$$P(\{w \in \overline{x_0, x_1, ..., x_n}\}) = P(X_0 = x_0, X_1 = x_1, ..., X_n = x_n).$$

Ainda, para $x_1, x_2, ..., x_n$ fixos em S denotamos

$$P(X_1 = x_1, ..., X_n = x_n) = \sum_{y_0 \in S} P(X_0 = y_0, X_1 = x_1, ..., X_n = x_n),$$

e assim por diante.

Assim, por exemplo, para x_0, x_1, x_2 fixos

$$P(X_2 = x_2 | X_0 = x_0, X_1 = x_1) = \frac{P(X_0 = x_0, X_1 = x_1, X_2 = x_2)}{P(X_0 = x_0, X_1 = x_1)}.$$

Fixado um vetor de probabilidade inicial $\pi = (\pi_1, \pi_2, ..., \pi_d)$ e uma matriz linha estocástica $\mathcal{P} = (P_{i,j}), i, j = 1, 2, ..., d$, defina uma Probabilidade P sobre cilindros

$$P(\overline{x_0, x_1, ..., x_n}\}) =$$

$$P(X_0 = x_0, X_1 = x_1, ..., X_n = x_n) = \pi_{x_0} P_{x_0,x_1} P_{x_1,x_2} ... P_{x_{n-1},x_n}. \qquad (2.1)$$

Esta probabilidade pode ser estendida a sigma algebra gerada pelos cilindros (ser Seção 5) e é o principal objeto da presente seção.

Definição 2.3 (**Processo de Markov**). *Seja $(X_n; n \geq 0)$ um processo estocástico com espaço de estados $S \subset \mathbb{R}$ finito ou enumerável. Dizemos que X_n é um processo estocástico de Markov com tempo discreto se vale a condição*

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) =$$

$$P(X_{n+1} = x_{n+1} | X_n = x_n), \qquad (*)$$

$$\forall n \geq 0,\ \forall x_0, x_1, \ldots, x_{n+1} \in S,$$

toda vez que $P(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) > 0$.

Da mesma forma, seja $(X_n; n \in \mathbb{Z})$ um processo estocástico com espaço de estados $S \subset \mathbb{R}$ finito ou enumerável. Dizemos que X_n é um processo estocástico de Markov com tempo discreto se vale para qualquer $n \in \mathbb{Z}$:

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) =$$

$$P(X_{n+1} = x_{n+1}|X_n = x_n), \qquad (*)$$
$$\forall x_0, x_1, \ldots, x_{n+1} \in S.$$

A grosso modo, processos deste tipo são tais que a probabilidade do valor $X_{n+1} = s \in S$ (na etapa $n+1$), vai depender apenas do valor $\tilde{s} \in S$ na etapa n, e não dos valores em S atingidos nos tempos anteriores, ou seja os valores $X_0, X_1, .., X_{n-1}$.

Definição 2.4 (**Probabilidade de Transição**). $P^{n,n+1}(i,j) = P(X_{n+1} = j|X_n = i)$ *é a probabilidade do processo estar no estado* j *no tempo (ou etapa)* $n+1$ *dado que o processo está no estado* i *no tempo* n.

Definição 2.5 (**Processo de Markov com Transições Estacionárias**). *Seja* $(X_n; n \geq 0)$ *um Processo de Markov, dizemos que ele tem probabilidades de transição estacionárias, se e somente se, existe uma matriz* $\mathcal{P} = (P(i,j))$ *de tipo* $\#S$ *por* $\#S$ *(ou seja* $i \in S, j \in S$*), tal que*

$$P(X_{n+1} = j|X_n = i) = P(X_1 = j \,|\, X_0 = i) = P(i,j), \;\; \forall n \geq 0, \forall i,j \in S$$

Em outras palavras, $P(X_{n+1} = j|X_n = i) = P(i,j)$ *não depende de* n.

A terminologia "Processo de Markov com transição Homogênea" também é utilizada na literatura.

Note que o Processo Estocástico $(X_n; n \geq 0)$ acima definido (com probabilidades de transição estacionárias) não precisa ser, necessariamente, estacionário.

Quando $S = \{1, 2, .., d\}$ é natural chamar de probabilidade inicial do processo $(X_n; n \geq 0)$ acima a matriz linha

$$\pi^0 = (\, P(X_0 = 1), P(X_0 = 2), ..., P(X_0 = d)\,)$$

Consideraremos no texto **apenas** a classe dos processos de Markov com transições estacionárias. Sendo assim, não poderá ocorrer, por exemplo, $P(X_4 =$

$2 \,|\, X_3 = 3) \neq P(X_2 = 2 \,|\, X_1 = 3)$. Ótimas referencias sobre Cadeias de Markov são [Kun] e [N].

Vamos mostrar mais tarde que todo processo de $(X_n; n \geq 0)$ com estados em S finito e com transições estacionárias, como definido abstratamente pela Definição 2.3, recai na forma descrita pela expressão (2.1) (ver Teorema 2.1). Mostraremos ainda que o processo definido pela expressão (2.1) tem transições estacionárias e satisfaz as propriedades descritas pela Definição 2.3 (ver Teorema 2.2).

Por exemplo, se $\#S = 3$ então a matriz $\mathcal{P}$ acima definida será uma matriz três por três, neste caso $i, j \in \{1, 2, 3\}$. Note que se $\#S$ for infinito então estaremos considerando uma matriz infinito por infinito.

Algumas vezes as entradas da matriz $\mathcal{P}$ serão denotadas também por $P_{i,j}$ ou $\mathcal{P}_{i,j}$.

Definição 2.6 (Matriz de Transição). *Dado um Processo de Markov com probabilidades de transição estacionária $(X_n; n \geq 0)$, a matriz $\mathcal{P} = (P(i, j))_{i,j \in S}$ tal que*

$$P(X_1 = j \,|\, X_0 = i) = P(i, j)$$

é chamada de matriz de transição sobre S associada ao processo e desempenhará um papel fundamental na teoria.

A matriz $\mathcal{P}$ da Definição 2.5 é linha estocástica e descreve o mecanismo aleatório das transições em uma etapa, ou seja, a entrada $\mathcal{P}_{x_i, x_j}$ de $\mathcal{P}$ descreve a probabilidade de se obter um $x_j \in S$, dado que no tempo imediatamente anterior tínhamos um $x_i \in S$.

Note que para cada $i \in S$ fixo, vale que $\sum_{j \in S} P(i, j) = 1$. Ou seja, a soma dos elementos de cada linha de $\mathcal{P}$ é igual a um.

Isto segue de

$$\sum_{j \in S} P(i, j) = \sum_{j \in S} P(X_1 = j \,|\, X_0 = i) = \sum_{j \in S} \frac{P(X_1 = j \,,\, X_0 = i)}{P(X_0 = i)} =$$

$$\frac{\sum_{j\in S} P(X_1=j\,,\,X_0=i)}{P(X_0=i)} = \frac{P(\cup_{j\in S}\{X_1=j\,,\,X_0=i\})}{P(X_0=i)} =$$

$$\frac{P(X_0=i)}{P(X_0=i)} = 1$$

Note que para i fixo a união $\cup_{j\in S}\{X_1=j\,,\,X_0=i\}$ considerada acima é disjunta.

Sendo assim para o processo Markoviano $(X_n; n\geq 0)$ com transições homogêneas, concluímos que se obtém uma matriz de transição $\mathcal{P}$ com entradas $(P(i,j))_{i,j\in S}$ satisfazendo

$$P(X_1=j\,|\,X_0=i) = P(i,j)$$

é estocástica.

Por exemplo, o caso geral quando $S=\{1,2,3\}$ seria:

$$\mathcal{P} = \begin{pmatrix} P(1,1) & P(1,2) & P(1,3)\\ P(2,1) & P(2,2) & P(2,3)\\ P(3,1) & P(3,2) & P(3,3)\end{pmatrix} =$$

$$\begin{pmatrix} P(X_1=1\,|\,X_0=1) & P(X_1=2\,|\,X_0=1) & P(X_1=3\,|\,X_0=1)\\ P(X_1=1\,|\,X_0=2) & P(X_1=2\,|\,X_0=2) & P(X_1=3\,|\,X_0=2)\\ P(X_1=1\,|\,X_0=3) & P(X_1=2\,|\,X_0=3) & P(X_1=3\,|\,X_0=3)\end{pmatrix}$$

Exemplo 2.3 (A Cadeia de Ehrenfest). O seguinte problema aparece em modelos simples de Mecânica Estatística (para uma analise mais completa ver [CM] Ex. 3.2.1 page 129). Considere um modelo em que temos 5 bolas e duas urnas, a da esquerda e a da direita. O número total de bolas nas duas urnas é cinco. A da esquerda pode ter de 0 a cinco bolas. Suponhamos que num dado momento tenhamos $r\geq 0$ bolas na urna da esquerda e $5-r$ na da direita. Por um mecanismo que não vem ao caso, sorteamos uma das cinco bolas e a passamos para a outra urna. Não há preferência na escolha da urna, o procedimento escolhe apenas uma bola ao acaso, sem se deter na questão em qual urna a bola está. Vamos denotar por X_n o número de bolas na urna da esquerda. Sendo assim, se $X_n = r$, teremos probabilidade $r/5$ de escolher uma bola na urna da esquerda. Isto fará com que $X_{n+1} = r - 1$. Se a bola escolhida estiver na urna da direita (fato que terá probabilidade $\frac{5-r}{5}$) então $X_{n+1} = r + 1$.

Fica assim definida uma matriz de transição de Markov tipo 6 por 6 em que $S = \{0, 1, 2, 3, 4, 5\}$ e que tem a forma

$$\mathcal{P} = \begin{array}{c c} & \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \begin{pmatrix} 0 & 5/5 & 0 & 0 & 0 & 0 \\ 1/5 & 0 & 4/5 & 0 & 0 & 0 \\ 0 & 2/5 & 0 & 3/5 & 0 & 0 \\ 0 & 0 & 3/5 & 0 & 2/5 & 0 \\ 0 & 0 & 0 & 4/5 & 0 & 1/5 \\ 0 & 0 & 0 & 0 & 5/5 & 0 \end{pmatrix} \end{array}$$

Fica definida assim uma cadeia de Markov natural a partir do modelo em consideração. Podemos considerar também o caso mais geral com duas urnas em que temos N bolas em vez de 5 bolas. Obteríamos neste caso uma matriz de transição $\mathcal{P}$ do tipo $N + 1$ por $N + 1$.

Este modelo foi introduzido para ajudar a explicar a segunda lei da termodinâmica (ver Wikipedia).

$\diamondsuit$

Fixado um processo estocástico $(X_n)_{n\in\mathbb{N}}$ tomando valores em S, então

$$\pi = (P(X_0 = s))_{s\in S} = (P(X_0 = 1), P(X_0 = 2), ..., P(X_0 = \#S)) = (\pi_s)_{s\in S},$$

é um exemplo de vetor de probabilidade. Isto porque $\sum_{s\in S} P(X_0 = s) = P(\cup_{s\in S}\{X_0 = s\}) = P(\Omega) = 1$. Tal π é denominado de **vetor de probabilidade inicial** do processo estocástico $(X_n)_{n\in\mathbb{N}}$

Este π será algumas vezes denotado por $\pi^0 = (\pi^0_s)_{s\in S}$. O índice superior zero vai indicar que estamos em $t = 0$.

Por exemplo, se $S = \{1, 2, 3\}$, então

$$\pi^0 = (\pi^0_1, \pi^0_2, \pi^0_3) = (P(X_0 = 1), P(X_0 = 2), P(X_0 = 3)).$$

O elemento

$$\pi^0 = (P(X_0 = s))_{s\in S}$$

é o que denominamos **vetor de probabilidade inicial**.

Para um processo estocástico de Markov X_n, $n \in \mathbb{Z}$, o sentido da probabilidade inicial não é exatamente da mesma natureza. Podemos assumir que $\pi^0 = (P(X_0 = s))_{s\in S}$ está fixado, e assim

$$P(X_0 = a_0, X_1 = a_1, ..., X_n = a_n) = \pi^0_{a_0}\, P_{a_0,a_1}\, P_{a_1,a_2} ... P_{a_{n-1},a_n}$$

de forma análoga.

Mas, observe que é preciso saber quem é $\pi^{-1} = (P(X_{-1} = s))_{s\in S}$ para calcular

$$P(X_{-1} = a_{-1}, X_0 = a_0, ..., X_n = a_n) = \pi^{-1}_{a_{-1}}\, P_{a_{-1},a_0}\, P_{a_0,a_1} ... P_{a_{n-1},a_n}$$

Fixado um processo estocástico de Markov X_n, $n \in \mathbb{N}$, (ou seja uma família de variáveis aleatórias sobre um espaço Ω, etc.) com probabilidades de transição estacionárias, então as distribuições conjuntas

$$P(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n)$$

ficam determinadas a partir de $\mathcal{P} = (P(i,j))_{i,j\in S}$ e $\pi^0 = (\pi^0_s)_{s\in S}$ como acima.

O próximo teorema quantifica tal afirmação.

Teorema 2.1. *Seja X_n, $n \in \mathbb{N}$, um processo de Markov com transições estacionárias onde $\mathcal{P} = (P(i,j))_{i,j \in S}$ e $\pi^0 = (\pi^0_s)_{s\in S}$ são dadas acima. Então,*

$$P(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) =$$

$$\pi^0(x_0)\, P(x_0,x_1)\, P(x_1,x_2) \ldots P(x_{n-1},x_n),$$

onde $\pi^0(x_0) = P(X_0 = x_0)$.

Demonstração: Ora,

$$P(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) =$$

$$P(X_0 = x_0)\frac{P(X_0 = x_0, X_1 = x_1)}{P(X_0 = x_0)}\frac{P(X_0 = x_0, X_1 = x_1, X_2 = x_2)}{P(X_0 = x_0, X_1 = x_1)} \ldots$$

$$\ldots \frac{P(X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = x_n)}{P(X_0 = x_0, \ldots, X_{n-1} = x_{n-1})} =$$

$$P(X_0 = x_0)\, P(X_1 = x_1 | X_0 = x_0)\, P(X_2 = x_2 | X_0 = x_0, X_1 = x_1) \ \ldots$$

$$\ldots\ P(X_n = x_n | X_0 = x_0, ..., X_{n-1} = x_{n-1}) =$$

$$P(X_0 = x_0)\, P(X_1 = x_1 | X_0 = x_0)\, P(X_2 = x_2 | X_1 = x_1)...$$

$$...P(X_n = x_n | X_{n-1} = x_{n-1}) =$$

$$\pi^0(x_0)\, P(x_0,x_1)\, P(x_1,x_2)...P(x_{n-1},x_n)$$

□

Exemplo 2.4. Considere $S = \{1,2\}$ e

$$\mathcal{P} = \begin{pmatrix} P(1,1) & P(1,2) \\ P(2,1) & P(2,2) \end{pmatrix} = \begin{pmatrix} 1/3 & 2/3 \\ 4/7 & 3/7 \end{pmatrix}$$

Considere também $\pi^0 = (\pi^0_1, \pi^0_2) = (1/3, 2/3)$. Neste caso,

$$P(X_0 = 2, X_1 = 1, X_2 = 1) = \pi^0_2\, P_{21}\, P_{11} = 2/3 \,\times\, 4/7 \,\times\, 1/3 = \frac{8}{63}.$$

Da mesma forma,

$$P(X_0 = 2, X_1 = 2, X_2 = 1) = \pi_2^0 \, P_{22} \, P_{21} = 2/3 \times 3/7 \times 4/7 = \frac{24}{147}.$$

Note que

$$P(X_2 = 1 \,|\, X_0 = 2, X_1 = 2) = \frac{P(X_0 = 2, X_1 = 2, X_2 = 1)}{P(X_0 = 2, X_1 = 2)} =$$

$$\frac{\pi_2^0 \, P_{22} \, P_{21}}{\pi_2^0 \, P_{22}} = P_{21} = P(X_2 = 1 \,|\, X_1 = 2)$$

que é a propriedade de Markov.

Ainda,

$$P(X_0 = 2, X_2 = 1) =$$

$$P(X_0 = 2, X_1 = 1, X_2 = 1) + P(X_0 = 2, X_1 = 2, X_2 = 1) =$$

$$\pi_2^0 \, P_{21} \, P_{11} + \pi_2^0 \, P_{22} \, P_{21} = \frac{8}{63} + \frac{24}{147}.$$

Se $i \in \{1, 2\}$,

$$\overline{i} = \{w \in \Omega \,:\, w = (i, w_2, w_3, ...), w_j \in S, j \geq 2\}.$$

Se $i \in \{1, 2\}$, $j \in \{1, 2\}$,

$$\overline{i, j} = \{w \in \Omega \,:\, w = (i, j, w_3, w_4, ...), w_k \in S, k \geq 3\}.$$

Desta forma P define uma probabilidade sobre $\{1, 2\}^{\mathbb{N}}$. Por exemplo,

$$P(X_0 = 2, X_2 = 1) = P(\overline{211}) + P(\overline{221}) = \frac{8}{63} + \frac{24}{147}.$$

$\diamondsuit$

Em resumo, sabendo que um processo X_n é de Markov, a partir da sua matriz de transição $\mathcal{P}$ e da distribuição inicial π^0, no tempo 0, fica determinado o valor

$$P(X_0 = x_0, X_1 = x_1, ..., X_n = x_n) = \pi^0(x_0)P(x_0, x_1).P(x_1, x_2)...P(x_{n-1}, x_n).$$

Ficará assim determinada uma probabilidade P sobre $\Omega = S^{\mathbb{N}}$. Vamos elaborar sobre isto.

Se $S = \{1, 2, ..., d\}$, então $\Omega = \{1, 2, ..., d\}^{\mathbb{N}}$, podendo-se tratar do caso de um S geral enumerável da mesma forma.

Dados $a_1, a_2, ..., a_n \in S$, $n \in \mathbb{N}$ fixos, considere um cilindro

$$\overline{a_1, a_2, .., a_n} = \{w \in \Omega \,|\, w = (a_1, a_2, ..., a_n, w_{n+1}, w_{n+2}, ...), w_j \in S, j \geq n+1\},$$

então
$$P(\overline{a_1, a_2, .., a_n}) = \pi^0(a_1)P(a_1, a_2).P(a_2, a_3).....P(a_{n-1}, a_n).$$

Fica assim determinada a probabilidade P sobre cilindros. Podemos, a partir disto, definir uma probabilidade estendida P sobre a sigma-algebra gerada pelos cilindros no conjunto $\Omega = S^{\mathbb{N}}$ (ver capítulo 5).

Teorema 2.2. *Seja* $(X_n; n \geq 0)$ *um processo estocástico com espaço de estados* $S = \{1, 2, ..., d\}$ *gerado pela expressão (2.1), obtido a partir da matriz estocástica* $\mathcal{P}$ *e* π *lá mencionada. Então* $(X_n; n \geq 0)$ *é um processo estocástico de Markov com tempo discreto e transições estacionarias no sentido da Definição 2.3*

Demonstração. Temos que mostrar que vale condição

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) =$$

$$P(X_{n+1} = x_{n+1} | X_n = x_n),$$

$$\forall n \geq 0,\ \forall x_0, x_1, \ldots, x_{n+1} \in S,$$

toda vez que $P(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) > 0$.

Faremos a demonstração para um caso particular que ilustra o caso geral (deixado a cargo do leitor):

$$P(X_2 = x_2 | X_0 = x_0, X_1 = x_1) = P(X_2 = x_2 | X_1 = x_1).$$

De fato,

$$P(X_2 = x_2 | X_0 = x_0, X_1 = x_1) = \frac{P(X_2 = x_2, X_0 = x_1, X_0 = x_0)}{P(X_0 = x_0, X_1 = x_1)} =$$

$$\frac{\pi_{x_0} P_{x_0,x_1} P_{x_1,x_2}}{\pi_{x_0} P_{x_0,x_1}} = P_{x_1,x_2} = \frac{P(X_2 = x_2, X_1 = x_1)}{P(X_1 = x_1)} = P(X_2 = x_2 | X_1 = x_1).$$

□

Lembre que I_A é a função tal que dá valor $I_A(x) = 1$, se $x \in A$, e $I_A(x) = 0$, se x não está em A.

Seja agora uma função $\phi : S^{\mathbb{N}} \to \mathbb{R}$. Dizemos que ϕ depende da primeira coordenada se

$$\phi = \sum_{i \in S} \alpha_i I_{\overline{i}},$$

onde $\alpha_i \in \mathbb{R}$ são constantes fixadas.

Dizemos que ϕ depende das duas primeiras coordenadas se

$$\phi = \sum_{i,j \in S} \alpha_{(i,j)} I_{\overline{i,j}},$$

onde $\alpha_{(i,j)} \in \mathbb{R}$ são constantes fixadas.

Dizemos que ϕ depende das n primeiras coordenadas se

$$\phi = \sum_{a_1,a_2,..,a_n \in S} \alpha_{(1,2,..,n)} I_{\overline{(a_1,a_2,...,a_n)}},$$

onde $\alpha_{(1,2,...,n)} \in \mathbb{R}$ são constantes fixadas. Em termos genéricos dizemos que uma ϕ de tal tipo depende de finitas coordenadas.

A função ϕ acima é mensurável pois cada cilindro é um elemento da sigma-algebra.

Para uma função ϕ da forma acima, como veremos no capítulo 5, a integral $\int \phi \, dP$, será

$$\mathbb{E}(\phi) \;=\; \int \phi \, dP = \sum_{a_1,a_2,...,a_n \in S} \alpha_{(1,2,...,n)} \, P(\overline{a_1, a_2, .., a_n}).$$

No contexto do exemplo acima, seja $\phi : \{1,2\}^{\mathbb{N}} \to \mathbb{R}$ tal que

$$\phi = 3\, I_{\overline{211}} + 7\, I_{\overline{221}}.$$

Assim, $\int \phi(w)\, dP(w) = 3\, P(\overline{211}) \;+ 7\, P(\overline{221}) \;=\; 3\, \frac{8}{63} \;+\; 7\, \frac{24}{147}$.
Desta forma, $\mathbb{E}(\phi) = 3\, \frac{8}{63} \;+\; 7\, \frac{24}{147}$.
Ainda,

$$\mathbb{E}(X_1) = \int X_1 \, dP = \int (1\, I_{\overline{1}} \;+ 2\, I_{\overline{2}}) = 1 \;\times 1/3 + \;2 \;\times 2/3 \;=\; 5/3.$$

Para uma função ϕ de uma forma mais geral a integral $\int \phi \, dP$ será definida no capítulo 5.

Destacamos uma questão puramente de notação. Muitos textos de Processos Estocásticos se baseiam no seguinte ponto de vista. Inicialmente nos é dada a informação de uma probabilidade $\tilde{P}$ sobre um espaço "não especificado", digamos $\tilde{\Omega}$, e uma sigma-algebra que não é dita qual é.

Denotemos por $\tilde{w}$ os elementos de $\tilde{\Omega}$. O Processo Estocástico é definido como uma sequência de variáveis aleatórias (mensuráveis) $\tilde{X}_n : \tilde{\Omega} \to S$ (que nunca se sabe quem são), indexados por n.

A informação básica (naquele contexto) é então, dados $x_0, x_1, ..., x_n \in S$, sabemos o valor

$$\tilde{P}(\tilde{X}_0(\tilde{w}) = x_0, \tilde{X}_1(\tilde{w}) = x_1, ..., \tilde{X}_n(\tilde{w}) = x_n).$$

Vamos considerar uma situação mais geral. Sejam $(Y_1, \mathcal{A}_1, P_1)$ e $(Y_2, \mathcal{A}_2)$, onde Y_1, Y_2 são conjuntos, $\mathcal{A}_1$, $\mathcal{A}_2$ respectivamente suas sigma-algebras, e P_1 probabilidade sobre $\mathcal{A}_1$. Dada $H : Y_1 \to Y_2$ mensurável definamos a probabilidade P_2 sobre $(Y_2, \mathcal{A}_2)$ da seguinte forma: para $A_2 \in \mathcal{A}_2$,

$$P_2(A_2) = P_1(H^{-1}(A_2)) = P_1(\{y_1 \in Y_1 \mid H(y_1) \in A_2\}).$$

Algumas vezes se chama P_2 a probabilidade "push forward"de P_1 via H.

Se considerarmos a função $H : \tilde{\Omega} \to S^{\mathbb{N}}$, tal que

$$H(\tilde{w}) = (\tilde{X}_0(\tilde{w}), \tilde{X}_1(\tilde{w}), ..., \tilde{X}_n(\tilde{w}), ...) = w,$$

as duas informações acima são as mesmas, mais exatamente

$$P(X_0 = x_0, X_1 = x_1, ..., X_n = x_n) =$$

$$P(\{w \in \overline{x_0, x_1, ..., x_n}\}) =$$

$$\tilde{P}\{\tilde{w} \mid H(\tilde{w}) \in \overline{x_0, x_1, ..., x_n}\}\} =$$

$$\tilde{P}(\tilde{X}_0(\tilde{w}) = x_0, \tilde{X}_1(\tilde{w}) = x_1, ..., \tilde{X}_n(\tilde{w}) = x_n).$$

Ou seja, P é o push forward da probabilidade $\tilde{P}$ via H.

Conforme foi dito anteriormente na introdução, se usa indistintamente a letra P para denotar tanto a probabilidade $\tilde{P}$ (P_1 na notação acima) sobre $\tilde{\Omega}$ (onde estão definidas as $\tilde{X}_n$) como a probabilidade associada sobre $S^{\mathbb{N}}$ (P_2 na notação acima).

Definição 2.7. *Uma cadeia de Markov (com estados em S enumerável) associada à matriz estocástica $\mathcal{P}$ fixada, tipo $\#S$ por $\#S$, é a classe de todos os processos estocásticos de Markov que podemos obter a partir de todos os possíveis vetores de probabilidade $\pi^0 = (\pi^0_s)_{s \in S}$.*

A definição acima não é usual mas queremos destacar aqui a diferença entre a informação dada por uma matriz estocástica e os Processos Estocásticos que dela podem ser originados ao se fixar distintas condições iniciais. O Processo assim obtido seria denominado de Processo Estocástico Markoviano ou Processo de Markov.

Como a cadeia de Markov fica caracterizada apenas pela matriz $\mathcal{P}$, muitas vezes vamos nos referir à cadeia como simplesmente a matriz estocástica $\mathcal{P}$ associada.

Fixado $i \in S$, é usual denotar por $e_i = (p_s)_{s \in S}$ o vetor de probabilidade sobre S tal que $p_i = 1$ e $p_s = 0$ para todo $s \neq i$.

Por exemplo, no caso $S = \{1, 2, 3, 4\}$, temos que $e_3 = (0, 0, 1, 0)$.

Seja uma cadeia de Markov fixada (ou seja, uma matriz $\mathcal{P}$ fixada) e $i \in S$ também fixado. Quando considerarmos a probabilidade inicial $\pi = e_i$, o Processo de Markov $(X_n)_{n \in \mathbb{N}}$ definido como acima (a partir da informação $\mathcal{P}$ e de tal $\pi = e_i$) será denominado de Processo de Markov condicionado a começar em i. Neste caso, $P(X_0 = i) = 1$. Vamos usar a notação P_i para a tal probabilidade sobre Ω (ou sobre $S^{\mathbb{N}}$).

Uma sutileza: quando nos referimos à expressão $P(X_1 = j \,|\, X_0 = i)$, ao mencionar a probabilidade P (sobre Ω) devemos considerar sempre uma probabilidade inicial π, mas como aparece a expressão $X_0 = i$, é porque estamos assumindo que $\pi = e_i$, ou seja, $P_i(X_1 = j) = P(X_1 = j \,|\, X_0 = i)$. Lembre que e_i é o vetor que tem todas as coordenadas nulas, menos a i-ésima que é igual a 1.

Definição 2.8. *Para $n \in \mathbb{N}$ fixo, denotamos por*

$$\pi^n = (\pi^n(s))_{s \in S} = (P(X_n = s))_{s \in S},$$

o vetor sobre S que define a probabilidade de $X_n = s$, para todo $s \in S$.

O vetor

$$\pi^n = (P(X_n = 1), P(X_n = 2), ..., P(X_n = \#S)) = (\pi^n_s)_{s\in S},$$

é um vetor de probabilidade e descreve a probabilidade do processo atingir o valor $s \in S$ no tempo n.

Se $S = \{1, 2, 3\}$, no tempo $t = 5$, teríamos

$$\pi^5 = (\pi^5_1, \pi^5_2, \pi^5_3) = (\, P(X_5 = 1), P(X_5 = 2), P(X_5 = 3)\,).$$

O índice 5 denota tempo $t = 5$, e não pode ser confundido, de forma alguma, com elevar a potência 5.

Note que em geral, $\pi^0 \neq \pi^n$. A análise das propriedades de π^n será objeto dos primeiros resultados que seguem. Se o Processo de Markov for também estacionário (conforme a definição apresentada no capítulo anterior), então para todo $n \in \mathbb{N}$, vale $\pi^0 = \pi^n$, conforme foi mostrado anteriormente. Lembre que, neste caso, para todo n e s vale que $P(X_0 = s) = P(X_n = s)$.

Passeio Aleatório: Seja $S = \{s_1, s_2, s_3, s_4, ..., s_d\} = \{a, a + 1, ..., b - 1, b\}$ um subconjunto de naturais ordenados com $a < b$. Dizemos que a e b são os extremos de S. Considere uma partícula que se move numa linha reta em passos de uma unidade para a direita ou para a esquerda. Cada passo é dado para a direita com probabilidade p e para esquerda com probabilidade q, de tal forma que $p + q = 1$. Ela se move em passos sucessivos (eventualmente alternando para a esquerda e direita) até atingir um dos dois extremos. As possibilidades para o seu comportamento nestes pontos determinam várias diferentes espécies de Cadeias de Markov. Os estados são as possíveis posições. Vários tipos de exemplos são descritos a seguir.

Exemplo 2.5. Suponha o caso com 5 estados, ou seja $S = \{s_1, s_2, s_3, s_4, s_5\}$ com s_1 e s_5 os estados extremos. Assuma que se a partícula atinge o estado s_1 ou s_5 ela fica lá para sempre (chamaremos tal estado de estado absorvente).

Neste caso, a matriz de transição é

$$\mathcal{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$\diamondsuit$

Exemplo 2.6. O estado é refletido quando atinge um estado limite e retorna ao ponto do qual ela partiu. Neste caso, a matriz é

$$\mathcal{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$\diamondsuit$

Exemplo 2.7. Sempre que a partícula atinge um dos pontos limites ela vai diretamente para o centro s_3.

$$\mathcal{P} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

$\diamondsuit$

Exemplo 2.8. Passeio aleatório sobre um conjunto S de 5 elementos:

Se o processo está em um dos três estados interiores, ele tem igual probabilidade de se mover para a esquerda, para a direita ou permanecer no mesmo estado. Considere barreiras absorventes em s_1 e s_5.

$$\mathcal{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$\diamondsuit$

Neste caso, a matriz descreve o fato que, neste processo, uma vez atingido o estado s_1 “ficamos parado nele". Mesma coisa se atingimos s_5. Estas afirmações mais ou menos informais serão tornadas rigorosas ao longo do texto.

Exemplo 2.9. Pode-se também considerar passeios aleatórios sobre $S = \mathbb{Z}$. O passeio aleatório mais comum é aquele em que estando no estado $s \in \mathbb{Z}$ então $P(s, s+1) = 1/2$, $P(s, s-1) = 1/2$, e $P(s, j) = 0$ para $j \neq s-1, s+1$. Fica assim definida uma cadeia de Markov. É usual considerar o processo estocástico (associado) condicionado a começar em 0, ou seja, tomar a distribuição inicial π^0 tal que $\pi^0(0) = 1$ e $\pi^0(i) = 0$ para $i \in \mathbb{Z} - \{0\}$.

Dito de uma maneira compacta, $\pi^0 = e_0 = (...., 0, 0, 1, 0, 0, ...)$. Este processo será denominado de passeio aleatório $(1/2, 1/2)$. Sua matriz de transição é

$$\mathcal{P} = \begin{pmatrix} . & . & . & . & . & . & . \\ . & 0 & \frac{1}{2} & 0 & 0 & 0 & . \\ . & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & . \\ . & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & . \\ . & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & . \\ . & 0 & 0 & 0 & \frac{1}{2} & 0 & . \\ . & . & . & . & . & . & . \end{pmatrix}.$$

$\diamondsuit$

Exemplo 2.10. Considere o seguinte passeio aleatório sobre $S = \mathbb{Z}$. Seja $p, q \geq 1$ tal que $p + q = 1$. Então, estando no estado $s \in \mathbb{Z}$, tem-se que $P(s, s-1) = q$, $P(s, s+1) = p$ e $P(s, j) = 0$ para $j \neq s-1, s+1$. Fica assim definida uma cadeia de Markov. É usual considerar o processo estocástico (associado) condicionado a começar em 0, ou seja, tomar a distribuição inicial $\pi^0 = e_0$.

Este processo será denominado de passeio aleatório (q, p).

Sua matriz de trasição será

$$\mathcal{P} = \begin{pmatrix} . & . & . & . & . & . & . \\ . & 0 & p & 0 & 0 & 0 & . \\ . & q & 0 & p & 0 & 0 & . \\ . & 0 & q & 0 & p & 0 & . \\ . & 0 & 0 & q & 0 & p & . \\ . & 0 & 0 & 0 & q & 0 & . \\ . & . & . & . & . & . & . \end{pmatrix}.$$

$\diamondsuit$

Após a apresentação dos exemplos acima vamos voltar aos resultados teóricos.

Definição 2.9. *Dadas duas matrizes $\mathcal{P} = (P_{i,j})$, $i, j \in S$, e $\mathcal{Q} = (Q_{i,j})$, $i, j \in S$, o produto $V = \mathcal{P}\,\mathcal{Q}$ é uma nova matriz $V = (V_{i,j})$, $i, j \in S$, tal que o elemento $V_{i,j}$ é dado por*

$$V_{i,j} = \sum_s P_{i,s} Q_{s,j},$$

quando o somatório convergir (S pode ser infinito).

Este produto não é necessariamente comutativo, ou seja, nem sempre vale que

$$\mathcal{P}\,\mathcal{Q} = \mathcal{Q}\mathcal{P}.$$

Denotaremos $\mathcal{P}^2 = \mathcal{P}\mathcal{P}$ e $\mathcal{P}^n$ o produto de $\mathcal{P}$ por si mesmo n vezes. A matriz $\mathcal{P}^0$ representa a matriz identidade I, ou seja, $I = (\delta(i,j))$, onde $\delta(i,j) = 1$, se $i = j$, e $\delta(i,j) = 0$, se $i \neq j$.

Note que $\mathcal{P}^{n+m} = \mathcal{P}^n \, \mathcal{P}^m = \mathcal{P}^m \, \mathcal{P}^n$.

Denotamos por $P^n(i,j)$ a entrada ij da matriz $\mathcal{P}^n$, com $i, j \in S$ e $n \geq 0$.

Seja $\mathcal{P}$ uma matriz $\#S$ por $\#S$, é fácil ver por indução que

$$P^{(n+m)}(i,k) = \sum_{r \in S} P^n(i,r) \, P^m(r,k).$$

Por exemplo, se $n = 1$ e $m = 1$ obtemos (ver última definição) a expressão para a entrada i, k da matriz $\mathcal{P}^{1+1} = \mathcal{P}^2$ através de

$$P^2(i,j) = \sum_{r \in S} P(i,r) \, P(r,j).$$

Observamos que $P^2(i,j)$ denota a entrada i,j da matriz $\mathcal{P}^2$, e não o número que resulta de tomar o quadrado da entrada $P(i,j)$ da matriz $\mathcal{P}$.

Ainda, para n fixo e $i, j \in S$ fixos, vale

$$P^n(i,j) = \sum_{s_1 \in S} \sum_{s_2 \in S} ... \sum_{s_{n-1} \in S} P(i,s_1) \, P(s_1,s_2) \, P(s_2,s_3) .... P(s_{n-1},j)$$

De maneira mais geral, uma matriz A da forma m_1 por m_2 pode ser multiplicada por uma matriz B da forma m_2 por m_3, obtendo assim uma matriz C da forma m_1 por m_3, denotada por $C = A\,B$, através de

$$C_{i,j} = \sum_{s \,\in\, \{1,2,3,...,m_2\}} A_{i,s} B_{s,j},$$

para cada $i \in \{1,2,3,...,m_1\}$ e $j \in \{1,2,3,...,m_3\}$.

Por exemplo, se $m_1 = 1$, $m_2 = 2$, $m_3 = 2$ e

$$A = (x_1 \quad x_2)$$

e

$$B = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

então

$$A\,B \,=\, (x_1 \quad x_2)\, \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} =$$

$$(x_1\, a_{11} \,+\, x_2\, a_{21} \qquad x_1\, a_{12} \,+\, x_2\, a_{22}\,) = C.$$

O vetor π^0 anteriormente definido é uma matriz do tipo 1 por $\#S$. Da mesma forma o vetor e_c é uma matriz do tipo 1 por $\#S$. Algumas vezes no texto o vetor e_c (ou outro qualquer) que é uma matriz do tipo 1 por $\#S$, dependendo da situação, poderá ser considerado como uma matriz do tipo $\#S$ por 1.

Iremos considerar no texto a seguir o produto $\pi\,\mathcal{P}$ e ainda $\pi\,\mathcal{P}^2 = \pi\,\mathcal{P}\,\mathcal{P}$, etc...

Note que se a matriz da transformação $\mathcal{P}$ é estocástica, então todas as suas entradas são não-negativas e desta forma a matriz $\mathcal{P}^n$ também tem todas as entradas não-negativas. Assim, se pode mostrar que $\mathcal{P}^n$ é também estocástica. Vamos demostrar este fato no caso três por três.

Ora, para $\mathcal{P}$ é válido

$$\mathcal{P} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Assim,

$$\mathcal{P}^2 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \mathcal{P}\,\mathcal{P} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \mathcal{P} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Assim, $\mathcal{P}^2$ é estocástica. A demonstração para $\mathcal{P}^n$ segue por indução.

Após estas questões e definições preliminares podemos voltar ao cálculo de algumas da informações fundamentais sobre Processos de Markov.

Como obter a distribuição da variável aleatória X_1, a de X_2, e ainda, a de uma genérica X_n ?

Ora,

$$P(X_1 = x_0) = P(X_1 = x_0, X_0 = s_0) + P(X_1 = x_0, X_0 = s_1) + \ldots =$$

$$\sum_{s_j \in S} P(X_1 = x_0, X_0 = s_j) =$$

$$\sum_{s_j \in S} P(X_0 = s_j)\, P(X_1 = x_0 | X_0 = s_j) = \sum_{s_j \in S} \pi^0_{s_j}\, P(s_j, x_0)$$

No caso $S = \{1, 2\}$ temos assim para cada $i \in \{1, 2\}$ fixo, que vale a expressão

$$P(X_1 = i) = \pi^0_1\, P(1, i) \,+\, \pi^0_2\, P(2, i).$$

Neste caso, em forma matricial, a expressão acima significa

$$\begin{pmatrix} \pi^0(1) & \pi^0(2) \end{pmatrix} \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} = \begin{pmatrix} P(X_1 = 1) & P(X_1 = 2) \end{pmatrix}.$$

Lembre que $\pi^1 = (\pi^1_s)_{s \in S}$ é a matriz do tipo 1 por $\# S$ (vetor linha) tal que sua ordenada s é igual a $P(X_1 = s)$.

Sendo assim, de forma compacta, a expressão acima significa

$$\pi^0\, \mathcal{P} \,=\, \pi^1.$$

Generalizando o que foi mostrado acima no caso $S = \{1, 2\}$, segue facilmente para um S qualquer (enumerável) que vale a expressão em forma compacta

$$\pi^0\, \mathcal{P} \,=\, \pi^1.$$

O seguinte exemplo é de grande importância - usa de forma essencial a propriedade Markoviana - para o que segue: seja $i \in S$ fixado

$$P(X_2 = i) = \sum_j \sum_k P(X_2 = i, X_1 = k, X_0 = j) =$$

$$\sum_j \sum_k P(X_0 = j) P(X_2 = i, X_1 = k | X_0 = j) =$$

$$\sum_j \sum_k P(X_0 = j) \, \frac{P(X_2 = i, X_1 = k, X_0 = j)}{P(X_0 = j)} =$$

$$\sum_j \sum_k P(X_0 = j) \, \frac{P(X_1 = k, X_0 = j)}{P(X_0 = j)} \, \frac{P(X_2 = i, X_1 = k, X_0 = j)}{P(X_1 = k, X_0 = j)} =$$

$$= \sum_j \sum_k P(X_0 = j) P(X_1 = k | X_0 = j) P(X_2 = i | X_1 = k) =$$

$$\sum_j \sum_k \pi_j^0 P(j,k) P(k,i).$$

Lembre que π^2 é o vetor tal que sua ordenada i é igual a $P(X_2 = i)$.

Utilizando a expressão da entrada $P^2(j,i)$ de $\mathcal{P}^2$ através de

$$P^2(j,i) = \sum_{k \in S} P(j,k) P(k,i),$$

da maneira similar a que procedemos antes, obtemos a expressão acima na forma compacta

$$\pi^2 = \pi^0 \, \mathcal{P}^2.$$

Ou seja, para passar em duas etapas de j a i, começando em j, passamos em uma etapa por todos os possíveis $s \in S$, e a seguir, passamos em uma etapa para i.

Mais geralmente, segue da mesma forma como anteriormente, por indução

$$P(X_n = i) = \sum_{i_0} \sum_{i_1} \ldots \sum_{i_{n-1}} \pi^0(i_0) P(i_0, i_1) P(i_1, i_2) \ldots P(i_{n-1}, i).$$

Lembre que π^n é o vetor tal que sua ordenada i é igual a $P(X_n = i)$.

Segue da expressão acima a forma compacta:

Proposição 2.1.

$$\pi^n = \pi^0 \, \mathcal{P}^n.$$

Desta forma, podemos calcular de maneira simples, a partir da matriz $\mathcal{P}^n$ e de π^0, a distribuição de X_n. Basta multiplicar por si mesma n vezes a matriz $\mathcal{P}$ e aplicar à esquerda o vetor π^0. Uma questão importante é saber da eventual existência de limite deste π_n qunado n vai a infinito.

O seguinte exemplo também é de grande importância.

Vamos calcular

$$P(X_2 = i, X_0 = j) = \sum_k P(X_2 = i, X_1 = k, X_0 = j) =$$

$$\sum_k P(X_0 = j) P(X_2 = i, X_1 = k | X_0 = j) =$$

$$\sum_k P(X_0 = j) \, \frac{P(X_2 = i, X_1 = k, X_0 = j)}{P(X_0 = j)} =$$

$$\sum_k P(X_0 = j) \, \frac{P(X_1 = k, X_0 = j)}{P(X_0 = j)} \, \frac{P(X_2 = i, X_1 = k, X_0 = j)}{P(X_1 = k, X_0 = j)} =$$

$$= \sum_k P(X_0 = j) \, P(X_1 = k | X_0 = j) \; P(X_2 = i | X_1 = k) =$$

$$\sum_k \pi^0_j \, P(j,k) \, P(k,i).$$

Em geral, segue da mesma forma como acima, por indução, que vale

$$P(X_n = i, X_0 = j) = \sum_{i_1} \dots \sum_{i_{n-1}} \pi_j^0 \, P(j, i_1) \, P(i_1, i_2) \dots P(i_{n-1}, i),$$

onde $\pi_j^0 = P(X_0 = j)$.

Esta expressão será utilizada várias vezes no texto. Note que acima i e j estão fixos e exaurimos no somatório todas as outras possibilidades intermediárias.

Por exemplo,

$$P(X_3 = i, X_0 = j) = \sum_{i_1 \in S} \sum_{i_2 \in S} \pi_j^0 \, P(j, i_1) \, P(i_1, i_2) \, P(i_2, i).$$

Proposição 2.2.

$$P(X_n = i | X_0 = j) = \sum_{i_1 \in S} \dots \sum_{i_{n-1} \in S} P(j, i_1) \, P(i_1, i_2) \dots P(i_{n-1}, i) = (P^n)_{j,i}.$$

Demonstração: Isto segue do fato que

$$P(X_n = i | X_0 = j) = \frac{P(X_n = i, X_0 = j)}{P(X_0 = j)}.$$

□

Por exemplo,

$$P(X_3 = i | X_0 = j) = \sum_{i_1 \in S} \sum_{i_2 \in S} P(j, i_1) \, P(i_1, i_2) \, P(i_2, i) = (P^3)_{j,i}.$$

Uma generalização fácil do que afirmamos acima é:

Proposição 2.3.

$$P(X_n = i | X_k = j) = P(X_{n-k} = i | X_0 = j).$$

Exemplo 2.11. Note que, a partir do que vimos acima

$$P(X_5 = i|X_0 = j) =$$

$$\sum_{i_1\in S}\sum_{i_2\in S}\sum_{i_3\in S}\sum_{i_4\in S} P(j,i_1)\, P(i_1,i_2)\, P(i_2,i_3)\, P(i_3,i_4)\, P(i_4,i) =$$

$$\sum_{i_1\in S}\sum_{i_2\in S}\sum_{i_4\in S} P(j,i_1)\, P(i_1,i_2)\, P(i_2,i_4)^2\, P(i_4,i) =$$

$$\sum_{i_1\in S}\sum_{i_2\in S} P(j,i_1)\, P(i_1,i_2)\, P(i_2,i)^3 \ =$$

$$\sum_{i_2\in S} P^2(j,i_2)\, P^3(i_2,i) \ = P^5(j,i)$$

Note também que

$$P(X_4 = i, X_2 = k, X_0 = j) =$$

$$\sum_{i_1\in S}\ \sum_{i_3\in S} P(X_4 = i, X_3 = i_3, X_2 = k, X_1 = i_1, X_0 = j)$$

$$\sum_{i_3\in S} \pi_j^0 P^2(j,k)\, P(k,i_3)\, P(i_3,i) =$$

$$\pi_j^0 P^2(j,k)\ P^2(k,i)\,.$$

$\diamondsuit$

A partir de um argumento como descrito no exemplo acima é fácil ver que vale mais geralmente:

Teorema 2.3.

$$P(\,\{w\in\Omega\,|\,X_0 = a_0, X_{t_1} = a_1, X_{t_2} = a_2, X_{t_3} = a_3, ..., X_{t_n} = a_n\}\,) =$$

$$\pi_{a_0}\ (\mathcal{P}^{t_1})_{a_0\,a_1}\,(\mathcal{P}^{t_2-t_1})_{a_1\,a_2}\,(\mathcal{P}^{t_3-t_2})_{a_2\,a_3}\ ...\,(\mathcal{P}^{t_n-t_{n-1}})_{a_{n-1}\,a_n},$$

onde $n\in\mathbb{N}$, $a_i\in S$ *e* $0 = t_0 < t_1 < t_2 < ... < t_n$.

Exemplo 2.12. Considere a seguinte matriz de transição $\mathcal{P}$ sobre o conjunto de estados $S = \{s_1, s_2, s_3, s_4, s_5\}$:

$$\mathcal{P} = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{array} \begin{array}{c} \begin{array}{ccccc} s_1 & s_2 & s_3 & s_4 & s_5 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

Para fixar um exemplo mais específico suponha que $p = q = 1/2$.

Suponhamos que o processo inicie no estado s_3, ou seja tome a probabilidade inicial igual a $\pi^0 = e_3$.

Na Figura 2.1 mostramos um gráfico tipo árvore em que mostramos na etapa 1, as probabilidades de transição que são não-negativas. Assim, s_3 pode passar a s_4 e s_2. Na etapa seguinte, s_2 pode passar a s_3 e s_1, e por sua vez, s_4 pode passar a s_5 e s_3. Nos ramos das árvores aparecem as respectivas probabilidades de transição. Assim, se pode ver que, começando em s_3, em três etapas existe probabilidade positiva de se atingir s_1, s_2, s_4, s_5, mas não s_3. A probabilidade de passar em três etapas de s_3 a s_1 é $1/2 \times 1/2 \times 1 = 1/4$. Isto pode ser obtido multiplicando as probabilidades nos ramos utilizados para passar de s_3 a s_1 em três etapas.

Vamos acompanhar neste caso os vetores π^j obtidos a partir de $\pi^0 = e_3$: é fácil ver que

$\pi^0 = (0, 0, 1, 0, 0)$
$\pi^1 = (0, \frac{1}{2}, 0, \frac{1}{2}, 0)$
$\pi^2 = (\frac{1}{4}, 0, \frac{1}{2}, 0, \frac{1}{4})$
$\pi^3 = (\frac{1}{4}, \frac{1}{4}, 0, \frac{1}{4}, \frac{1}{4})$

Vamos agora acompanhar o que acontece no caso geral p, q: é fácil ver que

$\pi^0 = (0, 0, 1, 0, 0)$

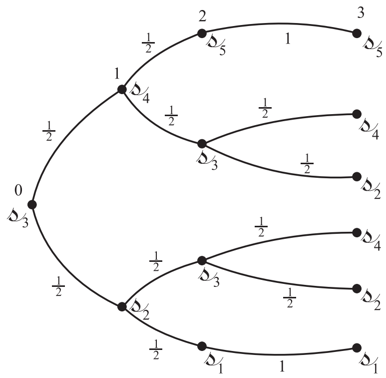


Figura 2.1:

$\pi^1 = (0, q, 0, p, 0)$
$\pi^2 = (q^2, 0, 2q\,p, 0, p^2)$
$\pi^3 = (q^2, 2q^2 p, 0, 2qp^2, p^2)$
Observe também que em qualquer caso $\sum_{j\in S} \pi_j^i = 1, \forall\, i = 0, 1, 2, 3.$

Para o terceiro caso note que: $2q^2p + 2qp^2 = 2q\,p\,(q + p)$.

É fácil ver que $\pi = (q\,0\,0\,0\,p)$ é um vetor estacionário para tal $\mathcal{P}$, isto é $\pi\mathcal{P} = \pi$ (Este tópico sera analisado com detalhe na Seção 2.2). Observe que $\pi_a = (1\,0\,0\,0\,0)$ e $\pi_b = (0\,0\,0\,0\,1)$ também satisfazem, respectivamente, $\pi_a\mathcal{P} = \pi_a$ e $\pi_a\mathcal{P} = \pi_a$. Segue que toda combinação convexa $\lambda\pi_a + (1-\lambda)\pi_b$, $\lambda \in [0, 1]$ também são vetores de probabilidade estacionários.

$\diamondsuit$

Após o exemplo acima vamos voltar a considerar resultados teóricos.

Proposição 2.4. *Seja $A_0 \subset S$, então*

$$P(X_n = j | X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \ldots, X_0 \in A_0) = P(X_n = j | X_{n-1} = i_{n-1})$$

Demonstração: Para simplificar a notação vamos mostrar o resultado acima no caso particular (o leitor pode generalizar facilmente tal prova) descrito a seguir: seja $S = \{1, 2, 3\}$ e $A_0 = \{1, 3\}$:

Neste caso, vamos mostrar que

$$P(X_2 = 1|X_1 = 3, X_0 \in A_0) = P(X_2 = 1|X_1 = 3, X_0 \neq 2) = P(X_2 = 1|X_1 = 3).$$

De fato,

$$P(X_2 = 1|X_1 = 3, X_0 \neq 2) =$$
$$\frac{P(X_2 = 1, X_1 = 3, X_0 \neq 2)}{P(X_1 = 3, X_0 \neq 2)} =$$
$$\sum_{j=1,3} \frac{P(X_2 = 1, X_1 = 3, X_0 = j)P(X_1 = 3, X_0 = j)}{P(X_1 = 3, X_0 \neq 2)P(X_1 = 3, X_0 = j)} =$$
$$\sum_{j=1,3} P(X_2 = 1|X_1 = 3, X_0 = j)\frac{P(X_1 = 3, X_0 = j)}{P(X_1 = 3, X_0 \neq 2)} =$$
$$\sum_{j=1,3} P(X_2 = 1|X_1 = 3)\frac{P(X_1 = 3, X_0 = j)}{P(X_1 = 3, X_0 \neq 2)} =$$
$$P(X_2 = 1|X_1 = 3).$$

□

O seguinte resultado mais geral segue facilmente do procedimento descrito acima.

Proposição 2.5. *Para $j, s \in S$ e $n > r$*

$$P(X_n = j|X_r = s, X_{r-1} \in A_{r-1}, \ldots, X_1 \in A_1, X_0 \in A_0) = P(X_n = j|X_r = s),$$

onde A_l, $l \in \{0, 1, 2, ..., r-1\}$ são subconjuntos de S.

Proposição 2.6. *Para $i, j, k \in S$ e $m, n \geq 0$*

$$P(X_{m+n} = j \,|\, X_m = k, X_0 = i) = P(X_{m+n} = j \,|\, X_m = k).$$

Demonstração: Segue direto da proposição anterior considerando $A_0 = \{i\}$ e os $A_i = S$, ou seja usando a expressão

$$P(X_{m+n} = j \mid X_m = k, X_0 = i) =$$

$$P(X_{m+n} = j | X_m = k, X_{m-1} \in S, \ldots, X_1 \in S, X_0 \in A_0) =$$

$$P(X_{m+n} = j | X_m = k).$$

□

Note que,

$$P(X_{m+n} = j | X_m = k) = P(X_n = j | X_0 = k).$$

Finalmente, podemos enunciar o resultado mais geral e que sintetiza todos os anteriores:

Proposição 2.7. *Sejam os conjuntos* $A_0, A_1, A_2, ..., A_{n-2}, A_{n-1} = \{a_{n-1}\}$, $A_n, ..., A_m \subset S$, *onde* $n < m$. *Então vale que*

$$P(X_m \in A_m, X_{m-1} \in A_{m-1}, \ldots, X_n \in A_n \mid X_{n-1} = a_{n-1}, X_{n-2} \in A_{n-2},$$

$$\ldots, X_1 \in A_1, X_0 \in A_0) =$$

$$P(X_m \in A_m, X_{m-1} \in A_{m-1}, \ldots, X_n \in A_n \mid X_{n-1} = a_{n-1}) =$$

$$P(X_{m-n+1} \in A_m, X_{m-n} \in A_{m-1}, \ldots, X_1 \in A_n \mid X_0 = a_{n-1}).$$

Este resultado segue de maneira rotineira da extensão dos resultados anteriores.

Proposição 2.8. *Considere um processo de Markov* $(X_n)_{n\in\mathbb{N}}$ *e seja* $\mathcal{P}$ *sua matriz de transição associada, ou seja, a entrada* ij *da matriz* $\mathcal{P}$ *é dada por* $P(i,j) = P(X_1 = j \mid X_0 = i)$. *Então para* $m, n \geq 0$ *e* $i, j \in S$ *fixos*

$$P_{ij}^{n+m} = \sum_{k\in S} P_{ik}^n P_{kj}^m,$$

Onde $P^r(i,j) = P(X_r = j|X_0 = i)$. *De maneira matricial a expressão acima significa*

$$\mathcal{P}^{n+m} = \mathcal{P}^n\,\mathcal{P}^m.$$

esta equação é conhecida como a equação de Chapman-Kolmogorov.

Demonstração: Sejam $i, j \in S$ e $m, n \in \mathbb{N}$ fixos. Condicionando obtemos

$$P(X_{m+n} = j|X_0 = i) = \frac{P(X_{m+n} = j, X_0 = i)}{P(X_0 = i)} =$$

$$\sum_{k\in S} \frac{P(X_{m+n} = j, X_m = k, X_0 = i)}{P(X_0 = i)} =$$

$$\sum_{k\in S} \frac{P(X_{m+n} = j, X_m = k, X_0 = i)}{P(X_m = k, X_0 = i)}\,\frac{P(X_m = k, X_0 = i)}{P(X_0 = i)} =$$

$$\sum_{k\in S} P(X_{m+n} = j\,|\,X_m = k, X_0 = i)\,P^m(i,k) =$$

$$\sum_{k\in S} P(X_{m+n} = j\,|\,X_m = k)\,P^m(i,k) =$$

$$\sum_{k\in S} P(X_n = j\,|\,X_0 = k)\,P^m(i,k) = \sum_{k\in S} P^m(i,k)\,P^n(k,j) = P^{m+n}(i,j).$$

Acima usamos a última proposição.

Finalmente, o resultado para matrizes segue de considerar i, j qualquer e a expressão de produto de matrizes.

$\square$

A proposição acima diz, de maneira literal, que para se calcular a probabilidade de ir de i a j no tempo r, podemos escolher um tempo intermediário $m < r$, condicionar no valor k do processo no tempo m e a seguir utilizar a probabilidade de se ir em tempo n de k ao valor final j, onde $r = m + n$.

Observação: Segue do último resultado que dado o vetor de probabilidade $\pi^1 = (\pi_1, \pi_2, ..., \pi_d)$, então para j fixo

$$\sum_{i=1}^{d} \pi_i P^{m+n}(i,j) = \sum_{i=1}^{d} \pi_i \sum_{k\in S} P^m(i,k)\, P^n(k,j) \, = q_j,$$

é a entrada j do vetor $(q_1, q_2, ..., q_d) = q = \pi P^{m+n}$.

2.2 Vetor de Probabilidade Estacionário e Processos Estacionários

Definição 2.10 (**Vetor de Probabilidade Estacionário**). *Um vetor de probabilidade π sobre o conjunto de estados S é dito fixo (ou estacionário) para a matriz estocástica $\mathcal{P}$ se $\pi\ \mathcal{P} = \pi$. Neste caso se costuma dizer que π é um vetor estacionário para a cadeia de Markov definida por $\mathcal{P}$.*

Por exemplo, considere $S = \{1, 2\}$ e a seguir vamos fixar

$$\mathcal{P} = \begin{pmatrix} 1/4 & 3/4 \\ 1/3 & 2/3 \end{pmatrix}.$$

Desejamos encontrar os vetores $\pi = (\pi_1 \ \ \pi_2)$, com $\pi_1 + \pi_2 = 1$, $\pi_1, \pi_2 \geq 0$, tais que a seguinte equação matricial seja verdadeira

$$(\pi_1 \ \ \pi_2) = \ (\pi_1 \ \ \pi_2) \begin{pmatrix} 1/4 & 3/4 \\ 1/3 & 2/3 \end{pmatrix}.$$

Para isto devemos resolver as equações:

$$\pi_1 \ 1/4 + \pi_2 \ 1/3 \ = \ \pi_1$$

$$\pi_1 \ + \ \pi_2 = 1.$$

Substituindo $\pi_2 = 1 - \pi_1$ na primeira equação obtemos

$$\pi_1\ 1/4 + (1 - \pi_1)\,1/3 \;=\; \pi_1.$$

Isolando π_1 nesta equação obtemos que $\pi_1 = 4/13$. Logo $\pi_2 = 9/13$. Desta forma o vetor $\pi = (4/13\ \ 9/13)$ satisfaz $\pi\,\mathcal{P} = \pi$, e assim $(4/13\ \ 9/13)$ é um vetor de probabilidade estacionário para a cadeia de Markov definida pela matriz de transição

$$\begin{pmatrix} 1/4 & 3/4 \\ 1/3 & 2/3 \end{pmatrix}.$$

Note que no presente exemplo (nem sempre ocorre isto em outros casos) o vetor π é único.

Observe que se $S = \{1, 2\}$ e

$$\mathcal{P} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

então qualquer vetor de probabilidade $\pi = (\pi_1\ \ \pi_2)$ é estacionário.

Note ainda que se $\pi = \pi\,\mathcal{P}$, então

$$\pi\,\mathcal{P}^2 = \pi\ \mathcal{P}\ \mathcal{P} = \ (\pi\,\mathcal{P}\,)\,\mathcal{P} = \pi\,\mathcal{P} = \pi.$$

Da mesma forma, neste caso, para qualquer $n > 0$ vale que $\pi\,\mathcal{P}^n = \pi$.

Lembre que um Processo Estocástico é estacionário, se para cada $n > 0$ e cada sequência $t_1 < t_2 < ... < t_n$, onde $t_i \in T$, e $t > 0$, e para cada sequência de conjuntos $A_1, A_2, ..., A_n$, onde $A_i \subset S$ vale que

$$P(X_{t_1} \in A_1,\, X_{t_2} \in A_2, ...,\, X_{t_n} \in A_n) =$$

$$P(X_{t_1+t} \in A_1,\, X_{t_2+t} \in A_2, ..., X_{t_n+t} \in A_n).$$

Proposição 2.9. *Uma condição necessária e suficiente para que o processo estocástico X_n (obtido através de uma matriz de transição $\mathcal{P}$ e um vetor de probabilidade inicial π) seja estacionário, é que π seja um vetor de probabilidade estacionário para $\mathcal{P}$.*

Demonstração: Se X_n for estacionário, então para cada $i \in S$ vale que $P(X_0 = i) = P(X_1 = i)$.

Como vimos acima $P(X_1 = i) = \sum_{j \in S} \pi_j P(j, i)$. Logo, se

$$\pi_i = P(X_0 = i) = P(X_1 = i),$$

para todo i, concluímos assim que $\pi \mathcal{P} = \pi$.

Vamos primeiro mostrar a recíproca no caso particular de $\{X_0 = a_2, X_1 = a_3, X_2 = a_4, X_3 = a_5\}$ e $\{X_2 = a_2, X_3 = a_3, X_4 = a_4, X_5 = a_5\}$, ou seja, $n = 4$, $t_i = i$, $i = 0, 1, 2, 3$, e $t = 2$. Estamos com isto evitando uma notação mais pesada. Se o leitor comprender o caso particular entãp entenderá o caso geral. Este será tratado mais tarde.

Vamos mostrar que

$$P(X_0 = a_2, X_1 = a_3, X_2 = a_4, X_3 = a_5) =$$

$$P(X_2 = a_2, X_3 = a_3, X_4 = a_4, X_5 = a_5).$$

Ora, se $\pi \mathcal{P} = \pi$, então, como vimos acima $\pi \mathcal{P}^2 = \pi$. Assim, considerando fixos a_2, a_3, a_4, a_5 obtemos

$$P(X_2 = a_2, X_3 = a_3, X_4 = a_4, X_5 = a_5) =$$

$$\sum_{i_0 \in S} \sum_{i_1 \in S} \pi^0_{i_0} P(i_0, i_1) P(i_1, a_2) P(a_2, a_3) P(a_3, a_4) P(a_4, a_5)$$

$$\Big(\sum_{i_0 \in S} \sum_{i_1 \in S} \pi^0_{i_0} P(i_0, i_1) P(i_1, a_2) \Big) \times P(a_2, a_3) P(a_3, a_4) P(a_4, a_5).$$

Ora como $\pi \mathcal{P}^2 = \pi$, então conforme observação após a Proposição 2.8 (tomando $m = 1 = n$) temos que

$$\sum_{i_0 \in S} \sum_{i_1 \in S} \pi^0_{i_0} \, P(i_0, i_1) P(i_1, a_2) = \pi^0_{a_2}.$$

Logo,

$$P(X_2 = a_2, X_3 = a_3, X_4 = a_4, X_5 = a_5) =$$

$$\pi^0_{a_2} \, P(a_2, a_3) P(a_3, a_4) \, P(a_4, a_5) = P(X_0 = a_2, X_1 = a_3, X_2 = a_4, X_3 = a_5).$$

Demonstramos assim, para $t = 2$ e $n = 4$ que vale a propriedade

$$P(X_{0+2} = a_2, X_{1+2} = a_3, X_{2+2} = a_4, X_{3+2} = a_5) =$$

$$P(X_0 = a_2, X_1 = a_3, X_2 = a_4, X_3 = a_5).$$

para $a_2, a_3, a_4, a_5 \in S$ quaisquer.

A mesma demonstração pode ser feita para o caso de $r > 0$ qualquer

$$P(X_{0+r} = a_2, X_{1+r} = a_3, X_{2+r} = a_4, X_{3+r} = a_5) =$$

$$P(X_0 = a_2, X_1 = a_3, X_2 = a_4, X_3 = a_5).$$

Para isto basta usar o fato que se $\pi P = \pi$, então $\pi P^r = \pi$, e a Proposição 2.8.

A demonstração para o caso geral, $z > 0$, $u > 0$ e $r > 0$, $a_u, a_{u+1}, ..., a_{u+z} \in S$, onde

$$P(X_{u+r} = a_u, X_{u+1+r} = a_{u+1}, X_{u+2+r} = a_{u+2}, ..., X_{u+z+r} = a_{u+z}) =$$

$$P(X_u = a_u, X_{u+1} = a_{u+1}, X_{u+2} = a_{u+2}, ..., X_{u+z} = a_{u+z}).$$

é semelhante ao caso descrito acima.

□

Segue do teorema acima que fixada uma matriz de transição $\mathcal{P}$ (de uma cadeia de Markov), então um certo π que define um processo estocástico de

Markov é estacionário, se e só se, π é vetor de probabilidade invariante para $\mathcal{P}$.

Por exemplo, considere $S = \{1, 2\}$ e o processo Markoviano obtido a partir da matriz

$$\mathcal{P} = \begin{pmatrix} 1/4 & 3/4 \\ 1/3 & 2/3 \end{pmatrix}$$

e do vetor de probabilidade inicial (4/13, 9/13). A partir do que calculamos no exemplo do começo desta seção obtemos que este Processo Estocástico Markoviano é estacionário.

Retornando ao caso geral descrito pelo teorema acima observamos que uma pergunta natural é : fixado um $\mathcal{P}$ sempre existe ao menos um π invariante? Vamos mostrar que a resposta a esta questão depende de S ser finito ou não. No caso em que S é finito sempre vai existir ao menos um π. No caso de S com cardinalidade infinita nem sempre isto ocorre. Como veremos mais adiante o passeio aleatório $(1/2, 1/2)$ vai ser um caso onde não existe vetor π de probabilidade estacionário. Note que neste caso $S = \mathbb{Z}$.

Definição 2.11. *Fixado S denotamos $\Sigma = \Sigma_S = \{(p_1, p_2, ..., p_s, ...) \, : \, s \in S$ tal que $p_i \geq 0$, $p_1 + p_2 + ... + p_s + ... = 1\}$.*

Proposição 2.10. *A função $T : \Sigma \to \Sigma$, tal que*

$$T((p_1, p_2, ..., p_d, ...)) = p \; \mathcal{P},$$

onde $p = (p_1, p_2, ..., p_d, ...)$ está bem definida.

Demonstração: Tal T é a restrição de uma função linear. Para ver que T está bem definida (de Σ em Σ) analisemos primeiro o caso em que $p = e_s$ (o vetor que é nulo a menos da posição $s \in S$) para um s fixo. Denote então $v^s = e_s\mathcal{P} = T(e_s)$. As coordenadas de v^s são dadas por $(v^s_j)_{j\in S} = v^s = e_s\,\mathcal{P} = (P(s,j))_{j\in S}$.

Note que $\sum_{j\in S} v^s_j = \sum_{j\in S} P(s,j) = 1$.

Logo $v^s \in \Sigma$.

Vamos agora ao caso geral: ora, um $p \in \Sigma$ qualquer é dado por $p = \sum_{s\in S} p_s e_s$, onde $\sum_{s\in S} p_s = 1$; o resultado então segue por linearidade: seja $(u_j)_{j\in S} = u = p\,\mathcal{P} = T(p)$. Então

$$T(\sum_{s\in S} p_s e_s) \,=\, \sum_{s\in S} p_s T(e_s) = \sum_{s\in S} p_s v^s \,=\, u \,=\, (u_j)_{j\in S}.$$

Agora

$$\sum_{j\in S} u_j = \sum_{j\in S}\sum_{s\in S} p_s\, v^s_j = \sum_{s\in S}\sum_{j\in S} p_s\, v^s_j = \sum_{s\in S} p_s \sum_{j\in S} v^s_j = \sum_{s\in S} p_s\, 1 = 1.$$

Note que como todos os fatores da soma dupla acima são positivos não importa a ordem de soma.

Logo, $u \in \Sigma$.

$\square$

Na Figura 5.5 mostramos, no caso em que $S = \{1,2,3\}$, o espaço Σ e sua imagem $T(\Sigma)$ sob a ação de T (que está contida em Σ).

Vamos agora apresentar um resultado extremamente importante:

Teorema 2.4. *Se S é finito, então para toda a matriz estocástica $\mathcal{P}$ existe ao menos um vetor π tal que é invariante para $\mathcal{P}$.*

Demonstração: Quando S é finito com cardinalidade d, o conjunto $\Sigma = \{(p_1,p_2,...,p_d)\,|\;$ tal que $p_i \geq 0,\, p_1+p_2+...+p_d = 1\}$ é um conjunto convexo fechado limitado dentro do espaço afim $\{x : x_1+x_2+...+x_d = 1\}$ (que tem dimensão $d-1$). Afirmação: a função $T : \Sigma \to \Sigma$, tal que

$$T((p_1,p_2,...,p_d)) = p\ \mathcal{P}$$

está bem definida e é contínua.

Ora, pela proposição anterior sabemos que T está bem definida. O fato que T é contínua segue de que $T(p)$ é definida a partir de produtos e somas finitas.

Logo, pelo Teorema do Ponto Fixo de Brouwer (ver [GP], ou [DL] Exercício 35 seção 6.5) temos que existe um ponto fixo π para T, ou seja, existe π tal que $T(\pi) = \pi$. Logo π é invariante para $\mathcal{P}$.

□

Sendo assim, quando S é finito sempre se pode obter a partir de $\mathcal{P}$ um certo processo estocástico Markoviano estacionário. Basta escolher $\pi^0 = (\pi^0_s)_{s\in S} = P(X_0 = s)_{s\in S}$ tal que $\pi^0\,\mathcal{P} = \pi^0$. Tal π^0 é autovetor (vetor de probabilidade) de $\mathcal{P}$ (à esquerda) associado ao autovalor 1.

O π fixo para $\mathcal{P}$ nem sempre é único. Quando S não é finito, dado $\mathcal{P}$, nem sempre existe π que é invariante para $\mathcal{P}$. Por exemplo, o passeio aleatório tipo $(1/2, 1/2)$ sobre $S = \mathbb{Z}$ definido anteriormente não possui π invariante a esquerda associado ao autovalor 1 como veremos em breve.

Exemplo 2.13. Considere a matriz estocástica

$$\mathcal{P} = \begin{pmatrix} p_1 & p_2 & p_3 & \dots & p_d \\ p_1 & p_2 & p_3 & \dots & p_d \\ . & . & . & \dots & . \\ p_1 & p_2 & p_3 & \dots & p_d \end{pmatrix},$$

onde $p_1 + p_2 + ... + p_d = 1$.

Seja um vetor de probabilidade $\pi = (\pi_1, \pi_2, ..., \pi_d)$ qualquer.

A equação $\pi\,\mathcal{P} = \pi$, nos determina d igualdades, mais exatamente, para cada i fixo temos

$$\pi_1\,p_i + \pi_2\,p_i + ... + \pi_d\,p_i = \pi_i,$$

mas como $\pi_1\,p_i + \pi_2\,p_i + ... + \pi_d\,p_i = p_i$, obtemos que o único π estacionário é $\pi = (p_1, p_2, ..., p_d)$.

Assumindo π desta forma, temos que

$$P(X_0 = a_0, X_1 = a_1, ..., X_n = a_n) \, = \, p_{a_0} \, P_{a_0 a_1} \, P_{a_1 a_2} ... P_{a_{n-1} a_n} =$$

$$p_{a_0} \, p_{a_1} \, p_{a_2} ... p_{a_n}.$$

Note que esta probabilidade P coincide em cada cilindro

$$\{X_0 = a_0, X_1 = a_1, ..., X_n = a_n\}$$

com aquela do processo independente e identicamente distribuído, em que cada elemento $i \in \{1, 2, 3, ..., d\} = S$, tem, respectivamente, probabilidade p_i. Logo, pelo teorema de Caratheodori-Kolmogorov os dois processos determinam as mesmas probabilidades na sigma-algebra determinada por estes cilindros(ver capítulo 5 em caso de necessidade de uma afirmação mais precisa).

Logo, os processos i. i. d. com a condição inicial $\pi = (p_1, p_2, ..., p_d)$ podem ser encarados como Processos Markovianos estacionários.

$\diamondsuit$

Exercício: : Uma urna contém duas bolas sem cor. Numa sequência de instantes uma bola é escolhida ao acaso de dentro da urna e pintada de vermelho ou preto e colocada novamente na urna. Se a bola não está pintada, a escolha da cor é feita aleatoriamente de forma igualmente distribuída (vermelho ou preto). Se ela está pintada, a sua cor é trocada. Este procedimento é feito em sequência várias vezes. A cada momento temos na urna duas bolas com as possibilidades individuais de vermelho, preta ou sem cor.

Qual a matriz de transição da cadeia de Markov $X_n, n \in \mathbb{N}$, com estados (x, y, z) onde x é o número de bolas sem cor, y é o número de bolas vermelhas, e, z é o número de bolas pretas?

Definição 2.12. *Uma matriz estocástica $\mathcal{P}$ é dita regular se existe $k > 0$ tal que todas as entradas da matriz $\mathcal{P}^k$ são positivas.*

Exemplo 2.14. 1) $A = \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix}$ é regular. De fato: $A^2 = \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix}$.

2) $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ não é regular. De fato: $I^m = I$ para qualquer m.

3) $B = \begin{pmatrix} 3/4 & 1/4 \\ 1 & 0 \end{pmatrix}$ é regular. De fato: $B^2 = \begin{pmatrix} 13/16 & 3/16 \\ 3/4 & 1/4 \end{pmatrix}$.

$\diamondsuit$

Teorema 2.5. *Suponha que a matriz de transição $\mathcal{P}$ seja finita e regular, então só existe um $\pi \in \Sigma$ tal que $\pi\mathcal{P} = \pi$. O vetor π tem todas as entradas positivas.*

Demonstração: Vamos supor primeiro que $\mathcal{P}$ tem todas as entradas estritamente positivas.

Vamos provar que só existe um $\pi \in \Sigma$ tal que $\pi\mathcal{P} = \pi$.

Note que isto é a mesma coisa que provar que só existe um $x \in \mathbb{R}^d$ com todas as coordenadas positivas, tal que $x\mathcal{P} = x$. De fato, dado x podemos obter a partir de x, o vetor $\pi = \frac{x}{\sum_{j\in S} x_j}$, e vice-versa.

Seja, $x = (x_1, x_2, ..., x_d) \in \Sigma$ e $y = (y_1, y_2, ..., y_d) \in \Sigma$ e suponha, por absurdo, que $x = x\mathcal{P}$ e $y = y\mathcal{P}$.

Note que as entradas de y são todas positivas, senão existiria i tal que

$$0 = y_i = \sum_{j\in S} P_{ji}\, y_j,$$

e isto implica que todos os y_j são nulos.

Logo, todas as entradas de um ponto fixo π são positivas.

Considere

$$t = \inf\{\frac{x_1}{y_1}, \frac{x_2}{y_2}, ..., \frac{x_d}{y_d}\} \leq 1.$$

Denote $\frac{x_{i_0}}{y_{i_0}}$ o valor mínimo.

Sendo assim, o vetor $x-ty$, tem todas as coordenadas x_i-ty_i não negativas. Ainda $x_{i_0} - ty_{i_0} = 0$.

Deste modo,

$$(x-t\,y)\,\mathcal{P} = (x-t\,y) = (x_1-t\,y_1, ..., x_{i_0-1}-t\,y_{i_0-1}, 0, x_{i_0+1}-t\,y_{i_0+1}, ..., x_d-t\,y_d).$$

Logo,

$$P_{1,i_0}\,(x_1 - t\,y_1) + ...+$$
$$P_{i_0-1,i_0}\,(x_{i_0-1} - t\,y_{i_0-1}) + 0 + P_{i_0+1,i_0}\,(x_{i_0+1} - t\,y_{i_0+1}) + ... + P_{d,i_0}\,(x_d - t\,y_d) = 0.$$

Logo, como todos os $P_{i,j} > 0$, e todos $x_i - ty_i \geq 0$, temos que todos os $x_j - t\,y_j = 0$. Ainda, como

$$0 = \sum_{i=1}^{d} x_i - t\,y_i = 1 - t.$$

Temos que $t = 1$. Sendo assim $x = y$.

No caso geral, quando existe $k > 0$ tal que todas as entradas da matriz $\mathcal{P}^k$ são positivas, o raciocínio é análogo. Basta considerar no argumento acima $(x - t\,y)\,\mathcal{P}^k = (x - t\,y)$ e proceder da mesma forma.

Desta forma obtemos para tal matriz $\mathcal{P}^k$ que existe apenas um vetor π em Σ tal que $\pi\,\mathcal{P}^k = \pi$.

Ressaltamos aqui o fato que desejamos fazer afirmações sobre a unicidade do vetor invariante para $\mathcal{P}$, e não para $\mathcal{P}^k$.

Já se sabe que $\mathcal{P}$ tem um vetor invariante. Se existissem dois distintos π_a e π_b, tais que $\pi_a\,\mathcal{P} = \pi_a$ e $\pi_b\,\mathcal{P} = \pi_b$, então, tambem é verdade que $\pi_a\,\mathcal{P}^k = \pi_a$ e $\pi_b\,\mathcal{P}^k = \pi_b$. Mas isto contraria a unicidade do vetor estacionário para $\mathcal{P}^k$.

Assim, concluímos, também neste caso, que o vetor estacionário em Σ para $\mathcal{P}$ é único.

$\square$

Exemplo 2.15. Dada a matriz estocástica $\mathcal{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{pmatrix}$

a) Verifiquemos que a matriz $\mathcal{P}$ é regular, i.é, $\exists m \geq 1$ tal que as entradas de $\mathcal{P}^m$ são positivas.

b) A seguir vamos determinar o seu único vetor fixo de probabilidade.

Para dar apoio a afirmação a) observe que

$$\mathcal{P}^2 = \begin{pmatrix} 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \end{pmatrix} ; \mathcal{P}^3 = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} ;$$

$$\mathcal{P}^4 = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/4 & 1/4 & 1/2 \\ 1/4 & 2/4 & 1/4 \end{pmatrix} ; \mathcal{P}^5 = \begin{pmatrix} 1/4 & 1/4 & 1/2 \\ 1/4 & 2/4 & 1/4 \\ 1/8 & 3/8 & 4/8 \end{pmatrix} .$$

Logo, $\mathcal{P}$ é regular.

Para resolver b) desejamos encontrar π tal que satisfaça a equação $\pi\mathcal{P} = \pi$, ou seja,

$$(x, y, 1 - x - y)\,\mathcal{P} = (x, y, 1 - x - y)$$

onde $0 \leq x \leq 1$ e $0 \leq y \leq 1$.

Vamos resolver o sistema

$$\left\{ \begin{array}{c} (1 - x - y)\frac{1}{2} = x \\ x + \frac{1}{2}(1 - x - y) = y \\ y = 1 - x - y \end{array} \right. ,$$

ou seja,

$$\left\{ \begin{array}{l} 1 - y = 3x \\ 1 + x = 3y \\ x + 2y = 1 \end{array} \right. ,$$

ou seja,

$$\begin{cases} 1-y=3(1-2y) \\ x+2y=1 \end{cases},$$

e finalmente

$$\begin{cases} y=\frac{2}{5} \\ x=\frac{1}{5} \end{cases}.$$

Portanto, $\pi=(\frac{1}{5},\frac{2}{5},\frac{2}{5})$ é o único vetor fixo de probabilidade de $\mathcal{P}$.

Se definirmos um processo estocástico X_n a partir da $\mathcal{P}$ e de tal $\pi = (\frac{1}{5},\frac{2}{5},\frac{2}{5})$, obteremos uma probabilidade P que torna o processo estacionário.

$\diamondsuit$

Observamos que existe um método bastante direto para se obter o vetor de probabilidade estacionário (em um grande número de casos).

Teorema 2.6. *Seja $\mathcal{P}$ matriz estocástica k por k, denote por $\mathbb{I}$ a matriz k por k, que tem todas as entradas iguais a 1, e I a matriz identidade, então se $I-\mathcal{P}+\mathbb{I}$ for inversível, temos que*

$$\pi \,=\, (1\,1\,...1\,1)(I-\mathcal{P}+\mathbb{I})^{-1}$$

satisfaz

$$\pi\,\mathcal{P}=\pi.$$

Demonstração: Sabemos que π existe, assim, note que

$$\pi\,(I-\mathcal{P}+\mathbb{I})=\pi-\pi+(1\,1\,...1\,1)=(1\,1\,...1\,1).$$

Aplicando em ambos os lados da igualdade acima (do lado direito) a matriz $(I-\mathcal{P}+\mathbb{I})^{-1}$ obtemos o resultado desejado.

$\square$

Existem variados pacotes de software que invertem matrizes de maneira bastante rápida, sendo assim, o método acima fornece uma maneira direta e eficiente de calcular (em muitos casos) o vetor estacionário.

Como exemplo de aplicação do resultado acima considere a matriz

$$\mathcal{P} = \begin{pmatrix} \frac{1}{7} & \frac{2}{7} & \frac{4}{7} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{5} & \frac{2}{5} & \frac{2}{5}. \end{pmatrix}$$

Neste caso o determinante da matriz $I - \mathcal{P} + \mathbb{I}$ é $\frac{513}{140}$ e $\pi = (1\,1\,...1\,1)(I - \mathcal{P} + \mathbb{I})^{-1} = (\frac{35}{171}, \frac{56}{171}, \frac{80}{171})$.

Sejam $r \in \mathbb{N}$ e $s \in \mathbb{N}$ fixos. Para cada $n \in \mathbb{N}$ considere matrizes A_n da forma r por s. Dizemos que a sequência de matrizes A_n converge para a matriz A, se cada entrada de A_n converge à respectiva entrada de A.

Por exemplo,

$$A_n = \begin{pmatrix} 3 + \frac{1}{n} & 5 & \text{sen}\,(1/n) \\ \cos(1/n) & 1 + \frac{1}{n^2} & 1/n \\ (\pi + \frac{1}{n^2})^2 & -7 & \sum_{i=0}^{n} \frac{1}{2^i} \end{pmatrix}$$

é tal que

$$\lim_{n \to \infty} A_n = \begin{pmatrix} 3 & 5 & 0 \\ 1 & 1 & 0 \\ \pi^2 & -7 & 2 \end{pmatrix}.$$

O conceito análogo quando as matrizes A_n e A são da forma r por infinito (onde $r \in \mathbb{N}$) ou mesmo da forma infinito por infinito pode ser igualmente considerado.

O teorema abaixo é de grande importância na teoria. A demonstração utiliza algumas propriedades um pouco mais sofisticadas de Álgebra Linear. Numa primeira leitura, o leitor que o desejar pode pular a demonstração, contanto que fique bem claro o que é afirmado no enunciado.

Teorema 2.7. *Seja $\mathcal{P}$ uma matriz estocástica regular em que S tem cardinalidade finita. Suponha que S seja da forma $S = \{1, 2, 3, .., d\}$.*

Então:

a) $\mathcal{P}$ tem um único vetor de probabilidade fixo π e os componentes de π são todos positivos, ou seja, é único o vetor de probabilidade π tal que $\pi\,\mathcal{P} = \pi$, $\sum_{s\in S} \pi_s = 1$ e $\pi_s > 0, \forall s \in S$. Os autovalores tem todos norma menor que 1.

b) Se p é qualquer vetor de probabilidade, então a sequência de vetores $p\mathcal{P}$, $p\mathcal{P}^2$, $p\mathcal{P}^3$, ... converge exponencialmente para o ponto fixo π, isto é,

$$\lim_{n\to\infty} p\,\mathcal{P}^n = \pi.$$

Deste fato segue:

c) As entradas das matrizes $\mathcal{P}$, $\mathcal{P}^2$, $\mathcal{P}^3$,...,$\mathcal{P}^n$ obtidas a partir de $\mathcal{P}$ convergem para as entradas correspondentes da matriz Q cujas linhas são todas iguais ao vetor fixo π. De outra forma

$$\lim_{n\to\infty} \mathcal{P}^n = Q = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 & ... & \pi_d \\ \pi_1 & \pi_2 & \pi_3 & ... & \pi_d \\ . & . & . & ... & . \\ \pi_1 & \pi_2 & \pi_3 & ... & \pi_d \end{pmatrix}$$

Demonstração: Considere como no último teorema $\Sigma = \{(p_1, p_2, ..., p_d),$ tal que $p_i \geq 0$, $p_1 + p_2 + ... + p_d = 1\}$ e a função $T : \Sigma \to \Sigma$, tal que $T(p) = p\,\mathcal{P}$, onde $p = (p_1, p_2, ..., p_d)$. Sabemos, pelo teorema anterior que existe ao menos um ponto π tal que $\pi\,\mathcal{P} = \pi$.

Se $\mathcal{P}$ tivesse autovalor real λ maior que 1 associado a v então, para qualquer k, temos que $\mathcal{P}^k$ também, pois $\mathcal{P}^k(v) = \lambda^k v$. Da mesma forma, $\mathcal{P}^{r\,k}(v) = \lambda^{kr} v$. Note que todas as entradas de v são positivas (mesmo argumento do teorema 2.5: senão existiria i tal que $0 = v_i \lambda = \sum P_{ji}\, v_j$, e isto implica que todos os v_j são nulos.). Isto significa que, tomando r grande, todas as entradas de

$\mathcal{P}^{r\,k}(v)$ são arbitrariamente grandes. Mas, o vetor $\mathcal{P}^{r\,k}(v)$ tem sempre todas as coordenadas menores que d. Assim, não existe autovalor real de norma maior que 1.

Como $\mathcal{P}$ é aperiódica pelo Teorema 2.20 (a ser enunciado no futuro) não existem autovalores complexos de norma 1. Apenas o valor 1 é autovalor com norma 1.

Ainda, como $\mathcal{P}^k$ tem todas as entradas positivas, pelo Teorema 2.5 o espaço dos autovetores de $\mathcal{P}^k$ associados ao autovalor 1 tem dimensão 1. Como $\mathcal{P}^k(v) = v$, caso v seja autovetor de $\mathcal{P}$ associado a 1, concluímos que o espaço dos autovetores de $\mathcal{P}$ associados ao autovalor 1 tem dimensão 1.

Assim, o π fixo para $\mathcal{P}$ é o único autovetor associado ao autovalor 1.

Desta forma, todo autovetor v de $\mathcal{P}$ tem autovalor com norma menor que 1.

Vamos mostrar que para qualquer $p \in \Sigma$ vale

$$\lim_{n\to\infty} p\,\mathcal{P}^n = \pi.$$

Vamos denotar por $\mathcal{P}^*$ a matriz transposta de $\mathcal{P}$, isto é,

$$\mathcal{P}^* = \begin{pmatrix} P_{11} & P_{21} & . & . & P_{d1} \\ P_{12} & P_{22} & . & . & P_{d2} \\ . & . & . & . & . \\ P_{1d} & P_{2d} & . & . & P_{dd} \end{pmatrix}.$$

É fácil ver que é válida a relação

$$< u\mathcal{P}, v >=< u, v\,\mathcal{P}^* >,$$

para todo $u, v \in \mathbb{R}^d$.

Considere o espaço

$$V = \{v \in \mathbb{R}^d \mid 0 = < v, (1,1,1,...,1) > = v_1 + v_2 + ... + v_d\}.$$

O espaço V tem dimensão $d-1$. O espaço gerado pelo vetor $(1,1,...,1)$ e o espaço vetorial V geram o $\mathbb{R}^d$.

Vamos mostrar agora que dado $v \in V$, temos que $v\,\mathcal{P} \in V$.

Como a matriz $\mathcal{P}$ é estocástica temos que vale

$$\mathcal{P}\begin{pmatrix}1\\1\\.\\.\\1\end{pmatrix} = \begin{pmatrix}1\\1\\.\\.\\1\end{pmatrix}.$$

Logo,

$$(1\ 1\ 1\ldots 1)\ \mathcal{P}^* = (1\ 1\ 1\ldots 1).$$

Então,

$$< v\,\mathcal{P}\,,\,(1,1,...,1) > = < v\,,\,(1,1,...,1)\,\mathcal{P}^* > = < v\,,\,(1,1,...,1) >= 0.$$

Assim $v\,\mathcal{P} \in V$.

Considere a transformação linear $K : V \to V$ agindo no espaço vetorial V, induzida por $\mathcal{P}$, isto é, $K(v) = v\,\mathcal{P}$. Os autovalores reais de K são todos com norma menor do que 1 (pois autovalores de K são autovalores de $\mathcal{P}$).

Afirmamos que não existem autovalores complexos de norma maior que 1. A demonstração deste fato está na observação que faremos ao fim da prova deste Teorema.

Seja $c < 1$ a norma do maior destes autovalores (reais ou complexos). Desta forma, pelo Lemma 9.10 [CD] (talvez com outra norma) $|K(v)| \leq c\,|v|$, para todo $v \neq 0$ em V .

Note que π não está em V pois $< \pi\,,\,(1,1,..,1) >= 1$. Assim, o subespaço gerado por π e V (que tem dimensão $d-1$) são dois espaços que juntos geram o $\mathbb{R}^d$.

Seja agora, $x, y \in \Sigma$, e escreva $x = v_1 + c_1\, \pi$, e $y = v_2 + c_2\, \pi$, onde $c_1, c_2 \in \mathbb{R}$, $v_1, v_2 \in V$.

Ora,

$$1 \;=< x, (1,1,...,1) > = < \,(x_1, x_2, ..., x_d), (1,1,...,1) > =$$

$$< v_1, (1,1,...,1) > \;+ c_1 \;< \pi, (1,1,...,1) > = \;0 \;+\; c_1.$$

Logo $c_1 = 1$, e assim pela mesma razão $c_2 = 1$.

Temos então que

$$x\,\mathcal{P} \;-\; y\,\mathcal{P} = (v_1 +\, \pi)\,\mathcal{P} - (v_2 +\, \pi)\,\mathcal{P} =$$

$$(K(v_1) +\, \pi) - (K(v_2) +\, \pi) = K(v_1) - K(v_2).$$

Logo,

$$|\, x\,\mathcal{P} \;-\; y\,\mathcal{P}\,| \;<\; c\,|\, v_1 \;-\; v_2\,| \;=$$

$$c|(v_1 + \pi) \;-\; (v_2 + \pi)\,| = c\,|\, x \;-\; y\,|.$$

Por indução,

$$|x\,\mathcal{P}^n \;-\; y\,\mathcal{P}^n| \;<\; c^n\,|x - y|$$

para $x, y \in \Sigma$ e onde $0 \leq c < 1$.

Considere agora $x \in \Sigma$ fixo e $y = \pi$. Temos então, que

$$|x\,\mathcal{P}^n \;-\; \pi\,| \;= |x\,\mathcal{P}^n \;-\; \pi\,\mathcal{P}^n| \;<\; c^n\,|x - \pi|.$$

Logo, para qualquer $p \in \Sigma$, vale que

$$\lim_{n \to \infty} \, p\,\mathcal{P}^n \;= \pi,$$

e a velocidade de convergência é exponencialmente rápida.

Observação: Se existisse autovalor complexo λ para K com norma maior que 1, então pela forma de Jordan [DL] existiria $x \in \mathbb{R}^d$ tal que para todo r vale $|K^r(x)| \geq |\lambda|^r |x|$ (ver começo da demonstração do Lema 9.10 em [DL]).

Ora, a matriz da transformação K^r é estocástica, e, assim todas as sua entradas são positivas e menores ou iguais a 1.

Assim, é impossível que $|K^r(x)| \to \infty$, quando $r \to \infty$, porque para qualquer x e $r \in \mathbb{N}$ temos que a norma de cada componente do vetor $K^r(x)$ é menor ou igual a $|x_1| + |x_2| + ... + |x_d|$. Assim, $|K^r(x)| \leq \sqrt{d}\,(|x_1| + |x_2| + ... + |x_d|)$.

$\square$

Na Figura 2.3 mostramos uma ilustração do resultado acima para matrizes regulares: a posição inicial é p_0 e exibimos a sua imagem através de sucessivas aplicações da transformação T. O exemplo é para o caso em que $S = \{1, 2, 3\}$. Observe que $T^n(p_0)$ se aproxima (quando n vai a infinito) de $\overline{p}$ que é o único vetor estacionário de probabilidade para $\mathcal{P}$.

Este último resultado é muito mais forte que o anterior. A partir dele obtemos o seguinte:

Teorema 2.8. *Seja $S = \{1, 2, ..., d\}$. Considere uma matriz estocástica $\mathcal{P}$ e um vetor $p \in \Sigma$ qualquer. Defina o processo estocástico markoviano X_n a partir do vetor de probabilidade inicial $p = (p_s)_{s\in S} = (P(X_0 = s))_{s\in S}$ e a matriz de transição $\mathcal{P}$. Suponha que $\mathcal{P}$ seja regular. Seja π o único vetor de Σ tal que $\pi\,\mathcal{P} = \pi$. O vetor $\pi^n = (P(X_n = s))_{s\in S}$ é tal que*

$$\lim_{n\to\infty} p\,\mathcal{P}^n =$$

$$\lim_{n\to\infty} (P(X_n = 1), P(X_n = 2), ..., P(X_n = d)) = \pi = (\pi_s)_{s\in\{1,2,...,d\}}.$$

Logo, para $s \in \{1, 2, ..., d\}$ fixo,

$$\lim_{n\to\infty} P(X_n = s) = \pi_s.$$

Note que a afirmação independe do p escolhido.

Demonstração: Segue direto do fato que $P(X_n = s)$ é a ordenada $s-$ésima do vetor $p\,\mathcal{P}^n$ e que

$$\lim_{n\to\infty} p\,\mathcal{P}^n = \lim_{n\to\infty} T^n(p) = \pi$$

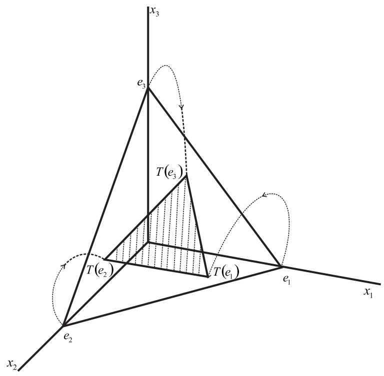


Figura 2.2: A ação de T em Σ

$\square$

Quando S é infinito, dada uma matriz estocástica regular $\mathcal{P}$ da forma S por S, nem sempre existe π tal que $\pi\mathcal{P} = \pi$.

Exemplo 2.16. Considere a matriz estocástica $\mathcal{P} = \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix}$.

Note que $\mathcal{P}^2 = \begin{pmatrix} 0,50 & 0,50 \\ 0,25 & 0,75 \end{pmatrix}$.

Logo a matriz $\mathcal{P}$ é regular.

Desejamos calcular $\pi = (x, 1-x)$ tal que $\pi\,\mathcal{P} = \pi$ e $0 \leq x \leq 1$.

Sendo assim,

$$\begin{cases} (1-x)\frac{1}{2} = x \\ x + (1-x)\frac{1}{2} = 1-x \end{cases} \Rrightarrow \begin{cases} 1-x = 2x \\ \frac{1}{2} + \frac{1}{2}x = 1-x \end{cases} \Rrightarrow \ 1 = 3x$$

Portanto, $\pi = (\frac{1}{3}, \frac{2}{3})$ é o único vetor fixo de probabilidade para a matriz $\mathcal{P}$.

Pelo teorema anterior, a sequência $\mathcal{P}$, $\mathcal{P}^2$, $\mathcal{P}^3$,...,$\mathcal{P}^n$ converge para a matriz

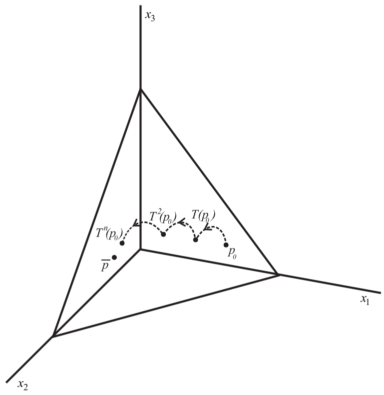


Figura 2.3: Os pontos $p_0, T(p_0)$, $T^2(p_0)$, $T^3(p_0)$,... que se aproximam de $\overline{p}$.

$$Q = \begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{pmatrix} = \begin{pmatrix} 0,33 & 0,67 \\ 0,33 & 0,67 \end{pmatrix}.$$

Observe que

$$\mathcal{P}^2 = \begin{pmatrix} 0,50 & 0,50 \\ 0,25 & 0,75 \end{pmatrix} ; \mathcal{P}^3 = \begin{pmatrix} 0,25 & 0,75 \\ 0,37 & 0,63 \end{pmatrix} ; \mathcal{P}^4 = \begin{pmatrix} 0,37 & 0,63 \\ 0,31 & 0,69 \end{pmatrix} ;$$
$$\mathcal{P}^5 = \begin{pmatrix} 0,31 & 0,69 \\ 0,34 & 0,66 \end{pmatrix}.$$

A sequência de matrizes acima nos confirma o que o último teorema afirma. Observe que os elementos da matriz $\mathcal{P}^n$ estão convergindo com $n \to \infty$, a certos valores e que são na verdade as entradas da matriz Q acima.

Note que a convergência é razoavelmente rápida (exponencial).

$\diamondsuit$

Exemplo 2.17. Considere a matriz estocastica

$$\mathcal{P} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Neste caso qualquer vetor inicial de probabilidade é estacionário e para cada π existe convergência de $\pi\mathcal{P}^n \to \pi$, quando $n \to \infty$.

Para a matriz estocastica

$$\mathcal{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

o vetor inicial estacionário é $(1/2, 1/2)$. O ponto fixo é único mas não existe convergência de $\pi\mathcal{P}^n$, quando $n \to \infty$, e $\pi \neq (1/2, 1/2)$.

$\diamondsuit$

Exemplo 2.18. Considere $S = \{0, 1, 2, 3, 4\}$ e a C.M. $(X_n)_{n\geq 0}$ cuja matriz de transição é dada por

$$\mathcal{P} = \begin{pmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1/9 & 0 & 3/9 & 0 & 5/9 \\ 1/9 & 0 & 0 & 5/9 & 3/9 \end{pmatrix}.$$

Note, a partir da matriz acima, que de qualquer ponto de S não se tem acesso a 1 em um passo (e assim também em n passos). Na Figura 2.4 mostramos um grafo (associado a $\mathcal{P}$) que conecta com uma seta dirigida de um estado a aqueles que se pode passar a partir dele (com probabilidade positiva). Em cada seta esta colocada a probabilidade de transição correspondente. Note que o estado 1 passa a 2 com probabilidade 1 e que nenhum estado em S pode passar ao estado 1. A partir do estado 4, em uma etapa, podemos atingir 0, 3 e continuar em 4. Olhando este grafo se pode ver que seguindo sucessivas etapas (basta percorrer as setas), em tempo finito, se pode passar de qualquer estado em $\{0, 2, 3, 4\}$ a qualquer estado em $\{0, 2, 3, 4\}$ com probabilidade positiva.

Vamos calcular a distribuição estacionária desta cadeia.

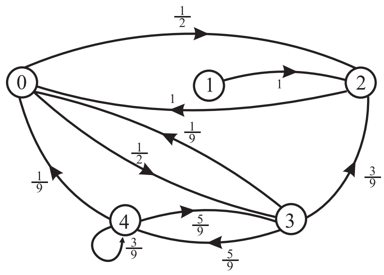


Figura 2.4:

Devemos encontrar π tal que satisfaz $\pi = \pi\mathcal{P}$.

Sendo assim, $\pi_2+\pi_3\frac{1}{9}+\pi_4\frac{1}{9} = \pi_0,\ 5\pi_3+3\pi_4 = 9\pi_4 \Rightarrow 5\pi_3 = 6\pi_4 \to \pi_4 = \frac{5}{6}\pi_3$

Ainda, $\pi_0\frac{1}{2} + \pi_1 + \pi_3\frac{3}{9} = \pi_2,\quad 9\pi_0 + 10\pi_4 = 18\pi_3 \Rightarrow 9\pi_0 + 10\pi_4 = 18\frac{6}{5}\pi_4 \Rightarrow 9\pi_0 = (18\frac{6}{5} - 10)\pi_4,\ \pi_0\frac{1}{2} + \pi_4\frac{5}{9} = \pi_3 \Rightarrow \pi_0 = \frac{58}{45}\pi_4$.

Procedendo de forma semelhante se obtém facilmente que $\pi_2 = \frac{47}{45}\pi_4$ e $\pi_1 = 0$.

Como $\sum_{i=0}^{4}\pi_i = 1$, então

$$\frac{58}{45}\pi_4 + 0 + \frac{47}{45}\pi_4 + \frac{6}{5}\pi_4 + \pi_4 = 1.$$

Sendo assim, segue que $\pi_4 = \frac{45}{204}$.

Portanto,

$$(\pi_0, \pi_1, \pi_2, \pi_3, \pi_4) \ = \ \pi \ = \ \left(\frac{58}{204}, 0, \frac{47}{204}, \frac{54}{204}, \frac{45}{204}\right),$$

é o vetor de probabilidade estacionário.

A partir do que se viu no grafo, não é de surpreender que a probabilidade π_1, correspondente ao estado 1, em π invariante é nula.

$\diamondsuit$

Exercício: Encontre o único vetor fixo de probabilidade da matriz estocástica regular

$$\mathcal{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix} \quad \mathcal{P}^2 = \begin{pmatrix} \frac{3}{8} & \frac{3}{8} & \frac{2}{8} \\ \frac{2}{8} & \frac{5}{8} & \frac{1}{8} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}; \quad \mathcal{P}^3 = \begin{pmatrix} \frac{12}{32} & \frac{11}{32} & \frac{9}{32} \\ \frac{7}{16} & \frac{3}{16} & \frac{6}{16} \\ \frac{2}{8} & \frac{5}{8} & \frac{1}{8} \end{pmatrix}.$$

Exercício: Por que uma matriz estocástica com um número 1 na diagonal principal não é regular?

Exemplo 2.19 (**Passeio Aleatório**). Considere a cadeia P sobre $S = \mathbb{Z}$ tal que para todo $i \in \mathbb{Z}$ fixo, temos para $P(i,i) = 0$, $P(i,i+1) = 1/2$, $P(i,i-1) = 1/2$ e $P(i,j) = 0$ para $j \neq i-1, i+1$. Neste caso não existe π tal que $\pi\, P = \pi$.

De fato, denote $\pi = (\pi_i)_{i \in \mathbb{Z}}$, então se π é tal que $\pi\, P = \pi$ temos que $\pi_0 = 1/2\ (\pi_{-1} + \pi_1)$.

Note que, por sua vez, $\pi_1 = 1/2\ (\pi_0 + \pi_2)$ e $\pi_{-1} = 1/2\ (\pi_{-2} + \pi_0)$.

Logo,

$$\pi_0 = 1/2\ (\pi_{-1} + \pi_1) = 1/2\ (\,1/2\ (\pi_{-2} + \pi_0) + 1/2\ (\pi_0 + \pi_2)\,).$$

Disto resulta que $\pi_0 = 1/2\ (\pi_2 + \pi_{-2})$.

Por indução é fácil ver que também vale para todo $n \in \mathbb{N}$, que $\pi_0 = 1/2\,(\pi_n + \pi_{-n})$. Se $\sum_{i \in \mathbb{Z}} \pi_i = 1$, então

$$\lim_{i \to \infty} \pi_i = 0 = \lim_{i \to -\infty} \pi_i.$$

Logo, $\pi_0 = 0$.

O mesmo raciocínio pode ser feito para qualquer π_j.

Assim, se $\pi P = \pi$ teríamos que $\pi_j = 0$ para todo $j \in \mathbb{Z}$. Logo, não existe $\pi \in \Sigma$ tal que $\pi\, P = \pi$.

$\diamondsuit$

2.3 Classificação de Estados de Cadeias de Markov

Desejamos agora classificar os estados $i \in S$ de uma C.M. definida por $\mathcal{P} = P(i,j)_{i,j \in S}$, matriz de transição, de acordo com a possibilidade de ir de um dado estado para outro. Esta análise será necessária para analisar cadeias de Markov mais gerais do que as obtidas a partir de matrizes estocásticas regulares.

Sabemos que se uma matriz estocástica tem todas as entradas estritamente positivas, então ela possui um único vetor estacionário. Muitas vezes somos levados a analisar cadeias que não possuem todas as entradas estritamente positivas (exemplos deste tipo surgem naturalmente quando consideramos cadeias de Markov de ordem superior, conforme seção 2.11).

Uma classe grande das cadeias de Markov em que algumas entradas são nulas tem também apenas um vetor estacionário. Um dos objetivos das próximas seções é obter resultados que nos assegurem esta unicidade em certos casos gerais.

Note que inicialmente não estamos analisando Processos de Markov, mas sim questões que envolvem apenas a matriz de transição $\mathcal{P}$, ou seja, a cadeia de Markov associada à probabilidade P do processo estocástico markoviano $(X_n)_{n\in\mathbb{N}}$. No seu devido momento voltaremos a considerar novamente os processos estocásticos associados a tal $\mathcal{P}$ e certos vetores de probabilidades iniciais π.

Definição 2.13. *Sejam $i, j \in S$*

a) Se $P^n(i,j) > 0$ para algum $n \geq 0$ dizemos que "i conduz a j", e denotamos tal fato por $i \to j$.

b) Se $i \to j$ e $j \to i$, dizemos que "i e j se comunicam" e denotamos tal fato por $i \leftrightarrow j$.

($i \leftrightarrow j \Leftrightarrow \exists\, n_1, n_2 \geq 0$ tal que $P^{n_1}(i,j).P^{n_2}(j,i) > 0$)

Note que da expressão

$$P_{ij}^n = P(X_n = j|X_0 = i) = \sum_{i_1} \dots \sum_{i_{n-1}} P(i,i_1)P(i_1,i_2)\dots P(i_{n-1},j),$$

segue que $i \to j$, se e só se, existem $i_1, i_2, ..., i_{n-1} \in S$ tal que

$$P_{ii_1} > 0, P_{i_1i_2} > 0, P_{i_2i_3} > 0, ..., P_{i_{n-2}i_{n-1}} > 0, P_{i_{n-1}j} > 0.$$

Proposição 2.11. *A relação "$\leftrightarrow$"é uma relação de equivalência, i.é, (1) $i \leftrightarrow i$; (2) $i \leftrightarrow j \Rightarrow j \leftrightarrow i$; (3) $i \leftrightarrow j$ e $j \leftrightarrow k \Rightarrow i \leftrightarrow k$.*

Demonstração:

(1) $P^0(i,i) = 1 > 0$;

(2) $i \leftrightarrow j \Rightarrow \exists n_1, n_2 \geq 0$ tal que $P^{n_1}(i,j)\, P^{n_2}(j,i) > 0 \Rightarrow \exists\, n_1, n_2 \geq 0$ tal que $P^{n_2}(j,i)\, P^{n_1}(i,j) > 0 \Rightarrow j \leftrightarrow i$;

(3) $i \leftrightarrow j$ e $j \leftrightarrow k \Rightarrow \exists n, m \geq 0$ tais que $P^n(i,j) > 0$ e $P^m(j,k) > 0 \Rightarrow P^{(n+m)}(i,k) = \sum_r P^n(i,r)P^m(r,k) > P^n(i,j).P^m(j,k) > 0 \Rightarrow i \leftrightarrow k$.

□

Segue deste último resultado que "$\leftrightarrow$"divide S em classes de equivalência. Uma classe de equivalência $U \subset S$ é um conjunto tal que

a) para todo $x, y \in U$ vale que $x \leftrightarrow y$

e

b) para todo $z \notin U$ e todo $x \in U$, não vale que $z \leftrightarrow x$.

Exemplo 2.20. Seja $S = \{0, 1, 2, 3, 4, \dots, n, \dots\} = \mathbb{N}$ e

$$\mathcal{P} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2^2} & \frac{1}{2^3} & . & \frac{1}{2^n} & . \\ 1 & 0 & 0 & 0 & . & 0 & . \\ 0 & 1 & 0 & 0 & . & 0 & . \\ . & . & . & . & . & . & . \\ 0 & 0 & . & 1 & . & 0 & . \\ . & . & . & . & . & . & . \end{pmatrix}.$$

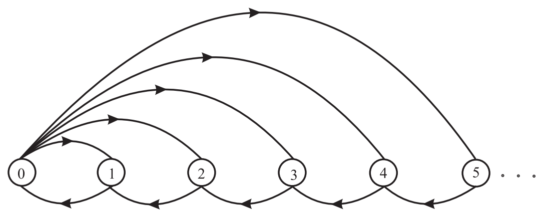


Figura 2.5:

Como dissemos antes, as setas representam comunicações em uma etapa entre dois estados. Quando não existe probabilidade positiva de ir em um estado de s_a a um estado s_b, não haverá seta dirigida apontando de s_a a s_b. O diagrama nos auxilia a enxergar se dois estados se comunicam.

Na Figura 2.5, por exemplo, mostramos o diagrama de setas correspondente a $\mathcal{P}$ acima. Desta forma, seguindo as setas podemos saber se um certo estado pode ser atingido ou não a partir de outro. Por exemplo, saindo de 0 podemos retornar a 0. De fato, seguimos de 0 para 2, depois de 2 a 1, e finalmente de 1 para 0. Não podemos, por exemplo, ir de 2 a 3 em uma etapa. Mas podemos, no entanto, ir de 2 a 3 em pelo menos três etapas. No presente exemplo só existe uma classe, ou seja, todos os estados se comunicam.

$\diamondsuit$

Exemplo 2.21. Considere $S = \{1, 2, 3, 4\}$ e uma C.M. com a seguinte matriz de transição.

$$\mathcal{P} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{3}{4} & \frac{1}{4} & 0 \\ \frac{1}{8} & \frac{7}{8} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{pmatrix}.$$

Neste caso, $1 \to 2$ e $1 \to 3$.

Ainda,

$2 \to 2$ e $2 \to 3$

e

$3 \to 1$ e $3 \to 2$

e

$4 \to 1$ e $4 \to 4$.

No presente exemplo, as classes de equivalência determinadas por "$\leftrightarrow$"são $\mathcal{C}_1 = \{1, 2, 3\}$ e $\mathcal{C}_2 = \{4\}$.

$\diamondsuit$

Definição 2.14. *Se existe somente uma classe de equivalência para a relação de comunicação "$\leftrightarrow$", a cadeia é dita irredutível. Diremos também que $\mathcal{P}$ é irredutível.*

Quando a matriz $\mathcal{P}$ for regular, então só existe uma classe de equivalência. De fato, se r for tal que P^r tem todas as entradas positivas, então para qualquer $i, j \in S$ vale que $P^r(i,j) > 0$.

Exemplo 2.22. Considere a matriz $\mathcal{P}$

$$\mathcal{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{8} & \frac{7}{8} \\ 0 & 0 & \frac{8}{11} & \frac{3}{11} \end{pmatrix}.$$

Denote por P_1 e P_2 as matrizes dois por dois

$$P_1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix},$$

e ainda,

$$P_2 = \begin{pmatrix} \frac{1}{8} & \frac{7}{8} \\ \\ \frac{8}{11} & \frac{3}{11} \end{pmatrix}.$$

Cada uma das matrizes P_1 e P_2 é regular.

Note que a evolução das probabilidades de transição em n etapas P^n é determinada pelas probabilidades de transição em n etapas P_1^n e P_2^n. De fato, de maneira compacta a matriz 4 por 4 definida por $\mathcal{P}^n$ satisfaz

$$\mathcal{P}^n = \begin{pmatrix} P_1^n & 0 \\ \\ 0 & P_2^n \end{pmatrix}.$$

Usamos a seguinte notação acima: o 0 na matriz $\mathcal{P}^n$ representa a matriz 2 por 2

$$\begin{pmatrix} 0 & 0 \\ \\ 0 & 0 \end{pmatrix}.$$

Neste sentido, podemos afirmar (de maneira mais ou menos informal) que o "mundo $\{1,2\}$"evolui de maneira independente do "mundo $\{3,4\}$".

Sendo assim, tal $\mathcal{P}$ não é irredutível. Existem neste caso duas classes, $\{1,2\}$ e $\{3,4\}$. É mais natural analisar em separado primeiro a matriz P_1 e depois a matriz P_2. As duas matrizes são regulares (e assim irredutíveis). Seja $p_1 \in \mathbb{R}^2$ (único) tal que $p_1 P_1 = p_1$ e $p_2 \in \mathbb{R}^2$ (único) tal que $p_2 P_2 = p_2$.

Dizemos que P_1 é a restrição de $\mathcal{P}$ ao conjunto $\{1,2\}$ e P_2 é a restrição de $\mathcal{P}$ ao conjunto $\{3,4\}$.

$\diamondsuit$

Esta idéia será explorada no futuro em nosso texto, ou seja, dada uma matriz $\mathcal{P}$ sobre S, vamos mostrar que (em muitos casos) S pode ser decomposta em classes de equivalência $C_r \subset S$ tais que a matriz induzida pela restrição de $\mathcal{P}$ a cada subconjunto C_r seja irredutível.

Note no exemplo acima que dado λ tal que $0 \leq \lambda \leq 1$, então se $p = \lambda\,(p_1^1, p_1^2, 0, 0) + (1-\lambda)\,(0, 0, p_2^1, p_2^2)$, onde $p_1 = (p_1^1, p_1^2)$ e $p_2 = (p_2^1, p_2^2)$, então p é invariante para $\mathcal{P}$. Note que isto vale para qualquer λ fixo como acima. Logo, para tal $\mathcal{P}$ o p invariante não é único.

Definição 2.15. *Se $P(i,i) = 1$, i é dito ser um estado absorvente.*

Nota: se i é estado absorvente, então $\{i\}$ é a sua classe de equivalência.

Exemplo 2.23. Considere $S = \{0, 1, 2, 3, 4\}$ e uma C.M. $(X_n)_{n\geq 0}$ com matriz de transição dada por

$$\mathcal{P} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Classes de equivalência: $\mathcal{C}_1 = \{0\}; \mathcal{C}_2 = \{1\}; \mathcal{C}_3 = \{2\}; \mathcal{C}_4 = \{3, 4\}$.

Note que o estado 2 é absorvente.

$\diamondsuit$

Exercício: Seja $(X_n)_{n\geq 0}$ uma C.M. com $S = \{0, 1, 2, 3, 4\}$, e matriz de transição dada por

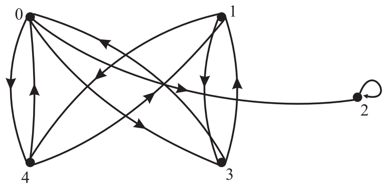


Figura 2.6:

$$\mathcal{P} = \begin{pmatrix} 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 & 0 & 0 \\ \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \end{pmatrix}.$$

As classes de equivalência são: $\mathcal{C}_1 = \{0, 1, 3, 4\}; \mathcal{C}_2 = \{2\}$.

Complete as probabilidades correspondentes nas setas do grafo da Figura 2.6 associado a tal $\mathcal{P}$.

Exemplo 2.24. Passeio Aleatório em $\mathbb{Z}$ ou Random walk.

Vamos elaborar com mais detalhes o que descrevemos anteriormente sobre passeio aleatório. Vamos considerar aqui um caso mais geral em que as probabilidades de saltar uma unidade para cima ou para baixo dependem agora do estado $s \in \mathbb{Z}$.

Sejam $Y_1, Y_2, \ldots, Y_t, \ldots$ independentes e identicamente distribuídas, ou seja, i.i.d., em notação condensada, onde $t \in T = \mathbb{N}$, assumindo valores em $\mathbb{Z}$. Vamos construir um novo processo estocástico X_t, $t \in T = \mathbb{N}$.

Vamos assumir que $X_0 = 0$,

$X_n = Y_1 + Y_2 + ... + Y_n$, $n \geq 1$

$X_{n+1} = X_n + Y_{n+1}$

$(X_n)_{n\geq 0}$ satisfaz a propriedade de Markov pois:

Y_{n+1} é independente de $Y_1, ..., Y_n$, então X_{n+1} depende apenas de X_n (ver (*) após a definição 1.9).

Mais precisamente, como Y_{n+1} é independente de $X_s = Y_1 + Y_2 + ... + Y_s$ para $s < n$ (ver (**) apos definição 1.9), então

$$P(X_{n+1} = s | X_n = x_n, X_{n-1} = x_{n-1}, ..., X_0 = 0) =$$

$$P(X_{n+1} = Y_{n+1} + X_n = s | X_n = x_n, X_{n-1} = x_{n-1}, ..., X_0 = 0) =$$
$$\frac{P(Y_{n+1} + X_n = s, X_n = x_n, X_{n-1} = x_{n-1}, ..., X_0 = 0)}{P(X_n = x_n, X_{n-1} = x_{n-1}, ..., X_0 = 0)} =$$
$$P(Y_{n+1} = s - x_n | X_n = x_n, X_{n-1} = x_{n-1}, ..., X_0 = 0) =$$
$$P(Y_{n+1} = s - x_n) =$$
$$P(Y_{n+1} = s - x_n \,|\, X_n = x_n) =$$
$$P(X_{n+1} = Y_{n+1} + X_n = s | X_n = x_n).$$

Vamos definir agora as probabilidades de transição. Considere $P(Y_1 = k) = a_k$, onde $a_k \geq 0$ e $\sum_{k\in\mathbb{Z}} a_k = 1$.

A partir desta informação desejamos calcular as probabilidades de transição:

$$P(i,j) = P(X_{n+1} = j | X_n = i) = \frac{P(X_{n+1} = j, X_n = i)}{P(X_n = i)} =$$
$$\frac{P(X_n + Y_{n+1} = j, X_n = i)}{P(X_n = i)} =$$
$$\frac{P(Y_{n+1} = j - i, X_n = i)}{P(X_n = i)} = \frac{P(Y_{n+1} = j - i).P(X_n = i)}{P(X_n = i)} =$$
$$P(Y_{n+1} = (j - i)) = a_{j-i}.$$

Portanto, $(X_n)_{n\geq 0}$ é uma C.M. com probabilidade de transição estacionárias e dadas por $P(i,j) = a_{j-i}$. Observamos que estas probabilidade só dependem do incremento $j-i$; isto é, $P(i,j)$ é homogênea no espaço

$P(i,j) = P(0, j-i)$.

Se todos os $a_i, i \in \mathbb{Z}$ são não nulos então a matriz $\mathcal{P}$ é irredutível.

$\diamondsuit$

Exemplo 2.25. Passeio Simples em $\mathbb{Z}$ (ou Passeio de Bernoulli).

Sejam $P(Y_1 = 1) = p$ e $P(Y_1 = -1) = 1 - p = q$. As variáveis aleatórias Y_i são $i.i.d$, em notação abreviada.

O passeio aleatório sobre $\mathbb{Z}$, denotado por X_n, é definido da seguinte forma:

$X_0 = 0$, $X_1 = Y_1$,..,$X_n = Y_1 + Y_2 + ... + Y_n$, $X_{n+1} = X_n + Y_{n+1}$,...

Neste caso $P(i, i+1) = p$ e $P(i, i-1) = q$, $p + q = 1$

A matriz de transição é dada por

$$\mathcal{P} = \begin{pmatrix} ... & . & . & . & . & . & ... \\ ... & 0 & p & 0 & 0 & 0 & ... \\ ... & q & 0 & p & 0 & 0 & ... \\ ... & 0 & q & 0 & p & 0 & ... \\ ... & 0 & 0 & q & 0 & p & ... \\ ... & . & . & . & . & . & .... \end{pmatrix}$$

Note que, dado que estamos em i, sempre podemos, com probabilidade positiva $P(i, i+1) > 0$ ir para a direita e com probabilidade $P(i, i-1) > 0$ ir para a esquerda. Assim, pela caracterização descrita após a definição de “i conduz a j", podemos ir de i para qualquer outro estado em um número finito de passos.

Concluímos que esta é uma cadeia irredutível pois só existe uma classe de equivalência.

Note que se $p = q = 1/2$, é fácil ver que a equação de

$$\pi = (..., \pi_{-n}, ..., \pi_{-1}, \pi_0, \pi_1, ..., \pi_n, ...),$$

tal que $\pi\mathcal{P} = \pi$, nos indica por indução, que π_n é constante.

Isto porque $\pi_n = \frac{1}{2}(\pi_{n-1} + \pi_{n+1})$, para todo n. Assim $\frac{1}{2}(\pi_{-2} + \pi_2) = \pi_0$, e por indução, $\frac{1}{2}(\pi_{-n} + \pi_n) = \pi_0$, para todo $n \in \mathbb{N}$.

Isto não pode ocorrer se $\sum_{n\in\mathbb{Z}} \pi_n = 1$ (pois π_n e $\pi_{-n} \to 0$, quando $n \to \infty$). Logo, para tal $\mathcal{P}$ não é possível encontrar uma probabilidade inicial π tal que determine um Processo Estacionário.

$\diamondsuit$

Exemplo 2.26. Passeio casual (sobe 1 com probabilidade p e desce 1 com probabilidade $q = 1 - p$, onde $p + q = 1$ e $p, q \geq 0$) em $\{0, 1, ..., d\}$ com barreiras absorventes em 0 e d, com $S = \{0, 1, 2, ..., d\}$, $d \in \mathbb{N}$, $d > 0$.

Neste caso a matriz de transição da cadeia de Markov é

$$\mathcal{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ q & 0 & p & 0 & \dots & 0 \\ 0 & q & 0 & p & \dots & 0 \\ 0 & 0 & q & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

Observamos que $\mathcal{C}_1 = \{0\}$, $\mathcal{C}_2 = \{1, 2, 3, \dots, d-1\}$ e $\mathcal{C}_3 = \{d\}$.

$\diamondsuit$

2.4 Tempo de Primeira chegada

Vamos agora introduzir alguns conceitos que são de fundamental importância para analisar cadeias de Markov mais gerais do que aquelas do tipo regular.

Definição 2.16. *Seja $A \subset S$ e $(X_n)_{n\geq 0}$ uma C.M. com matriz de transição $\mathcal{P}$. O tempo de primeira chegada (ou, de passagem por A) $T_A(\omega)$ (para o conjunto A) começando em $\omega \in \Omega$ é definido por*

$$\begin{cases} \min\{n \geq 1 \,:\, X_n(\omega) \in A\}, & \textit{no caso em que } \{n \geq 1 \,:\, X_n(w) \in A\} \neq \emptyset \\ +\infty\ , & \textit{caso contrário} \end{cases}$$

Em outras palavras $T_A(\omega)$ é o primeiro j tal que $w_j \in A$, onde

$$\omega = (w_0, w_1, w_2, w_3, \ldots, w_n, \ldots).$$

i) T_A é uma variável aleatória estendida (isto é, pode assumir valor $+\infty$) tomando valores em $\mathbb{N}^* \bigcup \{+\infty\}$, onde $\mathbb{N}^* = \mathbb{N} - \{0\}$.

Ou seja, $T_A : \Omega \to \mathbb{N}^* \bigcup \{+\infty\}$ é uma função mensurável. Pode, ou não, ser integrável em relação à probabilidade P sobre

$$\Omega = \{\omega = (w_0, w_1, w_2, w_3, \ldots, w_n, \ldots) \,:\, w_n \in S\}.$$

Note que $Z_A = \{\omega \,|\, T_A(w) = \infty\}$ é um conjunto mensurável. Se A não é vazio, então Z_A é não vazio. Em certos casos pode ter probabilidade positiva e em outros nula.

ii) Se $A = \{i\}$ denotamos $T_A = T_i$. Neste caso, $T_i(w) = \min\{n \geq 1\,, \text{tal que}, X_n(w) = i\}$. Note que não se pode dizer que T_i, $i \in S$ fixo, depende de finitas coordenadas.

O tempo de primeira chegada é um caso particular do que vai se chamar posteriormente de tempo de parada (ou tempo de Markov).

Exemplo 2.27. Seja $S = \{1, 2, 3\}$ e $T_2 : \Omega \to \mathbb{N} \bigcup \{+\infty\}$ definido como em ii) acima.

Sendo assim, se

$$\omega = (2, 3, 1, 3, 3, 2, 3, 3, 3, 1, 2, ...),$$

temos que $T_2(\omega) = 5$. Se $\omega = (w_0, w_1, w_2, w_3, ..., w_t, \dots)$ for o caminho amostral que começa em 2 e depois alterna 3 e 1 para sempre, ou seja,

$$\omega = (2, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1, ...),$$

então $T_2(\omega) = \infty$.

$\diamondsuit$

Definição 2.17. $f_{ij}^n = P(T_j = n | X_0 = i) = P_i(T_j = n)$ *é a probabilidade da cadeia começando em i atingir o estado j pela primeira vez no tempo n, ou, mais precisamente,*

$f_{ij}^n = P(X_r \neq j, r = 1, 2, \dots, n-1, X_n = j | X_0 = i).$

Ainda,

(1) Para $i \in S$ fixo, $f_{ii}^* = \sum_{n \geq 1} f_{ii}^n$

(2) Para i e j fixos em S, $f_{ij}^* = \sum_{n \geq 1} f_{ij}^n$.

Por definição $f_{ii}^0 = 0$ e também $f_{ij}^0 = 0$.

Por exemplo, se $S = \{1, 2, 3\}$, então se $w \in \overline{2, 3, 3, 3, 1}$, temos que seu tempo de atingir 1 pela primeira vez começando em 2 é 4. Se $w \in \overline{2, 3, 2, 1, 1}$, temos que seu tempo de atingir 1 pela primeira vez começando em 2 é 3.

Por exemplo, se $S = \{1, 2, 3\}$, então se $w \in \overline{2, 1, 3, 2}$, temos que seu tempo de atingir 2 pela primeira vez começando em 2 é 3.

Proposição 2.12. $P^n(i, j) = \sum_{k=1}^n f_{ij}^k P^{n-k}(j, j)$

Demonstração: Note que $A_k = \{\omega | T_j(\omega) = k, X_0 = i, X_n = j\} \subset \Omega$ onde $k \in \{1, 2, \ldots, n\}$ é uma partição de $\{X_0 = i, X_n = j\} \subset \Omega$.

Ainda,

$$P^n(i,j) = P(X_n = j | X_0 = i) = \frac{P(X_0 = i, X_n = j)}{P(X_0 = i)} =$$

$$\frac{P(X_n = j, \bigcup_{k=1}^n [T_j = k], X_0 = i)}{P(X_0 = i)} =$$

$$\sum_{k=1}^n \frac{P(X_n = j, T_j = k, X_0 = i)}{P(X_0 = i)} =$$

$$\sum_{k=1}^n \frac{P(X_n = j, X_k = j, X_{k-1} \neq j, X_{k-2} \neq j, ..., X_1 \neq j, X_0 = i)}{P(X_k = j, X_{k-1} \neq j, X_{k-2} \neq j, ..., X_1 \neq j, X_0 = i)} \times$$

$$\frac{P(X_k = j, X_{k-1} \neq j, X_{k-2} \neq j, ..., X_1 \neq j, X_0 = i)}{P(X_0 = i)} =$$

$$\sum_{k=1}^n P(X_n = j \,|\, X_k = j, X_{k-1} \neq j, X_{k-2} \neq j, ..., X_1 \neq j, X_0 = i) \times$$

$$P(X_k = j, X_{k-1} \neq j, X_{k-2} \neq j, ..., X_1 \neq j \,|\, X_0 = i) =$$

$$\sum_{k=1}^n P(X_n = j \,|\, X_k = j) \; P(X_k = j, X_{k-1} \neq j, X_{k-2} \neq j, ..., X_1 \neq j \,|\, X_0 = i) =$$

$$\sum_{k=1}^n f_{ij}^k P^{n-k}(j,j)$$

$\square$

Observação: Note que por definição

(1) Para $i \in S$ fixo, $f_{ii}^* = \sum_{n \geq 1} f_{ii}^n$

(2) Para i e j fixos em S, $f_{ij}^* = \sum_{n \geq 1} f_{ij}^n$.

Sendo assim,

$$f_{ii}^* = \sum_{n \geq 1} f_{ii}^n = \sum_{n \geq 1} P(T_i = n | X_0 = i) =$$

$$P(T_i = 1|X_0 = i) + P(T_i = 2|X_0 = i) + P(T_i = 3|X_0 = i) + \ldots$$
$$= P(T_i < \infty|X_0 = i).$$

Este valor, ou seja, f_{ii}^*, indica assim a probabilidade de um caminho amostral voltar ao estado i num tempo finito, dado que iniciou no estado i.

Do mesmo modo, para i e j fixos

$$f_{ij}^* = P(T_j < \infty|X_0 = i),$$

já que $T_j < \infty \Rightarrow \min\{n \geq 1 \,:\, X_n = j\} < +\infty$

Portanto, f_{ij}^* pode ser interpretada como a probabilidade condicional de que o processo visite o estado $\underline{j}$ pela primeira vez em algum tempo positivo finito $\underline{n}$,dado que ele começou no estado $\underline{i}$.

Note que para $i \neq j$, e n fixo, então $\{\omega \,|\, T_i(\omega) = n\} \cap \{\omega \,|\, T_j(\omega) = n\} = \emptyset$.

Seja i fixo. Se $f_{ii}^* < 1$ então a família de conjuntos $A_n = \{\omega \,| X_0(w) = i\,,\, T_i = n\}$ indexado por $n \geq 1$ são disjuntos, mas não formam uma partição do espaço Ω (em termos de probabilidade), ou seja, $P(\cup_{n=1}^{\infty} A_n\,) = \sum_{n=1}^{\infty} f_{ii}^n < 1$. Portanto, neste caso existe uma probabilidade positiva de que saindo de $X_0 = i$, o processo não volte nunca mais ao estado i. Isto é, $P_i(\,\{\omega \,:\, \omega = (w_0, w_1, w_2, w_3, ..., w_n, ...), w_n \neq i, \forall n > 1\}\,) > 0$.

Isto significa que i é transitório conforme definição abaixo.

Definição 2.18. *Se $f_{ii}^* < 1$ então* **i** *é dito estado transitório ou transiente. Se $f_{ii}^* = 1$ então* **i** *é dito estado recorrente.*

Logo, se i é recorrente, uma C.M. começando em i retorna a i com probabilidade 1.

Se i é transitório, uma C.M. começando em i tem probabilidade $1 - f_{ii}^*$ de nunca retornar a i.

Observação: Se **i** é um estado absorvente, então $P_i(T_i = 1) = P(T_i = 1|X_0 = i) = P(i, i) = 1$.

Portanto, $f^*_{ii} = 1$ o que mostra que todo estado absorvente é recorrente. Note que nem todo estado recorrente é absorvente.

Exemplo 2.28. Considere $S = \{1, 2, 3\}$ e uma C.M. $(X_n)_{n\geq 0}$ com matriz de transição dada por

$$\mathcal{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}.$$

Classes de equivalência: $\mathcal{C}_1 = \{1\}; \mathcal{C}_2 = \{2, 3\}$

Note que o estado 1 é transiente.

Para analisar tal cadeia devemos eliminar 1 de nossas considerações e considerar a matriz

$$\tilde{\mathcal{P}} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}.$$

Por exemplo, se (π_1, π_2) é vetor estacionário para $\tilde{\mathcal{P}}$, então, $(0, \pi_1, \pi_2)$ é vetor estacionário para $\mathcal{P}$.

O processo estocástico descrito por $\mathcal{P}$, de fato, "não *enxerga*"o estado 1.

$\diamondsuit$

Exemplo 2.29. Considere $S = \{1, 2\}$ e

$$\mathcal{P} = \begin{pmatrix} 0,2 & 0,8 \\ 0,3 & 0,7 \end{pmatrix}.$$

Observe que $P(i, i) \neq 1$ para todo $i \in S$. Logo, nenhum estado é absorvente. Vamos mostrar que todos os estados são recorrentes.

Primeiro note que a cadeia é irredutível pois trivialmente só tem uma classe de equivalência (a matriz é regular). Além disso, note que como só temos dois estados, se começamos em 1, então para $n > 2$ fixo temos que $f_{11}^n = P(T_1 = n|X_0 = 1) = (0.3)\,(0.7)^{n-2}\, 0.8$. Isto porque os caminhos $\omega = (w_t) \in \Omega$ possíveis em $\{T_1 = n \text{ e } X_0 = 1\}$ são sempre da forma: começa em $w_0 = 1$ salta em seguida para $w_1 = 2$ e fica lá $n-1$ vezes e, depois, volta pela primeira vez a $1 = w_n$ exatamente no tempo n. Não existem restrições sobre w_t para $t > n$.

Isto é, neste caso estamos calculando

$$P(\{X_0 = 1, X_1 = 2, X_3 = 2, ..., X_{n-1} = 2, X_n = 1\})$$

que tem o valor acima.

Logo,

$$f_{11}^* = \sum_{n\geq 1} f_{11}^n =$$

$$P(T_1 = 1|X_0 = 1) + P(T_1 = 2|X_0 = 1) + P(T_1 = 3|X_0 = 1)+$$

$$P(T_1 = 4|X_0 = 1) + \ldots =$$

$$0,2 + (0,3)(0,8) + (0,7)(0,3)(0,8) + (0,7)^2(0,3)(0,8) + \ldots =$$

$$0,2\ +\ 0,8\,[\sum_{k\geq 0}(0,7)^k]\,0,3 =$$

$$0,2\ +\ 0,8\,\frac{1}{1-0,7}\,0,3 = 0,2 + 0,8 = 1.$$

Logo, $f_{11}^* = 1$.

Da mesma forma,

$$f_{22}^* = \sum_{n\geq 1} f_{22}^n = P(T_2 = 1|X_0 = 2) + P(T_2 = 2|X_0 = 2) + P(T_2 = 3|X_0 = 2)+$$

$$P(T_2 = 4|X_0 = 2) + \ldots =$$

$$0,7 + (0,8)(0,3) + (0,8)(0,2)(0,3) + (0,8)(0,2)^2(0,3) + (0,8)(0,2)^3(0,3) + \ldots =$$

$$0,7\,+\,0,3\,[\sum_{k\geq 0}(0,2)^k]\,0,8 = 0,7+0,3\frac{1}{1-0,2}\,0,8 = 1.$$

Assim, $f^*_{22} = 1$.

Logo, todos os estados são recorrentes.

$\diamondsuit$

2.5 Critérios de Recorrência e Transiência

Vamos apresentar agora um critério de recorrência muito útil. Para isto é preciso definir para cada $i \in S$, a variável aleatória $N(i)$ = número total de visitas ao estado $\underline{i}$, onde $N(i) : \Omega \to \mathbb{N}\bigcup\{+\infty\}$.

$I_{\{i\}}$ denota o indicador do conjunto $\{i\} \subset S$.

Assim,

$$I_{\{i\}}(X_n) = \begin{cases} 1, \text{ se } X_n = i \,(\text{i.e, a cadeia está no estado } i \text{ no tempo } \mathbf{n}) \\ 0, \text{ se } X_n \neq i \,(\text{i.e, a cadeia não está em } i \text{ no tempo } \mathbf{n}) \end{cases}$$

Definimos $N(i) = \sum_{n\geq 1} I_{\{i\}}(X_n)$ como o número total de visitas ao estado **i** em qualquer tempo $\mathbf{n} \in \mathbb{N}$.

$N(i)$ é uma variável aleatória estendida, ou seja, é uma função mensurável $N(i) : \Omega \to \mathbb{N}\bigcup\{+\infty\}$.

Lembre que $P(\bullet|X_0 = i)$ define uma probabilidade em $(\Omega, \mathcal{A})$ e é denotada por P_i.

Mais precisamente, para C da forma $C = \overline{i, c_1, ..., c_n}$, temos que

$$P_i(C) = P_{ic_1} P_{c_1c_2} ... P_{c_{n-1}c_n}.$$

Se C é da forma $C = \overline{j, c_1, ..., c_n}$, com $j \neq i$, então $P_i(C) = 0$.

Desta forma, $P_i(N(j) < \infty)$ descreve a probabilidade de que o estado j foi atingido apenas um número finito de vezes, dado que se iniciou o processo com probabilidade 1 no estado i. Saber se este número é igual a 1 ou menor que 1 será um dos nossos principais objetivos na presente seção.

$\mathbb{E}_i(X)$ denota a esperança da função mensurável X em relação a P_i.

Para $X_n : \Omega \to \mathbb{R}$, $P_i(X_n \in A) = P(X_n \in A | X_0 = i)$.

Ainda, $\mathbb{E}_i(X_n) = \int X_n(w) dP_i(w) = \sum_{j \in S} j \, P_i\{X_n = j\}$.

Note que X_n é tal que para cada $i \in S$ o conjunto $\{X_n = i\}$ é uma união de cilindros. O valor de $\mathbb{E}_i(X_n)$ pode ser obtido portanto a partir do conhecimento de P_i apenas sobre cilindros C.

A seguir, $\mathbb{E}_i(N(j))$ descreve o número esperado de visitas a **j** para uma C.M. começando em **i**.

Considere a série de potências na variável real s (funções geradoras) dada por

$$F_{ij}(s) = \sum_{n=0}^{\infty} f_{ij}^n \; s^n,$$

e

$$P_{ij}(s) = \sum_{n=0}^{\infty} (\mathcal{P}^n)_{ij} \; s^n = \sum_{n=0}^{\infty} P_{ij}^n \; s^n.$$

Como os valores f_{ij}^n e P_{ij}^n são não negativos e menores ou iguais a 1 então as duas séries de potências estão definidas pelo menos no intervalo $(-1, 1)$.

Lembre que por hipótese $f_{ii}^0 = 0$.

Segue da Proposição 2.12 para o caso $i = j$ que $P^n(i,i) = \sum_{k=1}^n f_{ii}^k P^{n-k}(i,i)$, para todo $n \geq 1$. Vamos mostrar que desta expressão segue que

$$F_{ii}(s) P_{ii}(s) = P_{ii}(s) - 1.$$

De fato, como se sabe, se considerarmos duas séries de potências $\sum_{n=0}^{\infty} a_n s^n$ e $\sum_{m=0}^{\infty} b_m s^m$, então se fizermos o produto

$$\sum_{n=0}^{\infty} a_n s^n \times \sum_{m=0}^{\infty} b_m s^m,$$

obteremos uma nova série de potências

$$\sum_{k=0}^{\infty} c_k s^k,$$

onde o termo geral é

$$c_k = a_k\, b_0 + a_{k-1}\, b_1 + a_{k-2}\, b_2 + ... + a_2\, b_{k-1} + a_0\, b_k.$$

Por exemplo,

$$c_4 = a_4\, b_0 + a_3\, b_1 + a_2\, b_2 + a_1\, b_3 + a_0\, b_4.$$

Sendo assim, o termo geral de

$$\sum_{k=0}^{\infty} c_k s^k = F_{ij}(s)\, P_{ij}(s) =$$

$$\sum_{n=0}^{\infty} f_{ij}^n s^n \times \sum_{n=0}^{\infty} P_{ij}^n s^n,$$

é

$$c_k = f_{ij}^k\, P_{ij}^0 + f_{ij}^{k-1}\, P_{ij}^1 + ... + f_{ij}^1\, P_{ij}^{k-1} + f_{ij}^0\, P_{ij}^k =$$
$$= f_{ij}^k\, P_{ij}^0 + f_{ij}^{k-1}\, P_{ij}^1 + ... + f_{ij}^1\, P_{ij}^{k-1}.$$

Lembre que $P_{ij}^0 = 1$ se $i = j$ e $P_{ij}^0 = 0$ se $i \neq j$

Consideramos primeiro o caso $i = j$. Podemos comparar termo a termo as duas séries de potências

$$\sum_{k=0}^{\infty} c_k s^k = F_{ii}(s)\, P_{ii}(s)$$

e

$$\sum_{k=0}^{\infty} P_{ii}^k s^k = P_{ii}(s).$$

Seja $k > 1$. Como

$$P^k(i,i) = \sum_{r=1}^{k} f_{ii}^r P^{k-r}(i,i) = f_{ii}^k\, P_{ii}^0 + f_{ii}^{k-1}\, P_{ii}^1 + ... + f_{ii}^1\, P_{ii}^{k-1} = c_k,$$

então os termos que multiplicam s^k, para $k \geq 1$, são iguais nas duas séries.

O problema é o primeiro termo nas duas séries: para $P_{ii}(s)$ este termo é $1 = P(i,i)$ e para $F_{ii}(s)\, P_{ii}(s)$ este termo é $c_0 = 0$.

Para ter igualdade é preciso corrigir este termo e assim obtemos que

$$F_{ii}(s)P_{ii}(s) = P_{ii}(s) - 1$$

ou seja,

$$P_{ii}(s) = \frac{1}{1 - F_{ii}(s)}.$$

Ainda, da mesma forma como acima, segue da propriedade $P^n(i,j) = \sum_{k=1}^{n} f_{ij}^k P^{n-k}(j,j)$, para $n \geq 1$, que

$$F_{ij}(s)P_{ij}(s) = P_{ij}(s).$$

Note que $F_{ii}(1) = f_{ii}^*$ e $F_{ij}(1) = f_{ij}^*$.

Exemplo 2.30. Considere $S = \{1, 2\}$ e

$$\mathcal{P} = \begin{pmatrix} 0,2 & 0,8 \\ 0,3 & 0,7 \end{pmatrix}.$$

Mostramos em um exemplo anterior que para todo $n \geq 2$ vale $f_{11}^n = P(T_1 = n | X_0 = 1) = (0,3)\,(0,7)^{n-2}\ 0,8$.

Logo, neste caso,

$$F_{ii}(s) = 0,2\, s\ + \sum_{k=2}^{\infty} 0,3\,(0,7)^{k-2}\ 0,8\ s^k\ =$$

$$0,2\,s + s^2 \sum_{k=2}^{\infty} 0,3\,(0,7)^{k-2}\ 0,8\ s^{k-2} \ = \ 0,2\,s\ + 0,3\ \times 0,8\ \times \frac{s^2}{1-s\ \ 0,7}\ .$$

A partir de

$$P_{11}(s) = \frac{1}{1 - F_{11}(s)},$$

podemos obter a expressão analítica de $P_{11}(s)$, e desenvolvendo em série de potências (tomando derivadas de ordem superior de $P_{11}(s)$ em $s = 0$) se pode conseguir uma expressão para os distintos valores $P^n(1,1)$.

$\diamondsuit$

Lema 2.1 (Lema de Abel). *(a) Se* $\sum_{n=0}^{\infty} a_n$ *converge, então*

$$\lim_{s \to 1-} \sum_{n=0}^{\infty} a_n\, s^n = \sum_{n=0}^{\infty} a_n = \alpha.$$

(b) Suponha que $a_n \geq 0$ *para todo* n, *e* $\lim_{s\to 1-} \sum_{n=0}^{\infty} a_n\, s^n = \alpha \leq \infty$, *então*

$$\sum_{n=0}^{\infty} a_n \ = \ \lim_{m\to\infty} \sum_{n=0}^{m} a_n \ = \ \alpha.$$

Para uma demonstração deste resultado veja [KT1].

Proposição 2.13. $\mathbb{E}_i(N(j)) = \sum_{n\geq 1} P^n(i,j)$, *sempre que* $P_i(N(j) = +\infty) = 0$.

Demonstração:

$$\mathbb{E}_i(N(j)) = \mathbb{E}(N(j)|X_0 = i) = \mathbb{E}(\sum_{n\geq 1} I_{\{j\}}(X_n)|X_0 = i) = \mathbb{E}(\sum_{n\geq 1}(X_n = j)|X_0 = i).$$

Usando o Teorema da convergência monótona

$$\mathbb{E}(\sum_{n\geq 1} I_{\{j\}}(X_n)|X_0 = i) = \sum_{n\geq 1} \mathbb{E}(I_{\{j\}}(X_n)|X_0 = i) =$$

$$\sum_{n\geq 1} 1 \cdot P(X_n = j | X_0 = i) = \sum_{n\geq 1} P^n(i,j).$$

Observação: Se $P_i(N(j) = +\infty) > 0$, então $\mathbb{E}_i(N(j)) = +\infty$.

Teorema 2.9 (Critério de Recorrência). *i é recorrente* $\Leftrightarrow \sum_{n\geq 1} P_{ii}^n = +\infty$

Demonstração: Seja i recorrente. Sendo assim $\sum_{n=1}^{\infty} f_{ii}^n = 1$. Logo, pelo lema de Abel (a) temos

$$\lim_{s\to 1-} F_{ii}(s) = \lim_{s\to 1-} \sum_{n=0}^{\infty} f_{ii}^n \, s^n = \sum_{n=1}^{\infty} f_{ii}^n = 1.$$

Logo, da expressão

$$P_{ii}(s) = \frac{1}{1 - F_{ii}(s)},$$

temos que

$$\lim_{s\to 1-} P_{ii}(s) = \lim_{s\to 1-} \sum_{n=0}^{\infty} P_{ii}^n \, s^n = \infty.$$

Agora, usando o item b) do lema de Abel ($\alpha = \infty$) temos

$$\sum_{n=0}^{\infty} P_{ii}^n = \infty.$$

Isto prova a afirmação: i recorrente implica que $\sum_{n=0}^{\infty} P_{ii}^n = \infty$.

Suponha agora que i é transiente. Vamos mostrar que $\sum_{n=0}^{\infty} P_{ii}^n < \infty$.

Como i é transiente, então $\sum_{n=0}^{\infty} f_{ii}^n < 1$. Usando a parte a) do Lema de Abel temos que

$$\lim_{s\to 1-} F_{ii}(s) = \sum_{n=0}^{\infty} f_{ii}^n < 1.$$

Logo usando a expressão

$$P_{ii}(s) = \frac{1}{1 - F_{ii}(s)},$$

obtemos que

$$\lim_{s \to 1-} P_{ii}(s) < \infty.$$

Sendo assim, usando a parte b) do Lema de Abel concluímos que

$$\sum_{n=0}^{\infty} P_{ii}^n < \infty.$$

Fica assim demonstrado o que afirma o Teorema.

□

Corolário 2.1. *i é recorrente, se e só se, o número esperado de retornos a i começando em i, ou seja, $E_i(N(i))$ é infinito.*

Observe que: $\{N(j) \geq 1\} = \{T_j < +\infty\}$ pois o número de visitas ao estado **j** é pelo menos 1 se e só se o tempo da primeira visita a **j** é finito.

Então, $P_i(N(j) \geq 1) = P_i(T_j < +\infty) = f_{ij}^*$.

Teorema 2.10. *Seja $i \leftrightarrow j$. Então*

(a) j é transiente, se e só se, i é

(b) j é recorrente, se e só se, i é recorrente.

(c) No caso b), temos que $f_{ij}^* = 1 = f_{ji}^*$

Demonstração: Vamos mostrar o item (b). Suponha que j é recorrente. Seja i tal que $i \leftrightarrow j$. Vamos mostrar que i é recorrente.

Sejam n e m tais que $P^n(i,j) > 0$ e $P^m(j,i) > 0$.

Dado $t \geq 0$

$$P^{n+m+t}(i,i) = \sum_{k,s} P^n(i,k)\, P^t(k,s)\, P^m(s,i) \geq P^n(i,j)\, P^t(j,j)\, P^m(j,i) \geq 0.$$

Logo,

$$\sum_{t=0}^{\infty} P^{n+m+t}(i,i) \geq \sum_{t=0}^{\infty} P^n(i,j)\, P^t(j,j)\, P^m(j,i) = P^n(i,j)\, P^m(j,i) \sum_{t=0}^{\infty} P^t(j,j)\,.$$

Pelo Teorema 2.9, temos que $\sum_{t=0}^{\infty} P^t(j,j) = \infty$. Sendo assim, pelo mesmo Teorema 2.9 temos que $\sum_{t=0}^{\infty} P^{n+m}(i,i) = \infty$ e portanto i é recorrente.

Podemos intercambiar i e j no raciocínio acima. Isto prova (b).

Segue de imediato o item (a).

Isto prova a) e b).

Vamos agora mostrar o item c). Suponha que $i \to j$ e $f^*_{ji} < 1$. Vamos mostrar que então $f^*_{ii} < 1$.

Seja m o menor número natural tal que $P^m_{ij} > 0$. Logo existe uma sequência $i_1, i_2, .., i_{m-1}$ tal que

$$P(i,i_1)P(i_1,i_2)...P(i_{m-1},j) \,>\, 0.$$

Como m é mínimo, então $i_r \neq j$ para todo $r \in \{1,2,...,m-1\}$.

Se $f^*_{ij} = \sum_{n\geq 1} f^n_{ij} < 1$ então

$$P_i(T_i = \infty) \geq P(i,i_1)P(i_1,i_2)...P(i_{m-1},j)\,(1 - f^*_{j\,i}) \,>\, 0.$$

Observe que $(1 - f^*_{j\,i})$ é a probabilidade de começando em j não voltar em tempo finito a i. Note tambem que conjunto dos caminhos $w \in S^{\mathbb{N}}$ da forma $w = (i, i_1, i_2, ..., i_{m-1}, j, w_{m+2}, w_{m+3}, ...)$, onde $(j, w_{m+2}, w_{m+3}, ...)$ é um caminho que não volta a i em tempo finito, esta contido em $\{w \,|\, T_i(w) = \infty\}$.

Desta forma, $f^*_{ii} < 1$.

Portanto, se $f^*_{ii} = 1$, então $f^*_{ji} = 1$.

O resultado para f^*_{ij} segue da mesma maneira.

$\square$

O que o teorema acima está dizendo é que se uma classe contém um estado recorrente então todos os estados nesta classe são recorrentes. Sendo assim, para uma cadeia irredutível, se um estado é recorrente, todos os outros estados também o são.

Exemplo 2.31. Passeio Aleatório Simples em $\mathbb{Z}$ indexado por $0 < p < 1$.

$P(i, i+1) = p$, $P(i, i-1) = 1-p$, onde $0 < p < 1$. Ainda $P(i, j) = 0$ para todo $j \neq i-1$ ou $j \neq i+1$.

É fácil ver que a cadeia é irredutível. Resta saber se é transitória ou recorrente. Basta analisar o estado $i = 0$!!

Para isto, note que $P^n(0,0) = 0$, se **n** é ímpar.

Ainda, $P^{2m}(0,0) = \begin{pmatrix} 2m \\ m \end{pmatrix} p^m (1-p)^m$.

É fácil mostrar isto: considere um caminho amostral qualquer w; note que começando em 0 (ou seja $w_0 = 0$), para voltar a 0 exatamente em tempo $2m$, temos que, no total, m vezes o caminho vai para cima (sobe 1) e m vezes vai para baixo (desce 1).

O número $C_{2m}^m = \begin{pmatrix} 2m \\ m \end{pmatrix}$ descreve o número de possibilidades de se obter m eventos cara e m eventos coroa quando se lança uma moeda $2m$ vezes. As probabilidades p^m e $(1-p)^m$ surgem dos valores da matriz de transição.

A fórmula de **Aproximação de Stirling** afirma que: $n! \cong \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$.

Portanto,

$$P^{2m}(0,0) = \frac{(2m)!}{m!m!}[p(1-p)]^m \cong \frac{1}{\sqrt{2\pi}} \frac{(2m)^{2m+\frac{1}{2}}}{m^{2m+1}} [p(1-p)]^m \frac{e^{-2m}}{e^{-m}e^{-m}} =$$

$$\frac{1}{\sqrt{2\pi}} 2^{2m+1/2} \frac{m^{2m}.m^{\frac{1}{2}}}{m^{2m}.m} [p(1-p)]^m = \frac{2^{2m}}{\sqrt{\pi}} \frac{1}{\sqrt{m}} [p(1-p)]^m = \frac{1}{\sqrt{\pi m}} 4^m [p(1-p)]^m.$$

Se $p = 1/2$: $\sum_{m \geq 1} P^{2m}(0,0) = \sum_{m \geq 1} \frac{1}{\sqrt{\pi m}} 4^m \frac{1}{4^m} = \frac{1}{\sqrt{\pi}} \sum_{m \geq 1} \frac{1}{m^{1/2}} = +\infty$ então, neste caso, 0 é recorrente, pois $E_0(N(0)) = \sum_{n \geq 1} P^n(0,0) = \sum_{m \geq 1} P^{2m}(0,0) = \infty$. Sendo assim, $f_{00}^* = 1$, e como todos os estados se comunicam, pelo último Teorema, temos que $f_{ii}^* = 1$ e $f_{ij}^* = 1$, para todo $i, j \in \mathbb{N}$.

Se $p \neq 1/2$: então $p(1-p) < 1/4$. Assim, a soma em m com termos $\frac{1}{\sqrt{\pi m}} 4^m [p(1-p)]^m$ converge (teste da razão). Logo, $\sum_{m \geq 1} P^{2m}(0,0) < +\infty$, e então, neste caso, todos os estados são transitórios.

$\diamondsuit$

Proposição 2.14. *(i)*$P_i(N(j) \geq 2) = f^*_{ij} f^*_{jj}$;
(ii)$P_i(N(j) \geq n) = f^*_{ij}[f^*_{jj}]^{n-1}$;
(iii)$P_i(N(j) = n) = f^*_{ij}[f^*_{jj}]^{n-1}(1 - f^*_{jj})$.

Demonstração:

(i)

$$P_i(N(j) \geq 2) = \sum_{m,n \geq 1} P_i \begin{bmatrix} \text{a } 1^{\text{a}} \text{ visita a } \mathbf{j} \text{ ocorre na etapa m} \\ \text{e a } 2^{\text{a}} \text{ visita a } \mathbf{j} \text{ ocorre na etapa m + n} \end{bmatrix} =$$

$$= \sum_{m,n \geq 1} P_i[X_r \neq j, r \in \{1, 2, ..., m-1\}, X_m = j, X_t \neq j,$$

$$t \in \{m+1, ..., m+n-1\}, X_{m+n} = j\,] =$$

$$= \sum_{m,n \geq 1} P_i[X_r \neq j, r \in \{1, 2, \ldots, m-1\}, X_m = j] \times P_i[X_t \neq j,$$

$$t \in \{m+1, m+2, \ldots, m+n-1\}, X_{m+n} = j | X_r \neq j, r \in \{1, 2, ..., m-1\}, X_m = j\,] =$$

$$\sum_{m,n \geq 1} P_i(T_j = m) \times P(X_t \neq j,$$

$$t \in \{m+1, \ldots, m+n-1\}, X_{m+n} = j | X_0 = i, X_m = j$$

$$\text{e } \; X_r \neq j, \; r \in \{1, \ldots, m-1\}) =$$

$$\sum_{m,n \geq 1} P_i(T_j = m) \times P(X_t \neq j, \, t \in \{m+1, \ldots, m+n-1\}, X_{m+n} = j | X_m \neq j) =$$

$$\sum_{m,n\geq 1} P_i(T_j = m) \times P(X_t \neq j,\, t \in \{1,\ldots,n-1\}, X_n = j | X_0 = j) =$$

$$\sum_{m,n\geq 1} P_i(T_j = m) \cdot P_j(T_j = n) = \sum_{m,n\geq 1} f_{ij}^m f_{jj}^n = f_{ij}^* \cdot f_{jj}^*.$$

ii) demonstração análoga ao caso (i);

iii) $P_i(N(j) = n) = P_i(N(j) \geq n) - P_i(N(j) \geq n+1)$.

□

Observação:

1)$[N(j) \geq 1] \Leftrightarrow [T_j < +\infty]$.

2)$P_j(N(j) \geq n) = [f_{jj}^*]^n = [P_j(T_j < +\infty)]^n$.

Proposição 2.15. *Seja* $X : \Omega \to S \subset \mathbb{N}$. *Seja* $i \in S$, *então* $\mathbb{E}_i(X) = \sum_{m\geq 1} P_i(X \geq m)$.

Demonstração: Ora,

$$\mathbb{E}_i(X) = \sum_{n\geq 1} n\, P_i(X = n) = \sum_{n\geq 1} \underbrace{P_i(X = n) + P_i(X = n) + ... + P_i(X = n)}_{n \text{ vezes}}.$$

Como todos os termos envolvidos são positivos podemos alterar a ordem de soma, e assim se obtém

$$\mathbb{E}_i(X) = \sum_{m\geq 1} (\, P_i(X = m) + P_i(X = m+1) + ... + P_i(X = m+n) + ...\,) =$$

$$\sum_{m\geq 1} \sum_{n\geq 0} P_i(X = m+n) = \sum_{m\geq 1} P_i(X \geq m).$$

□

Teorema 2.11. *I) Se j é um* **estado transitório**, *ou seja $f^*_{jj} < 1$, então*

(1i) $P_i(N(j) < +\infty) = 1$, $\forall i \in S$;

(1ii) $\mathbb{E}_i(N(j)) = \frac{f^*_{ij}}{1-f^*_{jj}}$, $\forall i \in S$ *(neste caso $P^n(i,j) \to 0$ pela Prop 2.13).*

II) Se j é um **estado recorrente**, *então*

(2i) $P_j(N(j) = +\infty) = 1$;

(2ii) $P_i(N(j) = +\infty) = f^*_{ij}$, $\forall i \in S$;

(2iii) $\mathbb{E}_i(N(j)) = \begin{cases} +\infty, se\, f^*_{ij} > 0, \forall i \in S \\ 0, se\, f^*_{ij} = 0. \end{cases}$

Demonstração:

(1i) $P_i(N(j) = +\infty) = \lim_{n\to\infty} P_i(N(j) \geq n) = \lim_{n\to\infty} f^*_{ij}[f^*_{jj}]^{n-1} = 0$ pois $f^*_{jj} < 1$.

Logo, $P_i(N(j) < +\infty) = 1$.

(1ii) Utilizando a última proposição temos

$$P_i(N(j) = +\infty) = 0 \Rightarrow \mathbb{E}_i(N(j)) = \sum_{n\geq 1} P_i(N(j) \geq n) =$$

$$\sum_{n\geq 1} f^*_{ij}[f^*_{jj}]^{n-1} = f^*_{ij} \sum_{n\geq 1} [f^*_{jj}]^{n-1} = f^*_{ij} \frac{1}{1-f^*_{jj}} = \frac{f^*_{ij}}{1-f^*_{jj}} \; .$$

(2i)$P_j(N(j) = +\infty) = \lim_{n\to+\infty} P_j(N(j) \geq n) = \lim_{n\to\infty} f^*_{jj}[f^*_{jj}]^{n-1} = \lim_{n\to\infty} [f^*_{jj}]^n = lim_{n\to\infty} 1 = 1$, pois $f^*_{jj} = 1$. Logo, $P_j(N(j) = +\infty) = 1$.

(2ii)$P_i(N(j) = +\infty) = \lim_{n\to+\infty} P_i(N(j) \geq n) = \lim_{n\to\infty} f^*_{ij}[f^*_{jj}]^{n-1} = \lim_{n\to\infty} f^*_{ij} 1 = f^*_{ij}, \forall i \in S$.

(2iii)$\mathbb{E}_i(N(j)) = \sum_{n\geq 1} P^n(i,j) = \begin{cases} +\infty, \text{ se } P_i(N(j) = +\infty) > 0 \\ 0, \text{ se } P_i(N(j) = +\infty) = 0. \end{cases}$

□

Algumas considerações sobre o último Teorema:

a) se **j** é um **estado transitório** e $i \in S$, então existe um conjunto $K \subset S^{\mathbb{N}}$ tal que $P_i(K) = 1$ e para todo $\omega = (w_t)_{t\in\mathbb{N}} \in K$, o número de ocorrências $w_t = \mathbf{j}$ é finito. O número médio de visitas a **j** também é finito.

b) se **j** é um **estado recorrente** então existe um conjunto $K \subset S^{\mathbb{N}}$ tal que $P_j(K) = 1$ e para todo $\omega = (w_t)_{t\in\mathbb{N}} \in K$, o número de ocorrências $w_t = \mathbf{j}$ é infinito.

Dado S e $\mathcal{P}$, denotamos por S_R o conjunto dos elementos recorrentes e por S_T o conjunto dos elementos transientes.

Teorema 2.12. *Se uma cadeia de Markov irredutível possui um vetor invariante $\pi \in \Sigma$, então a cadeia é recorrente.*

Demonstração: Suponha que a cadeia não é recorrente (ou seja, possua elementos não recorrentes). Então, $\sum_n P_{ij}^n < \infty$ para todo i, j em S.

Em particular, para cada $i, j \in S$ temos que

$$\lim_{n\to\infty} P_{ij}^n = 0.$$

Como π é invariante, temos que

$$\pi_j = \sum_{i\in S} \pi_i \, P_{ij}^n,$$

para qualquer $n > 0$.

Segue da próxima proposição (intercambiar somatório e limite) que para cada j

$$\pi_j = \lim_{n\to\infty} \sum_{i=1}^{\infty} \pi_i \, P_{ij}^n = \sum_{i=1}^{\infty} \pi_i \, \lim_{n\to\infty} P_{ij}^n = \sum_{i=1}^{\infty} \pi_i \, 0 \, = 0.$$

Ora,

$$\sum_j \pi_j = 1.$$

Quando $S = \{1, 2, ..., d\}$ é finito, temos

$$1 = \sum_{j\in S} \pi_j = \lim_{n\to\infty} \sum_{j\in S} \sum_{i\in S} \pi_i \, P_{ij}^n = \sum_{j\in S} \sum_{i\in S} \pi_i \, \lim_{n\to\infty} P_{ij}^n = 0,$$

contradição. Quando S é infinito é preciso ter cuidado pois estamos lidando com um limite duplo em i e em n. Precisamos lançar mão do seguinte resultado:

Proposição 2.16. *Seja uma coleção de números reais* $\{x_{n,i}\}$ *com* $n, i \in \mathbb{N}$. *Suponha que*

$$\lim_{n\to\infty} x_{n,i} = a_i,$$

para cada i *fixo, onde* $a_i \in \mathbb{R}$. *Suponha ainda que*

$$\sum_{i=1}^{\infty} \sup_n |x_{n,i}| < \infty.$$

Então,

$$\lim_{n\to\infty} \sum_{i=1}^{\infty} x_{n,i} = \sum_{i=1}^{\infty} a_i.$$

Não apresentaremos a prova de tal proposição (uma prova elemementar aparece no apêndice A4 de [Ro]). Este resultado também pode ser obtido a partir do Teorema da convergência dominada da Teoria da Medida [Fe].

Seja agora j fixo, $a_i = 0$ e $x_{n,i} = \pi_i\, P^n_{ij} \leq \pi_i\, 1 = \pi_i$, assim

$$\pi_j = \lim_{n\to\infty} \sum_{i=1}^{\infty} \pi_i\, P^n_{ij} = \sum_{i=1}^{\infty} 0 = 0.$$

Logo, todo $\pi_j = 0$, o que contraria $\sum_{j=1}^{\infty} \pi_j = 1$.

□

Exemplo 2.32. Considere $S = \{0, 1, 2, 3, 4, 5\}$ e a C.M. com matriz de transição dada por

$$\mathcal{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 1/2 & 1/4 & 0 & 0 & 0 \\ 0 & 1/5 & 2/5 & 1/5 & 0 & 1/5 \\ 0 & 0 & 0 & 1/6 & 1/3 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1/4 & 0 & 3/4 \end{pmatrix}.$$

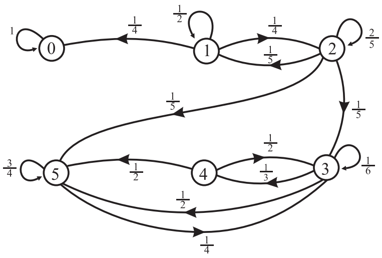


Figura 2.7:

Olhando o grafo associado a $\mathcal{P}$, que é exibido na Figura 2.7, pode-se facilmente descobrir as classes: basta seguir sucessivamente as setas em que existe passagem.

$C_0 = \{0\}$ é recorrente pois **0** é absorvente,

$C_1 = \{1, 2\}$ pois $1 \leftrightarrow 2$,

$C_2 = \{3, 4, 5\}$ pois $3 \leftrightarrow 4$ e $3 \leftrightarrow 5$.

Desejamos calcular f^*_{11}. Observe que para um dado $\omega = (w_t)$, se $w_{t_0} \in \{3, 4, 5\}$ para algum t_0 então $w_t \in \{3, 4, 5\}$ para $t > t_0$ e o caminho não retorna mais a 1 (com P-probabilidade 1). Logo no cálculo de f^*_{11} não entram os caminhos que passam por $\{3, 4, 5\}$.

$f^*_{11} = \sum_{n\geq 1} f^n_{11} = 0,5 + \frac{1}{4}\frac{1}{5} + \frac{1}{4}\frac{2}{5}\frac{1}{5} + \frac{1}{4}\frac{2}{5}^2\frac{1}{5} + \ldots = 0,5 + \frac{1}{20}\sum_{n\geq 0}(\frac{2}{5})^n = 0,5 + \frac{1}{20}\frac{1}{1-\frac{2}{5}} = 0,5 + \frac{1}{20}\frac{5}{3} = 0,5 + \frac{1}{12} = \frac{7}{12} < 1,$

Isto porque os ω a serem considerados, começam em 1, saltam para 2 e ficam lá até retornar pela primeira vez a 1.

Logo, $C_1 = \{1, 2\}$ é transitório. $\mathbb{E}_1(N(1)) = \frac{f^*_{11}}{1-f^*_{11}} = \frac{7/12}{5/12} = \frac{7}{5}; P_1(N(2) < +\infty) = 1$.

Observe que

$f^*_{33} = \sum_{n\geq 1} f^n_{33} = \frac{1}{6} + (\frac{1}{3}\frac{1}{2} + \frac{1}{2}\frac{1}{4}) + (\frac{1}{3}\frac{1}{2}\frac{1}{4} + \frac{1}{2}\frac{3}{4}\frac{1}{4}) + (\frac{1}{3}\frac{1}{2}\frac{3}{4}\frac{1}{4} + \frac{1}{2}\frac{3^2}{4}\frac{1}{4}) + \ldots = \frac{1}{6} + \frac{1}{6} + \frac{1}{24}\sum_{n\geq 0}(\frac{3}{4})^n + \frac{1}{8}\sum_{n\geq 0}(\frac{3}{4})^n = \frac{1}{3} + \frac{1}{24}4 + \frac{1}{8}4 = \frac{1}{3} + \frac{1}{6} + \frac{1}{2} = 1$

Logo, $C_2 = \{3, 4, 5\}$ é recorrente. $\mathbb{E}_3(N(3)) = +\infty$ já que $f^*_{33} = 1 > 0$;

$P_3(N(3) = +\infty) = 1$ e $P_2(N(3) = +\infty) = f^*_{23}$.

Temos então: estados Recorrentes $=\{0\} \cup \{3,4,5\} = S_R$ e estados Transientes $= \{1,2\} = S_T$. Sendo assim, $S = S_R \bigcup S_T$.

$\diamondsuit$

Definição 2.19. *Uma classe C é dita* **fechada** *se* $\sum_{j\in C} P(i,j) = 1$, $\forall i \in C$.

Observe no Exemplo 2.32 que $\sum_{j\in C_1=\{1,2\}} P(1,j) = P(1,1) + P(1,2) = \frac{1}{2} + \frac{1}{4} \neq 1$. Logo, C_1 não é fechada. Isto mostra que uma classe irredutível, pode não ser fechada.

Ainda, $\sum_{j\in C_2} P(3,j) = P(3,3) + P(3,4) + P(3,5) = \frac{1}{6} + \frac{1}{3} + \frac{1}{2} = 1$; $\sum_{j\in C_2} P(4,j) = P(4,3) + P(4,4) + P(4,5) = \frac{1}{2} + 0 + \frac{1}{2} = 1$ e $\sum_{j\in C_2} P(5,j) = P(5,3) + P(5,4) + P(5,5) = \frac{1}{4} + 0 + \frac{3}{4} = 1$.

Portanto, C_2 é fechada.

Mostraremos no Teorema 2.16 que toda classe irredutível recorrente é fechada.

Ainda C_0 é fechada pois 0 é absorvente.

Exercício: Seja $(X_n)_{n\geq 0}$ uma C.M. com $S = \{0,1,2,3,4\}$, e matriz de transição dada por

$$\mathcal{P} = \begin{pmatrix} 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 & 0 & 0 \\ \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \end{pmatrix}.$$

Mostre que a classe de equivalência $\mathcal{C}_1 = \{0,1,3,4\}$ não é fechada.

Observação: Seja $S \subset \mathbb{N}$ conjunto e $\mathcal{P}$ matriz estocástica. Suponha que $C \subset S$ é classe fechada. Neste caso fica bem definida uma nova matriz estocástica $\mathcal{P}_1$

do tipo $\#C$ por $\#C$ que é obtida a partir de $\mathcal{P}$ restringindo as entradas ij tais que $i, j \in S$.

No exemplo anterior, podemos restringir a matriz $\mathcal{P}$ sobre $\{0, 1, 2, 3, 4, 5\}$ à classe fechada $C_2 = \{3, 4, 5\}$. Obtemos assim a matriz

$$\mathcal{P}_1 = \begin{pmatrix} 1/6 & 1/3 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/4 & 0 & 3/4 \end{pmatrix}.$$

Uma classe fechada existe como se fosse independente do sistema maior, no presente caso sobre $S = \{0, 1, 2, 3, 4, 5\}$.

Muitas vezes estaremos interessados em saber o comportamento de caminhos típicos w a longo prazo, ou seja, o que acontece com w_n quando $n \to \infty$?

O conjunto dos caminhos w que em algum tempo entram em C_2, dali não saem (para tempo futuro) e seu comportamento a longo prazo, como será discutido em breve, será descrito pela matriz $\mathcal{P}_1$.

Por outro lado, os caminhos w que eventualmente atingem em algum tempo o estado 0 (que é absorvente), ficam, a partir deste tempo, para sempre em 0. O comportamento a longo prazo, neste caso, é trivial.

Finalmente, os estados $1, 2$, por serem transientes não serão observados a longo prazo. Isto será formalmente descrito em um teorema que vamos abordar em breve que afirma que, neste presente exemplo, vale para todo $i \in S$ fixado,

$$P_i(\omega = (w_t)_{t \in \mathbb{N}} | \text{ existe } t_0 \in \mathbb{N} \text{ tal que } \forall t > t_0 \text{ temos } w_t \neq 1 \text{ e } w_t \neq 2) = 1.$$

Se fixarmos uma probabilidade inicial π e considerarmos a correspondente probabilidade P, também vale que

$$P(\omega = (w_t)_{t \in \mathbb{N}} | \text{ existe } t_0 \in \mathbb{N} \text{ tal que } \forall t > t_0 \text{ temos } w_t \neq 1 \text{ e } w_t \neq 2) = 1.$$

Exemplo 2.33. Considere $(X_n), n \geq 0$ a C.M. com $S = \mathbb{N}$ e a matriz de transição $\mathcal{P}$ dada por

$$\mathcal{P} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & \dots \\ 1/2 & 0 & 1/2 & 0 & 0 & \dots \\ 1/2 & 0 & 0 & 1/2 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}.$$

a) Vamos calcular f_{00}^n e f_{00}^*.

b) A seguir vamos determinar f_{ii}^* para $i \neq 0$.

Vamos começar o procedimento de cálculo.

a) Desejamos primeiro calcular $f_{00}^* = \sum_{n\geq 1} f_{00}^n = \sum_{n\geq 1} P(T_0 = n|X_0 = 0)$.

Observe que

$f_{00}^1 = P(T_0 = 1|X_0 = 0) = 1/2 = (0,5)$

$f_{00}^2 = P(T_0 = 2|X_0 = 0) = 1/2 \cdot 1/2 = (0,5)^2$

$f_{00}^3 = P(T_0 = 3|X_0 = 0) = (0,5)^3$.

Logo, $f_{00}^* = \sum_{n\geq 1} f_{00}^n = \sum_{n\geq 1} (0,5)^n = \frac{0,5}{1-0,5} = 1$.

Desta forma, o estado 0 é recorrente.

b) Vimos em prova que, quando $0 < p < 1$, $C_0 = S$ é irredutível. Assim, todos os estados em S são recorrentes.

Observe que aqui $p = 0,5$ e assim

$$f_{11}^1 = P(T_1 = 1|X_0 = 1) = 0,$$

$$f_{11}^2 = P(T_1 = 2|X_0 = 1) = (0,5)(0,5),$$

$$f_{11}^3 = P(T_1 = 3|X_0 = 1) = (0,5)(0,5)(0,5) + (0,5)(0,5)(0,5) = 2(0,5)^3,$$

$$f_{11}^4 = P(T_1 = 4|X_0 = 1) = 3(0,5)^4.$$

Logo, $f_{11}^n = (n-1)(0,5)^n$, $n \geq 2$.

Portanto,

$$f_{11}^* = \sum_{n\geq 2} (n-1)(0,5)^n = \sum_{n\geq 2} n(0,5)^n - \sum_{n\geq 2} (0,5)^n =$$

$$= (0,5)\sum_{n\geq 2} n(0,5)^{n-1} - \sum_{n\geq 2}(0,5)^n = (0,5)(\sum_{n\geq 1}(0,5)^n)' - (\sum_{n\geq 0}(0,5)^n - 1 - 0,5) =$$

$$0,5(\frac{1}{(1-0,5)^2} - 1) - (\frac{1}{1-1/2} - 1,5) = 0,5(\frac{1}{0,25} - 1) - (2 - 1,5) = 1,5 - 0,5 = 1$$

Conclusão: $f^*_{ii} = 1$, $\forall i \neq 0$.

$\diamondsuit$

Teorema 2.13. *Toda cadeia de Markov em que S é finito possui ao menos um estado recorrente.*

Demonstração: Toda cadeia de Markov $\mathcal{P}$ com S finito possui uma autovetor p (associado ao autovalor 1) conforme Teorema 2.4, logo, se for irredutível, neste caso, cada estado é recorrente. Toda cadeia de Markov P com S finito vai possuir um subconjunto fechado e irredutíveis C_l (conforme futuro Teorema 2.18). Logo restringindo a cadeia P a tal subconjunto C_l obteremos também pelo mesmo Teorema 2.4, neste caso, a existência de estados recorrentes. Concluímos que se $\mathcal{P}$ é finita então sempre existem estados recorrentes.

$\square$

Note que no passeio aleatório com $p \neq 1/2$ (Exemplo 2.13) todos os estados são transientes (assim o teorema acima não se aplica).

Exemplo 2.34. Considere a C.M. (X_n), $n \geq 0$, sobre $S = \{0, 1, 2, 3, 4\}$, cuja matriz de transição é dada por

$$\mathcal{P} = \begin{pmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1/9 & 0 & 3/9 & 0 & 5/9 \\ 1/9 & 0 & 0 & 5/9 & 3/9 \end{pmatrix}.$$

Este caso já foi considerado antes no Exemplo 2.18. Note que 1 passa para 2 e depois não volta mais a 1.

É fácil ver que existem apenas duas classes: $C_0 = \{0, 2, 3, 4\}$ e $C_1 = \{1\}$, que C_0 é classe recorrente e C_1 é classe transiente.

Vamos analisar se os estados na classe C_0 são recorrentes.

Podemos considerar a matriz acima restrita ao conjunto C_0. Obtemos assim a matriz estocástica

$$\tilde{\mathcal{P}} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 1 & 0 & 0 & 0 \\ 1/9 & 3/9 & 0 & 5/9 \\ 1/9 & 0 & 5/9 & 3/9 \end{pmatrix}.$$

Já sabemos do Exemplo 2.18 que esta matriz tem o vetor

$$\pi = (\frac{58}{204}, \frac{47}{204}, \frac{54}{204}, \frac{45}{204})$$

como vetor estacionário para $\tilde{\mathcal{P}}$.

Sendo assim, pelo Teorema 2.12 temos que todo estado em $\{0, 2, 3, 4\}$ é recorrente.

$\diamondsuit$

Proposição 2.17. *Se j é um estado transiente, então $\lim_{n\to\infty} P^n(i, j) = 0$, $\forall i \in S$ fixo.*

Demonstração: $E_i(N(j)) = \sum_{n\geq 1} P^n(i, j) < \infty$, ou seja, a série é absolutamente convergente, e então, o termo geral da série converge a zero, i.é, $\forall i \in S$.

$$\lim_{n\to\infty} P^n(i, j) = 0.$$

$\square$

Logo, se j é **transiente**, então a probabilidade $P_i(X_n = j)$ de que o processo esteja no estado **j** converge a zero, qualquer que tenha sido o estado inicial i fixado.

O próximo resultado já foi obtido em um teorema anterior, mas vamos apresentar a seguir uma outra prova.

Teorema 2.14. *Se j é transiente então para $i \in S$ qualquer vale que*

$$P_i(\omega = (w_t)_{t\in\mathbb{N}} \; : \; \textit{existe } t_0 \in \mathbb{N} \textit{ tal que para } t > t_0 \textit{ temos } w_t \neq j \,) \;=\; 1.$$

Demonstração: Ora,

$$\cap_{m=1}^{\infty} \cup_{n=m}^{\infty} \{w_n = j\} =$$

$$\Omega - \{\omega = (w_t)_{t\in\mathbb{N}} \; : \; \text{existe } t_0 \in \mathbb{N} \text{ tal que para } t > t_0 \text{ temos } w_t \neq j\}.$$

Ainda, para m fixo

$$P_i(\cup_{n=m}^{\infty}\{w_n = j\}) \leq \sum_{n=m}^{\infty} P_i(X_n = j).$$

Como,

$$\sum_{n=1}^{\infty} P_i(X_n = j) < \infty,$$

então, para cada m a sequência

$$a_m = \sum_{n=m}^{\infty} P_i(X_n = j),$$

converge a zero.

Logo,

$$P_i(\cap_{m=1}^{\infty} \cup_{n=m}^{\infty} \{w_n = j\}) = \lim_{m\to\infty} P_i(\cup_{n=m}^{\infty}\{w_n = j\}) \leq$$

$$\lim_{m\to\infty} \sum_{n=m}^{\infty} P_i(X_n = j) = 0,$$

pois a sequência de conjuntos é decrescente.

□

Note que no caso em que S é infinito, dado um vetor inicial de probabilidade π e uma matriz $\mathcal{P}$ não é necessariamente verdade que a probabilidade P associada satisfaça

$$P(\omega=(w_t)_{t\in\mathbb{N}}\,:\,\text{existe } t_0\in\mathbb{N} \text{ tal que para } t>t_0 \text{ temos } w_t\neq j\,)\,=\,1.$$

Teorema 2.15. *Seja S finito, dado um vetor de probabilidade π e a matriz de transição $\mathcal{P}$. Então*

$$P(\omega=(w_t)_{t\in\mathbb{N}}\,:\,\textit{existe } t_0\in\mathbb{N} \textit{ tal que para } t>t_0 \textit{ temos } w_t\notin S_T\,)\,=\,1.$$

Demonstração: Como para cada $i\in S$ e $j\in S_T$ vale que $\sum_{n=1}^{\infty} P(X_n=j|X_0=i)<\infty$, então

$$\sum_{j\in S_T}\sum_{i\in S}\pi_i\sum_{n=1}^{\infty} P_i(X_n=j)<\infty.$$

Logo,

$$\sum_{n=1}^{\infty} P(X_n\in S_T)=\sum_{i\in S}\pi_i\sum_{n=1}^{\infty} P_i(X_n\in S_T)<\infty$$

Para cada m considere

$$a_m=\sum_{n=m}^{\infty} P(X_n\in S_T),$$

então a_m converge a zero.

O resto da prova é idêntico ao teorema anterior.

□

Teorema 2.16 (Teorema da Decomposição em Peças Irredutíveis). *Toda classe irredutível e recorrente é fechada. Ainda, considere a relação de equivalência $i\leftrightarrow j$, e denote por S_T o conjunto dos elementos transientes. Então S pode ser decomposto em*

$$S=\,S_T\,\cup\,C_1\,\cup\,C_2\ldots\cup C_n\,\cup\ldots$$

onde cada classe C_r é fechada, irredutível e recorrente.

Demonstração: Sejam as classes de equivalência obtidas pela relação de equivalência $i \leftrightarrow j$ e denote aquelas com elementos recorrentes por C_r. Os elementos transientes ficam todos colocados no conjunto S_T.

Cada classe C_r é irredutível, logo basta mostrar que toda classe irredutível e recorrente é fechada.

Suponha que $i \in C_r$ e $j \notin C_r$. Se por absurdo $P(i,j) > 0$, então como não existe n tal que $P^n(j,i) > 0$, temos que

$$P(X_n \neq i \text{ para todo } n > 1 \,|\, X_0 = i) =$$

$$\frac{P(X_n \neq i \text{ para todo } n > 1\,,\, X_0 = i)}{P(X_0 = i)} =$$

$$\frac{P(X_n \neq i \text{ para todo } n > 1\,, X_1 = j\,,\, X_0 = i)}{P(X_0 = i)} +$$

$$\frac{P(X_n \neq i \text{ para todo } n > 1\,, X_1 \neq j\,,\, X_0 = i)}{P(X_0 = i)} \geq$$

$$\frac{P(X_n \neq i \text{ para todo } n > 1\,, X_1 = j\,,\, X_0 = i)}{P(X_0 = i)} =$$

$$\frac{P(X_n \neq i \text{ para todo } n > 1\,, X_1 = j\,,\, X_0 = i)}{P(X_0 = i)} +$$

$$\frac{P(X_n = i \text{ para algum } n > 1\,, X_1 = j\,,\, X_0 = i)}{P(X_0 = i)} =$$

$$= \frac{P(X_1 = j\,,\, X_0 = i)}{P(X_0 = i)}$$

$$= P(X_1 = j \,|\, X_0 = i) \,=\, P_{ij} \,>\, 0,$$

isto é, o processo iniciando em i pode nunca voltar a i com probabilidade positiva.

Chegamos assim a uma contradição com o fato que i é recorrente.

Acima usamos o fato que

$$\frac{P(X_n = i \,\text{para algum}\; n > 1\,, X_1 = j\,,\, X_0 = i)}{P(X_0 = i)} \leq$$

$$\frac{\sum_{n>1} P(X_n = i\,, X_1 = j\,,\, X_0 = i)}{P(X_0 = i)} \;=\; 0,$$

que segue da hipótese de que não existe m tal que $P^m(j,i) > 0$.

□

Teorema 2.17. *Com probabilidade* 1 *os caminhos amostrais* ω *são tais que se* $w_{t_0} \in C_r$ *para algum* t_0 *e* r*, então* $w_t \in C_r$ *para todo* $t \geq t_0$*.*

Este resultado será demonstrado mais tarde.

Proposição 2.18. *O conjunto* S_T *de estados transientes em uma C.M. finita* $\mathcal{P}$ *é não fechado.*

Demonstração: Suponha por absurdo que S_T é fechada.

Dada a matriz $\mathcal{P}$ e o conjunto fechado S_T podemos considerar uma nova matriz $\mathcal{P}_1$ que é a restrição de $\mathcal{P}$ a S_T. Como vimos antes toda cadeia de Markov finita possui um estado recorrente, e isto é uma contradição pois $\mathcal{P}_1$ é finita.

□

Observações:

I) Toda C.M. com espaços de estados finito possui pelo menos um estado recorrente.

II) Seja C um conjunto de estados finito, fechado e irredutível. Então, todo estado em C é recorrente.

III) Se uma C.M. é tal que um $\omega = (w_t)$ começa no conjunto de estados transientes T (isto é, $w_0 \in S_T$), então ou ela permanece em S_T para sempre ou em algum momento ela entra em um dos conjuntos $C_r \subset S - S_T$ irredutíveis (isto é, existe t_0 tal que $w_t \in C_r$ para $t \geq t_0$) e permanece, a partir deste momento, visitando todos os estados em C_r infinitas vezes. Se S é finito em algum tempo t_0 finito o valor w_{t_0} vai atingir algum elemento em $S - S_T$. Se S é infinito isto pode não ocorrer.

Segue dos dois últimos resultados acima que quando S é finito, dada uma matriz $\mathcal{P}$ podemos restringí-la a cada C_r irredutível e analisar individualmente a matriz estocástica sobre o conjunto C_r. Deste modo a hipótese que faremos em vários resultados que seguem à matrizes irredutíveis não é nenhuma grande restrição. Isto porque para uma probabilidade qualquer P fixada, com probabilidade 1 os caminhos ω são tais que para t_0 grande $w_{t_0} \notin S_T$ e assim w_{t_0} entra em um certo conjunto irredutível C_r e portanto para $t > t_0$ temos que $w_t \in C_r$. Logo, se estamos interessados no comportamento de $\omega = (w_t)_{t \in \mathbb{N}}$ quando t é grande o que importa é o comportamento em cada C_r. Referimos a prova do próximo resultado a [I].

Teorema 2.18. *Seja S finito e considere $g(\lambda) = det\ (\lambda I - \mathcal{P})$ o polinômio característico da matriz $\mathcal{P}$. Sabemos que 1 é raíz de g pois $\lambda = 1$ é autovalor de $\mathcal{P}$ (quando S é finito sempre existe p tal que $p\mathcal{P} = p$). A multiplicidade de 1 como raíz de g é o número de classes irredutíveis fechadas de $\mathcal{P}$.*

Corolário 2.2. *Se 1 tem multiplicidade 1 para $g(\lambda) = det\ (\lambda I - \mathcal{P})$ então $\mathcal{P}$ é irredutível.*

Exemplo 2.35. A matriz

$$\mathcal{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/3 & 2/3 \end{pmatrix}$$

tem polinômio característico $g(\lambda) = (\lambda - 1)\,(\,\lambda^2 - \ 7/6\,\lambda + 1/6\,) = (\lambda - 1)^2\,(\lambda - (1/6)\,)$. Note que 1 tem multiplicidade 2 como raíz do polinômio g. Logo, existem duas classes irredutíveis.

Este resultado é muito útil pois permite detectar se $\mathcal{P}$ é irredutível apenas analisando as raízes de um polinômio de grau $d = \#S$, ou seja, analisando as raízes de $g(\lambda) = 0$.

$\diamondsuit$

2.6 Periodicidade e Aperiodicidade

No que segue, **mdc.** significa máximo divisor comum. Por exemplo, o m.d.c de $6, 8, 10, 12$ é 2.

Definição 2.20. *Seja $(X_n), n \geq 0$ uma C.M. com espaços de estados S. Seja $i \in S$.*

(a) O período de $\underline{i}$ é definido como

$$d(i) = \begin{cases} mdc\{n \geq 1 | P^n(i,i) > 0\}, \ se \ \{n \geq 1 | P^n(i,i) > 0\} \neq \emptyset \\ 0, \ se \ \forall n \geq 1, P^n(i,i) = 0 \end{cases}$$

(b) Se $d(i) > 1$, dizemos que $\underline{i}$ é periódico com período $d(i)$.

(c) Se $d(i) = 1$, dizemos que $\underline{i}$ é aperiódico.

Observamos que sempre $d(i) > 0$.

Note que se $P^r(i,i) > 0$, então, para qualquer $k > 0$ temos que $P^{r\,k}(i,i) > 0$.

Por exemplo, $P_{ii}^6 > 0$ então $P_{ii}^{12} > 0$. De fato, $P_{ii}^{12} = \sum_{j \in S} P_{ij}^6 P_{ji}^6 \geq P_{ii}^6 P_{ii}^6 > 0$. Da mesma forma $P_{ii}^{18} > 0$.

Note que se $i \in S$ é tal que $P_{ii} = P_{ii}^1 > 0$ então i tem período 1. Afinal, $d(i)$ tem que dividir 1.

Se $P(i,i) = 0$ mas $P^3(i,i) > 0$ e $P^5(i,i) > 0$, então necessariamente $d(i) = 1$.

Considere a matriz estocástica irredutível

$$\mathcal{P} = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}.$$

Note que $P_{11}^2 > 0$ e $P_{11}^3 > 0$.

Neste caso o período d de qualquer i é igual a 1 mas não existe i tal que $P(i,i) > 0$.

Considere agora a matriz estocastica irredutível

$$\mathcal{P} = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 1/4 & 1/4 \\ 1/2 & 1/2 & 0 \end{pmatrix}.$$

Neste caso o período d de qualquer i é igual a 1 e $P(2,2) > 0$. No entanto $P(1,1) = 0$.

Exemplo 2.36. A matriz

$$\mathcal{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

é irredutível e para cada $i \in \{1,2,3\}$ temos que $d(i) = 3$. Note que $f_{i,i}^* = f_{i,i}^3 = 1$, para qualquer i. O vetor $(1/3, 1/3, 1/3)$ é o único invariante para $\mathcal{P}$.

Se considerarmos a probabilidade inicial $(1,0,0)$ a probabilidade induzida pelo Processo Markoviano em $\{1,2,3\}^{\mathbb{N}}$ é

$$\frac{1}{3}\delta_{z_1} + \frac{1}{3}\delta_{z_2} + \frac{1}{3}\delta_{z_3},$$

onde

$$z_1 = (1,2,3,1,2,3,1,2,3,..),$$

$$z_2 = (2, 3, 1, 2, 3, 1, 2, 3, 1, ..)$$

e

$$z_3 = (3, 1, 2, 3, 1, 2, 3, 1, 2, ..).$$

$\diamondsuit$

Exemplo 2.37. Passeio Simples em $\mathbb{Z}$ indexado por $0 < p < 1$.

$P(i, i+1) = p$, $P(i, i-1) = 1 - p$, $\forall i$, onde $0 < p < 1$. Ainda, $P(i, j) = 0$ se $i \neq i - 1$ ou $i \neq i + 1$.

Neste caso, $P^n(i, i) = 0$, $\forall n$ ímpar e $P^n(i, i) > 0$ $\forall n$ par, $\{n \geq 1 | P^n(i, i) > 0\} = \{2, 4, 6, 8, 0, ...\}$, conforme vimos anteriormente.

O máximo divisor comum do conjunto acima é 2. Logo, 2 é o período da C.M.

$\diamondsuit$

Exemplo 2.38 (**A Cadeia de Ehrenfest**)**.** Neste caso, $S = \mathbb{N}$ e cada valor n descreve a população de um certo país ou cidade. Vamos supor que dado que a população é de i habitantes, então em uma unidade de tempo, existe a probabilidade q_i de uma pessoa morrer, r_i da população permanecer a mesma e p_i de nascer uma pessoa. Naturalmente, $q_i + r_i + p_i = 1$, e se $i = 0$, então q_0 não faz sentido e portanto $r_0 + p_0 = 1$.

Deste modo a cadeia de Markov, da forma $\mathbb{N}$ por $\mathbb{N}$, é dada por

$$\mathcal{P} = \begin{pmatrix} r_0 & p_0 & 0 & 0 & 0 & \dots \\ q_1 & r_1 & p_1 & 0 & 0 & \dots \\ 0 & q_2 & r_2 & p_2 & 0 & \dots \\ . & . & . & . & . & \dots \end{pmatrix}$$

Onde, $r_0 + p_0 = 1$, $p_0 > 0$, $r_0 \geq 0$, $q_i + r_i + p_i = 1$, $i \geq 1$, $r_i \geq 0$, $p_i > 0$, $q_i > 0$, $P(i, i+1) = p_i$, $P(i, i-1) = q_i$, $P(i, i) = r_i$, $i \geq 0$ e ainda $P(i, j) = 0$ nos outros casos.

Se algum $r_i > 0$ então i tem período 1 (é aperiódico)

É fácil ver que se $r_i = 0$ para todo i, a cadeia tem todos os estados periódicos (é também irredutível) e $d(i) = 2$. Se $r_i > 0$ para algum i, então todos os estados são aperiódicos.

$\diamondsuit$

Observação: A classe das Cadeias de Nascimento e Morte contém:

(a) Os passeios de Bernoulli $r_i = 0$, para todo $i \in S = \mathbb{N}$.

Exemplo 2.39. A cadeia de Ehrenfest é semelhante ao caso anterior, mas agora $S = 0, 1, 2, ..., d$. Neste caso a matriz $\mathcal{P}$ do tipo $d+1$ por $d+1$ é dada por

$$\mathcal{P} = \begin{pmatrix} r_0 & p_0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ q_1 & r_1 & p_1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & q_2 & r_2 & p_2 & 0 & \dots & 0 & 0 & 0 \\ . & . & . & . & . & \dots & . & . & . \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & q_d & r_d \end{pmatrix}$$

onde, $p_0 = 1$, $r_0 = 0$, $p_i = 1 - \frac{i}{d}$, $q_i = \frac{i}{d}$, $r_i = 0$, para $1 \leq i \leq d-1$, $P(0, 1) = p_0$, $P(d, d+1) = 1$ e $P(i, j) = 0$ nos outros casos.

Neste exemplo, $d(i) = 2$ para todo $i \in S$.

$\diamondsuit$

Note que se $\mathcal{P}$ for irredutível periódica então $P(i, i) = 0$ para todo $i \in S$. Vamos agora analisar algumas das principais propriedades de $d(i), i \in S$.

Proposição 2.19. *Se $i \leftrightarrow j$ então $d(i) = d(j)$.*

Demonstração: Sejam $A_i = \{n \geq 1 : P^n(i, i) > 0\}$ e $A_j = \{n \geq 1 : P^n(j, j) > 0\}$.

Precisaremos de duas propriedades.

Afirmação 1:

Seja $m \in A_i$. Então, $d(j)$ divide m.

Observe que da afirmação "$d(j)$ divide todo m" segue que, como, $d(i)$ é o máximo com tal propriedade, então vale $d(j) \leq d(i)$.

Afirmação 2:

Seja $m^* \in A_j$. Então, $d(i)$ divide m^*.

Observe que da afirmação segue que, como, $d(j)$ é o máximo com tal propriedade, então vale $d(i) \leq d(j)$.

$$d(i) \text{ divide } m^* \Rightarrow d(i) \text{ divide } d(j) \Rightarrow d(i) \leq d(j). \tag{2.2}$$

Das duas afirmações, segue que $d(i) = d(j)$.

Vamos agora provar as afirmações acima.

Prova da Afirmação 1: seja $m \in A_i$. Então $P^m(i,i) > 0$.

Como $i \leftrightarrow j$ então $\exists a > 0, b > 0$ tal que $P^a(i,j) > 0$ e $P^b(j,i) > 0$. Daí,

$$P^{b+m+a}(j,j) = \sum_k \sum_l P^b(j,k)P^m(k,l)P^a(l,j) \geq$$

$$P^b(j,i)P^m(i,i)P^a(i,j) > 0.$$

Então, $b + m + a \in A_j$ e assim $d(j)$ divide $b + m + a$.

Mas $P^m(i,i) > 0 \Rightarrow P^{2m}(i,i) > 0$. Logo pelo mesmo argumento acima, $d(j)$ divide $b + 2m + a$. Portanto, $d(j)$ divide $(b + 2m + a) - (b + m + a) = m$. Isto é, $d(j)$ divide $\underline{m}$.

A prova de Afirmação 2 é semelhante.

□

Desta forma podemos falar no período de um conjunto irredutível $C \subset S$. É o período de algum de seus elementos. Note que se um elemento numa classe é aperiódico então todos nesta classe também o são.

Diremos que uma **cadeia (uma matriz estocástica) irredutível é aperiódica** se algum estado é aperiódico

A demonstração do próximo teorema é razoavelemente complexa e o leitor pode saltá-la numa primeira leitura.

Teorema 2.19. *Seja S finito e $\mathcal{P}$ irredutível. Considere $i \in S$ e $d(i)$ o seu período (que não depende de i). Então existe $\tilde{n}_i > 0$ (dependendo apenas de i) tal que $\forall n \geq \tilde{n}_i$, $P^{n\,d(i)}(i,i) > 0$. Isto é, $n\,d(i)$ é um tempo de retorno a i (não aleatório). Ainda, existe $N > 0$ tal que $P^{m\,d}(i,i) > 0$, para todo $m > N$, independente de i.*

Demonstração:

Seja $A = A_i = \{n \in \{1,2,3,..\}; P^n(i,i) > 0\}$.

Se $A = \emptyset$, então $d(i) = 0$ e $P^{n\,d(i)}(i,i) = 1 > 0$.

Se $A \neq \emptyset$, então existe $n_0 \in A$, e assim, $P^{k\,n_0}(i,i) \geq P^{n_0}(i,i)\,...P^{n_0}(i,i) > 0$, $\forall k \in \mathbb{N}$, e assim A contém ao menos os multiplos de tal n_0.

Seja $B = \{b_1\,n_1 + b_2\,n_2 + ... + b_k\,n_k; b_1, b_2, ..., b_k \in \mathbb{Z};\ n_1, n_2, ..., n_k \in A,\ k \in \mathbb{N}\}$.

Como todo os $d(i)$ são iguais , então vamos denotá-lo por d.

Afirmação: $d\,\mathbb{Z} = B$.

De fato, primeiro note que $B \subset \mathbb{Z}$. Considere $B^+ = \{m \in B; m > 0\}$.

Então, pelo princípio da boa ordenação existe $m_0 \in B^+$ tal que, $\forall m \in B^+$, temos que $m_0 \leq m$.

Provaremos primeiro que $m_0\,\mathbb{Z} = B$.

1) Seja $z \in \mathbb{Z}$. Como $m_0 \in B$, então existem $b_1, b_2, ..., b_k \in \mathbb{Z}$; $n_1, n_2, ..., n_k \in A$, tais que $m_0 = b_1\,n_1 + b_2\,n_2 + ... + b_k\,n_k$. Logo $m_0\,z = (b_1\,z)\,n_1 + (b_2\,z)\,n_2 + ... + (b_k\,z)\,n_k \in B$. Assim, $m_0\,\mathbb{Z} \subset B$.

2) Seja $m \in B$, então existem $b_1, b_2, ..., b_k \in \mathbb{Z}$; $n_1, n_2, ..., n_k \in A$, tais que $m = b_1\,n_1 + b_2\,n_2 + ... + b_k\,n_k$. Logo, $-m = (-b_1)\,n_1 + (-b_2)\,n_2 + ... + (-b_k)\,n_k$. Desta forma, $|m| \in B^+$.

Assim, podemos assumir que m é positivo.

Pelo algoritmo da divisão, existem $q, r \in \mathbb{N}$ tais que $m = q\, m_0 + r$, $0 \leq r < m_0$. Desta forma, $0 \leq r = m - q\, m_0 < m_0$.

Observe que $r \in B$, porque $m, m_0 \in B$. Ora, como $0 \leq r < m_0$, temos que $r = 0$. Assim, $m = q\, m_0$. Logo $B \subset m_0\, \mathbb{Z}$. Concluímos assim que $B = m_0\, \mathbb{Z}$.

Agora vamos provar que $m_0 = d$ (o que vai implicar que $B = d\, \mathbb{Z}$).

Suponha que $m_0 = b_1\, n_1 + b_2\, n_2 + ... + b_k\, n_k$, onde $b_1, b_2, ..., b_k \in \mathbb{Z}$; $n_1, n_2, ..., n_k \in A$.

a) Como $d = \text{mdc } (A)$, temos que existem $r_1, r_2, ..., r_k \in \mathbb{N}$, tais que $n_j = r_j\, d$. Então,

$$m_0 = b_1\, n_1 + b_2\, n_2 + ... + b_k\, n_k = b_1\, (r_1\, d) + b_2\, (r_2\, d) + ... + b_k\, (r_k\, d) =$$

$$d(b_1\, r_1\ + b_2\, r_2 + ... + b_k\, r_k\ ).$$

Como $m_0, d > 0$, então, o número inteiro $(b_1\, r_1\ + b_2\, r_2 + ... + b_k\, r_k\ ) > 0$. Assim, ele é maior ou igual a 1. Desta forma, $m_0 \geq d$.

b) Note que $A \subset B = m_0\, \mathbb{Z}$. Assim, dado $n \in A$, existe $z \in \mathbb{Z}$, tal que $n = z\, m_0$. Desta forma, como $d = \text{mdc } (A)$ e m_0 é um divisor de todos os elementos de A, temos que $d \geq m_0$.

Concluímos assim, que $d = m_0$, e finalmente que $B = d\, \mathbb{Z}$.

Vamos mostrar agora que existe $\tilde{n}_i$, tal que, para todo $m > n_i$, temos que $P^{m\, d(i)}(i, i) > 0$.

Segue do resultado acima que $d = b_1\, n_1 + b_2\, n_2 + ... + b_k\, n_k$, onde $b_1, b_2, ..., b_k \in \mathbb{Z}$; $n_1, n_2, ..., n_k \in A$.

Podemos renomear os coeficientes b_j, $j = 1, 2, ..., k$, obtendo outra expressão para d, via $\tilde{b}_j$, $j = 1, 2..., k$, de tal forma que $b_j = \tilde{b}_j$, se $b_j > 0$ e $b_j = -\tilde{b}_j$, caso contrário. Assim,

$$d = (\tilde{b}_1\, n_1 + \tilde{b}_2\, n_2 + ... + \tilde{b}_j\, n_j) - (\tilde{b}_{j+1}\, n_{j+1} + \tilde{b}_{j+2}\, n_{j+2} + ... + \tilde{b}_k\, n_k) = \ N_1\ - N_2.$$

Onde, vale que $N_1, N_2 \in B^+ \subset d\,\mathbb{Z}$.

Segue que $N_2 = d\,z$, para algum $z \in \mathbb{Z}$.

Tome $\tilde{n}_i = \frac{(N_2)^2}{d} = \frac{(z\,d)^2}{d} = d\,z^2 \in \mathbb{N}$.

Seja $m > \tilde{n}_i$. Então existe $k \in \mathbb{N}$ tal que $m = \tilde{n}_i + k$. Agora, pelo algoritmo da divisão existem $p, s \in \mathbb{N}$, tais que $k = p\,\frac{N_2}{d} + s$, onde $0 \leq s < \frac{N_2}{d}$.

Portanto, $m = \tilde{n}_i + k = \frac{(N_2)^2}{d} + p\frac{N_2}{d} + s = \frac{1}{d}\,[N_2\,(N_2 + p) + s\,d]$.

Segue então que $m\,d = N_2\,(N_2 + p - s) + s\,(N_2 + d) = N_2\,l + s\,N_1$, onde $l = (N_2 + p - s)$.

Como $s < \frac{N_2}{d} \leq N_2$, então, $-s \geq -N_2$, e, assim $N_2 - s \geq 0$. Note que $l = N_2 + p - s \geq p \geq 0$.

Logo,

$$m\,d = sN_1 + l\,N_2 = (l\,\tilde{b}_{j+1}\,n_{j+1} + ... + l\,\tilde{b}_k\,n_k) + (s\,\tilde{b}_1\,n_1 + ... + s\,\tilde{b}_j\,n_j) =$$

$$(c_1\,n_1 + ... + c_j\,n_j) + (c_{j+1}\,n_{j+1} + ... + c_k\,n_k),$$

onde $c_1, c_2, ..., c_j$ são inteiros positivos.

Finalmente obtemos que para $m > \tilde{n}_i$, vale

$$P^{m\,d}(i,i) = P^{c_1\,n_1 + ... + c_j\,n_j + c_{j+1}\,n_{j+1} + ... + c_k\,n_k}(i,i) \geq$$

$$P^{\underbrace{n_1 + ... + n_1}_{c_1}}(i,i) ... P^{\underbrace{n_k + ... + n_k}_{c_k}}(i,i) \geq$$

$$\underbrace{P^{n_i}(i,i)...P^{n_i}(i,i)}_{c_1} ... \underbrace{P^{n_k}(i,i)...P^{n_k}(i,i)}_{c_k} > 0.$$

Se N for o valor supremo de todos os possiveis valores $\tilde{n}_i$, $i \in S$, temos que $P^{m\,d}(i,i) > 0$, para todo $m > N$, independente de i.

$\square$

Corolário 2.3. *Se $P^m(j,i) > 0$ então $(\mathcal{P}^{m\,+\,n\,d(i)})_{j,i} > 0$, $\forall n$ suficientemente grande.*

Demonstração: Segue de imediato de

$$\mathcal{P}^{m+n\,d(i)} = \mathcal{P}^{m}\ \mathcal{P}^{n\,d(i)}.$$

□

Corolário 2.4. *Seja $(X_n)_{n\geq 0}$ C.M. irredutível e aperiódica com S finito e matriz de transição $\mathcal{P}$. Então, $\exists n_0 \geq 1$ tal que se, $n \geq n_0$, então todas as entradas da matriz $\mathcal{P}^n$ são positivas.*

Demonstração: Os i e j são finitos, assim escolhemos $m = m(i,j)$ como acima (que são limitados por uma constante). O $d(i) = 1$ é o mesmo para todo i. Assim, existe um natural n_0 tal que vale $(\mathcal{P}^n)_{i,j} > 0$ para todo $n > n_0$.

□

Corolário 2.5. *Toda matriz $\mathcal{P}$ de transição com S finito que é irredutível e aperiódica é regular. Ainda, o autovalor 1 tem multiplicidade 1.*

Demonstração: Segue de imediato do teorema acima e do Teorema 2.7.

□

Proposição 2.20. *O espaço de estados de uma C.M. irredutível de período $d > 1$ pode ser particionado em d classes disjuntas $\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_{d-1}$ tal que algum elemento de $\mathcal{D}_j$ vai, no próximo passo, para $\mathcal{D}_{j+1}$, $j = 0, 1, \ldots, d-2$. Finalmente, algum elemento de $\mathcal{D}_{d-1}$ vai no passo seguinte para $\mathcal{D}_0$.*

Ainda, $\mathcal{P}^d|_{D_j}$ é irredutível e aperiódico, para todo j fixo.

Demonstração: Seja $i \in S$. Defina

$\mathcal{D}_m = \{j \,|\, \mathcal{P}^{nd+m}(i,j) > 0$, para algum $n\}$. Então, como a cadeia é irredutível, temos que $\bigcup_{m=0}^{d-1} \mathcal{D}_m = S$

Afirmamos que: $\mathcal{D}_{m_1} \bigcap \mathcal{D}_{m_2} = \emptyset$, $0 \leq m_1 < d$ e $0 \leq m_2 < d$, $m_1 \neq m_2$.

De fato, seja $j \in \mathcal{D}_{m_1} \bigcap \mathcal{D}_{m_2}$.

Então $j \in \mathcal{D}_{m_1}$ e $j \in \mathcal{D}_{m_2}$.

Ora, $j \in \mathcal{D}_{m_1} \Rightarrow \exists n_1$ tal que $P^{n_1\,d+m_1}(i,j) > 0$.

Ora, a cadeia irredutível $\Rightarrow \exists k > 0$ tal que $P^k(j,i) > 0 \Rightarrow P^{n_1\,d+m_1+k}(i,i) \geq P^{n_1\,d+m_1}(i,j)\,P^k(j,i) > 0$.

Logo, $P^k(j,i) > 0 \Rightarrow d$ divide $n_1\,d + m_1 + k \Rightarrow d$ divide $m_1 + k \Rightarrow m_1 + k$ é múltiplo de d.

Analogamente, $j \in \mathcal{D}_{m_2} \Rightarrow \exists n_2$ tal que $P^{n_2\,d+m_2}(i,j) > 0$.

Usando novamente o fato que a cadeia irredutível $\Rightarrow \exists k > 0$ tal que $P^k(j,i) > 0$.

Observação: O k acima pode ser escolhido como m.m.c. (k_1, k_2), onde $k_1 = \min\{s : P^s(i,j) > 0\}$ e $k_2 = \min\{s : P^s(j,i) > 0\}$.

De maneira semelhante ao caso anterior $m_2 + k$ é múltiplo de d.

Assim, d divide $(m_1 + k) - (m_2 + k) = m_1 - m_2$.

Como $0 \leq m_1 < d$ e $0 \leq m_2 < d$, concluímos que $m_1 = m_2$.

Logo, se $j \in \mathcal{D}_{m_1} \cap \mathcal{D}_{m_2}$, então $\mathcal{D}_{m_1} = \mathcal{D}_{m_2}$.

Sendo assim, $\{\mathcal{D}_0, \mathcal{D}_1, ..., \mathcal{D}_{d-1}\}$ forma uma partição de S.

Por construção, dado um elemento j de $\mathcal{D}_m$, se $P(j,k) > 0$, então k não está em $\mathcal{D}_m$.

De fato, se $m < d - 1$, então k é elemento de $\mathcal{D}_{m+1}$.

Isto porque $P^{nd+m+1}(i,k) = \sum_{s \in S} P^{nd+m}(i,s)P(s,k) \geq P^{nd+m}(i,j)\,P(j,k) > 0$.

Finalmente, se $j \in \mathcal{D}_{d-1}$ e $P(j,k) > 0$, então $k \in \mathcal{D}_0$.

Note que, conforme provamos antes no Teorema 2.19: se N for o valor supremo de todos os possíveis valores $\tilde{n}_i$, $i \in S$, temos que $P^{m\,d}(i,i) > 0$, para todo $m > N$.

Considere um $j \in \{1, 2, ..., d\}$ fixo. A cadeia $P^{m\,d}$ restrita a D_j, é fechada e irredutivel.

De fato, suponha $k, l \in D_j$. Então existe m tal que $\mathcal{P}^{md+j}(i,k) > 0$. Ainda, existe r tal que $\mathcal{P}^r(k,l) > 0$ (pela irredutibilidade de $\mathcal{P}$). Assim, $\mathcal{P}^{md+j+r}(i,l) > 0$. Portanto, r é múltiplo de d. Logo, $P^{m\,d}$ restrita a D_j é irredutível.

Por sua vez, se $k \in D_r$ e $l \in D_s$, com $r \neq s$, então $\mathcal{P}^{md}(k,l) > 0$, implicaria (semlhante ao acima) que k e l estão na mesma classe (contradição). Assim, cada classe é fechada para $\mathcal{P}^{md}$.

$\mathcal{P}^d|_{D_j}$ tem período 1 porque $\mathcal{P}^d(k,k) > 0$ para cada k em D_j.

A matriz estocástica $\mathcal{P}^{md}$ agindo no espaço de estados D_j, para um fixo $j = 0, ...d-1$ é irredutível e aperiódica. Esta matriz restrita é regular. Sendo assim podemos aplicar aqui os resultados já conhecidos para tal tipo de matriz (ver corolário 2.4).

Fica assim demostrada a última afirmação do Teorema.

$\square$

Em resumo: Se $\mathcal{P}$ é uma cadeia estocástica irredutível e de período d, então segue do que vimos acima que a matriz $\mathcal{R} = \mathcal{P}^d$ aplica cada um dos conjuntos acima $\mathcal{D}_j$ em si mesmo. Ainda, $P^d(k,l) > 0$, para qualquer $k, l \in \mathcal{D}_j$. Mais que isto, se $P^{d'}(k,l) > 0$, então d' é múltiplo de d.

Este é o procedimento canônico que devemos seguir quando tratamos com cadeias $\mathcal{P}$ que são irredutíveis e periódicas de período d, mais explicitamente, devemos tomar as subclasses $\mathcal{D}_j$, $j = 0, 1, .., d-1$, e a seguir analisar a matriz $\mathcal{P}^{md}$ restrita a uma destas classes.

Exemplo 2.40. Seja a cadeia de Markov com matriz de transição dada por

$$\mathcal{P} = \begin{pmatrix} 0 & 1/8 & 0 & 7/8 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/4 & 3/4 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 1/4 & 3/4 \end{pmatrix}.$$

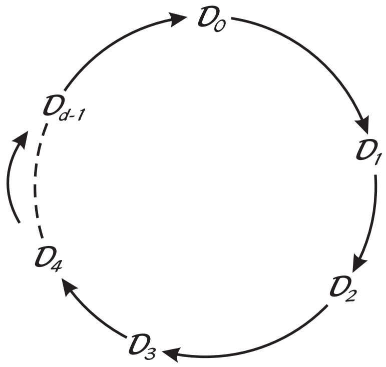


Figura 2.8:

a) Classes de equivalência.

$C_0 = \{0, 1, 2, 3\}$ e $C_1 = \{4, 5\}$.

Logo a cadeia não é irredutível e C_0 e C_1 são fechadas.

b) Classificação dos estados.

Vamos calcular apenas

$$f^*_{44} = \sum_{n\geq 1} f^n_{44} = \frac{1}{2} + \frac{1}{2}\,\frac{1}{4} + \frac{1}{2}\,\frac{3}{4}\,\frac{1}{4} + \frac{1}{2}\left(\frac{3}{4}\right)^2 \frac{1}{4} + \cdots =$$

$$= \frac{1}{2} + \frac{1}{8}\sum_{n\geq 0}\left(\frac{3}{4}\right)^n = \frac{1}{2} + \frac{1}{8}\,\frac{1}{1-\frac{3}{4}} = \frac{1}{2} + \frac{4}{8} = 1.$$

Logo, C_1 é classe recorrente.

c) Distribuição estacionária.

Podemos considerar a matriz $\mathcal{P}$ restrita aos estados $\{0,1,2,3\}$, obtendo assim, a matriz estocástica

$$\mathcal{P}_0 = \begin{pmatrix} 0 & 1/8 & 0 & 7/8 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 1/4 & 3/4 & 0 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix}.$$

Considere a equação $\pi P = \pi$, que equivale a

$$\begin{cases} \frac{1}{2}\pi(3) = \pi(0) \\ \\ \frac{1}{8}\pi(0) + \frac{1}{2}\pi(1) + \frac{1}{4}\pi(2) = \pi(1) \\ \\ \frac{3}{4}\pi(2) + \frac{1}{2}\pi(3) = \pi(2) \\ \\ \frac{7}{8}\pi(0) + \frac{1}{2}\pi(1) = \pi(3) \end{cases} \Longrightarrow \begin{cases} \pi(3) = 2\pi(0) \\ \\ \frac{1}{8}\pi(0) + \frac{1}{4}\pi(2) = \frac{1}{2}\pi(1) \\ \\ \pi(3) = \frac{1}{2}\pi(2) \\ \frac{7}{4}\pi(0) + \pi(1) = 2\pi(3) \end{cases}$$

Ora,
$\pi(3) = 2\pi(0)$
$\pi(3) = \frac{1}{2}\pi(2)$
$\pi(1) = \frac{9}{8}\pi(3)$
e $\pi(0) + \pi(1) + \pi(2) + \pi(3) = 1$

$$\pi = \left(\frac{1}{2}\ \pi(3),\ \ \frac{9}{8}\ \pi(3),\ \ 2\pi(3),\ \ \pi(3)\right),$$

então vale

$$\frac{1}{2}\pi(3) + \frac{9}{8}\pi(3) + 2\pi(3) + \pi(3) = 1 \Rightarrow \pi(3) = \frac{8}{37}.$$

Segue, assim que,

$$\pi = \left(\frac{4}{37},\ \frac{9}{37},\ \frac{16}{37},\ \frac{8}{37}\right).$$

C_0 tem uma única distribuição estacionária. Mas C_0 é aperiódica. De fato:

$d(0) \,=\, \text{mdc}\, \{2,\, 3,\, 5,\, 6\, \cdots\} \,=\, 1,$

$d(1) \,=\, \text{mdc}\, \{1,\, 3,\, 4,\, 5\, \cdots\} \,=\, 1$

Como C_0 é irredutível, aperiódica então tem uma única distribuição estacionária - disto concluiremos mais tarde que C_0 é recorrente positiva (a ser definido).

Observamos que

$d(4) \,=\, \text{mdc}\, \{1,\, 2,\, 3,\, 4,\, \cdots\} \,=\, 1 \,=\, d(5)$

Ainda, podemos considerar a matriz $\mathcal{P}$ restrita ao estados $\{4, 5\}$, obtendo assim,

$$\mathcal{P}_1 = \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix}.$$

$$\pi\, \mathcal{P}_1 \,=\, \pi \implies \begin{cases} \pi(4)\, \dfrac{1}{2} \,+\, \pi(5)\, \dfrac{1}{4} \,=\, \pi(4) \\ \\ \pi(4)\, \dfrac{1}{2} \,+\, \pi(5)\, \dfrac{3}{4} \,=\, \pi(5) \end{cases} \implies \begin{cases} \pi(5) \,=\, 2\, \pi(4) \\ \pi(4) \,+\, \pi(5) \,=\, 1 \end{cases}$$

Logo, $\pi \,=\, \left(\dfrac{1}{3}, \dfrac{2}{3}\right)$ é a única distribuição estacionária.

C_1 será um classe recorrente positiva (a ser definido mais tarde).

$\diamondsuit$

Exemplo 2.41. O seguinte exemplo é ilustrativo do que acontece com as iterações de uma cadeia de Markov periódica.

Na matriz abaixo colocamos o símbolo * para destacar blocos. A matriz é do tipo 7 por 7.

Seja a matriz $\mathcal{P}$ sobre $S = \{1, 2, 3, 4, 5, 6, 7\}$ de período 3; então

$$\mathcal{P} = \begin{pmatrix} 0 & 0 & * & 1/2 & 1/4 & 1/4 & * & 0 & 0 \\ 0 & 0 & * & 1/3 & 0 & 2/3 & * & 0 & 0 \\ * & * & * & * & * & * & * & * & * \\ 0 & 0 & * & 0 & 0 & 0 & * & 1/3 & 2/3 \\ 0 & 0 & * & 0 & 0 & 0 & * & 1/2 & 1/2 \\ 0 & 0 & * & 0 & 0 & 0 & * & 3/4 & 1/4 \\ * & * & * & * & * & * & * & * & * \\ 1/2 & 1/2 & * & 0 & 0 & 0 & * & 0 & 0 \\ 1/4 & 1/4 & * & 0 & 0 & 0 & * & 0 & 0 \end{pmatrix},$$

$$\mathcal{P}^2 = \begin{pmatrix} 0 & 0 & * & 0 & 0 & 0 & * & \frac{23}{48} & \frac{25}{48} \\ 0 & 0 & * & 0 & 0 & 0 & * & \frac{11}{18} & \frac{7}{18} \\ * & * & * & * & * & * & * & * & * \\ 1/3 & 2/3 & * & 0 & 0 & 0 & * & 0 & 0 \\ 3/8 & 5/8 & * & 0 & 0 & 0 & * & 0 & 0 \\ 7/16 & 9/16 & * & 0 & 0 & 0 & * & 0 & 0 \\ * & * & * & * & * & * & * & * & * \\ 0 & 0 & * & \frac{10}{24} & \frac{3}{24} & \frac{11}{24} & * & 0 & 0 \\ 0 & 0 & * & \frac{6}{16} & \frac{1}{16} & \frac{9}{16} & * & 0 & 0 \end{pmatrix},$$

$$\mathcal{P}^3 = \begin{pmatrix} \frac{71}{192} & \frac{121}{192} & * & 0 & 0 & 0 & * & 0 & 0 \\ \frac{29}{72} & \frac{43}{72} & * & 0 & 0 & 0 & * & 0 & 0 \\ * & * & * & * & * & * & * & * & * \\ 0 & 0 & * & \frac{14}{36} & \frac{3}{36} & \frac{19}{36} & * & 0 & 0 \\ 0 & 0 & * & \frac{38}{96} & \frac{9}{96} & \frac{49}{96} & * & 0 & 0 \\ 0 & 0 & * & \frac{26}{64} & \frac{7}{64} & \frac{31}{64} & * & 0 & 0 \\ * & * & * & * & * & * & * & * & * \\ 0 & 0 & * & 0 & 0 & 0 & * & \frac{157}{288} & \frac{131}{288} \\ 0 & 0 & * & 0 & 0 & 0 & * & \frac{111}{192} & \frac{81}{192} \end{pmatrix}.$$

Note que $\mathcal{P}^3$ tem todos os elementos da diagonal não nulos. Logo, todos elementos de S para $\mathcal{P}^3$ são aperiódicos.

Usando a notação do teorema anterior $\mathcal{D}_1 = \{1,2\}$, $\mathcal{D}_2 = \{3,4,5\}$ e $\mathcal{D}_3 = \{6,7\}$,

Note que

a) $1 \in \mathcal{D}_1$ vai na primeira etapa a $3 \in \mathcal{D}_2$,

b) $4 \in \mathcal{D}_2$ vai na primeira etapa a $6 \in \mathcal{D}_3$,

e

c) $6 \in \mathcal{D}_3$ vai na primeira etapa a $2 \in \mathcal{D}_1$.

Podemos considerar $\mathcal{P}^3$ como uma nova cadeia $\mathcal{R}$ sobre S e obtemos para esta cadeia $\mathcal{R}$ uma decomposição de $S = \mathcal{D}_1 \cup \mathcal{D}_2 \cup \mathcal{D}_3$, em conjuntos (fechados) irredutíveis, recorrentes e aperiódicos.

A recorrência segue do fato que a matriz é finita, e assim em cada classe irredutível existe um recorrente. Logo, todo elemento de cada classe é recorrente.

$\diamondsuit$

Proposição 2.21. *Fixada uma cadeia de Markov $\mathcal{P}$ aperiódica, então o mdc dos n tais que $f_{ii}^n > 0$ é igual a 1.*

Demonstração: Por hipótese o mdc dos n tais que $P_{ii}^n > 0$ é igual a 1. Seja v o mdc dos n tais que $f_{ii}^n > 0$.

Suponha que $v \neq 1$.

Se v divide todo n tal que $P_{ii}^n > 0$ então $v = 1$.

Seja n o menor inteiro tal que $P_{ii}^n > 0$ e v não divide n.

Seja agora q e r tal que $0 < r < v$ e $n = q\,v + r$. Sabemos que $f_{ii}^j = 0$ se j não é da forma $j = k\,v$, logo

$$P_{ii}^n = \sum_{j=1}^{n} f_{ii}^j\, P_{ii}^{n-j} =$$

$$\sum_{k=1}^{q} f_{ii}^{k\,v}\, P_{ii}^{(q\,v+r)-k\,v} = \sum_{k=1}^{q} f_{ii}^{k\,v}\, P_{ii}^{(q-k)\,v+r}.$$

Sabemos que $P_{ii}^{(q-k)\,v+r} = 0$, porque $(q-k)\,v + r < n$ e v não divide $(q-k)\,v + r$. Sendo assim $P_{ii}^n = 0$. Como isto é uma contradição, concluímos que v divide todo n tal que $P_{ii}^n > 0$. Logo, $v = 1$.

□

Os resultados abaixo são de grande utilidade mas não apresentaremos aqui as suas demonstrações (ver [I]).

Teorema 2.20. *Seja S finito e $\mathcal{P}$ irredutível de período d. Então as d raízes da unidade $e^{2\pi\,(r/d)\,i}$, $r \in \{0, 1, 2, ..., d-1\}$ são autovalores de $\mathcal{P}$. A multiplicidade de cada raiz complexa é igual a um.*

Teorema 2.21. *Seja S finito. Então as d raízes da unidade são autovalores de $\mathcal{P}$, se e só se, $\mathcal{P}$ tem uma classe recorrente de período d. A multiplicidade de cada raiz complexa é igual ao número de subclasses irredutíveis recorrentes de período d.*

Corolário 2.6. *Se não existem números complexos λ (além de $\lambda = 1$) de norma 1, tal que, para $g(\lambda) = det\,(\lambda I - \mathcal{P}) = 0$, então, $\mathcal{P}$ é aperiódica.*

Os resultados acima são muito úteis pois permitem detectar se $\mathcal{P}$ irredutível é aperiódica apenas analisando as raízes de um polinômio de grau $d = \#S$.

Exemplo 2.42. A matriz

$$\mathcal{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

é irredutível pois tem polinômio característico $g(\lambda) = \lambda^3 - 1$. De fato, as raízes λ de g são $1, e^{i\,2\pi\,1/3}, e^{i\,2\pi\,2/3}$. Logo, P tem período 3.

$\diamondsuit$

2.7 Estados Recorrentes Nulos e Positivos

Definição 2.21. *O tempo médio de primeiro retorno (ou, de recorrência) de um estado i é definido por*

$$\mu_i = \mathbb{E}_i(T_i) = \sum_{n \geq 1} n\, f_{ii}^n.$$

Acima T_i é o tempo de primeiro retorno ao estado $i \in S$ do processo começando em $i \in S$.

O tempo médio acima definido descreve o número médio de etapas necessárias para retornar ao estado inicial i dado que a cadeia começou em i.

Se $P_i(T_i = +\infty) > 0$ então $\mathbb{E}_i(T_i) = +\infty$.

Note que T_i, $i \in S$, é uma função mensurável mas não se pode dizer que depende de finitas coordenadas.

A definição de $\mathbb{E}_i(T_i)$, para i fixo em S, é

$$\mathbb{E}_i(T_i) = \int T_i\, dP_i \;= \sum_{n \in \mathbb{N}} \; n \;\, P_i(\{w \; tal \; que \;\; T_i(w) = n\}),$$

onde P_i é a probabilidade markoviana em $\Omega = S^{\mathbb{N}}$, obtida a partir de uma matriz fixada $\mathcal{P}$, e a probabilidade inicial π^0 que é igual a 1 na posição i (e zero fora). Portanto este valor depende de $\mathcal{P}$ e de $i \in S$.
Observações*:*

(a) **i** *é transitório* $\Rightarrow P_i(T_i = +\infty) > 0 \Rightarrow \mu_i = E_i(T_i) = \infty$,

(b) **i** *é recorrente* $\Rightarrow P_i(T_i = +\infty) = 0 \Rightarrow \mu_i = \sum_{n\geq 1} n\, f_{ii}^n$

Portanto, μ_i pode ser $+\infty$ ou $< +\infty$

Definição 2.22. *Se $i \in S$ é um estado recorrente então*

(a) **i** *é* **recorrente nulo** *quando* $\mu_i = \mathbb{E}_i(T_i) = +\infty$

(b) **i** *é* **recorrente positivo** *quando* $\mu_i = \mathbb{E}_i(T_i) < +\infty$

Proposição 2.22. *Ser recorrente nulo ou ser recorrente positivo é uma propriedade de classe, ou seja, se i possui tal propriedade e j é equivalente a i, então j também possui.*

Este resultado será demonstrado mais tarde na proposição 2.23.
Observações:

(1) Se j é transiente então $P^n(i,j) \underset{n\to\infty}{\longrightarrow} 0 \;, \forall\, i,$

(2) Se j é recorrente nulo então $P^n(i,j) \underset{n\to\infty}{\longrightarrow} 0 \;, \forall\, i.$

Teorema 2.22. *Seja $(X_n)_{n\geq 0}$ C.M. finita com matriz de transição P com uma classe C_1 de estados recorrentes positivos e aperiódica, sendo os outros estados transitórios. Então $(X_n)_{n\geq 0}$ tem uma única distribuição estacionária dada por*

$$\pi(j) = \lim_{n\to\infty} P^n(i,j) \quad se \quad j \in C_1, \; e$$
$$\pi(j) = 0 \quad se \quad j \notin C_1.$$

Este resultado será demonstrado mais tarde no Teorema 2.24.

Exemplo 2.43. Pela fórmula de Stirling temos que no caso do passeio aleatório em $\mathbb{Z}$ com $p = 1/2$, o valor $P_{00}^n \sim n^{-1/2}$, para n par e zero caso n seja ímpar. Logo, 0 é recorrente pois $\sum_{n\in\mathbb{N}} P_{00}^n = \infty$. O mesmo vale para qualquer $i \in \mathbb{Z}$

Portanto, neste caso, a cadeia é recorrente irredutível, mas 0 é recorrente nulo. De fato, um estado i é recorrente nulo se e só se $\lim_{n\to\infty} P_{ii}^n = 0$ conforme Teorema 2.23. Note que $P_{00}^n \sim n^{-1/2}$.

Note que quando S é finito toda cadeia irredutível e recorrente é recorrente positivo como veremos em breve.

$\diamondsuit$

Exemplo 2.44. Seja $S = \{0, 1, 2, 3, 4, 5\}$ e considere a matriz

$$\mathcal{P} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1/4 & 2/4 & 0 & 0 & 0 & 1/4 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix}.$$

Vamos a seguir obter o seguinte:

(a) determinar as classes de equivalência;

(b) estudar a periodicidade da cadeia;

(c) classificar os estados;

(d) obter os conjuntos cíclicos disjuntos ;

(e) obter as matrizes de transição associadas aos conjuntos críticos;

(f) mostrar que esta C.M. tem uma única distribuição estacionária.

a) Observe a partir do grafo exibido na Figura 2.9 que

$0 \leftrightarrow 3$, $1 \leftrightarrow 2$, $1 \leftrightarrow 4$, $2 \leftrightarrow 5$, $1 \leftrightarrow 5$, $1 \leftrightarrow 3$

Portanto, $C_0 = \{0, 1, 2, 3, 4, 5\}$ é classe irredutível e fechada.

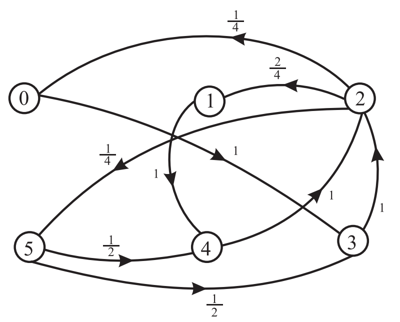


Figura 2.9:

b) Basta determinar o período para um estado. Observe que

$d(0) = \text{mdc}\ \{n \geq 1; P^n(0,0) > 0\} = \text{mdc}\ \{3, 6, 9, ...\} = 3$, porque $0 \to 3 \to 2 \to 0$.

Logo, a C.M. tem período 3.

c) Para classificar os estados, basta calcular f_{00}^*.

A cadeia é recorrente positiva tendo em vista que a $\mathcal{P}^3$ restrita a cada classe é irredutível, aperiódica e finita.

d) Desejamos obter D_0, D_1, D_2 conforme Porposição 2.20. Vamos mostrar que $D_0 = \{0, 1, 5\}$ $D_1 = \{3, 4\}$ $D_2 = \{2\}$

e) Para obtermos as matrizes de transição associadas a D_0, D_1, D_2, precisamos de $\mathcal{P}^3$.

$$\mathcal{P}^2\ \mathcal{P} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3/8 & 5/8 & 0 \\ 1/4 & 2/4 & 0 & 0 & 0 & 1/4 \\ 1/4 & 2/4 & 0 & 0 & 0 & 1/4 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1/4 & 2/4 & 0 & 0 & 0 & 1/4 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix} =$$

$$\begin{pmatrix} 1/4 & 2/4 & 0 & 0 & 0 & 1/4 \\ 1/4 & 2/4 & 0 & 0 & 0 & 1/4 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3/8 & 5/8 & 0 \\ 0 & 0 & 0 & 3/8 & 5/8 & 0 \\ 1/4 & 2/4 & 0 & 0 & 0 & 1/4 \end{pmatrix}.$$

Portanto,

$$\mathcal{P}_0 = \begin{pmatrix} 1/4 & 2/4 & 1/4 \\ 1/4 & 2/4 & 1/4 \\ 1/4 & 2/4 & 1/4 \end{pmatrix},$$

$$\mathcal{P}_1 = \begin{pmatrix} 3/8 & 5/8 \\ 3/8 & 5/8 \end{pmatrix},$$

e

$$\mathcal{P}_2 = (1).$$

Observe que as sub-matrizes de transição associadas aos conjuntos cíclicos são também estocásticas.

Desta forma concluímos que $D_0 = \{0, 1, 5\}$ $D_1 = \{3, 4\}$ $D_2 = \{2\}$.

f) Pelo Teorema 2.4 sabemos que esta C.M. restrita a cada D_i possui uma única distribuição estacionária. Desejamos encontrar ao menos um vetor $\pi = (\pi_0, \pi_1, \pi_2, \pi_3, \pi_4, \pi_5)$ tal que $\pi = \pi\mathcal{P}$.

Então, valem as relações
$\pi_2 \frac{1}{4} = \pi_0$,
$\pi_0 + \frac{1}{2}\pi_5 = \pi_3$,
$\pi_2 \frac{2}{4} = \pi_1$,
$\pi_1 + \frac{1}{2}\pi_5 = \pi_4$,
$\pi_3 + \pi_4 = \pi_2$

e

$\pi_2 \frac{1}{4} = \pi_5$.

Obtemos assim,

$\pi_2 = 4\pi_0$,

$\pi_5 = \frac{1}{4}\pi_2 = \frac{4}{4}\pi_0 = \pi_0$,

$\pi_1 = \frac{1}{2}\pi_2 = 2\pi_0$,

$\pi_4 = \frac{5}{2}\pi_0$

e

$\pi_3 + \pi_4 = \pi_2 = 4\pi_0$.

Ainda, como vale $\pi_3 + \pi_4 = 4\pi_0 \Rightarrow \pi_3 = 4\pi_0 - \pi_4 = 4\pi_0 - \frac{5}{2}\pi_0 \Rightarrow \pi_3 = \frac{3}{2}\pi_0$.

Como $\sum_{i=0}^{5} \pi_i = 1 \Rightarrow 12\pi_0 = 1$.

Portanto, $\pi = (\frac{2}{24}, \frac{4}{24}, \frac{8}{24}, \frac{3}{24}, \frac{5}{24}, \frac{2}{24})$.

Note que $(1/4, 2/4, 1/4)$ é invariante para $\mathcal{P}_0$, $(3/7, 5/7)$ é invariante para $\mathcal{P}_1$ e (1) é invariante para $\mathcal{P}_2$.

Sejam q_1, q_2, q_3 números não negativos que somam 1.

É fácil ver que

$$\pi = q_1\,(1/4, 2/4, 0, 0, 0, 1/4) + q_2(0, 0, 0, 3/7, 5/7, 0) + q_3(0, 0, 1, 0, 0, 0)$$

satisfaz $\pi\mathcal{P} = \pi$. Logo, tal π não é único.

$\diamondsuit$

Exemplo 2.45. Passeio Aleatório Simples em $\mathbb{Z}$.

Suponha $P(i, i+1) = p$, $P(i, i-1) = q$, onde $p + q = 1$, i $\in \mathbb{Z}$

e ainda que

$(X_n)_{n\geq 0}$ é irredutível e periódica com período d $= 2$.

(a) Desejamos obter os conjuntos cíclicos.

(b) Desejamos obter as duas C.M. irredutíveis e aperiódicas geradas por D_0 e D_1.

$$\mathcal{P} = \begin{array}{c} \\ \\ \vdots \\ -2 \\ -1 \\ 0 \\ 1 \\ 2 \\ \vdots \\ \\ \end{array} \begin{array}{c} \dots \quad -2 -1 \;\; 0 \;\;\; 1 \;\; 2 \dots \\ \left(\begin{array}{cccccccc} \dots & & & & & & \dots & \\ \dots & & & & & & \dots & \\ \dots q & 0 & p & 0 & 0 & 0 & 0 & \dots \\ \dots 0 & q & 0 & p & 0 & 0 & 0 & \dots \\ \dots 0 & 0 & q & 0 & p & 0 & 0 & \dots \\ \dots 0 & 0 & 0 & q & 0 & p & 0 & \dots \\ \dots 0 & 0 & 0 & 0 & q & 0 & p & \dots \\ \dots & & & & & & \dots & \\ \dots & & & & & & \dots & \end{array}\right) \end{array}.$$

$C_0 = \mathbb{Z}$ é irredutível, periódica com período d $= 2$.

a) $D_0 = \{\dots -3, -1, 1, 3, 5, \dots\}$ e $D_1 = \{\dots, -4, -2, 0, 2, 4, \dots\}$

pois $-1 \to \;\; 0$
$\qquad\searrow -2$ (0 e -2) devem estar no D_1 se $-1 \in D_0$ $\to -2 \to -3$
$\qquad\qquad\searrow -1$

$-2 \in D_1$, $(-3, -1) \in D_0$, $3 \to 2$
$\qquad\searrow 4$ $3 \in D_0$ $(2, 4) \in D_1$

Os conjuntos cíclicos devem ser disjuntos.

b) Queremos a C.M. gerada por D_0. Sua matriz de transição será denotada por P_0 e é uma parte da $\mathcal{P}^2$:

$$
\mathcal{P}^2 = \begin{array}{cc}
 & \begin{array}{cccccc} \ldots & -2 & -1 & 0 & 1 & 2\ldots \end{array} \\
\begin{array}{c} . \\ -2 \\ -1 \\ 0 \\ 1 \\ 2 \\ . \end{array} &
\begin{pmatrix}
\ldots. & . & . & . & . & . & \ldots \\
\ldots q & 0 & p & 0 & 0 & 0 & \ldots \\
\ldots 0 & q & 0 & p & 0 & 0 & \ldots \\
\ldots 0 & 0 & q & 0 & p & 0 & \ldots \\
\ldots 0 & 0 & 0 & q & 0 & p & \ldots \\
\ldots 0 & 0 & 0 & 0 & q & 0 & \ldots \\
\ldots. & . & . & . & . & . & \ldots
\end{pmatrix}
\end{array}
\begin{array}{cc}
 & \begin{array}{cccccc} \ldots & -2 & -1 & 0 & 1 & 2\ldots \end{array} \\
\begin{array}{c} . \\ -2 \\ -1 \\ 0 \\ 1 \\ 2 \\ . \end{array} &
\begin{pmatrix}
\ldots. & . & . & . & . & . & \ldots \\
\ldots q & 0 & p & 0 & 0 & 0 & \ldots \\
\ldots 0 & q & 0 & p & 0 & 0 & \ldots \\
\ldots 0 & 0 & q & 0 & p & 0 & \ldots \\
\ldots 0 & 0 & 0 & q & 0 & p & \ldots \\
\ldots 0 & 0 & 0 & 0 & q & 0 & \ldots \\
\ldots. & . & . & . & . & . & \ldots
\end{pmatrix}
\end{array} =
$$

$$
\begin{array}{cc}
 & \begin{array}{cccccc} \ldots & -2 & -1 & 0 & 1 & 2 \ \ldots \end{array} \\
\begin{array}{c} . \\ -2 \\ -1 \\ 0 \\ 1 \\ 2 \\ . \end{array} &
\begin{pmatrix}
\ldots & . & . & . & . & . & . & . & \ldots \\
\ldots & 0 & 2qp & 0 & p^2 & 0 & 0 & 0 & \ldots \\
\ldots & q^2 & 0 & 2qp & 0 & p^2 & 0 & 0 & \ldots \\
\ldots & 0 & q^2 & 0 & 2qp & 0 & p^2 & 0 & \ldots \\
\ldots & 0 & 0 & q^2 & 0 & 2qp & 0 & p^2 & \ldots \\
\ldots & 0 & 0 & 0 & q^2 & 0 & 2qp & 0 & \ldots \\
\ldots & . & . & . & . & . & . & . & \ldots
\end{pmatrix}
\end{array}.
$$

Portanto, P_0 (relativa a D_0) e P_1 (relativa a D_1) são, respectivamente,

$$P_0 = \begin{array}{c} \\ \vdots \\ -3 \\ -1 \\ 1 \\ 3 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccccc} ... & -3 & -1 & 1 & 3 & \dots \end{array} \\ \begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & 2qp & p^2 & 0 & 0 & \dots \\ \dots & q^2 & 2qp & p^2 & 0 & \dots \\ \dots & 0 & q^2 & 2qp & p^2 & \dots \\ \dots & 0 & 0 & q^2 & 2qp & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \end{array}$$

e

$$P_1 = \begin{array}{c} \\ \vdots \\ -2 \\ 0 \\ 2 \\ 4 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccccc} ... & -2 & 0 & 2 & 4 & \dots \end{array} \\ \begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & 2qp & p^2 & 0 & 0 & \dots \\ \dots & q^2 & 2qp & p^2 & 0 & \dots \\ \dots & 0 & q^2 & 2qp & p^2 & \dots \\ \dots & 0 & 0 & q^2 & 2qp & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \end{array}.$$

A matriz P_0 é indexada pelos ímpares em $\mathbb{Z}$ e a P_1 pelos pares em $\mathbb{Z}$.

Lembramos ao leitor que, no presente exemplo, se $p = 1/2$ então todo estado é recorrente, mas se $p \neq 1/2$, então todo estado é transiente.

$\diamondsuit$

O próximo exemplo ilustra a maneira como se pode obter resultados explícitos a partir da teoria descrita acima. Ele é bastante elaborado e pode ser omitido numa primeira leitura.

Escólio: Seja $\{S_n : n \geq 0\}$ um passeio aleatório em $\mathbb{Z}$ (probabilidade de transição 1/2 e 1/2) com $S_0 = 0$. Vamos mostrar que $X_n = |S_n|$ define uma cadeia de Markov e determinar as probabilidades de transição.

Solução: Seja $B = \{X_{n-1} = i_{n-1}, \cdots, X_1 = i_1, X_0 = i_0\}$, $i_r \in \mathbb{N}$, $\forall r \in \{0, 1, \cdots, n-1\}$.

Note que $X_0 = |S_0| = 0$. Como $i_0 \neq 0 \Rightarrow B = \emptyset$, consideramos $B = \{X_{n-1} = i_{n-1}, \cdots, X_1 = i_1, S_0 = 0\}$.

Queremos mostrar que, para quaisquer $i_{n+1}, i_n \in \mathbb{N}$,

$$P(X_{n+1} = i_{n+1} | X_n = i_n, B) = P(X_{n+1} = i_{n+1} | X_n = i_n). \tag{2.3}$$

Observe que, para todo $i \geq 1$

$$Se \;\; X_n = i \Rightarrow \left\{ \begin{array}{llll} S_n = i & \Rightarrow \;\; S_{n+1} = i+1 & \text{ou} & S_{n+1} = i-1 \\ \text{ou} & & & \\ S_n = -i & \Rightarrow \;\; S_{n+1} = -i+1 & \text{ou} & S_{n+1} = -i-1 \end{array} \right\}$$

$\Rightarrow X_{n+1} = |S_{n+1}| = i+1$ ou $X_{n+1} = i-1$.

Assim,

$$P(X_{n+1} \in \{i+1, i-1\}^c | X_n = i, B) = 0 = P(X_{n+1} \in \{i+1, i-1\}^c | X_n = i) \tag{2.4}$$

Logo,

$$1 = \frac{P(X_n = i, B)}{P(X_n = i, B)} = \frac{P(X_{n+1} \in S, X_n = i, B)}{P(X_n = i, B)}$$

$$= \frac{P(X_{n+1} \in \{i+1, i-1\}^c \cup \{i+1, i-1\}, X_n = i, B)}{P(X_n = i, B)}$$

$$= \frac{P(X_{n+1} \in \{i+1, i-1\}^c, X_n = i, B)}{P(X_n = i, B)} + \frac{P(X_{n+1} \in \{i+1, i-1\}, X_n = i, B)}{P(X_n = i, B)}$$

$$= P(X_{n+1} \in \{i+1, i-1\}^c | X_n = i, B) + \frac{P(X_{n+1} = i+1, X_n = i, B)}{P(X_n = i, B)}$$

$$+\ \frac{P(X_{n+1}=i-1, X_n=i, B)}{P(X_n=i, B)} = P(X_{n+1}=i+1|X_n=i, B)$$

$$+\ P(X_{n+1}=i-1|X_n=i, B),$$

onde a quinta igualdade acima é devido a expressão (2).
Isto implica que

$$P(X_{n+1}=i-1|X_n=i, B) = 1 - P(X_{n+1}=i+1|X_n=i, B). \tag{2.5}$$

Analogamente,

$$P(X_{n+1}=i-1|X_n=i) = 1 - P(X_{n+1}=i+1|X_n=i). \tag{2.6}$$

Desta forma, basta provar que
$P(X_{n+1}=i+1|X_n=i, B) = P(X_{n+1}=i+1|X_n=i)$, para todo $i \geq 0$.
De fato, se

$$i=0 \Rightarrow \left\{ \begin{array}{lcl} X_n=|S_n|=0 & \Leftrightarrow & S_n=0 \\ X_{n+1}=|S_{n+1}|=1 & \Leftrightarrow & S_{n+1}=1 \quad \text{ou} \quad S_{n+1}=-1 \end{array} \right\}.$$

Então,

$$P(X_{n+1}=1|X_n=0, B) = P(S_{n+1}=1 \,\text{ou}\, S_{n+1}=-1|S_n=0, B)$$

$$= P(S_{n+1}=1|S_n=0, B) + P(S_{n+1}=-1|S_n=0, B)$$

$$= P(S_{n+1}=1|S_n=0) + P(S_{n+1}=-1|Sn=0) = p+q=1$$

$$= P(S_{n+1}=1 \, ou \, S_{n+1}=-1|S_n=0) = P(X_{n+1}=1|X_n=0).$$

Suponha que $i \geq 1$. Primeiramente, vamos escrever $P(X_{n+1} = i+1|X_n = i, B)$ de outra maneira:

$$P(X_{n+1} = i+1|X_n = i, B) = \frac{P(X_{n+1} = i+1, X_n = i, B)}{P(X_n = i, B)}$$

$$= \frac{P(X_{n+1} = i+1, S_n = i, B) + P(X_{n+1} = i+1, S_n = -i, B)}{P(X_n = i, B)}$$

$$= \frac{P(X_{n+1} = i+1, S_n = i, B)}{P(S_n = i, B)} \times \frac{P(S_n = i, B)}{P(X_n = i, B)}$$

$$+ \frac{P(X_{n+1} = i+1, S_n = -i, B)}{P(S_n = -i, B)} \times \frac{P(Sn = -i, B)}{P(X_n = i, B)}$$

$$= P(X_{n+1} = i+1|S_n = i, B) \times \frac{P(S_n = i, B)}{P(S_n = i, B) \, + \, P(S_n = -i, B)}$$

$$+ \; P(X_{n+1} = i+1|S_n = -i, B) \times \frac{P(S_n = -i, B)}{P(S_n = i, B) \, + \, P(S_n = -i, B)}.$$

Calculemos:

$$P(X_{n+1} = i+1|S_n = i, B) = P(S_{n+1} = i+1|S_n = i, B)$$

$$+ \; P(S_{n+1} = -i-1|S_n = i, B) = P(S_{n+1} = i+1|S_n = i)$$

$$+ \; P(S_{n+1} = -i-1|S_n = i) = p \, + \, 0 = p$$

e

$$P(X_{n+1} = i+1|S_n = -i, B) = P(S_{n+1} = i+1|S_n = -i, B)$$

$$+ P(S_{n+1} = -i - 1 | S_n = -i, B) = P(S_{n+1} = i + 1 | S_n = -i)$$

$$+ P(S_{n+1} = -i - 1 | S_n = -i) = 0 \, + \, q = q.$$

Logo,

$$P(X_{n+1} = i + 1 | X_n = i, B) = \frac{pP(S_n = i, B) \, + \, qP(S_n = -i, B)}{P(S_n = i, B) \, + \, P(S_n = -i, B)}. \tag{2.7}$$

Falta encontrar $P(S_n = i, B)$ e $P(S_n = -i, B)$, para todo $i \geqslant 1$.

Seja $l = max\{r : i_r = 0\}$. Então, $S_k > 0, \forall_k \in (l, n]$ ou $S_k < 0, \forall_k \in (l, n]$

.

Observe que

$$P(S_n = i, B) = \sum_{l=0}^{n-1} P(S_n = i, S_{n-1} > 0, \cdots, S_{l+1} > 0, S_l = 0, X_{l-1} = i_{l-1}, \cdots,$$

$$X_1 = i_1, S_0 = 0) = \sum_{l=0}^{n-1} P(S_n = i, S_{n-1} > 0, \cdots, S_{l+1} > 0 | S_l = 0)$$

$$\times \;\; P(S_l = 0, X_{l-1} = i_{l-1}, \cdots, X_1 = i_1, S_0 = 0)$$

$$= \sum_{l=0}^{n-1} P(S_{n-l} = i, S_{n-l-1} > 0, \cdots, S_1 > 0 | S_0 = 0) \times \frac{P(S_l = 0, S_0 = 0)}{P(S_0 = 0)}$$

$$= \sum_{l=0}^{n-1} P(S_{n-l} = i, S_{n-l-1} > 0, \cdots, S_1 > 0 | S_0 = 0) \times P_{00}^l.$$

Observe que a terceira igualdade acima é devido à propriedade markoviana e a quarta igualdade segue do fato que $\mathbb{P}(S_0 = 0) = 1$.

Do Exemplo 2.31, temos que

$$P_{00}^{l} = \begin{cases} \binom{2k}{k} p^k q^k, & se\ l = 2k \\ 0, & se\ l = 2k+1. \end{cases}$$

Então,

$$P(S_{n-l} = i, S_{n-l-1} > 0, \cdots, S_1 > 0 | S_0 = 0)$$

$$= P(S_{n-l} = i | S_0 = 0) - P(S_{n-l} = i, S_j \leq 0, \text{ para algum } j \in \{1, \cdots, n-l-1\} | S_0 = 0),$$

onde

$$P(S_{n-l} = i | S_0 = 0) = \begin{cases} \binom{n-l}{\frac{n-l+i}{2}} p^{\frac{n-l+i}{2}} q^{\frac{n-l-i}{2}}, & \text{se } n-l-i \text{ é par} \\ 0, & \text{se } n-l-i \text{ é ímpar,} \end{cases}$$

e

$$P(S_{n-l} = i, S_j \leqslant 0, \text{ para algum j } \in \{1, \cdots, n-l-1\} | S_0 = 0)$$

$$= P(S_{n-l} = i, S_1 = -1 | S_0 = 0) + P(S_{n-l} = i, S_j \leqslant 0,$$

$$\text{para algum j } \in \{2, \cdots, n-l-1\}, S_1 = 1 | S_0 = 0).$$

Pelo Princípio da Reflexão, existe uma bijeção entre os caminhos que começam em 1 em tempo 1 e chegam a i em tempo $n-l$ que se anulam ou são negativos com os caminhos que saem de -1 em tempo 1 e chegam a i em tempo $n-l$. Ou seja,

$$P(S_{n-l}=i, S_j \leqslant 0, \text{ para algum j } \in \{2,\cdots,n-l-1\}, S_1=1|S_0=0)$$

$$= P(S_{n-l}=i, S_1=-1|S_0=0).$$

Logo,

$$P(S_{n-l}=i, S_j \leqslant 0, \text{ para algum j } \in \{1,\cdots,n-l-1\}|S_0=0)$$

$$= 2P(S_{n-l}=i, S_1=-1|S_0=0) = 2P(S_{n-l}=i|S_1=-1)\times P(S_1=-1|S_0=0)$$

$$= 2qP(S_{n-l}=i|S_1=-1) = 2qP(S_{n-l-1}=i+1|S_0=0)$$

$$= 2q\binom{n-l-1}{\frac{(n-l-1)+(i+1)}{2}} p^{\frac{(n-l-1)+(i+1)}{2}} q^{\frac{(n-l-1)+(i+1)}{2}}$$

$$= 2\binom{n-l-1}{\frac{n-l+i}{2}} p^{\frac{n-l+i}{2}} q^{\frac{n-l-i}{2}} \;,$$

se $n-l+i$ é par e $P(S_{n-l}=i, S_j \leqslant 0, \text{ para algum } j \in \{1,\cdots,n-l-1\}|S_0=0)=0$, se $n-l+i$ é ímpar. Observe que a terceira igualdade acima vale pela mudança de coordenadas no espaço dos caminhos.

Desta forma,

$$P(S_{n-l}=i, S_{n-l-1}>0, \cdots, S_1>0|S_0=0) = 0 \;,$$

se $n-l-i$ é ímpar e se $n-l-i$ é par, temos que, para todo $i \geqslant 1$,

$$P(S_{n-l}=i, S_{n-l-1}>0,\cdots,S_1>0|S_0=0)$$
$$=\binom{n-l}{\frac{n-l+i}{2}}p^{\frac{n-l+i}{2}}q^{\frac{n-l-i}{2}}-2\binom{n-l-1}{\frac{n-l+i}{2}}p^{\frac{n-l+i}{2}}q^{\frac{n-l-i}{2}}$$

$$=\left[\frac{(n-l)!}{\left(\frac{n-l+i}{2}\right)!\left(n-l-\frac{n}{2}+\frac{l}{2}-\frac{i}{2}\right)!}-\frac{2(n-l-1)!}{\left(\frac{n-l+i}{2}\right)!\left(n-l-1-\frac{n}{2}+\frac{l}{2}-\frac{i}{2}\right)!}\right]p^{\frac{n-l+i}{2}}q^{\frac{n-l-i}{2}}$$

$$=\left[\frac{(n-l)!}{\left(\frac{n-l+i}{2}\right)!\left(\frac{n-l-i}{2}\right)!}-\frac{2(n-l-1)!}{\left(\frac{n-l+i}{2}\right)!\left(\frac{n-l-i}{2}-1\right)!}\right]p^{\frac{n-l+i}{2}}q^{\frac{n-l-i}{2}}$$

$$=\frac{(n-l)!}{\left(\frac{n-l+i}{2}\right)!\left(\frac{n-l-i}{2}\right)!}\left[1-2\frac{\left(\frac{n-l-i}{2}\right)}{n-l}\right]p^{\frac{n-l+i}{2}}q^{\frac{n-l-i}{2}}$$

$$=\binom{n-l}{\frac{n-l+i}{2}}\left(\frac{n-l-n+l+i}{n-l}\right)p^{\frac{n-l-i}{2}}q^{\frac{n-l-i}{2}}$$

$$=\binom{n-l}{\frac{n-l+i}{2}}\frac{i}{n-l}\;\;p^{\frac{n-l+i}{2}}q^{\frac{n-l-i}{2}},$$

para todo $i\geqslant 1$, se $n-l+i$ é par.

Então,

$$P(S_n=i,B)=\sum_{l=0}^{n-1}P(S_{n-l}=i,S_{n-l-1}>0,\cdots,S_1>0|S_0=0)P_{00}^l$$

$$=\sum_{l=0}^{n-1}\binom{n-l}{\frac{n-l+i}{2}}\;\frac{i}{n-l}\;\;p^{\frac{n-l+i}{2}}\;q^{\frac{n-l-i}{2}}\;P_{00}^l$$

$$=\sum_{k=0}^{m}\binom{n-2k}{\frac{n-2k+i}{2}}\frac{i}{n-2k}\;p^{\frac{n+i}{2}-k}\;q^{\frac{n-i}{2}-k}\binom{2k}{k}\;p^kq^k$$

$$= p^{\frac{n+i}{2}} q^{\frac{n-i}{2}} \sum_{k=0}^{m} \binom{n-2k}{\frac{n-2k+i}{2}} \binom{2k}{k} \frac{i}{n-2k}, \tag{2.8}$$

se $n-i$ é par. Observe que na terceira igualdade acima temos que $l = 2k$ e

$$m = \begin{cases} \frac{n-1}{2}, & \text{se } n-1 \text{ é par} \\ \frac{n-2}{2}, & \text{se } n-1 \text{ é ímpar.} \end{cases}$$

Ainda, $P(S_n = i, B) = 0$, se $n-i$ é ímpar.

Cada caminho que sai de 0 em tempo 0 e chega a i em tempo $n-l$ sem nunca se anular, pode ser associado a um **único** caminho que sai de 0 em tempo 0 e chega a $-i$ em tempo $n-l$ sem nunca se anular: basta refletir o caminho “de cima” que obteremos o caminho “de baixo”.

Cabe salientar que, quando o caminho que é todo negativo, subir o caminho positivo correspondente a ele irá descer, então as probabilidades p e q serão invertidas, ou seja,

$$P(S_{n-l} = -i, S_{n-l-1} < 0, \cdots, S_1 < 0 | S_0 = 0)$$

$$= \begin{cases} \binom{n-l}{\frac{n-l+i}{2}} \frac{i}{n-l} \; p^{\frac{n-l-i}{2}} \; q^{\frac{n-l+i}{2}}, & se \;\; n-i \text{ é par} \\ 0, & se \;\; n-i \text{ é ímpar.} \end{cases}$$

Desta forma,

$$P(S_n = -i, B) =$$

$$\sum_{l=0}^{n-1} P(S_n = -i, S_{n-1} < 0, \cdots, S_{l+1} < 0, S_l = 0, X_{l-1} = i_{l-1}, \cdots, X_1 = i_1, S_0 = 0) =$$

$$\sum_{l=0}^{n-1} P(S_{n-l} = -i, S_{n-l-1} < 0, \cdots, S_1 < 0 | S_0 = 0)$$

$$P(S_l = 0, X_{l-1} = i_{l-1}, \cdots, X_1 = i_1, S_0 = 0)$$

$$= \sum_{l=0}^{n-1} \binom{n-l}{\frac{n-l-i}{2}} \frac{i}{n-l}\; p^{\frac{n-l-i}{2}}\; q^{\frac{n-l+i}{2}} P_{00}^{l}$$

$$= \sum_{k=0}^{m} \binom{n-2k}{\frac{n-2k-i}{2}} \frac{i}{n-2k}\; p^{\frac{n-2k-i}{2}}\; q^{\frac{n-2k+i}{2}} \binom{2k}{k} p^k q^k$$

$$= p^{\frac{n-i}{2}}\; q^{\frac{n+i}{2}} \sum_{k=0}^{m} \binom{n-2k}{\frac{n-2k-i}{2}} \binom{2k}{k} \frac{i}{n-2k}, \tag{2.9}$$

se $n-i$ é par. Observe que na quarta igualdade acima temos que $l = 2k$ e

$$m = \begin{cases} \frac{n-1}{2}, & \text{se } n-1 \text{ é par} \\ \frac{n-2}{2}, & \text{se } n-1 \text{ é ímpar.} \end{cases}$$

Ainda, $P(S_n = -i, B) = 0$ sempre que $n-i$ é ímpar.

Caso I) Suponha que $n-i \geq 0$ e $n-i$ é par. Das igualdades (2.7), (2.8) e (2.11) temos que

$$P(X_{n+1} = i+1 | X_n = i, B) = \frac{p p^{\frac{n+i}{2}} q^{\frac{n-i}{2}} + q p^{\frac{n-i}{2}} q^{\frac{n+i}{2}}}{p^{\frac{n+i}{2}} q^{\frac{n-i}{2}} + p^{\frac{n-i}{2}} q^{\frac{n+i}{2}}}$$

$$= \frac{p p^i + q q^i}{p^i + q^i} = \frac{p^{i+1} + q^{i+1}}{p^i + q^i},$$

para todo $i \geq 1$.

Como B é qualquer, considere

$$\widetilde{B} = \{X_{n-1} \in S, \cdots, X_1 \in S, X_0 \in S\}.$$

Então,

$$P(X_{n+1} = i+1 | X_n = i) = P(X_{n+1} = i+1 | X_n = i, \widetilde{B})$$

$$= \frac{p^{i+1} + q^{i+1}}{p^i + q^i} = P(X_{n+1} = i+1 | X_n = i, B), \text{ para todo } B.$$

Portanto, $(X_n)_{n \geq 0}$ é cadeia de Markov e suas *probabilidades de transição* são dadas por

$$P(X_{n+1} = 1 | X_n = 0) = 1$$

e

$$P(X_{n+1} = i+1 | X_n = i) = \frac{p^{i+1} + q^{i+1}}{p^i + q^i}, \text{ para todo } i \geq 1.$$

Caso II) Suponha que $n - i < 0$. Então, $n < i$. Observe que, saindo de 0 em tempo 0 não chegaremos a i em tempo n, ou seja, $P(X_n = i | S_0 = 0) = 0$. Logo, neste caso não faz sentido falar em $P(X_{n+1} = i+1 | X_n = i)$.

Caso III) Suponha que $n - i \geq 0$ e $n - i$ é ímpar. Então, $P(X_n = i | S_0 = 0) = 0$ e também não faz sentido falar em $P(X_{n+1} = i+1 | X_n = i)$.

2.8 Cadeias do Tipo Recorrente, Aperiódica e Irredutível

Seja $(X_n)_{n \geq 0}$ uma C.M. com espaço de estados S e matriz de transição $\mathcal{P}$. Lembre que uma distribuição de probabilidades $\pi = (\pi(s))_{s \in S}$ é dita **distribuição estacionária** para $(X_n)_{n \geq 0}$ se e só se $\pi = \pi\ \mathcal{P}$, ou seja, se para cada $j \in S$ fixo vale que $\pi(j) = \sum_{i \in S} \pi(i) P(i,j)$, $\forall j \in S$.

Questão Básica:

(1^{a}) Dada uma C.M. $\mathcal{P}$, quais são as condições necessárias (e suficientes) para que exista uma distribuição estacionária π para $\mathcal{P}$?

(2^{a}) Se existir, ela é única?

(3ª) Dado um vetor de probabilidade qualquer p sobre S, será que vale que

$$\lim_{n\to\infty} p\mathcal{P}^n = \pi,$$

onde π é único?

Observação: Anteriormente já exibimos exemplos em que:

1) $\mathcal{P}$ não é irredutível e não existe unicidade do π estacionário.

2) $\mathcal{P}$ é irredutível mas não recorrente e não existe π estacionário.

3) $\mathcal{P}$ é irredutível e aperiódica mas não existe π estacionário tal que para todo $p \in \Sigma$

$$\lim_{n\to\infty} p\mathcal{P}^n = \pi.$$

Isto nos leva a suspeitar que devemos supor que a cadeia é recorrente, irredutível e aperiódica para termos as propriedades desejadas acima em (1ª), (2ª) e (3ª).

Sabemos que para cada $n > 0$ e $i \in S$ fixos

$$P^n(i,i) = \sum_{k=0}^{n} f_{ii}^k P^{n-k}(i,i),$$

onde $f_{ii}^0 = 0$, $P^0(i,i) = 1$.

Então,

$$P^n(i,i) - \sum_{k=0}^{n} f_{ii}^k P^{n-k}(i,i) = \begin{cases} 0, & \text{se } n > 0; \\ 1, & \text{se } n = 0. \end{cases}$$

Ainda, para todo $n \geq 0$ e $i, j \in S$, vale que

$$P^n(i,j) = \sum_{k=0}^{n} f_{ij}^k \, P^{n-k}(j,j).$$

Teorema 2.23 (Equação da renovação em $\mathbb{N}$). *Sejam $(a_k)_{k\geq 0}, (b_k)_{k\geq 0}$ e $(c_k)_{k\geq 0}$ sequências de números reais tais que* *(a)* $a_k \geq 0, \quad \sum_{k\geq 0} a_k = 1,$

(b) $\sum_{k\geq 0} |\, b_k| < +\infty,$

(c) mdc $\{k : a_k > 0\} = 1.$

Se a equação $c_n - \sum_{k=0}^{n} a_k\, c_{n-k} = b_n, \quad \forall n \geq 0,$ *tiver uma solução* $(c_k)_k \geq 0$ *tal que* $sup|\, c_n| < +\infty$ *(i.é,* (c_n) *é limitada em* $\mathbb{R}$*), então*

(1) $\lim_{n\to\infty} c_n = \frac{\sum b_k}{\sum k\, a_k}$ *se* $\sum k\, a_k < +\infty$

(2) $\lim_{n\to\infty} c_n = 0$ *se* $\sum_k k\, a_k = +\infty.$

Demonstração: Referimos o leitor a [KT] para prova.

Teorema 2.24. *Considere* $(X_n)_{n\geq 0}$ *C.M. com espaço de estados* S *que seja irredutível, recorrente e aperiódica. Então,*

(a) $\lim_{n\to\infty} P^n(i,i) = \dfrac{1}{\sum_{k\geq 0} k\, f_{ii}^k} = \dfrac{1}{\mathbb{E}_i(T_i)} = \dfrac{1}{\mu_i} = \pi_i$ *se* $\sum_k k\, f_{ii}^k < +\infty$

(b) $\lim_{n\to\infty} P^n(i,i) = 0 = \pi_i$ *se* $\sum_k k\, f_{ii}^k = +\infty$

(c) $\forall i \in S, \quad \lim_{n\to\infty} P^n(i,j) = \lim_{n\to\infty} P^n(j,j) = \pi_j$

Demonstração:
(a) e (b) seguem da equação de renovação

$$P^n(i,i) - \sum_{k=0}^{n} f_{ii}^k \, P^{n-k}(i,i) = \begin{cases} 1, & se \quad n = 0 \\ 0, & se \quad n > 0 \end{cases} \qquad (*)$$

De fato, usando a notação do Teorema da Renovação, considere

$$b_n = \begin{cases} 1, & se \quad n = 0 \\ 0, & se \quad n > 0 \end{cases}$$

Note que $\sum_{n \geq 0} |b_n| = 1 < \infty$.

Tome $a_k = f_{ii}^k$. Observe que $\sum_{k \geq 0} a_k = \sum_{k \geq 0} f_{ii}^k = f_{ii}^* = 1$, pois i é recorrente.

Além disso, conforme Teorema 2.21 sabemos que o mdc de $\{n : f_{ii}^n > 0\} = 1$ (a cadeia é aperiódica).

Considere finalmente $c_n = P^n(i,i)$. Com estes valores a_n, c_n, b_n, a equação $*$ é a de renovação

$$c_n - \sum_{k=0}^{n} a_k \, c_{n-k} = b_n, \quad \forall n \geq 0.$$

Logo, pelo Teorema 2.23, temos

$$\lim_{n \to \infty} P^n(i,i) = \begin{cases} [\mathbb{E}_i(T_i)]^{-1}, & se \quad E_i(T_i) = \sum_{k \geq 0} k \, f_{ii}^k < +\infty \\ 0 \quad , & \text{caso contrário} \end{cases}$$

(c) Note que para todo $n \geq 0$ e $i, j \in S$, vale que

$$P^n(i,j) = \sum_{k=0}^{n} f_{ij}^k \, P^{n-k}(j,j).$$

Sejam i e j fixos.

Dos itens (a) e (b) sabemos que $\lim_{n\to\infty} P^{n-k}(j,j) = \pi_j$.

Seja $x_{n\,k} = f_{ij}^k P^{n-k}(j,j)$ se $n > k$ e $x_{n\,k} = 0$ caso contrário. Então, para k fixo

$$\lim_{n\to\infty} x_{n\,k} = f_{ij}^k \pi_j.$$

Ora,

$$\lim_{n\to\infty} P^n(i,j) = \lim_{n\to\infty} \sum_{k=0}^{n} f_{ij}^k\, P^{n-k}(j,j) = \lim_{n\to\infty} \sum_{k=0}^{\infty} x_{n\,k}.$$

Note que

$$\sum_{k=1}^{\infty} \sup_n |\, x_{n\,k} \,| \, \leq 1 < \infty.$$

Logo, pela Proposição 2.16

$$\lim_{n\to\infty} P^n(i,j) = \lim_{n\to\infty} \sum_{k=0}^{\infty} x_{n\,k} = \sum_{k=0}^{\infty} f_{ii}^k \pi_j = \pi_j,$$

pois i é recorrente.

$\square$

Teorema 2.25. *Seja $(X_n)_{n\geq 0}$ C.M. finita com matriz de transição $\mathcal{P}$ com uma classe C_1 de estados recorrentes e aperiódica, sendo os outros estados transitórios. Então $(X_n)_{n\geq 0}$ tem uma única distribuição estacionária dada por*

$$\begin{aligned} \pi(j) &= \lim_{n\to\infty} P^n(i,j) \quad \textit{se} \quad j \in C_1, \ e \\ \pi(j) &= 0 \quad \textit{se} \quad j \notin C_1 \end{aligned}$$

Demonstração: Seja $C_1 \subset S$ uma classe fechada, aperiódica e recorrente. Seja $\mathcal{P}_1$ a matriz correspondente às transições entre os elementos de C_1. Isto pode ser feito porque C_1 é fechado. Seja $P_1^n(i,j)$ a entrada ij da matriz $\mathcal{P}_1^n$. Como $\mathcal{P}_1$ é irredutível recorrente, então

$$\lim_{n\to\infty} P_1^n(j,j) = \frac{1}{\mu_j} = \frac{1}{\mathbb{E}_j\,(T_j)} \quad \forall\, j \in C_1 \text{ e}$$
$$\lim_{n\to\infty} P_1^n(i,j) = \frac{1}{\mu_j} = \frac{1}{\mathbb{E}_j\,(T_j)}, \quad \forall\, i,\, j \in C_1.$$

Além disso,

$\pi^1(j) = \dfrac{1}{\mathbb{E}\,(T_j)}$ define uma distribuição de probabilidade sobre C_1 satisfazendo

$$\pi^1 = \pi^1\, \mathcal{P}_1 \text{ e é única.}$$

Considere π o vetor de probabilidade sobre S que coincide com π^1 sobre C_1 e é zero sobre S_T. Sabe-se também (ver teorema 2.11 e proposição 2.13) que $\lim_{n\to\infty} P^n(i,j) = 0$ se j é transiente, $\forall\, i \in S$.

Basta provar que $\lim_{n\to\infty} P^n(i,j) = \pi(j)$ quando i é transiente e $j \in C_1$. Mas $P^n(i,j) = \sum_{k=0}^n f_{ij}^k\, P^{n-k}(j,j)$ e $\sum_{n\geq 1} f_{ij}^n = 1$. A demonstração segue agora a mesma linha que foi utilizada no item c) do último teorema.

□

Esclarecemos ao leitor que quando S é infinito, pode acontecer de $\mathcal{P}$ ser irredutível aperiódica recorrente, mas ter apenas estados recorrentes nulos. Estes exemplos podem ocorrer, por exemplo, em cadeias de nascimento e morte, que serão descritas ao fim deste capítulo.

Proposição 2.23. *Seja $\mathcal{P}$ aperiódico. A propriedade de ser recorrente nulo, recorrente positivo é uma propriedade de classe, ou seja, se i possui tal propriedade e j é equivalente a i, então j também possui.*

Demonstração: Seja $\mathcal{P}$ aperiódico. Vamos mostrar apenas que se i é positivo recorrente e j equivalente, então j é positivo recorrente.

Seja n e m tais que $P^n(i,j) > 0$ e $P^m(j,i) > 0$.

Considere agora um $r \in \mathbb{N}$ qualquer.

Ora,

$$P^{n+r+m}(i,i) \geq P^n(i,j)\, P^r(j,j)\, P^m(j,i).$$

Considere r variável e m, n fixos.

Se i é positivo recorrente então

$$\lim_{r\to\infty} P^{n+r+m}(i,i) = \frac{1}{\mu_i} < \infty.$$

Logo,

$$\infty > \lim_{r\to\infty} P^{n+r+m}(i,i) \geq$$
$$P^n(i,j)\,(\,\lim_{r\to\infty} P^r(j,j)\,)\,P^m(j,i).$$

Sendo assim,

$$\frac{1}{\mu_j} \;=\; \lim_{r\to\infty} P^r(j,j) \;<\; \infty.$$

Desta forma, j é recorrente positivo.

□

Teorema 2.26. *Se S é finito irredutível e aperiódico (portanto é recorrente) então todo estado é recorrente positivo.*

Demonstração: De fato, para cada i fixo, como a soma é finita

$$1 = \lim_{n\to\infty} \sum_{j\in S} P^n(i,j) = \sum_{j\in S} \lim_{n\to\infty} P^n(i,j).$$

Sendo assim, fixado i não pode valer para todo j

$$\lim_{n\to\infty} P^n(i,j) = 0.$$

□

Teorema 2.27. *Seja $(X_n)_{n\geq 0}$ com matriz de transição $\mathcal{P}$, irredutível, aperiódica, recorrente positiva. Então existe e é único o π tal que $\pi\,\mathcal{P} = \pi$. Ainda, se $\pi = (\pi_i)_{i\in S}$ então $\pi_i = \frac{1}{\mu_i}$, onde μ_i é o tempo de retorno médio ao estado i.*

Referimos o leitor para uma prova em [KT].

Exemplo 2.46. A matriz

$$\mathcal{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

é irredutível tem período 3 e tem $\pi = (1/3, 1/3, 1/3)$ como vetor invariante. Isto segue do fato que $\mu_i = 1/3$ para $i \in S$. Note que não existe o limite

$$\lim_{n\to\infty} P_{11}^n.$$

$\diamondsuit$

Observação: Quando P é irredutível e periódica devemos considerar a cadeia induzida associada a P^d, onde d é o período. Esta será aperiódica. Então se pode utilizar os resultados anteriores.

Teorema 2.28. *Se a cadeia de Markov definida por $\mathcal{P}$ é irredutível recorrente e de período d, então*

$$\lim_{n\to\infty} P_{ii}^{nd} = \frac{d}{\mu_i}.$$

Ainda $P_{ii}^m = 0$ quando m não é múltiplo de d.

Referimos o leitor para [KT] para uma prova desta afirmação.

Exemplo 2.47. 1) Seja $\mathcal{P} = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$

a) Vamos mostrar que a C.M. associada é irredutível, recorrente positiva e aperiódica.

b) Vamos obter a distribuição estacionária π.

a) É fácil ver que $C_0 = \{0, 1, 2\}$. Logo, a cadeia é irredutível

$d(0) =$ mdc $\{1, 3, 4, 6, 7, 9, \cdots\} = 1$

$d(1) =$ mdc $\{3, 4, 5, 6, 7, 8, 9, \cdots\} = 1 = d(2)$

$f^*_{00} = \sum_{n \geq 1} f^n_{00} = \frac{1}{2} + 0 + \frac{1}{2} 1.1 + 0 + 0 + \cdots = \frac{1}{2} + \frac{1}{2} = 1$

Logo, os estados são recorrentes.

Como $P_i(T_i = \infty) = 0$, concluímos que $\mu_i = \sum_{n \geq 1} n\, f^n_{ii}$ e

$$\mu_0 = \sum_{n \geq 1} n\, f^n_{00} = 1.\frac{1}{2} + 2\,.\,0 + 3.\,\frac{1}{2}.1.1 + 0 + 0 \cdots =$$

$$\frac{1}{2} + \frac{3}{2} = 2 < \infty.$$

Logo, a cadeia é irredutível, recorrente positiva e aperiódica.

b) Desejamos obter agora a distribuição invariante π.

A equação $\pi\ \mathcal{P} = \pi$ é equivalente a

$$(\pi(0)\ \pi(1)\ \pi(2)) \begin{pmatrix} 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = (\pi(0)\, \pi(1)\, \pi(2)),$$

$\pi(0)\,\frac{1}{2} + \pi(1) = \pi(0) \Rightarrow \pi(1) = \left(1 - \frac{1}{2}\right) \pi(0) \Rightarrow \pi(1) = \frac{1}{2}\,\pi(0)\ ;$

$\pi(2) = \pi(1)\ ;$

$\pi(0)\,\frac{1}{2} = \pi(2).$

Então, vale

$$\pi(0) = 2\pi(1) = 2\pi(2) \quad \text{e} \quad \pi(0) + \pi(1) + \pi(2) = 1.$$

Portanto,

$$\pi(0) + \pi(1) + \pi(1) = 1 \ \Rightarrow \ \pi(0) + 2\pi(1) = 1 \ .$$

Como, $\pi(0) = 2\pi(1)$ e $\pi(0) = 1 - 2\pi(1)$, temos que $4\pi(1) = 1 \ \Rightarrow \ \pi(1) = \frac{1}{4}$.

Logo,

$$\pi = \left(\frac{2}{4}, \frac{1}{4}, \frac{1}{4}\right)$$

Desta forma,

$$\pi(0) = \frac{1}{\mathbb{E}_0(T_0)} = \frac{1}{2}; \quad \pi(1) = \frac{1}{\mathbb{E}_1(T_1)} = \frac{1}{4}; \quad \pi(2) = \frac{1}{\mathbb{E}_2(T_2)} = \frac{1}{4}.$$

$\diamondsuit$

2.9 Tempo de Parada e a Propriedade Forte de Markov

Definição 2.23. *Dado um processo estocástico X_n, $n \in \mathbb{N}$, uma variável aleatória (extendida, ou seja, pode tomar o valor ∞)*

$$T : \Omega \to \{0, 1, 2, 3, ..., n, ...\} \cup \{\infty\},$$

é chamado de tempo de parada (para o processo X_n) se para cada $n \in \mathbb{N}$ fixo, o conjunto $\{w \,|\, T(w) = n\}$ depende apenas dos valores

$$X_0(w), X_1(w), X_2(w), ..., X_n(w).$$

É claro que assumimos que $Z = \{\omega \,:\, T(w) = \infty\}$ é um conjunto mensurável. Z pode ter ou não probabilidade nula.

Exemplo 2.48. O primeiro tempo de visita a $i \in S$ começando em $j \in S$

$$T_{ij} = \inf\{n \geq 1, X_n = i, X_0 = j\},$$

é um tempo de parada, pois para cada $n \geq 0$

$$\{T_{ij} = n\} = \{X_0 = j, X_1 \neq i, X_2 \neq i, ..., X_{n-1} \neq i, X_n = i\}.$$

$\diamondsuit$

Contra-exemplo: Seja $S = \{1, 2, 3\}$ e $T^* : \Omega \to \mathbb{N} \cup \{\infty\}$ tal que $T^*(\omega) = n$ se o tempo n é a primeira vez que aparece 1 seguido por 2 (no tempo $n+1$). Esta variável aleatória não é um tempo de parada.

Contra-exemplo: Fixe $i \in S$ e defina $\tilde{T}_i(w)$ como o último tempo n que $\omega = (w_t)_{t \in \mathbb{N}}$ visita i e $\tilde{T}_i(w) = \infty$ caso ω visite i infinitas vezes. Tal $\tilde{T}_i$ é uma variável aleatória estendida mas não é um tempo de parada pois para determinar se $\tilde{T}_i(w) = 3$ necessitamos conhecer o caminho $\omega = (w_t)$ em tempos t maiores do que 3.

Antes de apresentar a demonstração da propriedade forte de Markov necessitamos da seguinte versão da Proposição 2.7.

Proposição 2.24. *Fixe $i \in S$. Sejam os conjuntos $A_0, A_1, A_2, ..., A_{n-1}, A_n = \{i\}, A_{n+1}, ..., A_m \subset S$, onde $n < m$. Então vale que*

$$P(X_m \in A_m, X_{m-1} \in A_{m-1}, \ldots, X_{n+1} \in A_{n+1},\, X_n = i,\, X_{n-1} \in A_{n-1},$$

$$X_{n-2} \in A_{n-2}, \ldots, X_1 \in A_1, X_0 \in A_0) =$$

$$P_i(X_{m-n} \in A_m, X_{m-n-1} \in A_{m-1}, \ldots, X_1 \in A_{n+1}) \times$$

$$P(X_n = i, X_{n-1} \in A_{n-1}, X_{n-2} \in A_{n-2} \ldots, X_1 \in A_1, X_0 \in A_0).$$

Demonstração: Segue de imediato da Proposição 2.8. De fato, sejam os conjuntos $A_0, A_1, A_2, ..., A_{n-1}, A_n, A_{n+1}..., A_m \subset S$, onde $n < m$, então pela Proposição2.8

$$P(X_m \in A_m, X_{m-1} \in A_{m-1}, \ldots, X_{n+1} \in A_{n+1} | X_n \in A_n, X_{n-1} \in A_{n-1},$$

$$\ldots, X_1 \in A_1, X_0 \in A_0) =$$

$$P(X_m \in A_m, X_{m-1} \in A_{m-1}, \ldots, X_{n+1} \in A_{n+1} | X_n \in A_n) =$$

$$P(X_{m-n} \in A_m, X_{m-n-1} \in A_{m-1}, \ldots, X_1 \in A_{n+1} | X_0 \in A_n).$$

Isto porque, por definição

$$P(X_m \in A_m, X_{m-1} \in A_{m-1}, \ldots, X_{n+1} \in A_{n+1} \,|\, X_n \in A_n, X_{n-1} \in A_{n-1},$$

$$\ldots, X_1 \in A_1, X_0 \in A_0) =$$

$$P(X_m \in A_m, X_{m-1} \in A_{m-1}, \ldots, X_n \in A_n,\ X_{n-1} \in A_{n-1}, ..., X_1 \in A_1, X_0 \in A_0) \times$$

$$\frac{1}{P(X_n \in A_n, X_{n-1} \in A_{n-1}, \ldots, X_1 \in A_1, X_0 \in A_0)}\ .$$

□

Necessitamos na verdade de uma versão levemente mais sofisticada do resultado acima:

Proposição 2.25. *Fixe $i \in S$. Seja o conjunto $B \in \mathcal{F}$, onde $\mathcal{F}$ é a sigma-álgebra gerada pelos cilindros. Considere agora $n \in \mathbb{N}$, $i \in S$ e conjuntos $A_j \subset S$ fixos, $j \in \{1, 2, ..., n\}$. Assumimos que o conjunto $A_n = \{i\}$ e que o conjunto B não dependa das variáveis $X_0, X_1, ..., X_{n-1}$. Então, para n e $i \in S$ fixos vale que*

$$P(\,(X_{n+1}, X_{n+2}, X_{n+3}, ...) \in B\,,\ X_n = i,\ X_{n-1} \in A_{n-1},$$

$$X_{n-2} \in A_{n-2}, \ldots, X_1 \in A_1, X_0 \in A_0\,) =$$

$$P_i(\,(X_{n+1}, X_{n+2}, X_{n+3}, ...) \in B) \times$$

$$P(\,X_n = i,\, X_{n-1} \in A_{n-1},$$

$$X_{n-2} \in A_{n-2}, \ldots, X_1 \in A_1, X_0 \in A_0\,) =$$

$$P_i(\;(X_1, X_2, X_3, ...) \in B\,)\times$$

$$P(X_n = i, X_{n-1} \in A_{n-1}, X_{n-2} \in A_{n-2}, \ldots, X_1 \in A_1, X_0 \in A_0).$$

Demonstração: Ora, quando B é um cilindro da forma

$$B = (X_{n+1} = a_0, X_{n+2} = a_1, .., X_{n+r+1} = a_r),$$

com $r \in \mathbb{N}$ o resultado acima é válido pela Proposição anterior.

Sendo assim, para cilindros B podemos considerar

$$\hat{P}_a(B) = P(\,(X_{n+1}, X_{n+2}, X_{n+3}, ...) \in B\,,\, X_n = i,\, X_{n-1} \in A_{n-1},$$

$$X_{n-2} \in A_{n-2}, \ldots, X_1 \in A_1, X_0 \in A_0),$$

e

$$\hat{P}_b(B) = P_i(\,(X_1, X_2, X_3, ...) \in B)\times$$

$$P(X_n = i, X_{n-1} \in A_{n-1}, X_{n-2} \in A_{n-2}, \ldots, X_1 \in A_1, X_0 \in A_0).$$

Definimos assim duas probabilidades que coincidem sobre cilindros.

Com o mesmo tipo de resultado do Teorema da Extensão de Caratheodori-Kolmogorov podemos estender esta igualdade para a sigma algebra $\mathcal{F}$ gerada pelos cilindros. Ou seja, se $\hat{P}_a$ e $\hat{P}_b$ coincidem sobre cilindros, elas coincidem sobre a sigma-álgebra gerada pelos cilindros.

$\square$

O último teorema afirma que sejam quais forem as restrições após o tempo n, quando condicionamos em i no tempo n, o processo evolui de maneira independente do que aconteceu previamente ao tempo n. Ou seja, ele perde memória do que aconteceu antes do tempo n. Denominamos tal propriedade de **fraca de Markov**.

A proposição acima permite formalizar o seguinte exemplo de maneira apropriada.

Exemplo 2.49. Considere uma pessoa que possui um capital inicial $c > 0$, onde c é um número natural, e que vai participar de um jogo em que uma moeda (com probabilidade $1/2$ de sair cara e $1/2$ de sair coroa) é lançada sucessivamente.

O jogador ganha um real se sai cara e perde um real se sai coroa.

Ele gostaria de atingir um capital fixado $C > c > 0$. O jogo termina quando a fortuna do jogador atinge o valor C ou o valor 0. É natural supor que se o capital inicial c é grande e próximo de C, então existe maior probabilidade do jogador atingir seu objetivo de alcançar a fortuna C do que quando o capital inicial c for próximo de zero. Como quantificar tais probabilidades? Vamos denotar por $p(c)$ a probabilidade do jogador entrar em bancarrota se inicia com capital c, ou seja, atingir ao longo do jogo a fortuna 0 (antes de atingir o valor C).

Em princípio, apenas sabemos que $p(0) = 1$ e $p(C) = 0$.

Vamos modelar este problema através de um processo estocástico. Considere c fixado. Neste caso, tome o espaço de estados $S = \{0, 1, 2, ..., C\}$, $T = \mathbb{N}$, X_t é o valor da fortuna no tempo t. Considere $\Omega = \{1, 2, ..., C\}^{\mathbb{N}}$ e a probabilidade $P = P_c$ será descrita a seguir. O ponto fundamental é que vamos evitar dizer quem é P explicitamente.

Note que $P(\{w|X_0 = c\}) = 1$, $P(\{w|X_0 \neq c\}) = 0$. Ainda, $P(X_1 = c) = 0$.

É mais natural descrever P através de condicionais, mais exatamente,

$$P(X_{t+1} = d + 1 \; |X_t = d) = 1/2,$$

$$P(X_{t+1} = d - 1 \; |X_t = d) = 1/2,$$

para qualquer $t \geq 1$ e $1 \leq d \leq C-1$.

É claro que segue do estabelecido acima que fixado d, para qualquer $b < d-1$, ou $b > d+1$, ou mesmo $b = d$, vale que

$$P(X_{t+1} = b \,|X_t = d) = 0.$$

Ainda, é natural assumir que

$$P(X_{t+1} = 0 \,|\, X_t = 0) = 1,$$

$$P(X_{t+1} = C \,|\, X_t = C) = 1.$$

Isto é, o jogo termina ao ser alcançado um dos valores 0 ou C.

Note que estas informações implícitas e ainda a informação $P(X_0 = c) = 1$ são suficientes para calcular o valor de $P(A)$ para um conjunto qualquer $A \subset \Omega = \{1, 2, ..., C\}^{\mathbb{N}}$ que depende apenas de finitas restrições temporais.

De fato, considerando c fixado,

$$P(X_0 = c\,,\, X_1 = c+1) = P(X_1 = c+1 \,|\, X_0 = c)\, p(X_0 = c) = \frac{1}{2}\,.\,1 = \frac{1}{2}.$$

Note que é natural no presente exemplo que, para $i \geq 0$, e $d_i, d_{i+1}, d_{i+2} \in S$ valha

$$P(X_{i+2} = d_{i+2} \,|\, X_i = d_i,\, X_{i+1} = d_{i+1}) = P(X_{i+2} = d_{i+2} \,|\, X_{i+1} = d_{i+1}),$$

pois o valor de X_i não influencia na probabilidade de $X_{i+2} = d_{i+2}$. É claro que a probabilidade de $X_{i+2} = d_{i+2}$ é influenciada pelo valor $X_{i+1} = d_{i+1}$. Por exemplo,

$$P(X_{i+2} = 5 \,|\, X_{i+1} = 3) = 0,$$

mas

$$P(X_{i+2} = 5 \,|\, X_{i+1} = 4) = \frac{1}{2}.$$

Ainda, no caso em que $c = 1$,

$$P(\{w|X_0(w) = 1, X_1(w) = 2, X_2(w) = 1\}) = (\frac{1}{2})^2.$$

É usual denotar a probabilidade P quando assumimos que $X_0 = c$, ou seja, quando o capital inicial for c, por P_c. Este ponto de vista será importante a seguir. Vamos analisar a função $p(c)$ como função de c.

Para cada c, em função das condições de compatibilidade considere pelo teorema de Caratheodori-Kolmogorov o processo $X_t : \Omega \to S$, $t \in \mathbb{N}$, P_c, a probabilidade sobre $\Omega = \{0, 1, 2, ..., C\}^{\mathbb{N}}$ e a sigma-algebra $\mathcal{F}$ (de certos subconjuntos de Ω). A sigma-algebra $\mathcal{F}$ não depende de c.

Lembre que as famílias de variáveis aleatórias X_t que consideramos aqui sempre satisfazem a propriedade: se $w = (w_0, w_1, w_2, ..., w_n, ...)$ então $X_t(w) = w_t$.

Nosso objetivo é resolver o problema: quem é $p(c)$? Ou seja, calcular a probabilidade $P_c(B)$ do conjunto

$$B_c = B = \{w = (w_0, w_1, w_2, w_3, ..., w_t, ...) \text{ tal que } w_0 = c,\ w_t = 0$$

$$\text{para algum } t > 0, \text{ e ainda } w_r \neq C \text{ para todo } r \text{ tal que } 0 < r < t\ \} =$$

$$\{w\ :\ X_0(w) = c,\ X_t(w) = 0\ \text{ para algum } t \text{ e } X_r \neq C \text{ para } 0 < r < t\}.$$

Note que o conjunto B_c depende de infinitas informações.

Este conjunto B_c, está na sigma-algebra $\mathcal{F}$. Fica claro, deste modo, a necessidade de se considerar na teoria uma sigma-algebra na qual fique bem definido o conceito de probabilidade. Apenas as distribuições finito-dimensionais não bastam e conjuntos mais complexos, que dependem de infinitas coordenadas, aparecem de maneira natural em problemas concretos.

Existem eventos $w \in \Omega$ que nunca atingem qualquer dos valores C ou 0. Por exemplo, no caso $c = 2$ e $C = 5$, o evento $w = (w_0, w_1, w_2, w_3, ..., w_t, ..) = (2, 3, 2, 3, 2, 3, 2, 3, 2, 3, ...)$ possui tal propriedade.

Considere o conjunto

$$D = D_c = \{w = (w_0, w_1, w_2, w_3, ..., w_t, ...) \in \Omega = \{0, 1, 2, ..., C\}^{\mathbb{N}},$$

$$\text{tal que } w_0 = c,\ w_t \neq C \text{ e } w_t \neq 0 \text{ para todo } t \in \mathbb{N}\,\}.$$

O conjunto D está também em $\mathcal{F}$ e $P_c(D) = 0$. Este resultado foi obtido quando foi analisado o assim chamado passeio aleatório com $p = 1/2$. De fato, com probabilidade 1, começando em c, o processo atinge qualquer ponto em $\mathbb{Z}$. Logo, restrito a $S = \{0, 1, \ldots, C\}$, começando em c o processo atinge 0 ou C, com probabilidade 1.

Para ilustrar tal afirmação ao leitor vamos analisar um caso particular. Considere $C = 3$ e fixemos $c = 1$, então o conjunto D acima tem apenas um elemento

$$D = \{(1, 2, 1, 2, 1, 2, 1, 2, 1, 2, ...)\}.$$

Destacamos aqui o fato que um conjunto com apenas um elemento pode ter probabilidade positiva.

Note que

$$D = \cap_{n=1}^{\infty} \{w : X_0 = 1, X_1 = 2, X_2 = 1, X_3 = 2, ..., X_{2n-1} = 2, X_{2n} = 1\},$$

e portanto D é um conjunto mensurável em $\mathcal{F}$ (obtido por interseção de conjuntos de $\mathcal{F}$ indexados por um conjunto enumerável $n \in N$) e assim faz sentido perguntar pelo valor $P(D)$.

Como vale a propriedade que se $V \subset U$, então $P(V) \leq P(U)$ (pois, $P(V) \leq (P(V) + P(U - V)) = P(U)$), então, dado $n \in \mathbb{N}$,

$$P(D) \leq P(\{w : X_0 = 1, X_1 = 2, X_2 = 1, X_3 = 2, ...,$$

$$X_{2n-1} = 2, X_{2n} = 1\}) = (\frac{1}{2})^{2n-1},$$

para todo n. Note que tomando n grande o valor $(\frac{1}{2})^{2n-1}$ se torna arbitrariamente pequeno. Logo $P(D) = 0$.

$\diamondsuit$

Observe que a demonstração acima é bem geral, na verdade este procedimento mostra que dados uma sequência de conjuntos $A_0, A_1, A_2, A_3, ..., A_n, ..$ tais que

$$... \subset A_n \subset A_{n-1} \subset A_{n-2} \subset ... \subset A_4 \subset A_3 \subset A_2 \subset A_1 \subset A_0,$$

e

$$D = \cap_{n\in\mathbb{N}} A_n,$$

se $\lim_{n\to\infty} P(A_n) = 0$, então $P(D) = 0$.

Uma fácil generalização disto mostra que se $D = \cap_{n\in\mathbb{N}} A_n$, e

$$... \subset A_n \subset A_{n-1} \subset ... \subset A_2 \subset A_1 \subset A_0,$$

então $P(D) = \lim_{n\to\infty} P(A_n)$. Esta é a assim chamada **propriedade monótona** e que é válida para qualquer probabilidade P.

Esclarecidos estes pontos de grande importância, vamos voltar ao nosso problema.

Considere fixado o número natural c. Segue da regra de Bayes que para c tal que $C > c > 0$, vale que

$$p(c) = P_c(B_c) = P_c(\{w \in B_c\}) =$$

$$P_c(\{w \in B_c \,:\, X_1 = c+1\})\, P_c(\{X_1 = c+1\}) +$$

$$P_c(\{w \in B_c \,:\, X_1 = c-1\})\, P_c(\{X_1 = c-1\}) =$$

$$P_c(\{w \in B_c \,:\, X_1 = c+1\})\, \frac{1}{2} \,+\, P_c(\{w \in B_c \,:\, X_1 = c-1\})\, \frac{1}{2}.$$

O ponto fundamental agora é que para c tal que $0 < c < C$, vale que $p(c+1) = P_c(\{w \in B_c \,:\, X_1 = c+1\})$ e $p(c-1) = P_c(\{w \in B_c \,:\, X_1 = c-1\})$.

A afirmação acima, que é absolutamente intuitiva, requer uma prova, a qual será apresentada em breve.

Obtemos assim à equação de diferenças

$$P(\{w\,:\,X_0(w)=c,\,X_t(w)=0 \text{ para algum } t\})=p(c)=\frac{1}{2}(p(c-1)+p(c+1)),$$

com a condição inicial e final, respectivamente, $p(0)=1$ e $p(C)=0$. A solução desta equação é

$$p(c)=1-\frac{c}{C},$$

como pode ser confirmado por substituição na equação.

Esta fórmula dá o valor exato da dependência da probabilidade $p(c)$ em função da proximidade de c a C.

Vamos mostrar como se obtem a solução $p(c)=1-\frac{c}{C}$, da equação de diferenças

$$p(c)=\frac{1}{2}(p(c-1)+p(c+1)),$$

sujeita a condição de contorno $p(0)=1$ e $p(C)=0$.

Denote por $a_n=p(n)-p(n-1)$.

Ora,

$$p(n-1)+p(n-1)=2\,p(n-1)=(p(n-2)+p(n)),$$

então,

$$a_{n-1}=p(n-1)-p(n-2)=p(n)-p(n-1)=a_n.$$

Sendo assim, procedendo de maneira indutiva a partir de a_1, obtemos que $p(n)=n\,a_1\,+\,p(0)$.

Como $p(0)=1$ e

$$0=p(C)=\,Ca_1+1,$$

temos que

$$a_1=-\frac{1}{C}.$$

Logo,

$$p(c)=\,c\,\times(-\frac{1}{C})+1.$$

Em conclusão: utilizando a Regra de Bayes, ou seja condicionando, obtivemos uma equação de diferenças e a partir daí obtivemos a solução do problema que buscávamos. Não foi necessário calcular a probabilidade de nenhum conjunto específico! Variações destas ideias nos permitem obter soluções explícitas do valor de certas probabilidades que se deseja encontrar em inúmeros casos.

Uma das Regras de Ouro da Probabilidade: na dúvida, condicione!

Vamos agora apresentar o argumento que deixamos para depois nas considerações do exemplo acima. Usando a notação deste e da última proposição acima, considere sobre $S = \{0, 1, 2, ..., C\}$ o conjunto $B = B_c$ descrito acima.

Sendo assim, supondo $c+1 < C$, a partir da última proposição, temos que

$$P(\,(X_2, X_3, X_4, ...) \in B,\, X_1 = c+1,\, X_0 = c) =$$

$$P_{c+1}(\,(\,X_2,\, X_3,\, X_4, ...) \in B)\; \times P(\,X_1 = c+1,\, X_0 = c) =$$

$$P_{c+1}(\,(\,X_2,\, X_3,\, X_4, ...) \in B)\; \times\; P_{c\,,\,c+1}.$$

Logo,

$$P_c\,(\omega \in B \,|\, X_1 = c+1) = P_{c+1}(\,(c+1,\, X_2,\, X_3,\, X_4, ...) \in B)\; = \; p(c+1).$$

Exercício: Considere o mesmo problema acima, só que desta vez assuma que a probabilidade de ganhar é p, onde $1 > p > 0$, e não apenas $1/2$. De maneira análoga, denote por $p(c)$ a probabilidade de bancarrota, quando se começa com capital c. Condicionando, determine a equação

$$p(c) \; = \; p\, p(c+1) \; + \; (1-p)\, p(c-1).$$

A seguir resolva a equação de diferenças (com a condição de fronteira $p(0) = 1$ e $p(C) = 0$) e encontre

$$p(c) \; = \; \frac{[(1-p)/p]^c - [(1-p)/p]^C}{1 - [(1-p)/p]^C}.$$

Exercício: Considere dois jogadores 1 e 2 com o capital inicial do jogador 1 igual a $a > 0$ enquanto que $b > 0$ é o capital inicial do jogador 2, onde a e b são números naturais positivos. O total de capital é $d = a + b$. Em cada jogada o jogador 1 tem chance p de ganhar 1 real e tem chance q de perder 1 real, onde $p + q = 1$. O jogador 2, por sua vez, tem chance q de ganhar 1 e p de perder 1 real em cada jogada.

Seja X_n o capital do jogador 1 na n-ésima jogada.

O jogo termina quando um dos jogadores está arruinado. Perguntas de interesse:

1) O jogo acaba sempre? Isto é será que $K = \{w = (w_t) \in \Omega$, tal que existe t satisfazendo $w_t \in \{0, d\}\, \}$ satisfaz a propriedade $P(K) = 1$?

2) Qual é o valor P(o jogador 1 ficar arruinado)? Isto é, ao definir $V = \{w = (w_t) \in \Omega$, tal existe t satisfazendo $w_t \in \{0\}\, \}$, se pergunta: qual o valor $P(V)$?

3) Qual é o valor P(o jogador 2 ficar arruinado)? Isto é se $U = \{w = (w_t) \in \Omega$, tal que existe t tal que $w_t \in \{d\}\, \}$, qual o valor $P(U)$?

Dado um processo estocástico $X_n : \Omega \to \mathbb{R}$, (uma probabilidade sobre $S^{\mathbb{N}}$) onde para cada $w = (w_0, w_1, w_2, ..., w_n, ...) \in S^{\mathbb{N}}$, temos que $X_n(w) = w_n$, e um tempo de parada T. O processo estocástico X_T é naturalmente definido através das funções $(X_T)_n$, $n \in \mathbb{N}$:

$(X_T)_n\,(w) = X_n(w)$, para $n < T(w)$, e $(X_T)_n(w) = X_{T(w)}$ para $n \geq T(w)$.

Fica assim definida uma função $G : \Omega \to \Omega$ tal que

$$G(w) = (w_0, w_1, .., w_{T(w)-1}, w_{T(w)}, w_{T(w)}, ..., w_{T(w)}, ...).$$

A um processo estocstico está associada uma probabilidade P_T sobre $S^{\mathbb{N}}$.

Assim, precisamos definir P_T sobre cilindros.

Suponha que P seja a probabilidade sobre $S^{\mathbb{N}}$ associada ao processo estocástico $X_n : \Omega \to \mathbb{R}$.

Assim, estabelecemos que

$$P_T\left((X_T)_0 = a_0, (X_T)_1 = a_1, ..., (X_T)_k = a_k\right) =$$

$$P(\, G^{-1}\left(\{X_0 = a_0, X_1 = a_1, ..., X_k = a_k\,\}\,\right)\,).$$

Fica assim definida a probabilidade P_T sobre $S^{\mathbb{N}}$.

Seja T tempo de parada e X_n, $n \in \mathbb{N}$, processo de Markov, vamos mostrar um resultado geral a seguir que como caso particular diz: seja $S = \{1, 2, 3, 4\}$, então

$$P_T(X_{T+1} = 2, X_{T+2} = 3, X_{T+3} = 2 \,|\, X_T = 1) =$$

$$P(X_1 = 2, X_2 = 3, X_3 = 2 \,|\, X_0 = 1).$$

Teorema 2.29 (Propriedade Forte de Markov). *Seja X_n Processo de Markov obtido a partir de $\mathcal{P}$ e π. Considere T tempo de parada. Suponha que $P(T < \infty) = 1$. Então, condicionando em $X_T = i$, o Processo Estocástico $\{Y_m\}_{m\in\mathbb{N}}$ tal que $Y_m = X_{T+m}$ se torna um Processo de Markov (com probabilidade inicial e_i e matriz de transição $\mathcal{P}$).*

Mais precisamente,

$$P(X_{T+1} = s_1, ..., X_{T+m} = s_m | X_T = i) =$$

$$P(X_1 = s_1, ..., X_m = s_m \,|\, X_0 = i).$$

Demonstração: Seja n fixo.

Para facilitar o entendimento do leitor:

$$\{T = n\} \cap \{X_T = i\} = \{X_0 \in \Omega - \{i\}, X_1 \in \Omega - \{i\}, ..., X_{n-1} \in \Omega - \{i\}, X_n = i\}.$$

Agora seguimos com a demonstração.

Vamos considerar a seguir os conjuntos $A_r = \Omega - \{i\}$, onde $n > r \geq 0$, no resultado anterior.

Para n e m fixos, pela propriedade (fraca) de Markov descrita pela última proposição

$$P(\{X_T = s_0, X_{T+1} = s_1, ..., X_{T+m} = s_m\} \cap \{T = n\} \cap \{X_T = i\}) =$$
$$P(X_T = s_0, ..., X_{T+m} = s_m \,|\, T = n, X_T = i)\, P(\{T = n\} \cap \{X_T = i\}) =$$
$$P_i(\{X_0 = s_0, X_1 = s_1, ..., X_m = s_m\}\,)\, P(\{T = n\} \cap \{X_T = i\}\,).$$

Agora somando sobre $n \in \mathbb{N}$ obtemos

$$P(\{X_T = s_0, X_{T+1} = s_1, ..., X_{T+m} = s_m\} \mid \{T < \infty\} \cap \{X_T = i\}\,) =$$
$$\frac{P(\{X_T = s_0, X_{T+1} = s_1, ..., X_{T+m} = s_m\} \cap \{T < \infty\} \cap \{X_T = i\}\,)}{P(\{T < \infty\} \cap \{X_T = i\})} =$$
$$\sum_{n\in\mathbb{N}} \frac{P(\{X_T = s_0, X_{T+1} = s_1, ..., X_{T+m} = s_m\} \cap \{T = n\} \cap \{X_T = i\}\,)}{P(\{T < \infty\} \cap \{X_T = i\})} =$$
$$\sum_{n\in\mathbb{N}} P_i(X_0 = s_0, X_1 = s_1, ..., X_m = s_m) \times$$
$$\frac{P(\{T = n\} \cap \{X_T = i\}\,)}{P(\{T < \infty\} \cap \{X_T = i\}\,)} =$$
$$P_i(X_0 = s_0, X_1 = s_1, ..., X_m = s_m) \sum_{n\in\mathbb{N}} \frac{P(\{T = n\} \cap \{X_T = i\}\,)}{P(\{T < \infty\} \cap \{X_T = i\})} =$$
$$P_i(X_0 = s_0, X_1 = s_1, ..., X_m = s_m).$$

□

Corolário 2.7. *Propriedade Forte de Markov - Seja X_n Processo de Markov obtido a partir de $\mathcal{P}$ e π. Seja i fixo e suponha que T_i é o tempo de primeiro retorno a i começando em i. Então condicionado a T_i e $T_i < \infty$, o Processo Estocástico $\{Y_n\}_{n\in\mathbb{N}}$ dado por $Y_n = X_{T_i+n}$ satisfaz*

$$P(X_{T_i+1} = s_1, ..., X_{T_i+n} = s_n \mid X_{T_i} = i\,,\, T_i < \infty) =$$
$$P(X_0 = s_0, X_1 = s_1, ..., X_n = s_n \,|\, X_0 = i\,,\, T_i < \infty).$$

A demonstração é basicamente a mesma do teorema anterior.

Note que, o Teorema acima não vale para T que não seja tempo de parada. De fato, seja $S = \{1, 2, 3\}$ e $\mathcal{P}$ tal que tem todas as entradas positivas. Considere T^* tal que $T^*(\omega) = n$ se no tempo n é a primeira vez que aparece 1 seguido por 2 (no tempo $n + 1$).

Então, se valesse o Teorema obteríamos, em particular, que vale

$$0 = P(X_{T^*+1} = 3 \mid X_{T^*} = 1) =$$

$$P(X_1 = 3 \mid X_0 = 1) = P_{13} > 0,$$

o que é uma contradição.

Este resultado nos mostra que é necessário ser bastante cuidadoso no uso de nossa intuição. Olhando de maneira superficial, somos levados a crer pela propriedade "fraca de Markov" que após atingir no tempo aleatório T o valor 1, o processo começa a partir daí com completa perda de memória do passado e com a lei inicial determinada pela cadeia de Markov. Isto só vale se o tempo aleatório T é um tempo de parada. A denominação propriedade forte de Markov se deve ao fato de utilizarmos no resultado acima um tempo de parada.

Proposição 2.26. *Seja $\mathcal{P}$ cadeia de Markov sobre S e considere a decomposição $S = S_T \cup S_R = S_T \cup C_1 \cup C_2 \cup C_3 \cup ...$, onde C_r são conjuntos recorrentes irredutíveis fechados. Seja $T(w)$ o tempo de parada que é o primeiro tempo t tal que $\omega = (w_t)$ atinge $S - S_T$. Suponha que $P(T(\omega) < \infty) = 1$.*

Se ω deixa S_T então ele vai atingir um certo C_r. Para C_r fixo, vale

$$P(X_{T+n} \in C_r, \ \text{para todo } n \geq 1 \mid X_T \in C_r) = 1.$$

Demonstração: Primeiro, note que

$$P(X_1 \in C_r \mid X_0 \in C_r) = 1.$$

De fato,

$$P(X_1 \in C_r \,|\, X_0 \in C_r) =$$

$$\frac{P(X_1 \in C_r \,,\, X_0 \in C_r)}{P(X_0 \in C_r)} =$$

$$\frac{P(X_1 \in C_r \,,\, X_0 \in C_r) \,+\, P(X_1 \in S - C_r \,,\, X_0 \in C_r)}{P(X_0 \in C_r)} =$$

$$\frac{P(X_0 \in C_r)}{P(X_0 \in C_r)} = 1.$$

Usamos acima o fato que C_r é fechado e assim $P(X_1 \in S - C_r \,,\, X_0 \in C_r) = 0$.

Suponha agora que para k fixo

$$P(X_n \in C_r, \text{ para todo } n \text{ tal que } k \geq n \geq 1 \,|\, X_0 \in C_r) = 1.$$

Então,

$$P(X_n \in C_r, \text{ para todo } n \text{ tal que } k+1 \geq n \geq 1 \,|\, X_0 \in C_r) =$$

$$\frac{P(X_n \in C_r, \text{ para todo } n \text{ tal que } k+1 \geq n \geq 1 \;, X_0 \in C_r)}{P(X_0 \in C_r)} =$$

$$\frac{P(X_n \in C_r, \text{ para todo } n \text{ tal que } k+1 \geq n \geq 1 \;, X_0 \in C_r)}{P(X_0 \in C_r, X_1 \in C_r)} \times$$

$$\frac{P(X_1 \in C_r, X_0 \in C_r)}{P(X_0 \in C_r)} =$$

$$P(X_n \in C_r, \text{ para todo } n \text{ tal que } k+1 \geq n \geq 2 \,|\, X_0 \in C_r, X_1 \in C_r) \times$$

$$P(X_1 \in C_r \,|\, X_0 \in C_r) =$$

$$P(X_n \in C_r, \text{ para todo } n \text{ tal que}$$

$$k+1 \geq n \geq 2 \,|X_1 \in C_r) \times P(X_1 \in C_r \,|\, X_0 \in C_r) =$$

$$P(X_n \in C_r, \text{ para todo } n \text{ tal que } k \geq n \geq 1 \,|X_0 \in C_r) = 1.$$

Acima usamos a propriedade fraca de Markov.

Logo, por indução, para todo k vale que

$$P(X_n \in C_r, \text{ para todo } n \text{ tal que } k \geq n \geq 1 \,|X_0 \in C_r) = 1.$$

Interceptando sobre todos os $k \in \mathbb{N}$ obtemos

$$P(X_n \in C_r, \text{ para todo } n \text{ tal que } n \geq 1 \,|X_0 \in C_r) = 1.$$

Agora, utilizando a propriedade forte de Markov

$$P(X_{T+n} \in C_r, \text{ para todo } n \geq 1 \,|\, X_T \in C_r) =$$

$$P(X_n \in C_r, \text{ para todo } n \geq 1 \,|X_0 \in C_r) = 1.$$

□

O resultado acima nos afirma que se estivermos interessados no comportamento do caminho $\omega = (w_t)$, para valores grandes de t, podemos assumir que nosso sistema está determinado por uma cadeia de Markov $\mathcal{P}$ irredutível e recorrente. De fato, dada uma cadeia qualquer $\mathcal{P}$, se com probabilidade 1 vale que um caminho deixa os transientes, então o caminho entra num C_r e não sai mais de lá. Nem sempre vale que com probabilidade 1 um caminho deixa os transientes, como por exemplo, no passeio aleatório $(1/2, 1/2)$ sobre $\mathbb{Z}$, onde todos os estados são transientes. No entanto, se a cadeia for finita, com probabilidade 1, em algum momento t, o caminho $\omega = (w_t)$ deixa o conjunto dos transientes. Isto acontece porque o conjunto dos transientes, neste caso, não é fechado (ver Proposição 2.18).

Exemplo 2.50. Vamos analisar em todos os aspectos a cadeia de Markov sobre $S = \{1, 2, 3, 4\}$ com a seguinte matriz de transição:

$$\mathcal{P} = \begin{pmatrix} 1/2 & 0 & 1/2 & 0 \\ 0 & 1/3 & 2/3 & 0 \\ 1/4 & 0 & 3/4 & 0 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

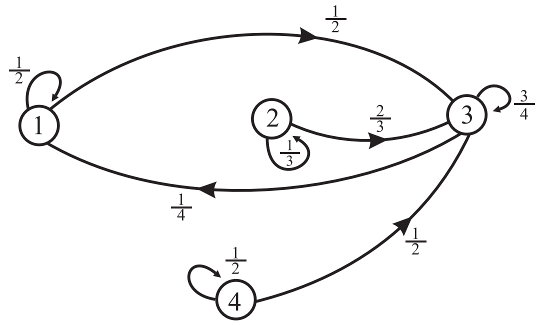


Figura 2.10:

a) Primeiro vamos analisar as Classes de equivalência:
$C_0 = \{1,3\} \quad C_1 = \{2\} \quad C_2 = \{4\}$

C_0 é fechada; C_1 e C_2 não são fechadas

$P(1,1) + P(1,3) = \dfrac{1}{2} + \dfrac{1}{2} = 1 = \dfrac{1}{4} + \dfrac{3}{4} = P(3,1) + P(3,3)$
Logo, a C.M. não é irredutível
b) Agora vamos analisar a Periodicidade.
$d(1) = mdc\{1, 2, 3, 4, \cdots\} = 1 = d(3)$

$d(2) = mdc\{1\} = 1 = d(4)$

Os estados são aperiódicos.

c) Desta vez consideramos a Classificação dos Estados.

$$f_{11}^* = \frac{1}{2} + \frac{1}{2}\frac{1}{4} + \frac{1}{2}\frac{3}{4}\frac{1}{4} + \frac{1}{2}\left(\frac{3}{4}\right)^2\frac{1}{4} + \cdots =$$
$$= \tfrac{1}{2} + \tfrac{1}{8}\textstyle\sum_{n\geq 0}\left(\tfrac{3}{4}\right)^n = \tfrac{1}{2} + \tfrac{1}{8}\tfrac{1}{1-\frac{3}{4}} = \tfrac{1}{2} + \tfrac{1}{8}\,4 = 1$$

Logo, C_0 é classe recorrente.
$f^*_{22} = \dfrac{1}{3} < 1, \qquad f^*_{44} = \dfrac{1}{2} < 1$
Logo, C_1 e C_2 são classes transientes.

d) Agora é a vez de analisar a Classificação dos Estados Recorrentes.

$$E_1(T_1) = \sum_{n\geq 1} n\ f^n_{11} = \frac{1}{2} + 2\,\frac{1}{2}\,\frac{1}{4} + 3\,\frac{1}{2}\,\frac{3}{4}\,\frac{1}{4} + 4\,\frac{1}{2}\left(\frac{3}{4}\right)^2\frac{1}{4}$$
$$+\,5\,\frac{1}{2}\left(\frac{3}{4}\right)^3\frac{1}{4} + \cdots = \frac{1}{2} + \frac{1}{8}\sum_{n=0}^{\infty}(n+2)\left(\frac{3}{4}\right)^n = \frac{1}{2} + \frac{3}{2} = 2$$

Observação: Se mostrarmos que $C_0 = \{1,\,3\}$ é irredutível e aperiódica com uma única distribuição estacionária, então ela é uma classe recorrente positiva.

Observe que:

$$\mathcal{P}_{1,3} = \mathcal{P} = \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix}.$$

$$\pi \times P_0 = \pi \Rightarrow \pi(1)\,\frac{1}{2} + \frac{1}{4}\,\pi(3) = \pi(1) \Rightarrow 2\pi(1) - 4\pi(1) + \pi(3) = 0$$
$$\Rightarrow 2\pi(1) = \pi(3)$$

e

$$\pi(1)\,\frac{1}{2} + \frac{3}{4}\,\pi(3) = \pi(3).$$

Como $\pi(1) + \pi(3) = 1$ temos $\pi(3) = 1 - \pi(1)$ e

$$2\pi(1) = 1 - \pi(1) \Rightarrow 3\pi(1) = 1 \Rightarrow \pi(1) = \frac{1}{3}.$$

Logo, $\pi = \left(\dfrac{1}{3}\,,\,\dfrac{2}{3}\right)$.

Concluímos assim que os estados em C_0 são recorrentes positivos.

e) Queremos agora obter a distribuição estacionária π. Ela será única? Sabemos que

$$\text{i recorrente} \iff \sum_{n\geq 1} P^n(i,j) = \infty$$

$$\text{i transiente} \iff \sum_{n\geq 1} P^n(i,j) < \infty$$

Como 2 e 4 são transientes temos que $\sum_{n\geq 1} P^n(i,j) < \infty \quad \forall\, i,\ j = 2,\, 4.$

e, portanto, $P^n(i,j) \underset{n\to\infty}{\to} 0$, i. é, $\pi(j) = 0$ sempre que j é transiente.

Portanto, $\pi(2) = 0 = \pi(4)$. Logo a distribuição estacionária é dada por

$$\boxed{\pi = (1/3,\ 0,\ 2/3,\ 0) \text{ e é único.}}$$

De outra maneira:

$$(\pi(1)\ \ \pi(2)\ \ \pi(3)\ \ \pi(4)) \begin{pmatrix} 1/2 & 0 & 1/2 & 0 \\ 0 & 1/3 & 2/3 & 0 \\ 1/4 & 0 & 3/4 & 0 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix} = (\pi(1)\ \ \pi(2)\ \ \pi(3)\ \ \pi(4)).$$

$$\Longrightarrow \begin{cases} \frac{1}{2}\,\pi(1) + \frac{1}{4}\,\pi(3) = \pi(1) \\[2ex] \frac{1}{3}\,\pi(2) = \pi(2) \\[2ex] \frac{1}{2}\,\pi(1) + \frac{2}{3}\,\pi(2) + \frac{3}{4}\,\pi(3) + \frac{1}{2}\,\pi(4) = \pi(3) \\[2ex] \frac{1}{2}\,\pi(4) = \pi(4) \end{cases} \Longrightarrow$$

$$\Longrightarrow \begin{cases} \frac{1}{2}\,\pi = \frac{1}{4}\,\pi(3) \Rightarrow 2\pi(1) = \pi(3) \\[2ex] \pi(2) = 0 \\[2ex] \pi(4) = 0 \end{cases}$$

Como $\sum_{j=1}^{4} \pi(j) = 1$ temos que $\pi = (1/3, \ 0, \ 2/3, \ 0)$.

$\diamondsuit$

2.10 Processos de Nascimento e Morte

Definição 2.24. *Uma cadeia (ou processo) de nascimento e morte é uma cadeia de Markov que tem a matriz de transição da forma*

$$\mathcal{P} = \begin{pmatrix} r_0 & p_0 & 0 & 0 & 0 & \cdots \\ q_1 & r_1 & p_1 & 0 & 0 & \cdots \\ 0 & q_2 & r_2 & p_2 & 0 & \cdots \\ 0 & 0 & q_3 & r_3 & p_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix},$$

onde q_i, r_i, $p_i, i = \{0, 1, 2, 3, \dots\}$, *são tais que*

$$p_i + q_i + r_i = 1.$$

Assumimos que $p_i > 0$ e $q_i > 0$ para todo i. Desta forma a cadeia é sempre irredutível.

Os p_i estão associados a taxa de nascimento e os q_i a taxas de morte (se a população é i). Uma pergunta natural é: o que vai acontecer com a população se o modelo é descrito por tal sistema de taxas? Em termos estatísticos, vai ocorrer extinção ou a população cresce sem limites?

É usual também denotar por cadeia de nascimento e morte uma cadeia tal que tem conjunto de estados $S = \{0, 1, 2, \dots, d\}$, satisfaz as propriedades correspondentes acima, e é da forma

$$\begin{pmatrix}
r_0 & p_0 & 0 & 0 & 0 & 0 & . & \cdots & 0 \\
q_1 & r_1 & p_1 & 0 & 0 & 0 & . & \cdots & 0 \\
0 & q_2 & r_2 & p_2 & 0 & 0 & . & \cdots & 0 \\
0 & 0 & q_3 & r_3 & p_3 & 0 & . & \cdots & 0 \\
. & . & . & . & . & . & . & . & . \\
0 & 0 & 0 & . & . & 0 & q_{d-1} & r_{d-1} & p_{d-1} \\
0 & 0 & 0 & . & 0 & . & 0 & q_d & r_d
\end{pmatrix},$$

A cadeia acima também é recorrente.

Se alguns dos r_i é não nulo , então a cadeia é aperiódica. A cadeia de nascimento e morte é um modelo bastante utilizado em aplicações. Vamos analisar a seguir tais cadeias, classificar o estado 1 e calcular a distribuição estacionária.

Vamos agora calcular a Distribuição Estacionária π para o Processo de Nascimento e Morte. Consideraremos a cadeia de Nascimento e Morte em $\{0,\, 1,\, \cdots,\, d\}$ ou $S \;=\; \mathbb{N}$.

Suponhamos que a cadeia é irredutível, i.é,

$$p_i > 0; \qquad 0 \leq i < d$$
$$q_i > 0; \qquad 0 < i \leq d,$$

ou, se $d = \infty$,

$$p_i > 0 \ \ para \ \ i \geq 0,$$
$$q_i > 0 \ \ para \ \ i > 0.$$

Supondo que $d = \infty$. Então

$$\sum_{i \geq 0} \pi(i) \, P(i, j) = \pi(j), \quad j \in S$$

é equivalente a

$$* \begin{cases} \pi(0) = \pi(0) \, r_0 + \pi(1) \, q_1 \\ \\ \pi(j) = \pi(j-1) \, p_{j-1} + \pi(j) \, r_j + \pi(j+1) \, q_{j+1}, \quad j \geq 1. \end{cases}$$

Como, $p_i + q_i + r_i = 1$ as equações em $*$ se reduzem a

$$** \begin{cases} q_1 \, \pi(1) - p_0 \, \pi(0) = 0 \\ \\ q_{j+1} \, \pi(j+1) - p_j \, \pi(j) = q_j \, \pi(j) - p_{j-1} \, \pi(j-1), \ j \geq 1. \end{cases}$$

Segue que

$q_{j+1} \, \pi(j+1) - p_j \, \pi(j) = 0 \qquad j \geq 0$

$\Longrightarrow \ \pi(j+1) = \dfrac{p_j}{q_{j+1}} \, \pi(j) \qquad j \geq 0$

$\Longrightarrow \ \pi(j) = \dfrac{p_0 \ p_1 \ \cdots \ p_{i-1}}{q_1 \ q_2 \ \cdots \ q_j} \qquad j \geq 1.$

Portanto,

$$\pi(j) = \begin{cases} \pi(0) \qquad se \ \ j = 0 \\ \\ \dfrac{p_o, \ p_i, \ \cdots, \ p_{j-1}}{q_1, \ \cdots \ q_j} \, \pi(0), \qquad se \ \ j \geq 1 \end{cases}$$

Como buscamos π tal que $\displaystyle\sum_{j \geq 0}^{\infty} \pi(j) = 1$, temos que

$$1 \;=\; \pi(0) \;+\; \sum_{j \geq 1}^{\infty} \frac{p_0 \cdots p_{j-1}}{q_1 \cdots q_j}\, \pi(0) \;\Longrightarrow$$

$$\pi(0) \;=\; \frac{1}{1 \;+\; \displaystyle\sum_{j \geq 1}^{\infty} \frac{p_0 \cdots p_{j-1}}{q_1 \cdots q_j}} \;=$$

$$=\; \frac{1}{\displaystyle\sum_{j \geq 0}^{\infty} \frac{p_0 \cdots p_{j-1}}{q_1 \cdots q_j}}$$

Obtemos assim a expressão do termo geral,

$$\pi(j) \;=\; \frac{\frac{p_0\, p_1 \cdots p_{j-1}}{q_1 \cdots q_j}}{\displaystyle\sum_{j \geq 0}^{\infty} \frac{p_0\, p_1 \cdots p_{j-1}}{q_1 \cdots q_j}} \tag{2.10}$$

Então, se $\displaystyle\sum_{j \geq 0}^{\infty} \frac{p_0 \cdots p_{j-1}}{q_1 \cdots q_j} \;<\; \infty$, a cadeia tem uma única distribuição estacionária dada pela expressão que obtivemos antes.

Se $\displaystyle\sum_{j \geq 0}^{\infty} \frac{p_0 \cdots p_{j-1}}{q_1 \cdots q_j} \;=\; \infty$, então qualquer solução π ou é identicamente nula ou tem soma infinita.

Se $d = \#S$ é finito, procedendo de maneira similar se obtem que a distribuição estacionária π é única e dada por

$$\pi(j) \;=\; \frac{\frac{p_0 \cdots p_{j-1}}{q_1 \cdots q_j}}{\displaystyle\sum_{i=0}^{d} \frac{p_0 \cdots p_{i-1}}{q_1 \cdots q_i}}, \qquad 0 \;\leq\; j \;\leq\; d.$$

Vamos agora abordar outros tópicos.

Fixado $i \in S$, sejam $a \in S$ e $b \in S$ tais que $a < i < b$. Por definição $T_a = T_{a,i}$ denota o tempo em que pela primeira vez, saindo de i em tempo 0, se atinge o estado a. A mesma coisa para $T_b = T_{b,i}$.

Lema 2.2. *Considere o intervalo de $a = 0$ a $b = n$, então,*

$$\lim_{n\to\infty} P(T_{0,1} < T_{n,1} \,|\, X_0 = 1) =$$

$$P(\,\{\omega \,|\, T_0(\omega) \,<\, T_n(\omega) \;\; \textit{para algum} \;\; n \;>\; 1 \,|\, X_0 = i\}\,).$$

Este lema será demonstrado após a apresentação de um exemplo que ilustra bem a teoria que está sendo desenvolvida nesta seção.

Note que se

$$P(\,\{\omega \,|\, T_0(\omega) \,<\, T_n(\omega) \;\; \text{para algum} \;\; n \;>\; 1 \,|\, X_0 = i\}\,) \,=\, 1,$$

então 1 é recorrente, pois dado um caminho $\omega = (w_t)$, para P-quase todo ω, existe n, tal que $T_0 < T_n$, $w_0 = 1$ e ainda existe $t < T_n$ tal que $w_t = 0$. Logo, para algum $s > 0$ temos que $w_s = 1$. De fato, como $p_0 > 0$,

$$P(\,\{\omega \,|\; \text{existe}\; k > 0,\; \text{tal que para todo} \;\; t > k \;\text{temos que}\; w_t = 0\}\,) \,=\, 0.$$

Logo, podemos usar o Teorema 2.18 para concluir que 1 é recorrente. Desta forma, neste caso, todo estado $i \in S$ é recorrente.

Assumindo a validade do lema 2.2 (demonstração em breve) podemos seguir. Vamos mostrar que sob a hipótese

$$\sum_{r=1}^{\infty} \frac{q_1\; q_2\; \cdots q_r}{p_1\; p_2\; \cdots\; p_r} = \infty,$$

então

$$P(T_{0,1} < \infty \,|\, X_0 = 1) = 1.$$

A hipótese acima, em que o somatório é infinito, indica que, neste caso, os q_i (que são as probabilidades que trazem cada estado a um de menor valor) agem coletivamente de forma que suplantam em intensidade o efeito dos p_j (que conduzem cada estado a um de valor maior).

Lema 2.3. *Assumindo que* $\sum_{r=1}^{\infty} \frac{q_1\, q_2\, \cdots q_r}{p_1\, p_2\, \cdots\, p_r} = \infty$, *então*

$$P(T_{0,i} < \infty \,|\, X_0 = i) = 1.$$

Demonstração: Sejam $a, b \in S$ com $a < b$. Definimos para i tal que $a < i < b$,

(a) $\mu\,(i) = P_i\,(T_a < T_b) = P\,(T_a < T_b | X_0 = i)$, para $a < i < b$

(b) $\mu\,(a) = 1$ e $\mu\,(b) = 0$.

É fácil ver, condicionando no tempo 1 a partir do tempo 0 em i, que esta função μ satisfaz a relação $\mu = \mathcal{P}\ \mu$, ou seja,

$$\mu\,(i) = q_i\,\mu\,(i-1) + r_i\,\mu\,(i) + p_i\,\mu\,(i + 1),$$

para todo i tal que $a < i < b$.

Vamos a seguir estimar $\mu\,(i) = P_i\,(T_0 < T_n)$, quando i está ente 0 e n.

Sejam a e b fixos e $a < i < b$,

Como $r_i = 1 - q_i - p_i$ temos que

$$\mu\,(i) = q_i\;\mu\,(i - 1) + (1 - q_i - p_i)\;\mu\,(i) + p_i\;\mu\,(i + 1),$$

isto é,

$$p_i\;\mu\,(i + 1) - p_i\;\mu\,(i) = \mu\,(i) - (1 - q_i)\;\mu\,(i) -$$

$$- q_i\;\mu\,(i - 1) = q_i\;[\mu\,(i) - \mu(i - 1)]\,.$$

Assim,

$$[\,\mu\,(i + 1) - \mu\,(i)\,] = \frac{q_i}{p_i}\,[\mu\,(i) - \mu(i - 1)].$$

Portanto, procedendo indutivamente quando $i > a$

$$\mu\,(i + 1) - \mu\,(i) = \frac{q_i\; q_{i-1}\; \cdots\; q_{a+1}}{p_i\; p_{i-1}\; \cdots p_{a+1}}\,[\mu\,(a + 1) - \mu\,(a)]$$

Seja

$$\alpha_r = \frac{q_1\, q_2\, \cdots q_r}{p_1\, p_2\, \cdots\, p_r} \qquad \text{e} \qquad \alpha_0 = 1.$$

Assim,

$$\mu(i+1) - \mu(i) = \frac{\alpha_i}{\alpha_a}\,(\,\mu(a+1) - \mu(a)\,)\ \ (*).$$

Somando a igualdade (*) sobre i entre a e $b-1$, através de cancelamentos, temos

$$0 - 1 = \mu(b) - \mu(a) = \frac{1}{\alpha_a}\,[\,\mu(a+1) - \mu(a)\,]\sum_{i=a}^{b-1}\alpha_i.$$

Desta forma,

$$\mu(a+1) - \mu(a) = -\,\frac{\alpha_a}{\sum_{i=a}^{b-1}\alpha_i}.$$

Substituindo este valor acima em (*) obtemos

$$\mu(i+1) - \mu(i) = -\,\frac{\alpha_i}{\sum_{j=a}^{b-1}\alpha_j},$$

para todo i tal que, $a \leq i < b$.

Somando sobre j, entre i e $b-1$, através de cancelamentos, temos

$$0 - \mu(i) = \mu(b) - \mu(i) = -\,\frac{\sum_{j=i}^{b-1}\alpha_j}{\sum_{j=a}^{b-1}\alpha_j}.$$

Concluímos assim que

$$P_i(T_a < T_b) = \mu(i) = \frac{\sum_{j=i}^{b-1}\alpha_j}{\sum_{j=a}^{b-1}\alpha_j}.$$

Para i e a fixos, $a < i$, seja $b = n$ e vamos fazer n tender a infinito. Suponha que $\sum_{j=a}^{\infty}\alpha_j = \infty$, então

$$\lim_{n\to\infty}\mu(i) = \lim_{n\to\infty}\frac{\sum_{j=i}^{n-1}\alpha_j}{\sum_{j=a}^{n-1}\alpha_j} = \lim_{n\to\infty}\frac{\sum_{j=i}^{n-1}\alpha_j}{\sum_{j=a}^{i-1}\alpha_j + \sum_{j=i}^{n-1}\alpha_j} =$$

$$= \lim_{n\to\infty} \frac{1}{\frac{\sum_{j=a}^{i-1}\alpha_j}{\sum_{j=i}^{n-1}\alpha_j} + \frac{\sum_{j=i}^{n-1}\alpha_j}{\sum_{j=i}^{n-1}\alpha_j}} = 1.$$

Logo, mostramos que para $a < i$ fixos,

$$\lim_{n\to\infty} P\,(T_a \,<\, T_n | X_0 \,=\, i) = \lim_{n\to\infty}\, \mu(i) = 1.$$

Segue do lema anterior que para $a = 0$ e i qualquer

$$P(T_{0,i} \,<\, \infty \,|\, X_0 = i) = 1.$$

$\square$

Note que se acontecer de $\sum_{j=a}^{\infty} \alpha_j < \infty$, então a cadeia é transiente. De fato, neste caso, é fácil ver, a partir da expressão acima, que

$$\lim_{n\to\infty}\, \mu(i) = \lim_{n\to\infty} \frac{\sum_{j=i}^{n-1}\alpha_j}{\sum_{j=a}^{n-1}\alpha_j} < 1.$$

$\diamondsuit$

Exemplo 2.51. Considere a Cadeia de Nascimento e Morte em

$$S = \{0, 1, 2, \dots\},$$

definida por

$$p_i \;=\; \frac{i\,+\,2}{2(i\,+\,1)}$$

e

$$q_i \;=\; \frac{i}{2(i\,+\,1)},$$

para todo $i \;\geq\; 0$.

Vamos mostrar que a cadeia é transiente. De fato, note que $\frac{q_i}{p_i} = \frac{i}{i+2}$. Logo,

$$\alpha_i \;=\; \frac{1\,2\,\cdots\,i}{3\,4\,\cdots\,(i\,+\,1)\,(i\,+\,2)} \;=\; \frac{2}{(i\,+\,1)\;(i\,+\,2)} \;=$$

$$= 2\,(\frac{1}{i+1} - \frac{1}{i+2}).$$

Então

$$\sum_{i\geq 1}^{\infty} \alpha_i = 2\sum_{i\geq 1}\left(\frac{1}{i+1} - \frac{1}{i+2}\right) =$$

$$= 2\left(\frac{1}{2} - \frac{1}{3} + \frac{1}{3} - \frac{1}{4} + \frac{1}{4} - \frac{1}{5} + \cdots\right) =$$

$$= 2.\frac{1}{2} = 1 < \infty.$$

Logo, a cadeia é transiente.

$\diamondsuit$

Vamos agora demonstrar o lema 2.2 que havia sido mencionado antes.

Demonstração: Vamos considerar o caso $i = 1$ (o caso geral é similar).

Vamos demonstrar que se

$$A = \{\omega \,|\, T_0(\omega) < T_n(\omega) \text{ para algum } n > 1\},$$

então,

$$P_1(A) = \lim_{n\to\infty} P_1(T_0 < T_n) = 1 - \frac{1}{\sum_{j\geq 0}^{\infty} \alpha_j}.$$

Sabemos que

$$\mu(1) = P_1(T_0 < T_n) = \frac{\sum_{j=1}^{n-1} \alpha_j}{\sum_{j=0}^{n-1} \alpha_j} = 1 - \frac{\alpha_0}{\sum_{j=0}^{n-1} \alpha_j} =$$

$$= 1 - \frac{1}{\sum_{j=0}^{n-1} \alpha_j}, \qquad (2.11)$$

já que $\alpha_0 = 1$.

Estamos considerando a cadeia começando no estado 1. Um dado caminho $\omega = (w_n)_{n\in\mathbb{N}}$, pode se mover no máximo um passo para a direita ao longo de $S = \mathbb{N}$, em cada unidade de tempo.

Então, para cada $\omega \in \Omega$,

$1 \leq T_2(\omega) < T_3(\omega) < \cdots \qquad (*)$,

logo,

$A_n = \{\omega \,|\, T_0(\omega) < T_n(\omega)\}, \; n > 1$, forma uma sequência não decrescente de conjuntos (ou, eventos).

Assim,

$\lim\limits_{n\to\infty} P_i\,(T_0 < T_n) = P_i\,(T_0 < T_n \;\; para \;\; algum \;\; n > 1)$.

Como a desigualdade $(*)$ implica em $T_n \geq n-1$, temos para cada ω fixo que $T_n(\omega) \underset{n\to\infty}{\longrightarrow} \infty$.

Logo,

$\lim\limits_{n\to\infty} P_i\,(T_0 < T_n) = P_i\,(A)$.

Segue que $P_1\,(A) = \lim\limits_{n\to\infty} P_1\,(T_0 < T_n) = \lim\limits_{n\to\infty} \mu\,(1) = \lim\limits_{n\to\infty} 1 - \frac{1}{\sum_{j=0}^{n-1} \alpha} = 1 - \frac{1}{\sum_{j\geq 0} \alpha_j}$

Se $\sum_{j \geq 0} \alpha_j = \infty$, então

$$P_1 \left(\{\omega \,|\, T_0(\omega) < T_n(\omega) \ \text{ para algum } \ n > 1\} \right) = 1.$$

□

Exemplo 2.52. Um jogador na roleta faz uma série de apostas de um real cada vez, iniciando com o capital de 10 reais. Ele tem probabilidade 9/19 de ganhar um real e 10/19 de perder um real em cada aposta. O jogador estabelece a princípio que vai parar quando seu capital atingir o valor 35 reais, ou que os resultados são tais que perde os dez reais iniciais, e assim atinge o capital zero. Vamos obter aqui soluções explícitas.

Primeiro vamos calcular a probabilidade do jogador atingir os 35 reais.

Considere X_n o capital do jogador no instante n, sendo assim temos que $X_0 = 10$.

Usando a notação acima, $(X_n)_{n \geq 0}$ forma uma cadeia de Markov de Nascimento e Morte com $S = \{0, 1, 2, \ldots, 35\}$,

$p_i = 9/19, \quad 0 < i < 35$, razão de nascimento;

$q_i = 10/19, \quad 0 < i < 35$, razão de morte;

$r_i = 0, \quad 0 < i < 35$.

Note que o jogador estabeleceu a priori que 0 e 35 são estados absorventes.

Neste caso, usando a notação acima, $a = 0$, $b = 35$ e $i = 10$.

Ora, $P_{10}\,(T_{35} \leq T_0) = 1 - \mu\,(10)$ e

$$\mu\,(10) = \frac{\sum_{i=10}^{34} \alpha_i}{\sum_{i=0}^{34} \alpha_i} \quad \text{onde} \quad \alpha_i = \frac{q_1 \cdots q_i}{p_1 \cdots p_i} = \left(\frac{10/19}{9/19}\right)^i = \left(\frac{10}{9}\right)^i.$$

Assim,

$$1 \; - \; \mu(10) \; = \; \frac{\sum_{i=0}^{9} \; \alpha_i}{\sum_{i=0}^{34} \; \alpha_i} \; = \; \frac{\sum_{i=0}^{9} (10/9)^i}{\sum_{i=0}^{34} (10/9)^i} \; = \; \frac{1 \; - \; (10/9)^{10}}{1 \; - \; (10/9)^{35}} \; = \; 0,047.$$

Desta forma, a probabilidade do jogador atingir o capital 35 reais é de $0,047$.

Com probabilidade 1 os caminhos atingem o valor 0 ou o valor 35. Logo, a probabilidade do jogador atingir o capital 0 é $0,953$.

Vamos calcular o ganho esperado.

Ora, ele perde dez reais, ou seja seus ganhos são -10 reais com probabilidade $0,953$. Seus ganhos são $+25$ reais com probabilidade $0,047$.

Valor esperado do ganho: $-10 \times 0,953 + 25 \times 0,047 = -9,53 + 1.17 = -8,36$.

$\diamondsuit$

2.11 Apêndice - Cadeias de Markov de Ordem Superior

Seja $S = \{1, 2, 3, .., d\}$ o espaço de estados e para cada 2-upla $(x_1, x_2) \in S \times S$, considere $p(x_1, x_2, x_3) = p_{x_1,x_2}(x_3) \in \mathbb{R}$, onde $x_3 \in S$, tal que $p_{x_1,x_2} : S \to \mathbb{R}$, é uma probabilidade sobre S, isto é,

$$\sum_{x_3 \in S} p_{x_1,x_2}(x_3) = 1,$$

e

$$0 \leq p_{x_1,x_2}(x_3) \leq 1.$$

Um processo estocástico X_t, onde $t \in T \subset \mathbb{N}$, tal que para todo n, e todo $x_n, x_{n+1}, x_{n+2} \in S$ vale que

$$P(X_{n+2} = x_{n+2} \,|\, X_n = x_n, X_{n+1} = x_{n+1}) = p(x_n, x_{n+1}, x_{n+2}) =$$

$$p_{x_n, x_{n+1}}(x_{n+2}), \qquad (*)$$

é chamado de uma 2-cadeia de Markov.

Generalizando o que vimos antes, é fácil ver que uma vez fixado, uma matriz 2 por $\#S$

$$(p^0_{x_0,x_1})_{x_0,x_1 \in S},$$

tal que

$$\sum_{x_0,x_1 \in S} p^0_{x_1,x_2} = 1,$$

se assumirmos que

$$p^0_{x_0,x_1} = P(X_0 = x_0, X_1 = x_1),$$

fica determinado de maneira única o processo X_t, ou seja, podemos calcular

$$P(X_0 = x_0, X_1 = x_1, ..., X_m = x_m),$$

para todo $m \in \mathbb{N}$ e $x_0, x_1, ..., x_m \in S$.

Este processo será chamado de um 2-processo estocástico de Markov com estados S.

Em algumas situações do mundo real é mais natural assumir que o que acontece num momento depende não somente do estado anterior (no tempo), mas sim (apenas) dos dois estados anteriores (no tempo). Neste caso, é preferível modelar o problema através de um 2-processo estocástico de Markov.

Ressaltamos que para análisar as propriedades deste processo estocástico com espaço de estados S e da 2-cadeia de Markov associada, podemos fazer recair o problema no estudo de uma cadeia de Markov como anteriormente descrito. De fato, basta considerar uma cadeia de Markov em que o novo espaço S^* dos estados é constituído por pares $(x_1, x_2) \in S \times S$. O que se faz é rebatizar os símbolos através de pares de elementos de S.

Vamos ilustrar esta afirmação através de um exemplo específico: suponha $S = \{1, 2\}$ e assuma que nos foi dada a informação

$$p_{1,1} : S \to \mathbb{R},\ p_{1,2} : S \to \mathbb{R},\ p_{2,1} : S \to \mathbb{R},\ p_{2,2} : S \to \mathbb{R}.$$

Neste caso são um total de 8 números reais, e vamos supor para simplificar que são todos positivos.

Considere $S^* = \{1, 2, 3, 4\}$, onde identificamos 1 com $(1, 1)$, 2 com $(1, 2)$, 3 com $(2, 1)$ e finalmente 4 com $(2, 2)$.

Seja agora a matriz $\mathcal{P} = (p(i, j))_{i,j \in \{1,2,3,4\}}$, da forma 4 por 4, tal que traduz a informação que nos foi concedida. Por exemplo, $p(2, 3) = p_{1,2}(1)$ ou seja, do par $(1, 2)$ se passa para 1 com probabilidade $p_{1,2}(1)$. Ainda, $p(3, 4) = 0$, pois não se pode passar de $(2, 1)$ para $(2, 2)$, visto que a segunda entrada $(2, 1)$ não é igual a primeira de $(2, 2)$.

A partir da informação acima vamos descrever o processo através de uma matriz estocástica $\mathcal{P}$ quatro por quatro onde identificamos

$$1 \sim 11, \quad 2 \sim 12 \quad 3 \sim 21 \;\text{ e }\; 4 \sim 22.$$

Neste caso,

$$\mathcal{P} = \begin{pmatrix} p(1,1,1) & p(1,1,2) & 0 & 0 \\ 0 & 0 & p(1,2,1) & p(1,2,2) \\ p(2,1,1) & p(2,1,2) & 0 & 0 \\ 0 & 0 & p(2,2,1) & p(2,2,2) \end{pmatrix},$$

ou seja,

$$\mathcal{P} = \begin{pmatrix} p(1,1) & p(1,2) & p(1,3) & p(1,4) \\ p(2,1) & p(2,2) & p(2,3) & p(2,4) \\ p(3,1) & p(3,2) & p(3,3) & p(3,4) \\ p(4,1) & p(4,2) & p(4,3) & p(4,4) \end{pmatrix}.$$

Note que se os oito números $p(i,j,k)$ inicialmente dados forem todos positivos, então $\mathcal{P}^2$, da forma 4 por 4, tem todas as entradas positivas e assim $\mathcal{P}$ é regular.

Pode-se considerar também 3-cadeias de Markov ou até n-cadeias de Markov. Todos estes casos recaem no estudo de cadeias de Markov através de um processo de rebatizar n-uplas de símbolos, em tudo similar ao descrito acima. Dizemos que a cadeia de Markov tem ordem n no caso correspondente. O caso inicialmente analisado por nós eram as 1-cadeia de Markov. Estamos afirmando acima que toda n-cadeia de Markov pode ser analisada através de uma 1-cadeia de Markov.

Voltando agora ao exame de problemas concretos (com conjunto de estados S) oriundos do mundo real, podemos nos perguntar num problema específico se o modelo correto a ser considerado é uma 1-cadeia de Markov, ou uma 2-cadeia de Markov, ou então uma 3-cadeia de Markov, etc. Quanto maior for a ordem da cadeia, maior será a a dimensão da matriz da nova 1-cadeia de Markov associada $\mathcal{P}$ e mais complexa será a análise do problema.

Referimos o leitor a um exemplo muito interessante sobre a precipitação de chuva em Snoqualmie Falls nos EUA apresentado na Seção (2.8) em P. Guttorp [Gu]. Após um detalhado estudo dos dados colhidos ao longo dos anos, utilizando estimação de “maximum likelihood”, se compara a análise do problema através de uma 1-cadeia de Markov, uma 2-cadeia de Markov ou uma 3-cadeia de Markov. A melhor escolha ficou em modelar o problema através de uma 1-cadeia de Markov. Já num outro exemplo sobre a força do vento na Irlanda, na mesma seção do livro, a escolha melhor foi para a modelagem através de uma 2-cadeia.

Uma vez determinado o melhor modelo (em função dos dados reais) se pode fazer previsões sobre a evolução temporal do processo estocástico associado ao problema em consideração.

2.12 Exercícios

1. Considere um jogador que ganha ou perde uma unidade com probabilidades p e q, respectivamente. Suponha que o seu capital inicial é **x** e que o seu adversário tem um capital de **a-x**, $a \geq x$ (portanto, o capital inicial total é **a**). O jogo continua até que um dos jogadores fique arruinado, ou seja, até que o primeiro jogador aumente seu capital até **a** ou perde **x**.

   a. Determine o espaço de estados S.

   b. Defina a Processo de Markov $(X_n)_{n\geq 0}$ associado.

   c. Determine a matriz de transição para esta C.M..

2. Defina a cadeia de Markov e determine seu espaço de estados e suas matrizes de transição de primeira e segunda ordem para as seguintes situações:

   a. Seis pontos são marcados em um círculo no sentido horário. O processo se move de um dos pontos para um de seus vizinhos com probabilidade $\frac{1}{2}$.

   b. Escolhe-se ao acaso um número X_1 entre os inteiros de 1 a 7. Para $n > 1$, X_n é o número escolhido ao acaso entre os inteiros $1, 2, 3, \cdots, X_{n-1}$.

   c. A área de atuação de um vendedor é constituída de três cidades, A, B e C. Ele nunca vende na mesma cidade em dias sucessivos. Se vende na cidade A, no dia seguinte vende na cidade B. Se vende em B ou em C, então no dia seguinte é duas vezes mais provável que ele venda em A do que em outra cidade.

   d. Considere lançamentos independentes de um dado honesto. Defina X_n o menor dos números que aparecem nos $\underline{n}$ primeiros lançamentos.

3. Suponha que o nível sócio-econômico de uma pessoa é classificado em três categorias: classe alta (A), classe média (M) e classe baixa (B). Suponha que dos filhos de um homem da classe alta, 85% são de classe alta e 15% são de classe média. Já um indivíduo de classe média, 10% são da classe alta, 70% da classe média e 20% são da classe baixa. Dos filhos de um homem da classe baixa, 25% são da classe média e 75% são da classe baixa. Supondo que cada indivíduo tem um filho, podemos formar uma C.M. observando uma família através de gerações sucessivas.

   a. Defina a C.M. $(X_n)_{n\geq 1}$.

   b. Determine o espaço de estados desta cadeia.

   c. Determine a matriz de transição P da C.M..

   d. Calcule P^2 e interprete seu resultado.

   e. Suponha que um homem não tenha necessariamente um filho e que é 0.85 a probabilidade de um indivíduo qualquer ter um filho. Neste caso, determine a matriz de transição da nova C.M.

   f. O que voce pode dizer a respeito do estado introduzido no item (e)?

4. Três tanques lutam entre si. O tanque A atinge seu alvo com probabilidade 1/2, o tanque B com probabilidade 2/3 e o tanque C com probabilidade 1/3. Tiros são dados simultaneamente e quando um tanque é atingido, ele fica fora de ação. Como conjunto de estados escolha o conjunto de tanques ainda em ação. Suponha que em cada passo (ou etapa), cada tanque atira em seu oponente mais forte.

a. Verifique se a seguinte matriz de transição descreve o processo corretamente:

$$\begin{array}{cccccccc} E & A & B & C & AB & AC & BC & ABC \end{array}$$

$$\mathcal{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1/3 & 1/6 & 1/3 & 0 & 1/6 & 0 & 0 & 0 \\ 1/6 & 1/3 & 0 & 1/6 & 0 & 2/6 & 0 & 0 \\ 2/9 & 0 & 4/9 & 1/9 & 0 & 0 & 2/9 & 0 \\ 0 & 0 & 0 & 8/18 & 0 & 4/18 & 4/18 & 2/18 \end{pmatrix}.$$

O estado AC significa que os tanques A e C estão em ação; o estado A significa que apenas o tanque A está em ação; o estado E significa que nenhum tanque está em ação.

b. Modifique a matriz de transição do item (a) assumindo que, quando todos os tanques estão em ação, A atira em B, B atira em C e C em A.

5. Seja $X_1, X_2, \cdots$ uma seqüência de v.a.'s independentes (está claro o sentido?). Defina $S_n = X_1 + X_2 + \cdots + X_n$. Mostre que $(S_n)_{n\geq 1}$ é um processo Markoviano.

6. <u>Cadeia de Ehrenfest</u>. Suponha que existam **d** bolinhas que são enumeradas como $1, 2, \cdots, d$. Inicialmente estas **d** bolinhas estão distribuídas em duas urnas. Um número entre 1 e **d** é escolhido aleatoriamente e a

bolinha com esse número é transferida da urna em que se encontra para a outra. Seja X_n o número de bolinhas na urna 1 no instante de tempo **n**. $(X_n; n \geq 0)$ é uma C.M. com matriz de transição dada por

$$P(i,j) = \begin{cases} \frac{i}{d}, & \text{j} = i-1 \\ 1 - \frac{i}{d}, & j = i+1 \\ 0, & j \neq i-1 \ \ e \ \ j \neq i+1. \end{cases}$$

a. Descreva o espaço de estados.

b. Determine a matriz de transição para este tipo de cadeia quando **d**=4.

7. No país de OZ, não existem dois dias bons sucessivamente. Se ocorrer um dia bom, no próximo estará nevando ou chovendo. Se estiver chovendo (ou nevando) a chance é de $1/2$ de permanecer no mesmo estado no dia seguinte e é de $1/4$ de ser um dia com tempo diverso. Considere a C.M. com estados N (nevando), C (chovendo) e B (dia bom).

   a. Obtenha a matriz de transição.

   b. Se π^0= (1/5, 2/5, 2/5), obtenha π^1 e π^2.

   c. Qual é a distribuição π^n ?

8. Considere uma seqüência de experimentos da seguinte forma: primeiro uma moeda honesta é lançada. Então, se no $(n-1)$-ésimo ensaio ocorre cara, lançamos uma moeda com probabilidade $1/k$ de sair coroa; se ocorre coroa, lançamos novamente a moeda honesta. Quais são as probabilidades de transição ?

9. Um jogador tem 2.00 dólares. Ele aposta 1,00 dólar de cada vez e ganha 1,00 dólar com probabilidade $1/2$. Ele pára de jogar se perder os 2,00 dólares ou ganhar 4.00 dólares.

a. Determine a matriz de transição da C.M.

b. Qual é a distribuição inicial?

c. Qual é a probabilidade de que ele perca seu dinheiro após no máximo 5 jogadas?

d. Qual é a probabilidade de que o jogo dure mais do que 7 jogadas?

10. Um psicólogo faz os seguintes assentamentos a respeito do comportamento de camundongos submetidos a um programa particular de alimentação: para qualquer ensaio particular, 80% dos camundongos que se dirigiram para a direita no experimento anterior, irão dirigir-se para a direita neste ensaio, e 60% dos que se dirigiram para a esquerda no experimento anterior, irão dirigir-se para a direita neste ensaio.

    a. Descreva o espaço de estados da C.M.

    b. Determine a matriz de transição.

    c. Se 50% se dirigiram para a direita no primeiro ensaio, o que o psicólogo poderia prever com respeito ao segundo ensaio?

    d. Em relação ao item (c), o que ele poderia prever com respeito ao terceiro ensaio?

    e. Em relação ao item (c), o que ele poderia prever com respeito ao milésimo ensaio?

11. Um homem se encontra em algum ponto inteiro do eixo da abscissa compreendido entre a origem e o ponto 6. Em cada etapa ele anda para o ponto imediatamente à esquerda com probabilidade q, ou para o ponto imediatamente à direita com probabilidade p, ou permanece no ponto onde se encontra com probabilidade e (assim $p+q+e=1$), a menos que esteja na origem 0 ou no ponto 6. Neste último caso ele caminhará para o ponto imediatamente à direita ou à esquerda, respectivamente.

a. Descreva a v.a. X_n, $n \geq 1$.

b. Descreva o espaço de estados S.

c. Determine a matriz de transição de $(X_n; n \geq 1)$.

12. *Observação*: Este exercício é um exemplo do Passeio Aleatório. Considere a cadeia de Ehrenfest com d=3 (veja Exercício 6) Sabe-se que a matriz de transição desta cadeia é dada por

$$P(i,j) = \begin{cases} \frac{i}{3}\,, \;\; j = i-1 \\ 1 - \frac{i}{3}\,, \;\; j = 1+i \\ 0\,, \;\; j \neq i-1 \;\, e \;\, j \neq i+1. \end{cases}$$

a. Encontre $\mathcal{P}$, $\mathcal{P}^2$, $\mathcal{P}^3$.

b. Supondo que a distribuição inicial é dada por $\pi^0 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. encontre as distribuições π^1, π^2, π^3.

13. Mostre que se a é um estado absorvente da C.M. $(X_n)_{n\geq 0}$ então $P^n(i,a) = P(X_n = a \,|\, X_0 = i) = P_i(T_a \leq n), n \geq 1$.
Sugestão: Use a definição de um estado absorvente e o fato que

$$P^n(i,j) = \sum_{k=1}^{n} f_{ij}^k P^{n-k}(j,j).$$

14. Prove que

$$P_i(T_j = 3) = \sum_{k \neq j} P(i,k) P_k(T_j = 2).$$

15. Suponha que três fabricantes de automóveis mantiveram os seguintes dados com respeito a compras dos clientes:

Compra Presente (n=0)	Próxima Compra (n=1)		
	%Ford (0)	%Chevrolet (1)	%Volkswagen (2)
Ford (0)	40	30	30
Chevrolet (1)	20	50	30
Volkswagen (2)	25	25	50

a. Descreva a situação acima através de uma matriz de transição.

b. Interprete P(0,2), P(1,2) e P(2,0).

c. Se um cliente acabou de comprar um Ford, qual é a probabilidade dele comprar um Chevrolet daqui a 2 anos?

d. Suponha distribuição uniforme para X_0. Determine π^1.

e. Calcule $\mathcal{P}^2 = \mathcal{P} \times \mathcal{P}$, $\pi^0 \times \mathcal{P}^2$, $\pi^1 \times \mathcal{P}$.

f. Verifique se a cadeia é irredutível ou não.

16. Considere a C.M. $(X_n)_{n\geq 1}$ cuja matriz de transição é dada por

$$\mathcal{P} = \begin{pmatrix} 0.4 & 0.6 & 0 \\ 0.3 & 0 & 0.7 \\ 0 & 0 & 1.0 \end{pmatrix}$$

a. Determine as classes de equivalência do espaço de estados.

b. Calcule f^*_{ii}, para $i \in S$.

c. Determine π^1, π^2 e π^3 considerando $\pi^0 = (1, 0, 0)$.

17. Para suas férias anuais, um executivo seleciona um dos três lugares: Europa (E), Havaí (H) ou Grécia (G) usando a seguinte regra.
Se ele foi à Europa no ano passado, escolherá o Havaí com probabilidade 1/3 e Grécia com probabilidade 2/3.
Se ele foi ao Havaí no ano anterior, escolherá Europa, Havaí novamente

e Grécia com probabilidades 1/2, 1/8 e 3/8, respectivamente.
Se ele foi à Grécia no ano anterior, Europa e Havaí são igualmente prováveis de serem escolhidas este ano.

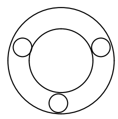
a. Descreva o espaço de estados.

b. Determine a matriz de transição.

c. Calcule $\mathcal{P}^2$.

d. Como você classificaria as preferências do executivo após um longo tempo? Isto é, quem é a distribuição estacionária?

18. Suponha que uma partícula se move entre três posições 1, 2, e 3. Em qualquer tempo, a probabilidade da partícula fazer um movimento no sentido horário é p e a probabilidade da partícula se mover no sentido anti-horário é 1-p, sempre provindo de uma posição prévia.



a. Considere Y_n a posição da partícula no n-ésimo tempo.

b. Determine a matriz de transição da C.M. $(Y_n)_{n\geq 1}$ e o seu espaço de estados.

c. Com distribuição inicial uniforme, isto é, $\pi^0 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, calcule π^1 e π^2.

d. O que você pode dizer da distribuição de π^n de Y_n?

e. Determine as classes de equivalência de S.

f. Calcule f^*_{11}. Como você classifica os estados desta C.M.?

19. Seja $(X_n)_{n\geq 0}$ uma C.M. com espaço de estados {1,2,3,4} e matriz de transição

$$\mathcal{P} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1/3 & 2/3 & 0 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix}.$$

a. Calcule f_{ii}^*, $\forall i \in S$.

b. Descreva os estados recorrentes e transientes.

c. Obtenha as classes de equivalência.

20. Passeio aleatório em Z: Sejam $Y_1, Y_2, Y_3, ...$ variáveis aleatórias independentes e identicamente distribuídas assumindo valores em Z. Sejam
$X_0 = 0$
$X_n = Y_1 + Y_2 + ... + Y_n,\ n \geq 1$
$X_{n+1} = X_n + Y_{n+1}$
Observe que $(X_n)_{n\geq 0}$ satisfaz a propriedade de Markov pois Y_{n+1} é independente de $Y_1, ..., Y_n$ então X_{n+1} depende apenas de X_n.
Sejam $\mathrm{P}(Y_1 = k) = a_k,\ a_k \geq 0$ e $\sum_{k\in Z} a_k = 1$.

a. Calcule as probabilidades de transição $P(i,j) = P(X_{n+1} = j/X_n = i)$.

b. Mostre que $P(i,j) = P(0, j-i)$.

c. Escreva a matriz de transição $\mathcal{P}$.

21. Seja $(X_n)_{n\geq 0}$ uma Cadeia de Markov com dois estados (0 e 1) cuja probabilidade de transição de 0 para 1 é $\underline{p}$ e de 1 para 0 é $\underline{q}$.

a. Prove que

$$\pi^n(0) = (1-p-q)^n \pi^0(0) + q \sum_{j=0}^{n-1} (1-p-q)^j.$$

Sugestão: Use indução matemática.

b. Para $p = 0.3$ e $q = 0.9$ determine π^n a partir de π^0.

c. Para a situação do item (b) determine $\mathcal{P}^5$ e a distribuição conjunta de (X_0, X_1).

22. Cada projétil disparado pelo canhão principal de um tanque, atinge ou erra o alvo com uma probabilidade que depende do sucesso da descarga anterior. Se o tiro anterior atingiu o alvo, há uma probabilidade relativamente alta de que o próximo tiro também atinja o alvo. Se o tiro anterior não atingiu o alvo, é preciso corrigir a posição do canhão, e a probabilidade de acertar será menor no próximo tiro. O tanque e o alvo podem estar ambos em movimento de modo que o conhecimento do sucesso ou fracasso nos tiros anteriores (a menos do imediatamente anterior) é de pouca importância quanto a predizer o sucesso ou fracasso do próximo tiro. Assim, é razoável supor que se esteja observando um processo Markoviano estacionário com espaço de estados $S = \{0, 1\}$ onde $0 =$ “atinge o alvo" e $1 =$ “não atinge o alvo". Suponha que a partir de observações no campo de batalha foi possível obter estimativas das probabilidades de transição de 1 passo:

$$P(0,0) = \frac{3}{4}, \quad P(1,0) = \frac{3}{8}.$$

a. Prove que

$$P^{n+1}(0,0) = P^n(0,0)\alpha + \beta, \tag{2.12}$$

onde $\alpha = P(0,0) - P(1,0)$ e $\beta = P(1,0)$.

b. Aplique a Equação 2.12 repetidas vezes para obter

$$P^{n+1}(0,0) = P^1(0,0)\alpha^n + \beta(\frac{1-\alpha^n}{1-\alpha}),$$

sempre que $\alpha \neq 1$.

c. Obtenha as fórmulas para $P^n(1,1)$, $P^n(0,1)$ e $P^n(1,0)$ em termos de α, β e $\gamma = P(0,1)$ mantendo similaridade com (1).

d. Calcule o limite de $\mathcal{P}^n$.

23. Considere $(X_n)_{n\geq 0}$ uma C.M. com matriz de transição dada por

$$\mathcal{P} = \begin{pmatrix} 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1/9 & 0 & 3/9 & 0 & 5/9 \\ 1/9 & 0 & 0 & 5/9 & 3/9 \end{pmatrix}.$$

a. Obtenha as classes de equivalência.

b. Classifique os estados.

c. Obtenha o período de cada estado.

d. Calcule $\mathrm{P}(T_3 = +\infty | X_0 = 3)$.

24. Prove que $P(i,j) > 0 \implies f^*_{ij} > 0, \forall i,j \in \mathrm{S}$, mas o contrário não é verdadeiro.

25. Cadeia de Nascimento e Morte : S=**N**

$$\mathcal{P} = \begin{pmatrix} r_0 & p_0 & 0 & 0 & \dots & \dots \\ q_1 & r_1 & p_1 & 0 & \dots & \dots \\ 0 & q_2 & r_2 & p_2 & 0 & \dots \\ 0 & 0 & q_3 & r_3 & p_3 & 0\dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix},$$

onde $r_0 \geq 0$, $p_0 > 0$, $r_i \geq 0$, $p_i > 0$, $q_i > 0$.

a. Verifique se a C.M. $(X_n)_{n\geq 0}$ é irredutível.

b. Verifique que a C.M. é aperiódica se e só se algum $r_i > 0$.

c. Considere o caso particular em que p_n=p, $\forall$ n, q_n=1-p, r_n=0, n$\geq$1 e r_0=1-p. Conclua que

$$p < q \longrightarrow \text{a cadeia é recorrente positiva;}$$

$$p = q \longrightarrow \text{a cadeia é recorrente nula;}$$

$$p > q \longrightarrow \text{a cadeia é transiente.}$$

d. Existe distribuição estacionária? Se existe, é única?

26. Seja $(X_n)_{n\geq 0}$ uma C.M. com espaço de estados S={1,2,3,4,5} e matriz de transição dada por

$$\mathcal{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 2/3 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 3/8 & 0 & 3/8 & 0 & 2/8 \end{pmatrix}.$$

a. Classifique os estados.

b. Obtenha a distribuição limite quando n$\longrightarrow +\infty$.

c. Encontre o tempo médio de recorrência para todos os estados recorrentes.

27. Considere a C.M. $(X_n)_{n\geq 0}$ cuja matriz de transição é dada por

$$\mathcal{P} = \begin{pmatrix} 1/4 & 3/4 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/4 & 3/4 \end{pmatrix}$$

a. Verifique se a C.M. é irredutível.

b. Calcule f_{00}^*.

c. Utilize os itens (a) e (b) para classificar todos os estados.

28. Seja $(X_n)_{n\geq 0}$ uma C.M. com espaço de estados $S = \mathbf{N} - \{0\}$ e matriz de transição dada por

$$\mathcal{P} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & \dots \\ 1/2 & 0 & 1/2 & 0 & \dots \\ 1/2 & 0 & 0 & 1/2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{pmatrix}.$$

a. Obtenha f_{11}^n e f_{11}^*.

b. Determine f_{ii}^*, para todo $i \in S - \{1\}$.

29. Considere a matriz estocástica

$$\mathcal{P} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1/4 & 0 & 3/4 \end{pmatrix}.$$

a. Classifique os estados da C.M. $(X_n)_{n\geq 0}$ cuja matriz de transição é a matriz $\mathcal{P}$ acima e cujo espaço de estados é $S = \{0, 1, 2, 3\}$.

b. Calcule f_{ij}^n, para $n \in \{1, 2, 3\}$ e $i, j \in \{0, 2\}$.

c. Calcule $P^2(1, 3)$.

d. Calcule $E_i(N(j))$, para $i, j \in \{0, 1\}$.

30. Suponha que existam duas urnas 1 e 2 e $\underline{2d}$ bolas das quais $\underline{d}$ são brancas e $\underline{d}$ são azuis. Inicialmente $\underline{d}$ bolas são colocadas na urna 1 e as restantes na urna 2. A cada ensaio, uma bola é escolhida ao acaso de cada uma das urnas e elas são trocadas de lugar. Defina X_n = número de bolas azuis na urna 1 após $\underline{n}$ ensaios.

a. O processo $(X_n)_{n\geq 0}$ é uma C.M.? Justifique sua resposta.

b. Obtenha a matriz de transição, se ela existir.

31. Um equipamento pode estar em uma das três seguintes situações: funcionando (F), em conserto (C) ou parado (P) esperando por mais trabalho. Este equipamento é observado sempre que há uma mudança de estados (situações). Considere X_n = o estado após a n-ésima mudança. Suponha que

$$\mathcal{P} = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

a. Calcule $\mathcal{P}^2$, $\mathcal{P}^3$, $\mathcal{P}^4$ e $\mathcal{P}^5$.

b. Com base no item a, obtenha $\mathcal{P}^{2n}$ e $\mathcal{P}^{2n-1}$.

c. Calcule $P_0(T_0 = n)$ para $n = 2, 3, 4$ e 5, onde 0 é o estado F.

d. Qual é a probabilidade $P_0(T_0$ ser um número ímpar)?

32. Considere uma C.M. com espaço de estados $S = \{0, 1, \cdots, 4\}$ e matriz de transição

$$\mathcal{P} = \begin{pmatrix} 0 & 3/8 & 0 & 5/8 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

a. Determine as classes de equivalência de S.

b. Classifique os estados.

c. Calcule $P_0(T_4 = 4)$ e $P_3(T_3 = 5)$.

d. Determine f^*_{00} e f^*_{11}.

e. Calcule $E_0(N(2))$.

f. Calcule $P_0(N(4) = +\infty)$.

33. Seja $(X_n)_{n\geq 0}$ uma C.M. com espaço de estados $S = \{1, 2, 3, 4\}$ e matriz de transição

$$\mathcal{P} = \begin{pmatrix} 1/2 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

a. Mostre que esta cadeia é recorrente, irredutível e aperiódica.

b. Calcule $\mathcal{P}^2$, $\mathcal{P}^4$ e $\mathcal{P}^8$. Use estes resultados para estimar $E_j(T_j)$, para $j \in S$.

c. Obtenha a distribuição estacionária $(\pi(j))_{j\in S}$ resolvendo o sistema $\pi \times \mathcal{P} = \pi$.

d. Compare os itens b e c.

34. Seja $(X_n)_{n\geq 0}$ uma C.M. com espaço de estados $S = \{1,2,3,4,5\}$ e matriz de transição

$$\mathcal{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 1/3 & 0 & 2/3 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1/4 & 0 & 1/4 & 0 & 1/2 \end{pmatrix}.$$

a. Classifique os estados.

b. Obtenha a distribuição limite de X_n quando $n \to \infty$.

c. Encontre o tempo médio de recorrência para todos os estados recorrentes.

35. Seja $(X_n)_{n\geq 0}$ uma C.M. com espaço de estados $S = \{0,1,2,3,\cdots\}$ e matriz de transição

$$\mathcal{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \dots \\ 1/4 & 0 & 3/4 & 0 & 0 & \dots \\ 1/4 & 0 & 0 & 3/4 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}.$$

a. Classifique os estados.

b. Encontre a distribuição limite, se existir.

36. Considere duas urnas U_1 e U_2. A soma do número de bolas que estão nas duas urnas é N. Uma pessoa escolhe ao acaso uma bola (todas as bolas tem a mesma probabilidade de serem escolhidas) da totalidade destas N bolas (independente da urna em que estão). Então se joga uma moeda com probabilidade p de cair 1 (cara) e com probabilidade $1-p$ de sair 2 (coroa). A bola selecionada anteriormente é depositada na urna U_1 ou U_2,

dependendo se saiu respectivamente 1 ou 2 quando jogamos a moeda. O procedimento é repetido indefinidamente. Tome o conjunto de estados S como $S = \{0, 1, 2, 3, ...N\}$ e assuma que cada $s \in S$ descreve o número de bolas na urna U_1. Determine a cadeia de transição de Markov associada ao procedimento.

37. Seja uma matriz estocástica $\mathcal{P}$ do tipo n por n. Mostre que se o estado i pode ser atingido a partir de j (isto é, existe $k > 0$ tal que $(\mathcal{P}^k)_{j,i} > 0$), então existe $r \leq (n-1)$, tal que $(\mathcal{P}^r)_{j,i} > 0$).

38. Considere o passeio aleatório sobre os inteiros tal que para todo $i \in \mathbb{Z}$ vale $\mathcal{P}_{i,i+1} = p$, $\mathcal{P}_{i,i-1} = 1-p$ e $\mathcal{P}_{i,j} = 0$ para os outros casos de $j \in \mathbb{Z}$.

    Calcule $(\mathcal{P}^m)_{0,0}$ e a seguir determine a função geradora

$$F(s) \, = \, \sum_{m=0}^{\infty} (\mathcal{P}^m)_{0,0} \; s^m.$$

39. Nas mesmas hipóteses do último exercício, para

$$\omega = (w_0, w_1, w_2, ...),$$

tal que $w_i = 0$, seja $T_0(\omega) = i$, se i é a primeira vez que $w_j = 0$ (no caso de tal não ocorrer $T_0(\omega) = \infty$.) Considere o processo tal que $P(X_0 = 0) = 1$. Calcule a função geradora

$$P(s) \, = \, \sum_{m=0}^{\infty} P(T_0 = m) \; s^m.$$

3

Convergência de Variáveis Aleatórias

Neste capítulo vamos descrever alguns dos vários sentidos em que uma sequência de variáveis aleatórias pode convergir a uma outra fixada. Diversos problemas em Probabilidade e Estatística dependem do correto entendimento destas questões.

Lembre que neste texto para analisar um Processo Estocástico tomando valores em um conjunto $S \subset \mathbb{R}$ consideramos uma probabilidade P sobre o conjunto $\Omega = S^{\mathbb{N}}$. Ainda, se $\omega = (w_0, w_1, w_2, ..., w_n, ...) \in \Omega = S^{\mathbb{N}}$, então, como dissemos antes, assumiremos que $X_n(w) = w_n$.

O valor esperado de ϕ, denotado por $\mathbb{E}(\phi)$, é a integral $\int \phi(w) dP(w)$. Se o leitor desejar pode ver a formulação rigorosa no capítulo 5 mas isto não será necessário para o presente capítulo.

Lembre que quando $\phi : \Omega \to \mathbb{R}$ toma valores num conjunto finito este valor $E(\phi)$ é obtido via a definição 1.13 (que é consistente com o capítulo 5). No caso em que Ω é o espaço de Bernoulli $\Omega = \{1, 2, ..., d\}^{\mathbb{N}}$ e ϕ é constante, digamos nos cilindros de tamanho k fixo, então ϕ só toma finitos valores, e se pode usar o caso mais simples da definição e que é mencionado acima.

Por exemplo, se P é uma probabilidade de Markov, $k = 3$ (acima), e ϕ tem o valor ϕ_{a_0,a_1,a_2} no cilindro $\overline{a_0, a_1, a_2}$, $a_j \in \{1, 2, ..., d\}$, $j = 0, 1, 2$, então

$$\int \phi \, dP = \sum_{(a_0,a_1,a_2)\in\{1,2,\ldots,d\}^3} \phi_{a_0,a_1,a_2} \, P(\overline{a_0, a_1, a_2}).$$

Exemplos numéricos aparecem após o exemplo 3.4.

3.1 Lei dos Grandes Números

Sejam $X_1, X_2, \cdots, X_n, \cdots$ variáveis aleatórias independentes e identicamente distribuídas (ver definição na seção inicial). Assim, se X_1 é integrável, todas elas serão e ainda $\mathbb{E}(X_n) = \mathbb{E}(X_1)$, para todo $n \in \mathbb{N}$.

Queremos mostrar que em algum sentido

$$\lim_{n\to\infty} \frac{1}{n} \left(X_1(\omega) + \cdots + X_n(\omega)\right) = \mathbb{E}(X_1).$$

Recomendamos o leitor a [GS] e [HPS] para uma descrição mais detalhada do assunto.

O seguinte exemplo é ilustrativo.

Exemplo 3.1. Jogue uma moeda honesta n vezes, de maneira independente e conte o número de caras obtidas. O espaço amostral é $\Omega = \{(w_1, w_2, \cdots) : w_n = \text{cara ou } w_n = \text{coroa}, n \in \mathbb{N}\}$. Associe a cara o número 1 e a coroa o número 0. Neste caso $\Omega = \{0, 1\}^{\mathbb{N}}$.

$$X_n(w) = X(w_n) = \begin{cases} 1, & \text{se no n-ésimo lançamento ocorre cara} \\ 0, & \text{se no n-ésimo lançamento ocorre coroa} \end{cases} = I_A,$$

onde $A =$ cara no n-ésimo lançamento.

Logo, $X_1, X_2, \cdots$, são variáveis aleatórias independentes e identicamente distribuídas, onde $\mathbb{E}(X_n) = \frac{1}{2}$. Então, é natural esperar que, para ω num conjunto de probabilidade 1, quando $n \to \infty$,

$$S_n(\omega) = \frac{1}{n}\,(X_1(\omega) + \cdots + X_n(\omega)) \longrightarrow \mathbb{E}(X_1) = \frac{1}{2}.$$

Isto porque sabemos, intuitivamente, que se jogarmos a moeda muitas vezes, mais ou menos metade das vezes obteremos cara. Nesta seção desejamos fazer afirmações que sejam válidas, por exemplo, para caminhos ω num conjunto $K \subset S^{\mathbb{N}} = \{0,1\}^{\mathbb{N}}$, tal que $P(K) = 1$. Esta seria a maneira Matemática de formalizar o que indica a nossa intuição. Outras formas mais fracas desta afirmação também são úteis.

$\diamondsuit$

Vamos agora formalizar o que estamos afirmando acima. Primeiro é necessário descrever distintos tipos de convergência.

Note que os resultados obtidos no presente contexto tem natureza diversa do que foi analisado no capítulo anterior. Uma coisa é falar do limite de $P(X_n = j)$, para j fixo, quando n vai a infinito, outra é fazer afirmações para um conjunto K de caminhos $\omega \in K \subset S^{\mathbb{N}}$. No primeiro caso nos bastaria o conhecimento das distribuições finito-dimensionais de P sobre $\Omega = S^{\mathbb{N}}$, mas no segundo caso não.

Definição 3.1 (Convergência em Probabilidade). *Denotamos* $Y_n \stackrel{p}{\to} Y$, *se* $\forall \varepsilon > 0$, *temos que*

$$\lim_{n\to\infty} P(\{\omega \text{ tal que } |\,Y_n(\omega) - Y(\omega)\,| \geqslant \varepsilon\}\,) = 0,$$

onde Y, Y_n *são variáveis aleatórias em* $(\Omega, \mathcal{A}, P)$.

O conceito acima descreve o fato que para valores grandes de n, de alguma forma, as variáveis aleatórias Y_n e Y são aproximadamente iguais.

Definição 3.2 (**Convergência Quase Certamente**). *Denotamos* $Y_n \overset{q.c.}{\to} Y$,

$$P(\{\omega \mid \lim_{n\to\infty} Y_n(\omega) = Y(\omega)\}) = 1.$$

O último conceito acima descreve de outra forma o fato que para valores grandes de n, as variáveis aleatórias Y_n e Y são aproximadamente iguais. Os dois conceitos nem sempre coincidem.

Teorema 3.1. $Y_n \to Y$ *quase certamente* $\Rightarrow Y_n \overset{p}{\to} Y$.

Demonstração: Suponha que $Y_n \to Y$ quase certamente. Queremos mostrar que

$$P\left(|Y_n - Y| \geqslant \varepsilon\right) \to 0, \forall \varepsilon > 0.$$

Seja $A_0 = \{w \in \Omega : Y_n(w) \to Y(w)\}$. Sabemos que $P\left(A_0\right) = 1$. Para todo $w \in A_0$, $|Y_n(w) - Y(w)| < \varepsilon$, para todo n suficientemente grande. Considere

$$A_n = \{\omega \mid \text{para todo}\, k \geq n,\, |Y_k(\omega) - Y(\omega)| < \varepsilon\,\} = \bigcap_{k=n}^{\infty} [|Y_k - Y| < \varepsilon].$$

Como $A_n \subset A_{n+1}$ e como $w \in A_0$, então $w \in A_n$ para algum n, temos que

$$A_0 \subset \underset{n\geq 1}{\cup} A_n = \lim_{n\to\infty} A_n \quad \text{isto é,}\;\; A_n \uparrow \cup_{n\geq 1} A_n.$$

Logo,

$$1 = P\left(A_0\right) \leqslant P\left(\underset{n\geqslant 1}{\cup} A_n\right) \text{, e, então, } P\left(A_n\right) \uparrow 1.$$

Mas $A_n \subset [|Y_n - Y| < \varepsilon]$. E, então, $P\left(|Y_n - Y| \geq \varepsilon\right) = 1 - P\left(|Y_n - Y| < \varepsilon\right) \longrightarrow 1 - 1 = 0$.

Logo, $Y_n \overset{p}{\to} Y$.

$\square$

Exemplo 3.2. $Y_n \overset{p}{\to} Y$ mas $Y_n \overset{q.c.}{\nrightarrow} Y$. Considere a sequência de conjuntos

$$I_1 = [0,1],\; I_2 = \left[0, \frac{1}{2}\right],\; I_3 = \left[\frac{1}{2}, 1\right],\; I_4 = \left[0, \frac{1}{4}\right], \cdots, I_7 = \left[\frac{3}{4}, 1\right]$$

e mais geralmente,

$$I_{2^m+i} = \left[\frac{i}{2^m}, \frac{i+1}{2^m}\right].$$

Defina

$$Y_n = \begin{cases} 0, & \text{se } X \notin I_n \\ 1, & \text{se } X \in I_n. \end{cases}$$

(i) $Y_n \xrightarrow{p} 0$.
De fato: $P\left(|Y_n - 0| \geq \varepsilon\right) = P\left(Y_n = 1\right) = P\left(X \in I_n\right) = \frac{i+1}{2^n} - \frac{i}{2^n} = \frac{1}{2^n} \underset{n\to\infty}{\longrightarrow} 0$.

(ii) $Y_n \overset{q.c.}{\nrightarrow} 0$

De fato: $P\left(Y_n \nrightarrow 0\right) = 1$, já que $Y_n = 0$ ou 1, para infinitos n.

$P\left(\{w \in \Omega; Y_n(w) \nrightarrow 0\}\right) = P\left(\{w \in \Omega; Y_n(w) = 0 \,\text{ou}\, 1\,, \text{para infinitos } n\}\right) = P(\Omega) = 1$.

$\diamondsuit$

Sejam $X_1, X_2, \cdots, X_n, \cdots$ variáveis aleatórias integráveis em $(\Omega, \mathcal{A}, P)$.

Considere $S_n(\omega) = \sum_{i=1}^{n} X_i(\omega)$.

Definição 3.3. *A sequência $X_1, X_2, \cdots, X_n, \cdots$ satisfaz a Lei Fraca dos Grandes Números, se $\frac{S_n - \mathbb{E}(S_n)}{n} \to 0$ em probabilidade, ou equivalentemente, se*

$$P\left(\left|\frac{S_n - \mathbb{E}(S_n)}{n}\right| \geq \varepsilon\right) \to 0, \forall \varepsilon > 0.$$

Definição 3.4. *A sequência $X_1, X_2, \cdots, X_n, \cdots$ satisfaz a Lei Forte dos Grandes Números, se $\frac{S_n - \mathbb{E}(S_n)}{n} \to 0$ quase certamente, isto é,*

$$P\left(\frac{S_n - \mathbb{E}(S_n)}{n} \to 0\right) = 1.$$

Note que a afirmação da Lei Fraca envolve apenas as distribuições finito-dimensionais.

Se a sequência $X_1, X_2, \cdots, X_n, \cdots$ satisfaz a Lei Forte, então satisfaz a Lei Fraca, pelo teorema acima.

Teorema 3.2. *A desigualdade de Chebyshev afirma o seguinte: seja* $g : (0,\infty) \to (0,\infty)$, *estritamente crescente, e* $X : \Omega \to \mathbb{R}$, *variavel aleatória. Suponha que* $g(X)$ *seja* P*-integrável. Então, para qualquer* $a \in \mathbb{R}$, $a > 0$, *vale que*

$$P(X \geq a) \leq \frac{E(g(X))}{g(a)}.$$

Demonstração: Segue de imediato de

$$g(a)\, P(\{X \geq a\}) = E(g(a)\, I_{\{X \geq a\}}) \leq E(g(X)\, I_{\{g(X) \geq g(a)\}}) \leq E(g(X)).$$

□

Se aplicarmos a afirmação acima para $g(x) = x^2$ obtemos para cada a real fixado

$$P(X \geq a) \leq \frac{E(X^2)}{a^2}.$$

Teorema 3.3 (Lei Fraca de Chebyschev). *Sejam* $X_1, X_2, \cdots, X_n, \cdots$ *variáveis aleatórias, independentes com variâncias finitas e uniformemente limitadas (isto é,* $\exists\, c \in \mathbb{R}$ *tal que* $Var\ (X_n) \leqslant c, \forall n$*). Então,*

$$\frac{S_n - \mathbb{E}(S_n)}{n} \xrightarrow{p} 0.$$

Demonstração: Usando a desigualdade de Chebyshev acima para $X = |S_n - \mathbb{E}(S_n)|$, $g(x) = x^2$, e $a = n\varepsilon$, obtemos

$$P\left(\left|\frac{S_n - \mathbb{E}(S_n)}{n}\right| \geq \varepsilon\right) = P\left(|S_n - \mathbb{E}(S_n)| \geq \varepsilon n\right) \leqslant$$
$$\leqslant \frac{\mathrm{Var}\,(S_n)}{n^2\varepsilon^2} \leqslant \frac{nc}{n^2\varepsilon^2} = \frac{c}{n\varepsilon^2} \underset{n\to\infty}{\to} 0.$$

□

Teorema 3.4 (Lei Fraca de Khintchin). *Sejam $X_1, X_2, \cdots, X_n, \cdots$ variáveis independentes, identicamente distribuídas e integráveis com média comum μ. Então,*

$$\frac{S_n}{n} \xrightarrow{p} \mu.$$

Teorema 3.5 (Lei Forte de Kolmogorov). *Seja $(X_n)_{n\geq 1}$ sequência de variáveis aleatórias independentes e integráveis tais que $\sum\limits_{n\geq 1} \frac{Var\,(X_n)}{n^2} < \infty$. Então, vale a Lei Forte dos Grandes Números, isto é, $\frac{S_n - \mathbb{E}(S_n)}{n} \xrightarrow{q.c.} 0$.*

Os resultados acima não serão demonstrados.

No Capítulo 5 vamos enunciar o Teorema Ergódico de Birkhoff, resultado de grande generalidade, e que tem como consequência a Lei Forte do Grandes Números. Maiores detalhes serão fornecidos naquela parte.

3.2 Lema de Borel-Cantelli

Seja $A_1, A_2, \cdots, A_n, \cdots$ sequência de conjuntos mensuráveis (eventos). Definimos:

limite superior: $\lim\limits_{n\to\infty} \sup A_n = \bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k=n}^{\infty} A_k$

limite inferior: $\lim\limits_{n\to\infty} \inf A_n = \bigcup\limits_{n=1}^{\infty} \bigcap\limits_{k=n}^{\infty} A_k$.

Idéia Intuitiva:

1) O evento $\underline{\lim\sup A_n}$ é o evento "ocorrência de um número infinito dos A_n", já que $w \in \lim\sup A_n \Rightarrow w \in \bigcap\limits_{n\geq 1} \bigcup\limits_{k\geq n} A_k \Rightarrow w \in \bigcup\limits_{k=n}^{\infty} A_k, \forall n \Rightarrow w \in \bigcup\limits_{k=1}^{\infty} A_k \Rightarrow w \in A_{k_1}$ para algum k. Mas $w \in \bigcup\limits_{k=k_1+1}^{\infty} A_k$, logo $w \in A_{k_2}$ para algum $k_2 > k_1$. Continuando, $w \in \bigcup\limits_{k=k_2+1}^{\infty} A_k$, logo $w \in A_{k_3}$ para algum $k_3 > k_2$, etc. Desta maneira, temos uma sequência crescente de números inteiros positivos

$k_1 < k_2 < k_3 < \cdots$ que dependem de w tais que $w \in A_{k_n}$, $\forall n$. Logo, w pertence a um número infinito dos A_n's então $w \in \bigcup\limits_{k \geq n} A_k$, $\forall n$. Logo,

$$w \in \limsup A_n.$$

Notação: $\limsup A_n = [A_n \text{ infinitas vezes}] = [A_n \text{ i.v.}]$.

2) $\liminf A_n =$[ocorrência de A_n para todo n suficientemente grande].

$w \in \liminf A_n \Rightarrow w \in \bigcap\limits_{k \geq n_0} A_k$ para algum $n_0 = n_0(w) \Rightarrow w \in A_k$ para todo k suficientemente grande, isto é, $k \geq n_0$.

Definição 3.5. *Se* $\lim \sup A_n = \lim \inf A_n = A$ *então dizemos que* A *é o limite dos* A_n*, e denotamos* $A_n \to A$.

Lema 3.1 (Lema de Borel-Cantelli). *Seja* $(A_n)_{n \geq 1}$ *sequência de eventos aleatórios em* $(\Omega, \mathcal{A}, P)$*, isto é,* $A_n \in \mathcal{A}, \forall n$.

(a) *Se* $\sum\limits_{n \geq 1} P(A_n) < \infty$ *então* $P(\lim \sup A_n) = P(x \in A_n$ *infinitas vezes*$)=0$.

(b) *Se* $(A_n)_{n \geq 1}$ *é uma sequência de eventos independentes tais que* $\sum\limits_{n \geq 1} P(A_n) = \infty$, *então* $P(x \in A_n$ *infinitas vezes*$)=1$.

Demonstração:

(a) Se $\sum\limits_{n \geq 1} P(A_n) < \infty \Rightarrow \lim_{n \to \infty} \sum\limits_{k=n}^{\infty} P(A_k) \to 0$.

Observe que

$$P(A_n \text{ infinitas vezes}) = P\left(\bigcap_{n \geq 1} \bigcup_{k=n}^{\infty} A_k\right) = P\left(\lim_{n \to \infty} \bigcup_{k=n}^{\infty} A_k\right) = \lim_{n \to \infty} P\left(\bigcup_{k=n}^{\infty} A_k\right) \leqslant$$

$$\leqslant \lim_{n \to \infty} \sum_{k=n}^{\infty} P(A_k) = 0, \text{ pois} \sum_{n \geq 1} P(A_k) < \infty \Longleftrightarrow \lim_{n \to \infty} \sum_{k=n}^{\infty} P(A_k) = 0.$$

Logo, $P(A_n \text{ i.v.})=0$.

(b) Se $A = \lim\limits_{n\to\infty} \sup A_n = \lim\limits_{n\to\infty} \bigcup\limits_{k=n}^{\infty} A_k = \bigcap\limits_{n\geq 1} \bigcup\limits_{k=n}^{\infty} A_k$ então, $A^c = \bigcup\limits_{n\geq 1} \bigcap\limits_{k=n}^{\infty} A_k^c$ de tal modo que

$$P(A^c) = P\left(\bigcup_{n\geq 1} \bigcap_{k=n}^{\infty} A_k^c\right) = P\left(\lim_{n\to\infty} \bigcap_{k=n}^{\infty} A_k^c\right) = \lim_{n\to\infty} P\left(\bigcap_{k=n}^{\infty} A_k^c\right).$$

Para $n_0 > n$ observamos que $\bigcap\limits_{k=n}^{\infty} A_k^c \subset \bigcap\limits_{k=n}^{n_0} A_k^c$ e, portanto,

$$P\left(\bigcap_{k=n}^{\infty} A_k^c\right) = \lim_{n_0\to\infty} P\left(\bigcap_{k=n}^{n_0} A_k^c\right) = \lim_{n_0\to\infty} \prod_{k=n}^{n_0} P(A_k^c) = \lim_{n_0\to\infty} \prod_{k=n}^{n_0} (1 - P(A_k))$$

já que $(A_n)_{n\geq 1}$ é a sequência de eventos independentes.
Como para todo $x > 0$ vale a desigualdade $1 - e^{-x} \leq x$, então $e^{-x} \geq 1 - x$, e assim, para cada k vale $1 - P(A_k) \leq e^{-P(A_k)}$. Desta forma, tomando o produto $\prod\limits_{k=n}^{n_0} (1 - P(A_k))$, concluímos que

$$P\left(\bigcap_{k=n}^{\infty} A_k^c\right) \leqslant \lim_{n_0\to\infty} \exp\left(-\sum_{k=n}^{n_0} P(A_k)\right) = \exp\left(-\lim_{n_0\to\infty} \sum_{k=n}^{n_0} P(A_k)\right) = 0,$$

pois a série $\sum\limits_{n\geq 1} P(A_n)$ diverge. Portanto, $P(A^c) = \lim\limits_{n\to\infty} P\left(\bigcap\limits_{k=n}^{\infty} A_k^c\right) = 0$, o que implica em $P(A) = 1$.

$\square$

Note que se $(A_n)_{n\geq 1}$ é uma sequência de eventos independentes então $P(x \in A_n$ infinitas vezes $)$ será ou igual a 0 ou igual a 1. Esta é uma forma simples da assim chamada Lei zero ou um (que aparece em alguns processos ergódicos).

Exemplo 3.3. Seja uma cadeia de Markov com conjunto de estados S. Seja i em S e suponha que $P(X_0 = i) = 1$. Para cada n, considere o subconjunto de Ω, dado por $A_n = \{\omega | X_n(\omega) = i\}$.

Se $\sum_n P(A_n) < \infty$, pelo Lema de Borel-Cantelli (a), temos que

$$P\left(A_n \text{ infinitas vezes}\right) = 0.$$

Desta forma, se $\sum_n P(A_n) < \infty$, então i é transiente.

$\diamondsuit$

Exemplo 3.4. Seja X_n processo independente e identicamente distribuído sobre $S = \{1, 2, ..., d\}$, definido por $P(X_0 = i) = p_i > 0$, para todo i.

Fixe $i_0 \in S$. Para cada n seja $A_n = \{\omega | X_n(\omega) = i_0\}$. Os conjuntos A_n são independentes e $P(A_n) = p_{i_0}$. Logo $\sum_n P(A_n) = \infty$, e assim, pelo Lema de Borel-Cantelli (b), temos que $P\left(A_n \text{ infinitas vezes}\right) = 1$. Neste caso, todos os estados são recorrentes.

$\diamondsuit$

Exemplo 3.5. Pode acontecer que $Y_n \overset{p}{\to} 0$, mas $Y_n \overset{q.c.}{\nrightarrow} 0$.

Considere $X_1, X_2, \cdots$ independentes e identicamente distribuídas tais que $X_n \sim \mathcal{E}(1)$.
Defina $Y_1 = X_1$ e $Y_n = \frac{X_n}{\ln(n)}$, para $n > 1$.

(i) $Y_n \overset{p}{\to} 0$ já que $P\left(|Y_n| \geq \varepsilon\right) = P\left(|X_n| \geq \varepsilon |\ln(n)|\right) = P\left(X_n \geq \varepsilon \ln(n)\right) = 1 - P\left(X_n \leqslant \varepsilon \ln(n)\right) = 1 - (1 - e^{-\varepsilon \ln(n)}) = e^{-\ln(n)^{\varepsilon}} = e^{\ln(n)^{-\varepsilon}} = n^{-\varepsilon} = \frac{1}{n^{\varepsilon}} \underset{n\to\infty}{\longrightarrow} 0$.

(ii) $Y_n \overset{q.c.}{\nrightarrow} 0$. De fato:

Vamos mostrar que $P\left(Y_n \geq \varepsilon \text{ infinitas vezes}\right)$=1 para algum $\varepsilon > 0$.
Isto implica que $Y_n \geq \varepsilon$ infinitas vezes com probabilidade 1. Logo, Y_n não converge para zero.
Observe que tomando
$A_n = [Y_n \geq \varepsilon]$, então, $w \in [A_n$ i.v.]$\Rightarrow$ $Y_n(w) \geq \varepsilon$ para um número infinito de n's

$\Rightarrow Y_n(w)$ não converge a zero.

Se $P(A_n \text{ i.v.}) = 1 \Rightarrow P(Y_n \nrightarrow 0) = 1$.

Como os eventos $[Y_n \geq \varepsilon]$ são independentes já que os Y_n's o são,

$$\sum_n P(Y_n \geq \varepsilon) = \sum_n P(X_n \geq \varepsilon \ln(n)) = \sum_n e^{-\varepsilon \ln(n)} = \sum_n \frac{1}{n^\varepsilon} = \infty,$$

se $0 < \varepsilon < 1$.

Pelo Lema de Borel-Cantelli, parte **(b)**, $P(Y_n \geq \varepsilon \text{ i.v.})=1$, se $0 < \varepsilon < 1$. Portanto $Y_n \overset{q.c.}{\nrightarrow} 0$.

$\diamondsuit$

Teorema 3.6. *Seja $(A_n)_{n\geq 1}$ sequência de eventos independentes e $A_n \to A$. Então, $P(A) = P(x \text{ esta em } A_n \text{ infinitas vezes }) = 0$ ou $P(A) = 1$.*

Demonstração: O teorema segue do fato que a série $\sum_{n\geq 1} P(A_n)$ ou converge ou diverge. Quando diverge, isto implica em $P(A) = 1$ e quando converge, implica em $P(A) = 0$.

$\square$

Como uma simples aplicação do Lema de Borel-Cantelli, obteremos uma versão da Lei Forte dos Grandes Números.

Teorema 3.7. *Se $(X_n)_{n\geq 1}$ é sequência de variáveis aleatórias independentes e identicamente distribuídas com média comum μ e quarto momento finito, ou seja, $\int (X_1 - \mu)^4 dP < \infty$, então,*

$$P\left(\lim_{n\to\infty} \frac{S_n}{n} = \mu\right) = 1.$$

Demonstração:

Defina

$$\binom{n}{a_1, \ldots, a_k} = \frac{n\,!}{a_1\,! \ldots a_k\,!}$$

Observe que

$$\left[\sum_{i=1}^{n}(X_i-\mu)\right]^4 =$$

$$\sum_{i=1}^{n}(X_i-\mu)^4 + \binom{4}{3,1}\sum_{i\neq j}(X_i-\mu)^3(X_j-\mu)+$$

$$\binom{4}{2,2}\sum_{i\neq j}(X_i-\mu)^2(X_j-\mu)^2+$$

$$\binom{4}{1,1,2}\sum_{i\neq j\neq k \text{ e } \; i\neq k}(X_i-\mu)\,(X_j-\mu)(X_k-\mu)^2+$$

$$\binom{4}{1,1,1,1}\sum_{i,j,k,s \text{ distintos}}(X_i-\mu)\,(X_j-\mu)(X_k-\mu)\,(X_s-\mu).$$

Como as v.a.'s X_i são independentes tais que $\mathbb{E}(X_i-\mu)=0$ e Var $X_i=\sigma^2$, então $\mathbb{E}[\sum(X_i-\mu)]^4 = \mathbb{E}[\sum(X_i-\mu)^4] + \binom{4}{2}\,\mathbb{E}\left(\sum_{i\neq j}(X_i-\mu)^2(X_j-\mu)^2\right) = n\,\mathbb{E}(X_1-\mu)^4 + 6\sum_{i\neq j}\mathbb{E}(X_i-\mu)^2\,\mathbb{E}(X_j-\mu)^2 = n\,\mathbb{E}(X_1-\mu)^4 + \frac{6\,n(n-1)}{2}\sigma^2\sigma^2 = n\,\mathbb{E}(X_1-\mu)^4 + 3\,n(n-1)\sigma^4 \leqslant Cn^2$.

Portanto, $\mathbb{E}[\sum(X_i-\mu)]^4 \leqslant Cn^2$.

A desigualdade de Markov afirma que para $a>0$ e $\beta>0$ dados

$$P(\{|X|>a\}) = P(\{|X|^\beta > a^\beta\}) \leq \frac{\int |X|^\beta\, dP}{a^\beta}.$$

A demostração deste fato segue de considerar a função indicador de $\{|X|^\beta > a^\beta\}$.

Pela Desigualdade de Markov, temos que

$$P\left\{\left|\sum_{i=1}^{n}(X_i-\mu)\right| > n\,\varepsilon\right\} \leqslant \frac{\mathbb{E}\{\sum(X_i-\mu)\}^4}{n^4\varepsilon^4} \leqslant \frac{Cn^2}{n^4\varepsilon^4} = \frac{C}{n^2\varepsilon^4} = \frac{c}{n^2}.$$

Neste caso, $P\left(|S_n - n\mu| > n\varepsilon\right) \leqslant \frac{c'}{n^2}$ e então

$$\sum_{n\geq 1} P\left(|S_n - n\mu| > n\varepsilon\right) \leqslant \sum_{n\geq 1} \frac{c'}{n^2} < \infty \text{ , pois a série } \sum_{n\geq 1} n^{-2} \text{ converge.}$$

Pelo Lema de Borel-Cantelli, item **(a)**, com probabilidade 1 somente um número finito de eventos $\{w : |\frac{S_n}{n} - \mu| > \varepsilon\}$ ocorrem, isto é, $P\left(A_\varepsilon\right) = 0$ onde

$$A_\varepsilon = \lim_{n\to\infty} \sup \left\{ \left| \frac{S_n}{n} - \mu \right| > \varepsilon \right\}.$$

Observe ainda que os conjuntos A_ε crescem, à medida que $\varepsilon \to 0$, para o conjunto $\{w; \frac{S_n}{n} \nrightarrow \mu\}$. Fazendo $\varepsilon \to 0$ através de um conjunto enumerável de valores temos que

$$P\left(\frac{S_n}{n} - \mu \nrightarrow 0\right) = P\left(\bigcup_{k\in\mathbb{N}^*} A_{\frac{1}{k}}\right) = 0.$$

Portanto,

$$P\left(\frac{S_n}{n} - \mu \to 0\right) = 1, \text{ ou seja, } P\left(\{w \,|\, \lim_{n\to\infty} \frac{S_n(w)}{n} = \mu\}\right) = 1.$$

□

Teorema 3.8. *Se $X_1, X_2, \cdots$ são variáveis aleatórias independentes e identicamente distribuídas tais que $P\left(|X_n| < K\right) = 1$ para todo n, onde $K \in \mathbb{R}$ e $K > 0$, então $\left(\frac{S_n}{n} \overset{q.c.}{\to} \mu\right)$.*

Demonstração: Se a sequência $(X_n)_{n\geq 1}$ é formada por variáveis uniformemente limitadas, então todos os seus momentos de qualquer ordem existem. Logo, valem as hipóteses do Teorema anterior e o resultado segue.

□

Exemplo 3.6. Fixe $\epsilon > 0$. Assim se jogarmos uma moeda que tem probabilidade $\frac{1}{3}$ de sair cara (associada ao número 1) e $\frac{2}{3}$ de sair coroa (associada ao número 2) uma quantidade n de vezes, então, a probabilidade que a média de vezes que sai cara (nas n jogadas) fique distante ϵ do valor $\frac{1}{3}$ vai a zero com $n \to \infty$. Esta afirmação é a Lei Fraca dos Grandes Números. Isto porque jogar uma moeda várias vezes seguidas é descrito por um processo independente.

Dito de outra forma

$$\lim_{n\to\infty} P(\{\, |\, \frac{\text{número de vezes que saiu cara em } n \text{ jogadas}}{n} - \frac{1}{3}\, | > \epsilon\,\}\,) = 0.$$

Ou, ainda,

$$\lim_{n\to\infty} P(\{w = (w_0, w_1, ..., w_k, ...) \,\text{tal que}\, |\frac{\sum_{j=0}^{n-1} I_{\overline{1}}\,(w_j, w_{j+1}, ...)}{n} - \frac{1}{3}\,| > \epsilon\,\})\; =\; 0,$$

onde $I_{\overline{1}}$ é a função indicador do cilindro $\overline{1}$ (lembre que $I_{\overline{1}}(w) = 1$, se e só se $w = (w_0, w_1, ..)$ começa com 1).

Note que os conjuntos da forma

$$\{w \in \Omega\,,\; |\, \frac{\text{número de vezes que saiu cara em } n \text{ jogadas}}{n} - \frac{1}{3}\,| > \epsilon\,\}$$

com n fixo, dependem só das n primeiras coordenadas de cada w.

Podemos ainda ver o enunciado acima da seguinte forma: seja $X_j : \{1, 2\}^{\mathbb{N}} \to \{0, 1\} \subset \mathbb{R}$, $j \geq 0$, dado por $X_j(w) = I_{\overline{1}}(w_j, w_{j+1}, ..., w_k, ...)$.

Então vale

$$\lim_{n\to\infty} P(\{w \,\text{tal que}\, |\frac{\sum_{j=0}^{n-1} X_j(w)}{n} - \frac{1}{3}\,| \geq \epsilon\,\}\,) \;=\; 0.$$

A Lei Forte dos Grandes Números afirma que existe um conjunto $K \subset \{1, 2\}^{\mathbb{N}}$ de probabilidade 1 tal que para qualquer $w = (w_0, w_1, ..., w_k, ...) \in K$ vale que

$$\lim_{n\to\infty} \frac{\sum_{j=0}^{n-1} I_{\overline{1}}\,(w_j, w_{j+1}, ..., w_k, ...)}{n} = \frac{1}{3}.$$

Ou, ainda

$$\lim_{n\to\infty} \frac{\sum_{j=0}^{n-1} X_j(w)}{n} = \frac{1}{3}.$$

Este resultado segue do Teorema 3.6 ou 3.7 acima. Ele também pode ser obtido do Teorema Ergódico (ver Exemplo 5.22) que será discutido na seção 5. Isto porque, como veremos, um processo independente é ergódico.

Note que o conjunto

$$\{w = (w_0, w_1, w_2, w_3, ..., w_k..) \in \Omega \,|\, \lim_{n\to\infty} \frac{\sum_{j=0}^{n-1} I_{\overline{1}}\,(w_j, w_{j+1}, ..., w_k..)}{n} = \frac{1}{3}\} =$$

$$= \{\, w \in \Omega \,|\, \lim_{n\to\infty} \frac{\sum_{j=0}^{n-1} X_j(w)}{n} = \frac{1}{3}\,\},$$

em que necessitamos explicitar a sua probabilidade **não é um cilindro** (depende das infinitas coordenadas de cada w).

$\diamondsuit$

Note que se X_n, $n \in \mathbb{N}$, é uma sequencia de funções, então w é tal que $\lim_{n\to\infty} X_n\,(w) = 0$, se e só se, $w \in \cap_k \cup_N \cap_{n\geq N} \{X_n(w) \,\leq\, 1/k\}$.

Teorema 3.9. *Seja $(X_n)_{n\geq 1}$ sequência de variáveis aleatórias independentes. Então,*

$$X_n \overset{q.c.}{\to} 0 \iff \sum_{n\geq 1} P\,(|X_n| > \varepsilon) < \infty, \text{ para todo } \varepsilon > 0.$$

Demonstração: $(\Longrightarrow)$

Consideremos os eventos $A_n = \{|X_n| > \varepsilon\}$. Por hipótese, $(A_n)_{n\geq 1}$ é uma sequência de eventos independentes.

Suponha que $X_n \overset{q.c.}{\to} 0$. Então, $X_n \to 0$ em um conjunto E^c com $P\,(E) = 0$. Um ponto $w \in E^c$ pertence somente a um número finito de eventos A_n. Segue então que

$$\lim_{n\to\infty} \sup A_n \subset E.$$

Portanto, $P\,(x \text{ está em } A_n \text{ infinitas vezes }) \leqslant P\,(E) = 0$, ou seja, $P\,(x \text{ está em } A_n$ infinitas vezes $) = 0$. Pelo Lema de Borel-Cantelli, parte **(b)**, então $\sum_{n\geq 1} P\,(A_n) <$

∞, pois se $\sum_{n\geq 1} P(A_n) = \infty$, então $P(A_n \text{ i.o.}) = 1$. Mas então $\sum_{n\geq 1} P(|X_n| > \varepsilon) < \infty$, para todo $\varepsilon > 0$.

($\Longleftarrow$)

Considere os eventos $A_{\frac{1}{k}} = \lim_{n\to\infty} \sup \{|X_n| > \frac{1}{k}\}$.
Usando os mesmos argumentos na prova do Teorema 3.1 concluímos que $X_n \stackrel{q.c.}{\to} 0$.

$\square$

Teorema 3.10. *Seja $(X_n)_{n\geq 1}$ sequência de variáveis aleatórias independentes. Então,*

$$X_n \stackrel{q.c.}{\to} X \Longleftrightarrow \sum_{n\geq 1} P(|X_n - X| > \varepsilon) < \infty, \text{ para todo } \varepsilon > 0.$$

Demonstração: Imediata, pois

$$X_n \stackrel{q.c.}{\to} X \Longleftrightarrow X_n - X \stackrel{q.c.}{\to} 0.$$

$\square$

Teorema 3.11. *Seja $(X_n)_{n\geq 1}$ sequência de variáveis aleatórias quaisquer.*
Se $\sum_{n\geq 1} P(|X_n| > \varepsilon) < \infty$, para todo $\varepsilon > 0$, então $X_n \xrightarrow{q.c.} 0$

Demonstração: Seja $\varepsilon = \frac{1}{k}$, $k \in \mathbb{N}^*$. Então, pelo Lema de Borel-Cantelli

$$\sum_{n\geq 1} P\left(|X_n| > \frac{1}{k}\right) < \infty \Longleftrightarrow$$

$$P\left(\left\{w/\exists\ \mathcal{N} > 0, \forall n > \mathcal{N}, |X_n(w)| < \frac{1}{k}\right\}\right) = P(A_k) = 1.$$

Considere $A = \bigcap_{k\geq 1} A_k$. Observe que para todo $w \in A$ e todo $\xi > 0$, existe $k \in \mathbb{N}^*$ tal que $\frac{1}{k} < \xi$. Como $w \in A_k$ para todo $k \in \mathbb{N}^*$ segue que $1 = P\left(\left\{w/\exists\ k \in \mathbb{N}^* \text{ tal que } |X_n(w)| < \frac{1}{k} < \xi\right\}\right)$, ou seja, $X_n \xrightarrow{q.c.} 0$.

3.3 Teorema Central do Limite

Lembre que dada uma variável aleatória

$$X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{R}),$$

chamamos de distribuição de X, a probabilidade μ_X sobre $(\mathbb{R}, \mathcal{R})$ tal que

$$\mu_X(A) \,=\, P(X^{-1}(A)),$$

para todo boreleano $A \in \mathcal{R}$ (mais detalhes na definição 5.5).

Note que para calcular $\int X^2 dP$, basta calcular $\int x^2 d\mu_X(x)$.

Da mesma forma para calcular $\int e^X dP$ basta calcular $\int e^x d\mu_X(x)$.

De uma maneira mais geral, dada uma variável aleatória da forma $G(X)$ onde $G : \mathbb{R} \to \mathbb{R}$ é contínua, temos que $\int G(X) dP = \int G(x) d\mu_X(x)$. Este fato é demonstrado no Capítulo 5 logo após o Exemplo 5.13.

Dada uma função contínua $G : \mathbb{R} \to \mathbb{C}$, podemos escrevê-la da forma $G = G_1 + i\, G_2$, $G_i : \mathbb{R} \to \mathbb{R}$, $i = 1, 2$. Temos então, por definição, que $\int G(x) d\mu_X(x) = \int G_1(x) d\mu_X(x) + i \int G_2(x) d\mu_X(x)$.

Finalmente, lembre que a função de distribuição $F : (\mathbb{R}, \mathcal{R}) \to (\mathbb{R}, \mathcal{R})$ de X é aquela que satisfaz

$$F_X(x) = F(x) \,=\, P(X \leq x) \,=\, \mu_X\,(\,(-\infty, x]\,).$$

Note que F_X é monótona não decrescente.

É usual a notação $\int g(x)\, d\mu_X(x) \,=\, \int g(x)\, d\, F_X(x)$, ou seja, muitas vezes se usa a distribuição F_X da variável X na integração em vez da probabilidade μ_X. Resulta ser a mesma coisa. A integral de Stieltjes, ou seja, a integração da distribuição $d\, F_X$ é descrita com detalhes em [Ba2].

Por exemplo, dizemos que uma variável aleatória $X : \Omega \to \mathbb{R}^+$, onde $(\Omega, \mathcal{A}, P)$ é um espaço de probabilidade, tem distribuição exponencial de parâmetro λ se

$$F_X(x) \,=\, P(\{w\,|X(w) \,\leq x\}\,) \,=\, \int_0^x \lambda\, e^{-\lambda\, y}\, dy.$$

Se μ_X for absolutamente contínua em relação a medida de Lebesgue, então, pelo Teorema de Radon-Nikodyn (ver Capítulo 5), existe $f : (\mathbb{R}, \mathcal{R}) \to (\mathbb{R}, \mathcal{R})$, onde $f(y) \geq 0$, tal que

$$\int_{-\infty}^{x} f(y)\, dy \;=\; \mu_X\,(\,(-\infty, x)\,) = \; F(x),$$

para todo x.

Se F é diferenciável em x, então $F'\,(x) = f(x)$.

Se F_X tem derivada, dizemos que f_X, dada por $F_X'\,(x) = f_X(x)$, descreve a densidade da variável X.

No caso de uma Variável Aleatória X que é descrita por uma exponencial de parâmetro λ, temos que $f_X(x) = \lambda\, e^{-\lambda\, x}$.

Neste caso, por exemplo,

$$\int X^2 dP \;=\; \int\, x^2\, \lambda\, e^{-\lambda\, x}\, dx \;=\int\, x^2\, f_X(x)\, dx.$$

É usual na literatura expressar $\int\, x^2\, f_X(x)\, dx$ como $\int\, x^2\, d\, F_X(x)\,$.

Vamos considerar a seguir o caso geral em que a probabilidade μ_X não é necessariamente absolutamente continua em relação a medida de Lebesgue.

Definição 3.6. *Seja $X : \Omega \to \mathbb{R}$ variável aleatória definida no espaço de probabilidade $(\Omega, \mathcal{A}, \mathbf{P})$. A função $\varphi_X(t)$ (tomando valores complexos) dada por*

$$\varphi_X(t) = \mathbb{E}(e^{itX}) = \int e^{itx} dF_X(x) = \int \cos(t\,x) dF_X(x) + \, i \int \sin(t\,x) dF_X(x),$$

para todo $t \in \mathbb{R}$, é dita ser **função característica de X**, *onde $i = \sqrt{-1}$.*

Conforme exemplo 3.20 a função caraterística de uma variável aleatória X que possui distribuição Gaussiana com média μ e variância σ é

$$\varphi_X(t) = e^{i\,t\,\mu - \frac{1}{2}\sigma^2\,t^2}.$$

Um caso de particular interesse é quando a variável aleatória X é Gaussiana tem média zero e a variância igual a 1. Isto é sua densidade f é descrita por $f(x) = \frac{e^{-\frac{1}{2}x^2}}{c}$, onde c é uma constante para normalização. Neste caso

$$\varphi_X(t) = \frac{1}{c}\,e^{-\frac{1}{2}t^2}.$$

Ao fim da seção 3.4 apresentamos algumas das propriedades básicas da função característica de uma variável aleatória X.

Teorema 3.12. *Suponha que $f : \mathbb{R} \to \mathbb{R}$ seja a função densidade de distribuição da variável $X : \Omega \to \mathbb{R}$ e que φ seja a função característica de X. Então,*

$$f(x) \,=\, \frac{1}{2\,\pi} \int_{-\infty}^{\infty} e^{-\,i\,t\,x}\,\varphi(t)\,dt,$$

em todo o ponto em que f é diferenciável.

O resultado acima segue de propriedades clássicas de transformada de Fourier. Referimos o leitor interessado numa prova deste fato para [Ru].

A função característica, como veremos, possui propriedades operacionais que a tornam particularmente útil em diversos cálculos.

Teorema 3.13. *Sejam X e Y duas variáveis aleatórias e suponha que existam f e g suas funções de densidade. Então, $f = g$, (Lebesgue quase toda parte) se e só se, as funções características φ_X e φ_Y são iguais (Lebesgue quase toda parte). Mais geralmente, duas variáveis tem a mesma distribuição, se e só se, tem a mesma função característica.*

O resultado acima não será demonstrado. Referimos o leitor interessado na prova para a Seção 5.9 em [GS].

O teorema acima afirma que podemos recuperar a distribuição a partir da função característica.

Vamos considerar a seguir as variáveis aleatórias X_n, $n \in \mathbb{N}$, de um Processo Estocástico, as respectivas funções de distribuição F_n e as respectivas funções características φ_{X_n}.

Definição 3.7. *Dizemos que a sequência de funções de distribuição F_n converge simplesmente à função F, se*

$$F(x) = \lim_{n\to\infty} F_n(x),$$

para todo ponto x em que F é contínua.

Uma pergunta natural é se podemos tratar da convergência das funções de distribuição a partir da convergência de funções características.

Definição 3.8. *Dizemos que a sequência de variáveis aleatórias X_n, $n \in \mathbb{N}$, converge em distribuição à variável aleatória X, se a sequência de funções de distribuição F_n (respectivamente, de cada X_n) converge simplesmente à função F (a distribuição de X). Isto é, se*

$$F(x) = \lim_{n\to\infty} F_n(x),$$

para todo ponto x em que F é contínua.

O resultado fundamental neste tópico é o seguinte:

Teorema 3.14 (**Teorema da Continuidade**)**.** *Considere uma sequência de variáveis aleatórias X_n, $n \in \mathbb{N}$, a sequência respectiva de funções de distribuição F_n e a sequência de funções características φ_{X_n}.*

(1) Se F_n converge simplesmente a uma função F, que é distribuição de uma variável aleatória X, e que possui função característica φ_X, então φ_{X_n} converge simplesmente a φ_X.

(2) Suponha que φ_{X_n} converge simplesmente à ϕ, e que ϕ seja contínua em $t=0$, então existe uma variável aleatória X, com $\phi = \varphi_X$, que possui função de distribuição F, e ainda, vale que F_n converge simplesmente à F. Ou seja, $\phi = \varphi_X$ para alguma $X : \Omega \to \mathbb{R}$, e X_n converge em distribuição à X.

Referimos o leitor à [GS] para a prova do resultado acima.

Segue do resultado acima que se X_n, $n \in \mathbb{N}$, é tal que $\varphi_{X_n}(t)$, converge simplesmente à $\phi(t) = e^{i\,t\,\mu - \frac{1}{2}\sigma^2\,t^2}$, quando $n \to \infty$, então X_n converge em distribuição a uma variável aleatória que tem distribuição Gaussiana de média μ e variância σ.

No próximo teorema vamos usar o item (2) acima.

Teorema 3.15 (**Teorema Central do Limite**). *Seja X_n, $n \in \mathbb{N}$, um Processo Estocástico independente identicamente distribuído com média μ e variância não-nula σ^2. Seja $S_n = X_1 + X_2 + ... + X_n$, então*

$$\frac{S_n - n\mu}{\sqrt{n\,\sigma^2}}$$

converge em distribuição a uma variável gaussiana com média 0 *e variância* 1

De outra forma, para qualquer intervalo (a,b) vale

$$\lim_{n\to\infty} P(\omega | \frac{S_n - n\mu}{\sqrt{n\,\sigma^2}} \in (a,b)) = \int_a^b f(x)dx = \frac{1}{\sqrt{2\,\pi}} \int_a^b e^{-\frac{x^2}{2}}\,dx,$$

onde f é a densidade Gaussiana de média 0 *e variância* 1.

Demonstração: Vamos assumir, sem perda de generalidade que $E(X_1) = 0$ e $E(X_1^2) = 1$. O caso geral pode ser obtido a partir deste caso.

A grosso modo a ideia principal da prova é mostrar que se $X = X_1$, então

$$\varphi_X(\frac{t}{\sqrt{n}}) \sim (1 - \frac{t^2}{2\,n}).$$

Pela Fórmula de Taylor de ordem 2 sabemos que

$$e^{y\,i} = 1 + i\,y + \frac{1}{2}\,(i\,y\,)^2 + r(y),$$

tal que $\lim_{y\to 0} \frac{r(y)}{y^2} = 0$.

Logo, dado $\epsilon > 0$, existe $\delta > 0$ tal que se $|y| < \delta$, então $\frac{r(y)}{y^2} < \epsilon$.

Disto segue que,

$$\varphi_{X_1}(s) = 1 + iE(X_1)\,s + i^2 \frac{E(X_1^2)\,s^2}{2} + R(s) = 1 - \frac{s^2}{2} + R(s),$$

onde $\lim_{s\to 0} \frac{R(s)}{s^2} = 0$.

A prova da afirmação acima será feita ao fim da demonstração do teorema.

Todas as X_i tem distribuição comum de uma determinada X.

Como as variáveis X_i são independentes então, como veremos no Teorema 3.18, se $S_n = (X_1 + X_2 + ... + X_n)$ temos que

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) = (\,\varphi_X(\frac{t}{\sqrt{n}})\,)^n.$$

Desejamos mostrar que

$$\lim_{n\to\infty} (\,\phi_X(\frac{t}{\sqrt{n}})\,)^n = e^{-\frac{t^2}{2}},$$

pois então seguirá do teorema da continuidade que a distribuição de $\frac{S_n}{\sqrt{n}}$ converge a uma normal de média zero e variância 1.

Para t fixo, vamos denotar $\theta_t(n) = \theta(n) = R(\frac{t}{\sqrt{n}})$. Segue da fórmula de Taylor de ordem dois que

$$\lim_{n\to\infty} \frac{\theta(n)}{\frac{t^2}{n}} = 0.$$

Vamos provar que para qualquer t fixado vale que

$$\lim_{n\to\infty}(\,1-\frac{t^2}{2\,n}+\theta(n))^n=e^{-\frac{t^2}{2}}.$$

Ora, sabemos que

$$\lim_{n\to\infty}(\,1-\frac{t^2}{2\,n})^n=e^{-\frac{t^2}{2}}.$$

O resultado segue então de

$$\lim_{n\to\infty}(\,\frac{(\,1-\frac{t^2}{2\,n}+\theta(n))}{(\,1-\frac{t^2}{2\,n})}\,)^n=1,$$

o qual é equivalente a

$$0=\lim_{n\to\infty}\frac{1}{1/n}\log(\,\frac{(\,1-\frac{t^2}{2\,n}+\theta(n))}{(\,1-\frac{t^2}{2\,n})}\,)=\lim_{n\to\infty}\frac{1}{1/n}\log(\,1+\frac{\theta(n)}{(\,1-\frac{t^2}{2\,n})}\,).$$

Como,

$$\lim_{n\to\infty}\frac{\theta(n)}{\frac{t^2}{n}}=0,$$

dado $\epsilon>0$, para n suficientemente grande, temos que

$$\theta(n)<(\frac{1}{n})\;\epsilon\;t^2,$$

ou seja,

$$\frac{1}{(\frac{1}{n})}<\frac{\epsilon\;t^2}{\theta(n)}.$$

Ora, como para t fixo, $1-\frac{t^2}{2\,n}\to 1$ quando $n\to\infty$, então

$$0\leq\liminf_{n\to\infty}\frac{1}{1/n}\log(\,1+\frac{\theta(n)}{(\,1-\frac{t^2}{2\,n})}\,)\leq\liminf_{n\to\infty}\log(\,1+\frac{\theta(n)}{(\,1-\frac{t^2}{2\,n})}\,)\frac{1}{\theta(n)}\;\;\epsilon\;t^2=$$

$$=\liminf_{n\to\infty}\log(\,1+\frac{\theta(n)}{(\,1-\frac{t^2}{2\,n})}\,)\frac{1}{(\frac{\theta(n)}{(\,1-\frac{t^2}{2\,n})})}\;\;\epsilon\;t^2.$$

Pela Regra de L'Hopital vale que

$$\lim_{x\to 0} \frac{\log(1+x)}{x} = 1.$$

Sendo assim para todo $\epsilon > 0$ vale que

$$0 \leq \liminf_{n\to\infty} \frac{1}{\frac{1}{n}} \log(1 + \frac{\theta(n)}{(1-\frac{t^2}{2n})}) \leq \epsilon t^2.$$

Logo,

$$\liminf_{n\to\infty} \frac{1}{\frac{1}{n}} \log(1 + \frac{\theta(n)}{(1-\frac{t^2}{2n})}) = 0.$$

O mesmo procedimento pode ser feito para o $\limsup$. Sendo assim

$$\lim_{n\to\infty} \frac{1}{\frac{1}{n}} \log(1 + \frac{\theta(n)}{(1-\frac{t^2}{2n})}) = 0,$$

e isto demonstra o resultado desejado.

Lembre que assuminos que $E(X_1) = 0$ e $E(X_1^2) = 1$

Vamos agora mostrar que

$$\varphi_{X_1}(s) = 1 + iE(X_1)\, s + i^2 \frac{E(X_1^2)\, s^2}{2} + R(s) = 1 - \frac{s^2}{2} + R(s),$$

onde $\lim_{s\to 0} \frac{R(s)}{s^2} = 0$.

Devemos estimar,

$$\varphi_{X_1}(s) = \int e^{sX_1(\omega)i} dP(\omega) = \int e^{sxi} dF(x).$$

Ora,

$$\int x^2 dF(x) = 1 < \infty,$$

por hipótese.

Dado $\epsilon > 0$, seja $a > 0$ tal que

$$\frac{1}{a^4} = \frac{E(X_1^2)}{a^4} < \epsilon/2.$$

Ora

$$\varphi_{X_1}(s) = \int e^{i\,s\,X_1(\omega)}\,dP(\omega) = \int_{|X_1|<a} e^{i\,s\,X_1(\omega)}\,dP(\omega) + \int_{|X_1|\geq a} e^{i\,s\,X_1(\omega)}\,dP(\omega).$$

Pela fórmula de Taylor, dado $\frac{\epsilon}{2\,a^2}$, seja $\delta > 0$ tal que se $y < \delta$ então $\frac{r(y)}{y^2} < \frac{\epsilon}{2\,a^2}$.

Tomando $\delta_1 = \frac{\delta}{a}$, obtemos que se $0 < s < \delta_1$, então $s\,|X_1(\omega)| < s\,a < \delta$, para ω tal que $|X_1(\omega)| < a$.

Isto dá conta do termo $\int_{|X_1|<a} e^{i\,s\,X_1(\omega)}\,dP(\omega) < \epsilon/2$, em função da expressão acima (formula de Taylor).

De fato, para cada ω fixo em $|X_1(\omega)| < a$, aplicamos a formula de Taylor para $y = sX_1(\omega)$ e obtemos

$$\frac{r(s\,X_1(\omega))}{(a\,s)^2} \leq \frac{r(s\,X_1(\omega))}{(|X_1(\omega)|\,s)^2} \leq \frac{\epsilon}{2\,a^2}.$$

Logo,

$$\frac{R(s)}{s^2} = \frac{r(s\,X_1(\omega))}{s^2} \leq \frac{\epsilon}{2}.$$

O resultado segue por integração em P sobre o conjunto $\{|X_1(\omega)| < a\}$.

Finalmente, pela desigualdade de Chebyschev,

$$|\int_{X_1\geq a} e^{i\,s\,X_1(\omega)}\,dP(\omega)| < \int_{|X_1|\geq a} |e^{i\,s\,X_1(\omega)}|\,dP(\omega) = P(|X_1| \geq a) =$$

$$P(X_1^2 > a^2) < \frac{1}{a^4} < \epsilon/2.$$

□

3.4 Funções Geradoras de Probabilidade e Funções Características

Iniciaremos esta seção pelas funções geradoras de probabilidade válidas apenas para as variáveis aleatórias discretas. Depois analisaremos as funções geradoras de momentos e características.

Recomendamos o leitor a [GS] e [HPS] para uma descrição mais detalhada do assunto.

Definição 3.9 (**Função Geradora de Probabilidade**). *Seja X variável aleatória discreta com função massa de probabilidade $f_X(\cdot)$. Definimos a função real $\Phi_X(\cdot)$ dada por*

$$E(t^X) \;=\; \Phi_X(t) = \sum_{k\geq 0} \mathbf{P}(X=k)t^k = \sum_{k\geq 0} f_X(k)t^k,$$

para todo $t \in [-1,1]$, como a **função geradora de probabilidades de** X *se $\Phi_X(\cdot)$ for convergente para todo $|t| \leq 1$.*

Exemplo 3.7. Considere a variável aleatória X que tem distribuição Poisson de parâmetro λ. Determinaremos $\Phi_X(\cdot)$.

Observe que

$$\begin{aligned} \Phi_X(t) &= \sum_{k\geq 0} f_X(k)t^k = \sum_{k\geq 0} \frac{e^{-\lambda}\lambda^k}{k!} t^k \\ &= e^{-\lambda} \sum_{k\geq 0} \frac{(\lambda\, t)^k}{k!} = e^{-\lambda} e^{\lambda\, t} = e^{-\lambda(1-t)}, \end{aligned} \tag{3.1}$$

para todo $|t| \leq 1$.

$\diamondsuit$

Exemplo 3.8. Considere a variável aleatória X, tal que para $0 < p < 1$, vale $f_X(k) = pq^k$, onde $q = 1-p$. Determinaremos $\Phi_X(\cdot)$.

Observe que

$$\begin{aligned} \Phi_X(t) &= \sum_{k\geq 0} f_X(k)t^k = \sum_{k\geq 0} pq^k t^k = p \sum_{k\geq 0} (qt)^k \\ &= p\,\frac{1}{1-qt}, \end{aligned} \tag{3.2}$$

para todo $|t| \leq 1$.

$\diamondsuit$

Observação: $\Phi_X(1) = 1$ e $\Phi_X(\cdot)$ é absolutamente e uniformemente convergente em $|t| \leq 1$.

Através da função geradora de probabilidades $\Phi_X(\cdot)$ podemos determinar a esperança e a variância de X. De fato:

(i) $\Phi'_X(t) = \sum_{k\geq 1} k f_X(k) t^{k-1}$ e $\Phi'_X(1) = \mathbb{E}(X)$.

(ii) $\Phi''_X(t) = \sum_{k\geq 2} k(k-1) f_X(k) t^{k-2} = \sum_{k\geq 2} k^2 f_X(k) t^{k-2} - \sum_{k\geq 2} k f_X(k) t^{k-2}$,

e então $\Phi''_X(1) = \mathbb{E}[X(X-1)]$.

Portanto,

$$Var(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \Phi''_X(1) + \Phi'_X(1) - [\Phi'_X(1)]^2. \tag{3.3}$$

Exemplo 3.9. Por exemplo, $\Phi'''_X(1) = \mathbb{E}[X(X-1)(X-2)]$ já que

$$\begin{aligned} \Phi'''_X(t)|_{t=1} &= \sum_{k\geq 3} k(k-1)(k-2) f_X(k) t^{k-3}|_{t=1} \\ &= \sum_{k\geq 0} (k^3 - 3k^2 + 2k) f_X(k) = \sum_{k\geq 3} k^3 f_X(k) - 3\sum_{k\geq 3} k^2 f_X(k) \\ &+ 2\sum_{k\geq 3} k\, f_X(k). \end{aligned} \tag{3.4}$$

Portanto, $\mathbb{E}X^3 = \Phi'''_X(1) + 3\Phi''_X(1) - 2\Phi'_X(1)$.

Ainda,

$$\mathbb{E}X^r = \Phi^{(r)}_X(1) + \begin{pmatrix} r \\ 1 \end{pmatrix} \Phi^{(r-1)}_X(1) - \begin{pmatrix} r \\ 2 \end{pmatrix} \Phi^{(r-2)}_X(1) + \cdots + (r-1)\Phi'_X(1).$$

$\diamondsuit$

A partir da função geradora de probabilidades conseguimos obter a função massa de probabilidade de uma variável aleatória X. Observe que se X tem distribuição Poisson de parâmetro λ, então

$$\begin{aligned} e^{-\lambda(1-t)} &= \sum_{k\geq 0} f_X(k)t^k \\ e^{-\lambda(1-t)} &= e^{-\lambda}e^{\lambda t} = e^{-\lambda}\sum_{k\geq 0}\frac{(\lambda\, t)^k}{k!} = e^{-\lambda}\sum_{k\geq 0}\frac{\lambda^k}{k!}t^k. \qquad (3.5) \end{aligned}$$

Comparando as duas séries de potências na expressão (3.5) obtemos a função $f_X(\cdot)$, ou seja, conseguimos calcular $f_X(k)$, para todo k.

Exemplo 3.10. Voltando ao exemplo 3.7 desta seção, vamos determinar $\mathbb{E}(X)$ e $Var(X)$ à partir da função $\Phi_X(\cdot)$. Como $\Phi'_X(t) = e^{-\lambda(1-t)}\lambda$, temos que $\Phi'_X(1) = \mathbb{E}(X) = \lambda$.

Como $\Phi''_X(t) = \lambda\, e^{-\lambda(1-t)}\lambda$, temos que $\Phi''_X(1) = \lambda^2$. Então, pela expressão (3.3) obtemos $Var(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$.

$\diamondsuit$

Exemplo 3.11. Voltando ao exemplo 3.8, determinaremos $\mathbb{E}(X)$ e $Var(X)$ à partir da função $\Phi_X(\cdot)$. Como $\Phi'_X(t) = -\frac{p}{(1-qt)^2}(-q) = \frac{pq}{(1-qt)^2}$, temos que $\Phi'_X(1) = \mathbb{E}(X) = \frac{q}{p}$. Como $\Phi''_X(t) = -\frac{pq}{(1-qt)^3}(-2q)$, temos que $\Phi''_X(1) = 2\frac{pq^2}{p^3} = 2\frac{q^2}{p^2}$. Portanto, pela expressão (3.3) obtemos $Var(X) = 2\frac{q^2}{p^2} + \frac{q}{p} - \frac{q^2}{p^2} = \frac{q}{p^2}$.

$\diamondsuit$

Definição 3.10. *Seja X variável aleatória qualquer definida no espaço de probabilidade $(\Omega, \mathcal{A}, \mathbf{P})$. A função real $M_X(\cdot)$ dada por*

$$M_X(t) = \mathbb{E}(e^{tX}) = \int e^{tx} dF_X(x) =$$

$$\sum_{n=0}^{\infty}[\,\frac{\int x^n\, d\, F_X(x)}{n\,!}\,]\, t^n = \sum_{n=0}^{\infty}[\,\frac{\int X^n\, d\, P}{n\,!}\,]\, t^n, \quad (*)$$

é dita ser a função geradora de momentos de X se a esperança acima existe em alguma vizinhança de zero.

Note que esta função captura, via os coeficientes da serie de potencias, a informação de todos os momentos $\int X^n \, dP$, $n \in \mathbb{N}$.

Algumas vezes

$$s \to E(s^X) \quad (**)$$

é também denominada **função geradora de momentos**.

Exemplo 3.12. Mostraremos que a variável aleatória X, cuja função massa de probabilidade é dada por $f_X(x) = 6(\pi^2 x^2)^{-1}$, para todo $x \in \mathbb{N}$, não possui função geradora de momentos. Observe que

$$\sum_{x\geq 1} f_X(x) = \frac{6}{\pi^2} \sum_{x\geq 1} \frac{1}{x^2} = \frac{6}{\pi^2}\frac{\pi^2}{6} = 1.$$

No entanto,

$$M_X(t) = \mathbb{E}(e^{tX}) = \sum_{x\geq 1} e^{tx} f_X(x) = \frac{6}{\pi^2} \sum_{x\geq 1} e^{tx} \frac{1}{x^2} = \infty,$$

sempre que $t > 0$. Portanto, não existe função geradora de momentos para X.

$\diamondsuit$

Exemplo 3.13. Considere a variável aleatória X (com valores sobre $\mathbb{N}$) que tem distribuição exponencial de parâmetro λ. Determinaremos $M_X(\cdot)$.

Observe que

$$\begin{aligned} M_X(t) &= \mathbb{E}(e^{tX}) = \sum_{x\geq 0} e^{tx} f_X(x) \\ &= e^{-\lambda} \sum_{x\geq 0} \frac{(e^t \lambda)^x}{x!} = e^{-\lambda} e^{\lambda e^t} = e^{-\lambda(1-e^t)}, \end{aligned} \tag{3.6}$$

para todo $t \in \mathbb{R}$.

$\diamondsuit$

Exemplo 3.14. Considere a variável aleatória $X \sim \mathcal{E}(\lambda)$, $\lambda > 0$ (exponencial com valores sobre $\mathbb{R}^+$). Determinaremos $M_X(\cdot)$. Observe que

$$\begin{aligned} M_X(t) &= \mathbb{E}(e^{tX}) = \int_0^\infty e^{tx} f_X(x)dx \\ &= \lambda \int_0^\infty e^{-(\lambda-t)x}dx = \lambda \left[-\frac{1}{\lambda - t} e^{-(\lambda-t)x} \right]_0^\infty = \frac{\lambda}{\lambda - t}, \end{aligned} \tag{3.7}$$

para todo $t < \lambda$.

$\diamondsuit$

Exemplo 3.15. Considere a variável aleatória $X \sim \Gamma(\alpha, \beta)$, tomando valores reais positivos, onde $\alpha, \beta > 0$. Isto significa que $f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$, onde Γ denota a função gamma [Ru]. Determinaremos $M_X(\cdot)$.

Observe que

$$\begin{aligned} M_X(t) &= \mathbb{E}(e^{tX}) = \int_0^\infty e^{tx} f_X(x)dx = \int_0^\infty e^{tx} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\beta-t)x}\, dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha)}{(\beta-t)^\alpha} \\ &= \left(\frac{\beta}{\beta - t} \right)^\alpha, \end{aligned} \tag{3.8}$$

para todo $t < \beta$.

$\diamondsuit$

Observações:

1. A função geradora de momentos determina unicamente a função de distribuição de X.

2. Se $M_X(\cdot)$ existe, então ela é única.

3. A função geradora de momentos $M_X(\cdot)$ nada mais é do que a Transformada de Laplace.

4. Se a função geradora de momentos $M_X(\cdot)$ de uma variável aleatória X existe para todo $t \in (-t_0, t_0)$, com $t_0 > 0$, então existem as derivadas de $M_X(\cdot)$ de todas as ordens em $t = 0$. Neste caso,

$$M_X^{(k)}(t)|_{t=0} = \mathbb{E}(X^k),$$

para todo $k \in \mathbb{N}$.

Exemplo 3.16. Voltando ao último exemplo, vamos determinar a esperança e a variância de X, utilizando a função $M_X(\cdot)$. Observe que

$$\begin{aligned}
\mathbb{E}(X) &= M_X'(t)|_{t=0} = \left[\left(\frac{\beta}{\beta - t}\right)^{\alpha}\right]'\Bigg|_{t=0} = \alpha \left(\frac{\beta}{\beta - t}\right)^{\alpha - 1} \frac{\beta}{(\beta - t)^2}\Bigg|_{t=0} \\
&= \alpha \left(\frac{\beta}{\beta}\right)^{\alpha-1} \frac{\beta}{\beta^2} = \frac{\alpha}{\beta} \qquad (3.9)
\end{aligned}$$

e ainda

$$\begin{aligned}
\mathbb{E}(X^2) &= M_X''(t)|_{t=0} = \Bigg[\alpha(\alpha - 1)\left(\frac{\beta}{\beta - t}\right)^{\alpha-2} \frac{\beta^2}{(\beta - t)^4} + \alpha \left(\frac{\beta}{\beta - t}\right)^{\alpha-1} \times \\
&\times \quad \frac{-\beta\, 2\,(-1)}{(\beta - t)^3}\Bigg]\Bigg|_{t=0} = \frac{\alpha(\alpha-1)}{\beta^2} + \frac{2\,\alpha}{\beta^2} = \frac{\alpha^2 + \alpha}{\beta^2}. \qquad (3.10)
\end{aligned}$$

Portanto,

$$Var(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \frac{\alpha^2 + \alpha}{\beta^2} - \frac{\alpha^2}{\beta^2} = \frac{\alpha}{\beta^2}.$$

$\diamondsuit$

Teorema 3.16. *Seja $(X_n)_{n \geq 1}$ sequência de variáveis aleatórias independentes e suponha que a função geradora de momentos de X_n existe, para cada $n \in \mathbb{N}$. Então, a função geradora de momentos de $S_n = \sum_{i=1}^n X_i$ existe e é dada por*

$$M_{S_n}(t)=\prod_{i=1}^{n} M_{X_i}(t). \tag{3.11}$$

Demonstração: Observe que

$$\begin{aligned} M_{S_n}(t) &= \mathbb{E}(e^{tS_n})=\mathbb{E}\left(e^{t\sum_{i=1}^{n}X_i}\right)=\mathbb{E}\left(\prod_{i=1}^{n}e^{tX_i}\right)\\ &= \prod_{i=1}^{n}\mathbb{E}(e^{t\,X_i})=\prod_{i=1}^{n}M_{X_i}(t), \end{aligned} \tag{3.12}$$

onde a quarta igualdade acima é válida pela hipótese de independência das variáveis aleatórias.

□

Exemplo 3.17. Considere $(X_m)_{m\geq 1}$ sequência de variáveis aleatórias independentes e identicamente distribuídas tais que $X_1\sim\mathcal{B}(n,p)$ (binomial). Consideremos $S_m=\sum_{i=1}^{m}X_i$ e determinemos $M_{S_m}(\cdot)$.

Observe que

$$\begin{aligned} M_{X_1}(t)=\mathbb{E}(e^{t\,X_1})&=\sum_{x=0}^{n}e^{t\,x}\binom{n}{x}p^x\,q^{n-x}\\ &=\sum_{x=0}^{n}\binom{n}{x}(e^t\,p)^x q^{n-x}=\left(e^t p+q\right)^n, \end{aligned}$$

para todo $t\in\mathbb{R}$. Então, $M_{S_m}(t)=(e^tp+q)^{mn}$, para todo $t\in\mathbb{R}$, pela expressão (3.11).

♢

Desejamos agora definir a função geradora de momentos para um vetor aleatório (X,Y) em $\mathbb{R}^2$, ou seja

$$(X,Y):\Omega\to\mathbb{R}^2.$$

Nosso interesse é na probabilidade conjunta, ou seja, dado um conjunto $C \subset \mathbb{R}^2$, qual o valor

$$P(\, \{w \,|\, (X,Y)(w) \,\in\, C\}\,)\,?$$

Dizemos, quando existir, que $d\,F_{X,Y}$ é a densidade conjunta se para qualquer conjunto da forma $C = (a,b) \times (c,d) \subset \mathbb{R}^2$, vale

$$P(\, \{w \,|\, (X,Y)(w) \,\in\, C\}\,) \,=\, \int_a^b (\int_c^d d\,F_{X,Y}\,(x,y)\,dy)\; dx\,.$$

Definição 3.11. *Seja (X,Y) um vetor aleatório em $\mathbb{R}^2$. Se $\mathbb{E}(e^{t_1 X + t_2 Y})$ existe para todo $|t_1| \leq h_1$ e $|t_2| \leq h_2$, onde $h_1, h_2 \in \mathbb{R}\,,\; h_1, h_2 > 0$, então a função geradora de momentos de (X,Y) é definida por*

$$M_{X,Y}(t_1,t_2) = \mathbb{E}(e^{t_1 X + t_2 Y}) = \int e^{t_1 x + t_2 y}\, dF_{X,Y}(x,y). \tag{3.13}$$

Observações:

1. A função geradora de momentos $M_{X,Y}(\cdot,\cdot)$ determina unicamente a função de distribuição conjunta de (X,Y).

2. Se $M_{X,Y}(\cdot,\cdot)$ existe, então ela é única.

3. Observe que

$$\begin{aligned} M_{X,Y}(t_1,0) &= \mathbb{E}(e^{t_1\, X}) = M_X(t_1) \\ M_{X,Y}(0,t_2) &= \mathbb{E}(e^{t_2\, Y}) = M_Y(t_2) \end{aligned} \tag{3.14}$$

Portanto, $M_{X,Y}(\cdot,\cdot)$ determina completamente as distribuições marginais de X e de Y.

4. Se $M_{X,Y}(\cdot,\cdot)$ existe, então também existem os momentos de (X,Y) de todas as ordens, e eles podem ser obtidos através da igualdade,

$$\left.\frac{\partial^{m+n} M_{X,Y}(t_1,t_2)}{\partial t_1^m\,\partial t_2^n}\right|_{t_1=0=t_2} = \mathbb{E}(X^m\,Y^n), \tag{3.15}$$

para quaisquer $m, n \in \mathbb{N}$.

Observe ainda que

$$\left.\frac{\partial M_{X,Y}(t_1,0)}{\partial t_1}\right|_{t_1=0} = \mathbb{E}(X); \qquad \left.\frac{\partial M_{X,Y}(0,t_2)}{\partial t_2}\right|_{t_2=0} = \mathbb{E}(Y);$$

$$\left.\frac{\partial^2 M_{X,Y}(t_1,0)}{\partial t_1^2}\right|_{t_1=0} = \mathbb{E}(X^2); \qquad \left.\frac{\partial^2 M_{X,Y}(0,t_2)}{\partial t_2^2}\right|_{t_2=0} = \mathbb{E}(Y^2);$$

$$\left.\frac{\partial^2 M_{X,Y}(t_1,t_2)}{\partial t_1\,\partial t_2}\right|_{t_1=0=t_2} = \mathbb{E}(XY). \tag{3.16}$$

Teorema 3.17. *Sejam X e Y variáveis aleatórias definidas no espaço de probabilidade $(\Omega, \mathcal{A}, \mathbf{P})$. Então, X e Y são independentes se e somente se*

$$M_{X,Y}(t_1,t_2) = M_{X,Y}(t_1,0)\cdot M_{X,Y}(0,t_2),$$

para todo $t_1, t_2 \in \mathbb{R}$.

Demonstração: $(\Longrightarrow)$

Observe que

$$\begin{aligned} M_{X,Y}(t_1,t_2) &= \mathbb{E}(e^{t_1\,X+t_2\,Y}) = \mathbb{E}(e^{t_1\,X})\cdot\mathbb{E}(e^{t_2\,Y}) \\ &= M_X(t_1)\cdot M_Y(t_2) = M_{X,Y}(t_1,0)\cdot M_{X,Y}(0,t_2). \end{aligned} \tag{3.17}$$

$(\Longleftarrow)$

A prova será feita para o caso em que as variáveis aleatórias X e Y são contínuas. Se $M_{X,Y}(t_1,t_2) = M_{X,Y}(t_1,0)\cdot M_{X,Y}(0,t_2)$ então temos que

$$\begin{aligned} \int\!\!\int e^{t_1\,x+t_2\,y} f_{X,Y}(x,y)\,dx\,dy &= \left(\int e^{t_1\,x} f_X(x)\,dx\right)\left(\int e^{t_2\,y} f_Y(y)\,dy\right) \\ &= \int\!\!\int e^{t_1\,x+t_2\,y} f_X(x)\,f_Y(y)\,dx\,dy. \end{aligned} \tag{3.18}$$

Portanto, $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$, para todo $(x,y) \in \mathbb{R}^2$, e concluímos que X e Y são variáveis aleatórias independentes.

$\square$

Exemplo 3.18. Considere (X,Y) vetor aleatório contínuo com função densidade conjunta dada por

$$f_{X,Y}(x,y) = \begin{cases} e^{-(x+y)}, & \text{se } x > 0 \text{ e } y > 0 \\ 0, & \text{c.c .} \end{cases}$$

(a) Determine $M_{X,Y}(\cdot,\cdot)$. (b) Calcule $\mathbb{E}(X)$, $\mathbb{E}(Y)$, $Var(X)$, $Var(Y)$ e $\mathbb{E}(XY)$.

Para resolver o item (a), observe que

$$\begin{aligned} M_{X,Y}(t_1,t_2) = \mathbb{E}(e^{t_1 X + t_2 Y}) &= \int_0^\infty \int_0^\infty e^{t_1 x + t_2 y} e^{-x-y} dx\, dy \\ &= \int_0^\infty e^{(t_1-1)x} dx \int_0^\infty e^{(t_2-1)y} dy = \left(\int_0^\infty e^{-(1-t_1)x} dx \right) \cdot \left(\int_0^\infty e^{-(1-t_2)y} dy \right) \\ &= \left(\frac{1}{1-t_1} \right) \left(\frac{1}{1-t_2} \right), \end{aligned} \tag{3.19}$$

para todo $t_1 < 1$ e $t_2 < 1$. Para resolver o item (b), considere as igualdades (3.16), obtendo

$$\begin{aligned} \mathbb{E}(X) &= \left. \frac{\partial M_{X,Y}(t_1,0)}{\partial t_1} \right|_{t_1=0} = \left. \frac{(-1)(-1)}{(1-t_1)^2} \left(\frac{1}{1-t_2} \right) \right|_{(t_1,t_2)=(0,0)} = 1 = \mathbb{E}(Y) \\ \mathbb{E}(X^2) &= \left. \frac{\partial^2 M_{X,Y}(t_1,0)}{\partial t_1^2} \right|_{t_1=0=t_2} = \left. \frac{(-1)(-2)}{(1-t_1)^3} \left(\frac{1}{1-t_2} \right) \right|_{(t_1,t_2)=(0,0)} = 2 = \mathbb{E}(Y^2) \\ \mathbb{E}(XY) &= \left. \frac{\partial^2 M_{X,Y}(t_1,t_2)}{\partial t_1 \partial t_2} \right|_{(t_1,t_2)=(0,0)} = \left. \frac{1}{(1-t_1)^2} \frac{1}{(1-t_2)^2} \right|_{(t_1,t_2)=(0,0)} = 1. \end{aligned}$$

Concluímos que $Var(X) = 1 = Var(Y)$ e que $cov(X,Y) = 0$. De fato, as variáveis aleatórias são independentes já que

$$f_{X,Y}(x,y) = e^{-x-y} = e^{-x} \cdot e^{-y} = f_X(x) \cdot f_Y(y),$$

para todo $(x, y) \in \mathbb{R}^2$.

$\diamondsuit$

A função característica é definida para qualquer variável aleatória X e ela existe sempre.

Definição 3.12. *Seja X variável aleatória qualquer definida no espaço de probabilidade $(\Omega, \mathcal{A}, \mathbf{P})$. A função $\varphi_X(\cdot)$ (tomando valores complexos) dada por*

$$\varphi_X(t) = \mathbb{E}(e^{itX}) = \int e^{itx}\, dF_X(x) = \int \sum_{n=0}^{\infty} \frac{(i\,t\,x)^n}{n!}\, dF_X(x),$$

para todo $t \in \mathbb{R}$, é dita ser a função característica de X, onde $i = \sqrt{-1}$.

Exemplo 3.19. Considere $Z \sim \mathcal{N}(0,1)$ (normal), ou seja, Z possui distribuição normal com média 0 e variância 1. Determinaremos $\varphi_Z(\cdot)$. Observe que

$$\begin{aligned}
\varphi_Z(t) &= \mathbb{E}(e^{itZ}) = \int_{-\infty}^{\infty} e^{itz} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2 - 2itz)} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} e^{-\frac{1}{2}(z-it)^2} dz \\
&= e^{-\frac{t^2}{2}}, \qquad (3.20)
\end{aligned}$$

para todo $t \in \mathbb{R}$.

$\diamondsuit$

Observações:

1. A função característica $\varphi_X(\cdot)$ determina unicamente a função de distribuição de X através das fórmulas de inversão (transformada de Fourier inversa).

2. A função característica $\varphi_X(\cdot)$ nada mais é do que a Transformada de Fourier da função $f_X(\cdot)$. Portanto, a função $\varphi_X(\cdot)$ sempre existe.

3. A função característica $\varphi_X(\cdot)$ é limitada por 1. De fato:

$$\begin{aligned} |\varphi_X(t)| &= |\mathbb{E}(e^{itX})| \leq \mathbb{E}|e^{itX}| = \mathbb{E}\left(\sqrt{\cos^2(tX) + \text{sen}^2(tX)}\right) \\ &= \mathbb{E}(1) = 1, \end{aligned} \tag{3.21}$$

para todo $t \in \mathbb{R}$, onde a desigualdade acima é devido à de Jensen.

A desigualdade de Jensen afirma: se g é convexa, $g : \mathbb{R}^+ \to \mathbb{R}$, então $= \int g(X)dP = E(g(X)) \geq g(E(X)) = g(\int XdP)$, onde $X : \mathbb{R} \to \mathbb{R}$ é uma função P integrável (ver [Fe] para prova).

4. $\varphi_X(0) = 1$.

5. $\overline{\varphi_X(t)} = \varphi_X(-t)$.

6. φ_X é uniformemente contínua na reta.

7. A v.a. X tem distribuição simétrica em torno de zero se e só se $\varphi_X(t)$ é real para todo t.

Teorema 3.18. *Sejam X e Y v.a.'s independentes. Então,*

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t),$$

para todo $t \in \mathbb{R}$. Ainda, se $X_1, X_2, ..., X_n$ são independentes, então,

$$\varphi_{X_1+X_2+...+X_n}(t) = \varphi_{X_1}(t)\,\varphi_{X_2}(t)...\varphi_{X_n}(t).$$

Demonstração: Observe que

$$\varphi_{X+Y}(t) = \mathbb{E}(e^{it(X+Y)}) = \mathbb{E}(e^{itX}e^{itY}) = \mathbb{E}(e^{itX})\,\mathbb{E}(e^{itY}) = \varphi_X(t) \cdot \varphi_Y(t),$$

onde a terceira igualdade é devido à independência de X e Y.

Por indução, $\varphi_{S_n}(t) = \prod_{i=1}^{n} \varphi_{X_i}(t)$ sempre que X_i são v.a.'s independentes, onde

$$Sn = \sum_{i=1}^{n} X_i.$$

□

Teorema 3.19. *Seja X v.a. qualquer. Considere $Y = aX + b$, para a, $b \in \mathbb{R}$ e $a \neq 0$. Então, $\varphi_Y(t) = e^{itb}\varphi_X(at)$, para todo $t \in \mathbb{R}$.*

Demonstração: Observe que

$$\varphi_Y(t) = \mathbb{E}(e^{itY}) = \mathbb{E}(e^{it(aX+b)}) = \mathbb{E}(e^{itb}e^{iatX}) = e^{itb}\varphi_X(at), \forall\ t \in \mathbb{R}.$$

□

Exemplo 3.20. Considere $X \sim \mathcal{N}(\mu, \sigma^2)$ (a distribuição normal ou Gaussiana). Determinaremos $\varphi_X(\cdot)$.

Observe que

$$X = \mu + \sigma Z \quad e \quad \varphi_Z(t) = e^{-\frac{t^2}{2}}, \text{ onde } Z \sim N(0,1).$$

Usando resultados anteriores obtemos

$$\varphi_X(t) = e^{it\mu}\varphi_Z(\sigma t) = e^{it\mu}e^{-\frac{\sigma^2 t^2}{2}}.$$

♢

Observação: Se X é v.a. qualquer tal que $M_X(\cdot)$ exista, então

$$\varphi_X(t) = \mathbb{E}(e^{itX}) = M_X(it), \text{ para todo } t \in \mathbb{R}.$$

Exemplo 3.21. Seja $X \sim \epsilon(\lambda)$ (exponencial). Determine $\varphi_X(\cdot)$.
Observe que $M_X(t) = \frac{\lambda}{\lambda - t}$, para todo $t < \lambda$. Então,

$$\varphi_X(t) = M_X(it) = \frac{\lambda}{\lambda - it}, \text{ para todo } t \in \mathbb{R}.$$

$\diamondsuit$

Teorema 3.20. *Seja X v.a. qualquer tal que $\mathbb{E}|X|^n <\propto$. Então, $\varphi_X(\cdot)$ possui n derivadas contínuas e ${\varphi_X}^{(k)}(t) = \int (ix)^k e^{itx} dF_X(x)$, para todo $t \in \mathbb{R}$ e $1 \leq k \leq n$. Em particular, ${\varphi_X}^{(k)}(0) = i^k\, \mathbb{E}(X^k)$, onde ${\varphi_X}^{(k)}(0) = \frac{d^k \varphi_X(t)}{dt^k}|_{t=0}$.*

Demonstração: Observe que

$$
{\varphi_X}^{(k)}(t) = \frac{d^k}{dt^k} \int e^{itx} dF_X(x) = \int \frac{d^k}{dt^k} e^{itx} dF_X(x) =
$$

$$
= \int (ix)^k e^{itx} dF_X(x) = i^k \int x^k e^{itx} dF_X(x), \text{ para todo } t \in \mathbb{R}.
$$

Então,

$$
{\varphi_X}^{(k)}(0) = \frac{d^k \varphi_X(t)}{dt^k}|_{t=0} = i^k \int x^k dF_X(x) = i^k\, \mathbb{E}\,(X^k).
$$

$\square$

Exemplo 3.22. Considere $X \sim P(\lambda_1)$, $\lambda_1 > 0$, ou seja, X tem distribuição de Poisson.

(a) Calcularemos $\varphi_X(\cdot)$.
(b) Obteremos $\mathbb{E}(X)$ e $\mathbb{E}(X^2)$.
(c) Considere $Y \sim P(\lambda_2)$ independente de X. Mostraremos que $X + Y \sim P(\lambda_1 + \lambda_2)$ utilizando $\varphi_{X+Y}(\cdot)$.

Vamos mostrar a validade das afirmações acima.

(a) Observe que $\varphi_X(t) = M_X(it) = e^{\lambda(e^{it}-1)}$, para todo $t \in \mathbb{R}$.
(b) Como $\mathbb{E}(X) = \frac{1}{i}\varphi_X^{'}(0)$ e como $\varphi'_X(t) = e^{\lambda(e^{it}-1)}.\, \lambda e^{it} i$,

temos que

$$
\mathbb{E}(X) = \frac{1}{i}\, \varphi'_X(0) = \frac{1}{i}\, \lambda i\, (e^{\lambda(e^{it}-1)} e^{it})|_{t=0} = \lambda.
$$

Como $\mathbb{E}(X^2) = \frac{1}{i^2}\,\varphi''_X(0)$ e como $\varphi''_X(t) = \lambda\, i\, e^{it}\, i\, e^{\lambda(e^{it-1})} + \lambda\, i\, e^{it} e^{\lambda(e^{it-1})} \lambda\, i\, e^{it} =$

$$= -\lambda\, e^{it} e^{\lambda(e^{it-1})} - \lambda^2\, e^{2it} e^{\lambda(e^{it-1})},$$

temos que

$$\mathbb{E}(X^2) = (\lambda\, e^{it}\, e^{\lambda(e^{it}-1)} + \lambda^2\, e^{2it}\, e^{\lambda(e^{it}-1)})\big|_{t=0} = \lambda + \lambda^2.$$

(c) Observe que

$$\varphi_{X+Y}(t) = \varphi_X(t)\,\varphi_Y(t) = e^{\lambda_1(e^{it}-1)}\, e^{\lambda_2(e^{it}-1)} = e^{(\lambda_1+\lambda_2)(e^{it}-1)},$$

para todo $t \in \mathbb{R}$. Pelo Teorema da Unicidade, temos que $X+Y \sim P(\lambda_1+\lambda_2)$.

$\diamondsuit$

Definição 3.13. *Seja $X = (X_1, X_2, ..., X_k)$ um vetor aleatório $k-$ dimensional, ou seja, $X : \Omega \to \mathbb{R}^k$, e $X(w) = (X_1(w), X_2(w), ..., X_k(w))$. A função*

$$\varphi_X : \mathbb{R}^k \longrightarrow \mathbb{C}$$

definida por

$$\varphi_X(t_1, \cdots, t_k) = \mathbb{E}\left(exp\,\left\{ i\sum_{j=1}^{k} t_j\, X_j \right\}\right) = \mathbb{E}\left(e^{i<t,X>}\right),$$

onde $< t, X >= \sum_{j=1}^{k} t_j\, X_j$ representa o produto interno dos vetores X e $t = (t_1, \cdots, t_k)$, é dita ser função característica de X.

Observações:

1. Valem todas as propriedades enunciadas para $\varphi_X(\cdot)$ onde X é uma v.a. unidimensional.

2. A partir da função $\varphi_X(\cdot)$ podemos sempre obter a função característica de qualquer uma das v.a. X_j, $1 \leqslant j \leqslant k$. De fato:

$$\varphi_{X,Y}(t,u) = \mathbb{E}\,(e^{itX + iuY}) = \mathbb{E}\,(e^{itX + iuY + i0Z}) = \varphi_{X,Y,Z}(t,u,0),$$

para todo $(t, u) \in \mathbb{R}^2$.

Exemplo 3.23. Considere o vetor aleatório (X, Y) dado por

$X \backslash Y$	0	1	2	$f_X(x)$
1	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{3}$
2	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{2}{3}$
$f_Y(y)$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{1}{6}$	1

Calcularemos $\varphi_{X,Y}(t_1, t_2)$, para todo $(t_1, t_2) \in \mathbb{R}^2$.
Observe que

$$\varphi_{X,Y}(t_1, t_2) = \mathbb{E}\,(e^{it_1\,X + it_2\,Y}) = \sum_{x=1}^{2}\ \sum_{y=0}^{2}\ e^{\,it_1\,x}\ \ e^{\,it_2\,y}\,p_{X,Y}{}^{\,(x,y)}$$

$$e^{\,it_1}\,e^{\,it_2\cdot 0}\,\frac{1}{6} + e^{\,it_1}\,e^{\,it_2}\,\frac{1}{6} + e^{\,it_1}\,e^{\,it_2\cdot 2}\cdot 0 + e^{\,i2t_1}\,e^{\,it_2 0}\,\frac{1}{6} + e^{\,i2t_1}\,e^{\,it_2}\,\frac{2}{6} +$$

$$+\,e^{\,i2t_1}\,e^{\,i2t_2}\,\frac{1}{6} = e^{\,it_1}\,\frac{1}{6} + e^{\,it_1}\,e^{\,it_2}\,\frac{1}{6} + e^{\,i2t_1}\,\frac{1}{6} + e^{\,i2t_1}\,e^{\,it_2}\,\frac{2}{6} + e^{\,i2t_1}\,e^{\,i2t_2}\,\frac{1}{6}$$

$$= \frac{1}{6}(e^{\,it_1} + e^{\,it_1}\,e^{\,it_2} + e^{\,i2t_1} + e^{\,i2t_1}\,e^{\,it_2}2 + e^{\,i2t_1}\,e^{\,i2t_2})$$

$$= \frac{1}{6}(e^{\,it_1}\,(1 + e^{\,it_2} + e^{\,it_1}) + e^{\,i2t_1}(2\,e^{\,it_2} + e^{\,i2t_2})).$$

Concluímos que

$$\varphi_{X,Y}(t_1, t_2) = \frac{1}{6}\,(e^{it_1}\,(1 + e^{it_2} + e^{it_1}) + e^{i2t_1}(2\,e^{it_2} + e^{i2t_2})), \text{para todo } t_1, t_2 \in \mathbb{R}.$$

$\diamondsuit$

Exemplo 3.24. Calcularemos $\varphi_{X,Y}$, onde X, Y são v.a. i.i.d. $\mathcal{E}(1)$, ou seja, X e Y tem distribuição exponencial de parâmetro 1

Observe que

$$\varphi_{X,Y}(t_1, t_2) = \mathbb{E}(e^{it_1 X}\,e^{it_2 Y}) = \mathbb{E}(e^{it_1 X})\,\mathbb{E}(e^{it_2 Y}) =$$

$$= \int_0^\infty e^{it_1 x}\, e^{-x} dx \int_0^\infty e^{it_2 y}\, e^{-y} dy = \int_0^\infty e^{-(1-it_1)x} dx \int_0^\infty e^{-(1-it_2)y} dy =$$

$$= \frac{1}{1-it_1} \cdot \frac{1}{1-it_2} = \frac{1}{(1-it_1)(1-it_2)},$$

para quaisquer t_1 e $t_2 \in \mathbb{R}$.

Portanto,

$$\varphi_{X,Y}(t_1,t_2) = \frac{1}{(1-it_1)(1-it_2)}, \text{para quaisquer } t_1 \text{ e } t_2 \in \mathbb{R}.$$

$\diamondsuit$

3.5 Exercícios

1. Determine a função geradora de momentos de $X \sim \mathcal{U}([a,b])$ (uniforme). Calcule a esperança e a variância de X através da $M_X(\cdot)$.

2. **a.** $\mathcal{B}(n,p)$ é a distribuição de Bernoulli de parâmetros n, p ; é a probabilidade sobre $\{0, 1, ..., n\}$ tal que a probabilidade do evento k é $\begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k}$. Se $X \sim \mathcal{B}(n,p)$, qual a função característica de X?

   b. Mostre, usando funções características, que se $X \sim \mathcal{B}(m,p)$ e $Y \sim \mathcal{B}(n,p)$ e X e Y são independentes, então $X+Y \sim \mathcal{B}(m+n,p)$.

3. Considere $X \sim \mathcal{N}(\mu, \sigma^2)$.

   a. Calcule $E(X^3)$.

   b. Calcule $E(X^4)$.

Sugestão: Calcule primeiro para $\mathcal{N}(0,1)$ e use linearidade.

4. **a.** Suponha que $X \sim \mathcal{E}(\lambda)$. Mostre que a função característica de X é

$$\varphi_X(t) = \frac{\lambda}{\lambda - it} = \frac{\lambda^2 + it\lambda}{\lambda^2 + t^2}.$$

b. Seja Y variável aleatória com distribuição exponencial dupla cuja densidade é dada por

$$f_Y(y) = \frac{\lambda}{2} e^{-\lambda|y|}, \quad \text{para todo} \quad y \in \mathbb{R}.$$

Calcule a função característica de Y.

c. Demonstre: Se Z e W são independentes e identicamente distribuídas, com $Z \sim \mathcal{E}(\lambda)$, então $Z - W$ é exponencial dupla.

5. Use a função característica do Exercício 4 desta Lista para mostrar que se $X \sim \Gamma(n, \beta)$, $n \in \mathbb{N}$, então

$$\varphi_X(t) = \left(\frac{\beta}{\beta - it}\right)^n.$$

6. Considere a variável aleatória X com distribuição Laplace com parâmetro λ cuja densidade é dada por

$$f_X(x) = \frac{1}{2\lambda} e^{-|x-\mu|/\lambda}, \quad x, \mu \in \mathbb{R} \quad \text{e} \quad \lambda > 0.$$

a. Mostre que a função geradora de momentos existe e coincide com

$$M_X(t) = (1 - \lambda^2 t^2)^{-1} e^{\mu t}, \quad |t| < 1/\lambda.$$

7. Seja (X, Y) vetor aleatório contínuo com função de densidade de probabilidade conjunta dada por

$$f_{X,Y}(x,y) = \begin{cases} x^2 + \dfrac{xy}{3}, & 0 < x < 1 \quad \text{e} \quad 0 < y < 2 \\ 0, & \text{c.c..} \end{cases}$$

a. Encontre a função geradora de momentos de (X, Y).

b. Determine todos os momentos de ordem 2 do vetor (X, Y).

8. Considere X uma variável aleatória positiva com primeiro momento finito. Mostre que

a. $\mathbb{E}(\sqrt{X}) \leq \sqrt{\mathbb{E}(X)}$.

b. $\mathbb{E}\left(\frac{1}{X}\right) \geq \frac{1}{\mathbb{E}(X)}$.

9. Considere o vetor aleatório (X, Y) com função densidade conjunta dada por

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{4}[1 + xy(x^2 - y^2)], & |x| \leq 1 \quad \text{e} \quad |y| \leq 1 \\ 0, & \text{c.c..} \end{cases}$$

a. Encontre a função geradora de momentos de (X, Y).

b. X e Y são independentes?

10. Calcule a função característica da variável aleatória X quando

a. X é geometricamente distribuída com parâmetro p, $0 < p < 1$.

Nota: A série geométrica

$$\sum_{n=0}^{\infty} a^n$$

onde a é uma constante real, converge sempre que $|a| < 1$ e a soma da série é dada por $(1 - a)^{-1}$.

b. $S_n = X_1 + ... + X_n$, onde $X_i \sim \mathcal{E}(\lambda)$, independente, para todo $1 \leq i \leq n$. Sugestão: Use o Exercício 4 desta lista.

11. Suponha que X e Y sejam variáveis aleatórias independentes e identicamente distribuídas. Mostre que $\varphi_{X-Y}(t) = |\varphi_X(t)|^2$.

12. Seja X uma variável aleatória contínua tendo função densidade dada por

$$f_X(x) = \frac{1}{2}e^{-|x|}, x \in \mathbb{R}.$$

a. Mostre que $M_X(t) = (1-t^2)^{-1}, \quad -1 < t < 1$.

b. Use esta função geradora de momentos para encontrar os momentos de ordem r de X.

13. Considere a variável aleatória contínua X do Exercício 12 desta Lista.

a. Mostre que $\varphi_X(t) = (1+t^2)^{-1}$.

b. Mostre que

$$e^{-|x|} = \int_{-\infty}^{\infty} e^{-itx} \frac{1}{\pi(1+t^2)} dt.$$

14. Use o Exercício 13 desta Lista para mostrar que

$$\varphi_X(t) = e^{-|t|}, t \in \mathbb{R},$$

quando X é variável aleatória com distribuição Cauchy com parâmetros $0, 1$.

15. Considere a variável aleatória X com função geradora de momentos $M_X(\cdot)$. Calcule $M_Y(\cdot)$, em termos de $M_X(\cdot)$, onde $Y = a + bX$ com $a, b \in \mathbb{R}$, e $b \neq 0$.

16. Sejam X e Y v.a. independentes tais que $X \sim \mathcal{C}(M_1, b_1)$ e $Y \sim \mathcal{C}(M_2, b_2)$.

a. Calcule a função característica das variáveis $X+Y$ e $\frac{X+Y}{2}$.

b. As variáveis aleatórias $X+Y$ e $\frac{X+Y}{2}$ têm distribuição de Cauchy?

Sugestão: Lembre que se $Z \sim \mathcal{C}(0,1)$ e $X \sim \mathcal{C}(M, b)$, então $X = bZ + M$.

17. Sejam X e Y v.a. independentes cuja função densidade conjunta é dada por

$$f_{X,Y}(x,y) = \frac{1}{8}\,(6 - x - y)\; I_{(0,2)}(x)\; I_{(2,4)}(y)$$

a. Determine a função característica $\varphi_{X,Y}(\cdot\,,\cdot)$ do vetor (X,Y).

4

Cadeias de Markov em Tempo Contínuo

4.1 Introdução e Propriedades Gerais

Vamos analisar agora as Cadeias de Markov em tempo contínuo. Será necessário (em algum momento) para o seu bom entendimento algum breve conhecimento da Teoria das Equações Diferencias. Vamos apresentar uma breve introdução ao tópico na Seção 4.5. O material que lá se encontra é basicamente o que se precisa neste capítulo. Ao leitor interessado em mais detalhes sobre este tópico recomendamos [DL] Capítulo 2.

O conjunto $\mathbb{R}^+$ vai denotar a seguir o conjunto dos números reais t tal que $t \geq 0$.

Seja $(\hat{\Omega}, \hat{\mathcal{A}}, \hat{P})$ um espaço de probabilidade, $(S, \mathcal{G})$ um espaço mensurável e ainda uma família de variáveis aleatórias X_t indexadas por um parâmetro $t \in \mathbb{R}^+$, e onde cada $X_t : (\hat{\Omega}, \hat{\mathcal{A}}, \hat{P}) \to (S, \mathcal{G})$ é mensurável, $\hat{\mathcal{A}} \subset \mathfrak{p}(\hat{\Omega})$ é uma sigma-algebra. Dizemos que tal família $(X_t)_{t \in \mathbb{R}^+}$ é um processo estocástico com espaço de parâmetros temporais $t \in \mathbb{R}^+$. Vamos assumir aqui que S é finito ou enumerável e assim podemos supor, sem perda de generalidade que $S \subset \mathbb{Z}$. A sigma-álgebra sobre S será $\mathcal{G} = \mathfrak{p}(S)$, ou seja, o conjunto das partes de S.

Como exemplo, o leitor pode ter em mente o seguinte caso: uma central telefônica recebe telefonemas durante o dia. Digamos que X_t, denote o número de chamadas recebidas até o tempo t. A medida que o tempo passa este número pode ficar igual ou aumentar. A qualquer momento $t+s$, depois de t, poderá ocorrer uma nova chamada. Em princípio, não parece natural indexar o parâmetro tempo pelo conjunto dos naturais.

O espaço S de estados seria o conjunto dos números $s = 0, 1, 2, 3, 4, ..., n, ...$. Assumimos que $X_0 = 0$ com probabilidade 1.

Esta central poderia estar localizada em uma cidade com mais frequencia de telefonemas ou em uma com menos. Os modelos teriam que ser diferentes para levar em conta este fato. Suponha, por exemplo, que em certo momento t fixado temos que $X_t = 142$. Seja s fixo, é claro que no primeiro caso teremos maior probabilidade de se ter um número mais elevado de telefonemas no tempo $t+s$, do que no segundo. Ou seja, X_{t+s} deveria ser maior no primeiro caso (em termos probabilísticos) do que no outro. A primeira cidade tem maior frequencia de telefonemas. Deveria existir menor intensidade de telefonemas na segunda. Seria natural supor a existência de um parâmero λ que determinasse tal intensidade. Neste caso, na primeira cidade λ seria maior.

Pode-se mostrar que neste modelo teríamos que para t fixo, e s fixo, $s \in \{0, 1, 2, 3, 4, ..., n, ...\}$,

$$P(X_t = s) = \frac{(\lambda\, t)^s}{s\,!}\, e^{-\lambda\, t}.$$

Note que para s fixo, a medida que t cresce, o valor $P(X_t = s)$ decresce. Isto traduz o fato que a medida que o tempo passa, o número de telefonemas recebidos vai aumentando até que com grande probabilidade vai ficar maior que s.

Observe que se λ é grande então a intensidade deste decrescimento (da probabilidade) com t aumenta dramaticamente.

Este processo é conhecido pelo nome de Processo de Poisson com parâmetro λ. A razão de se ter os valores das probabilidades das X_t dadas desta forma será explicado posteriormente.

Alertamos o leitor que este é um texto introdutório sobre Processos Estocásticos e optamos nesta seção por evitar certas tecnicalidades para tornar o livro acessível a uma audiência maior. Acreditamos, de qualquer forma, que as ideias centrais ficarão claras do ponto de vista matemático (embora não totalmente formalizadas).

Definição 4.1 (**Processo de Markov**). *Seja $(X_t; t \geq 0)$ um processo estocástico com espaço de estados $S \subset \mathbb{Z}$ finito ou enumerável como definido acima. Dizemos que X_t é um processo estocástico de Markov com tempo contínuo $t \geq 0$ se vale a condição*

$$P(X_t = j | X_0 = i_0, X_{t_1} = i_1, \ldots, X_{t_n} = i_n) =$$
$$= P(X_t = j | X_{t_n} = i_n), \qquad (*)$$
$$\forall t \geq 0,\ \forall j, i_0, i_1, \ldots, i_n \in S,$$

toda vez que $0 < t_1 < t_2 < t_3 < \ldots < t_n < t$ *e*

$$P(X_0 = i_0, X_{t_1} = i_1, \ldots, X_{t_n} = i_n) > 0.$$

A cada $\hat{w}$ em $\hat{\Omega}$, podemos associar o caminho amostral $\omega = (w_t) = X_t(\hat{w})$. Logo, $\omega : \mathbb{R}^+ \to S$. Considere $S^{\mathbb{R}^+} = \{\omega \,|\, \omega : \mathbb{R}^+ \to S\}$. Seja $U : \hat{\Omega} \to S^{\mathbb{R}^+}$ que associa a cada $\hat{w} \in \hat{\Omega}$ o elemento $U(\hat{w}) = \omega = (w_t)_{t \in \mathbb{R}^+} \in S^{\mathbb{R}^+}$ tal que $w_t = X_t(\hat{w})$. Nosso ponto de vista, mais uma vez, será considerar prioritariamente a sigma-álgebra $\mathcal{A} = U^{-1}(\hat{\mathcal{A}})$ e a probabilidade P induzida por U, tal que $P(B) = \hat{P}(U^{-1}(B))$

Desta forma, para nós, P é uma probabilidade que associa valores reais não negativos a certos subconjuntos $B \subset S^{\mathbb{R}^+}$ (os elementos da sigma-algebra em consideração).

Sendo assim, as variáveis X_t que vamos nos ater, serão as induzidas por U, ou seja, vamos supor que para cada $t \geq 0$ fixado, $X_t : S^{\mathbb{R}^+} \to S$, será tal que $X_t(\omega) = w_t$, se $\omega = (w_t)_{t\in\mathbb{R}}$. Ainda, de maneira análoga ao caso de tempo discreto, um cilindro é um conjunto da forma

$$\{X_{t_1} = s_1, X_{t_2} = s_2, ..., X_{t_n} = s_n\},$$

com $t_1 < t_2 < ... < t_n \in \mathbb{R}^+$, $s_i \in S$, $i \in \{1, 2, ..., n\}$, ou seja, o conjunto

$$\{\, \omega = (w_t)_{t\in\mathbb{R}^+} \text{ tal que } w_{t_1} = s_1, w_{t_2} = s_2, ..., w_{t_n} = s_n \,\} \subset S^{\mathbb{R}^+}.$$

A sigma-álgebra natural a ser considerada aqui é a gerada pelos conjuntos cilindros (ver [EK] ou capítulo 5). Não vamos elaborar muito sobre este ponto que é um pouco mais complexo do que o correspondente a tempo discreto (conforme capítulo 2).

Note que o caminho ω não será em geral contínuo, pois toma valores num subconjunto de $\mathbb{Z}$.

Vamos descrever a seguir um exemplo de Processo Markoviano que será basicamente o único tipo de processo estocástico com tempo contínuo e estado discreto que será considerado neste texto.

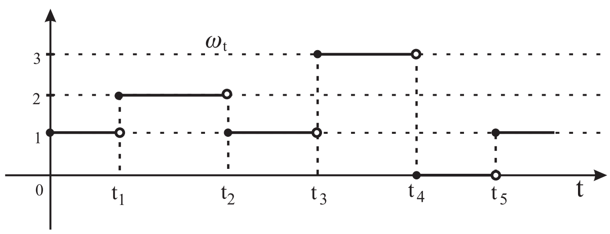


Figura 4.1:

Considere $S \subset \mathbb{Z}$ como o conjunto de estados. Relembrando o caso com tempo discreto, ou seja, $T = \mathbb{N}$, o que determina a probabilidade P sobre $S^{\mathbb{N}}$, associada a um Processo Markoviano é a matriz estocástica $\mathcal{P}$ e um vetor de probabilidade inicial π sobre S. Dito de outra forma P fica determinada pela família de matrizes $\mathcal{P}^n$, que satisfaz a propriedade: $\forall n, m \in \mathbb{N}$

$$\mathcal{P}^n \, \mathcal{P}^m = \mathcal{P}^{n+m},$$

e um vetor de probabilidade inicial π sobre S.

Como vimos antes, para qualquer $n \in \mathbb{N}$, a matriz $\mathcal{P}^n$ é estocástica. Ainda, $\mathcal{P}^0 = I$.

Para descrever um Processo Estocástico Markoviano em que o espaço de estados é $S \subset \mathbb{Z}$ e o conjunto de parâmetros temporais T é igual a $\mathbb{R}^+$, necessitamos de algumas propriedades semelhantes.

Seja uma família de matrizes estocásticas $\mathcal{P}^t$, da forma $\#S$ por $\#S$, indexadas por $t \in \mathbb{R}^+$ tal que vale para todo $s, t \in \mathbb{R}^+$,

$$\mathcal{P}^s \, \mathcal{P}^t = \mathcal{P}^{s+t}.$$

Supomos ainda que $\mathcal{P}^0 = I$, onde I é a matriz identidade da forma $\#S$ por $\#S$ (ou seja a matriz I que é 1 na diagonal e 0 fora dela).

Diremos que a família $\mathcal{P}^t, t \in \mathbb{R}^+$, define um semigrupo a tempo contínuo sobre o conjunto S.

Fixado t, então $(\mathcal{P}^t)_{ij} = \mathcal{P}^t_{ij}$ determina o elemento $i, j \in S$ da matriz $\mathcal{P}^t$.

$\mathcal{P}^t_{ij}$ vai nos dizer qual a probabilidade de estar em j no tempo t, dado que se estava em i no tempo 0, ou seja, para todo $i, j \in S$ e $t \geq 0$, vale que

$$(\mathcal{P}^t)_{ij} = P(X_t = j \,|\, X_0 = i).$$

Da mesma forma, como no caso Markoviano com tempo discreto, vamos supor no nosso modelo que uma vez fixado r, o número $\mathcal{P}^t_{ij}$ vai determinar também qual a probabilidade de estar em j no tempo $r + t$, dado que se estava em i no tempo $r \geq 0$. Ou seja, vamos supor que o processo é homogêneo no tempo.

Seja agora $\pi = (\pi_s)_{s\in S}$, $\sum_s \pi_s = 1$, um vetor de probabilidade inicial sobre S, isto é, $P(X_0 = s) = \pi_s$, $s \in S$.

Com a informação acima gostaríamos de definir uma probabilidade P sobre $S^{\mathbb{R}^+}$. Na verdade vamos definir a probabilidade P sobre um certo subconjunto de $S^{\mathbb{R}^+}$, mas isto explicamos mais tarde.

Fixado o semigrupo $\mathcal{P}^t$ e π vamos definir primeiro a probabilidade de um cilindro contido em $S^{\mathbb{R}^+}$.

Por definição,

$$P(\,X_0 = a_0, X_{t_1} = a_1, X_{t_2} = a_2, X_{t_3} = a_3, ..., X_{t_n} = a_n\,) =$$

$$= P(\,\{w \in \Omega \,|\, w(0) = a_0, w(t_1) = a_1, w(t_2) = a_2, w(t_3) = a_3, ..., w(t_n) = a_n\}\,) =$$

$$\pi_{a_0}\, \mathcal{P}^{t_1}_{a_0\, a_1}\, \mathcal{P}^{t_2-t_1}_{a_1\, a_2}\, \mathcal{P}^{t_3-t_2}_{a_2\, a_3} ... \mathcal{P}^{t_n-t_{n-1}}_{a_{n-1}\, a_n}.$$

Ainda,

$$P(\,X_0 = a_0) = \pi_{a_0}.$$

Por exemplo, se $S = \{1, 2, 3\}$, $\pi = (\pi_1, \pi_2, \pi_3) = (\frac{1}{5}, \frac{2}{5}, \frac{2}{5})$ é o vetor de probabilidade inicial e $\mathcal{P}^t$ é o semigrupo com $t \in \mathbb{R}$ tal que para $t = 1.37$ temos que $\mathcal{P}^{1.37}$ é a matriz estocástica

$$\mathcal{P}^{1.37} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{4}{7} & \frac{2}{7} & \frac{1}{7} \\ \frac{2}{5} & \frac{2}{5} & \frac{1}{5} \end{pmatrix},$$

então

$$P(X_0 = 1, X_{1.37} = 2) = \pi_1 \times (\mathcal{P}_{12})^{1.37} = \frac{1}{5} \times \frac{1}{3} = \frac{1}{15}.$$

Algumas vezes, denotaremos $w_t \in S$ por $w(t)$.

Na verdade é necessário (por razões que não vamos comentar aqui) restringir o espaço $S^{\mathbb{R}^+}$ a um subconjunto Ω. Seja Ω o subconjunto de funções $\omega : \mathbb{R}^+ \to S$, tal que são contínuas à direita e ainda que possuem limite à esquerda (embora não sejam necessariamente contínuas à esquerda). Isto é, para todo t vale que

$$\lim_{u \to t,\, u \geq t} w(u) = w(t),$$

e ainda, existe o limite

$$\lim_{u \to t,\, u \leq t} w(u).$$

Como S é discreto (toma valores em $\mathbb{N}$), tal função ω será constante em intervalos $[t_i, t_{i+1})$ (fechados a esquerda e abertos a direita) onde $0 = t_0 < t_1 < t_2 < t_3 < ... < t_n < ...$

No caso em que $S = \{0, 1, 2, 3\}$, um típico elemento de Ω seria ω como descrito na Figura 4.1.

Um cilindro sobre Ω é definido da mesma forma que foi introduzido antes em $S^{\mathbb{R}^+}$ (só trocando $S^{\mathbb{R}^+}$ por Ω).

Finalmente, afirmamos que existe uma sigma-álgebra $\mathcal{F}$ sobre Ω que contém todos os possíveis cilindros em Ω. Ainda, podemos definir $P(A)$ para qualquer elemento em $A \in \mathcal{F} \subset \mathfrak{p}(\Omega)$ de tal modo que

a) P é uma probabilidade sobre $\mathcal{F}$,

b) $P(B)$ coincide com a definição acima quando B for um cilindro.

c) duas probabilidades sobre $\mathcal{F}$ que coincidem nos cilindros são iguais.

Esta sigma-álgebra $\mathcal{F}$ será denominada de sigma-álgebra gerada pelos cilindros. Vamos analisar nesta seção exclusivamente este espaço de Probabilidade $(\Omega, \mathcal{F}, P)$.

A maioria dos Processos Estocásticos X_t com espaço de estados S e parametro temporal $t \in \mathbb{R}^+$, que aparecem em aplicações tem a propriedade que os caminhos em $S^{\mathbb{R}^+}$, tem limites à esquerda e à direita para quase todo ponto. Em geral, se pode assumir que o processo é tal que todos os caminhos amostrais são contínuos à direita, sem que isto interfira nas distribuições finito-dimensionais (conforme [EK]).

Após a apresentação dos exemplos do Processo de Poisson e de nascimento e morte (nas Seções 4.2 e 4.3) vamos voltar a fazer algumas considerações gerais sobre a Teoria dos Processos Markovianos a tempo contínuo e estado discreto na Seção 4.4.

O leitor que desejar mais detalhes da formalização rigorosa do que foi descrito acima pode encontrar ótimas exposições em [KT], [KT2], [N], [EK] e [GS].

O texto [CD] é uma apresentação em português de caráter mais elementar mas bastante didática.

Em resumo, se existe uma família de matrizes estocásticas $\mathcal{P}^t$, onde $t \geq 0$, do tipo $\#S$ por $\#S$ tal que

1) para todo $t, u \in \mathbb{R}^+$, vale que $\mathcal{P}^{t+u} = \mathcal{P}^t \, \mathcal{P}^u$,

2) para cada $t \in \mathbb{R}^+$ a matriz $\mathcal{P}^t$ é estocástica,

3) $\mathcal{P}^0 = I$,

então, a partir de qualquer vetor de probabilidade $\pi = (\pi_s)_{s \in S}$ sobre S podemos definir uma probabilidade P no conjunto Ω (de caminhos contínuos à direita).

Prioritariamente vai nos importar aqui definir P sobre conjuntos tipo cilindro.

Exemplo 4.1. Por exemplo, seja π vetor de probabilidade inicial, então

$$P(X_0 = s_1, X_{7,5} = s_2) \, = \, \pi_{s_1} \, (\mathcal{P}^{7,5})_{s_1, s_2},$$

$$P(X_t = 3) \, = \sum_{s \in S} \pi_s \, (\mathcal{P}^t)_{s,3}.$$

$$P(X_{2.3} = s_1, X_{7,5} = s_2) \, = \sum_{s \in S} \pi_s \; (\mathcal{P}^{2,3})_{s, s_1} \; (\mathcal{P}^{5,2})_{s_1, s_2}.$$

Seja, por exemplo $S = \{1, 2, 3\}$. Note que

$$(P(X_t = 1), P(X_t = 2), P(X_t = 3)) \, = \, (\pi_1, \pi_2, \pi_3) \, \mathcal{P}^t =$$

$$(\pi_1, \pi_2, \pi_3) \begin{pmatrix} (\mathcal{P}^t)_{11} & (\mathcal{P}^t)_{12} & (\mathcal{P}^t)_{13} \\ (\mathcal{P}^t)_{21} & (\mathcal{P}^t)_{22} & (\mathcal{P}^t)_{23} \\ (\mathcal{P}^t)_{31} & (\mathcal{P}^t)_{32} & (\mathcal{P}^t)_{33} \end{pmatrix}.$$

Note que

$$P(X_t = 3 \,|\, X_0 = 2) \,=\, \frac{\pi_2\, (\mathcal{P}^t)_{2\,3}}{\pi_2} \,=\, (\mathcal{P}^t)_{2\,3}.$$

Ainda,

$$(P(X_t = 1 \,|\, X_0 = 2),\ P(X_t = 2 \,|\, X_0 = 2),\ P(X_t = 3 \,|\, X_0 = 2)) \,=$$

$$\Big(\frac{P(X_t = 1\,,\, X_0 = 2)}{P(X_0 = 2)},\ \frac{P(X_t = 2\,,\, X_0 = 2)}{P(X_0 = 2)},\ \frac{P(X_t = 3\,,\, X_0 = 2)}{P(X_0 = 2)}\Big) \,=$$

$$\Big(\,\frac{\pi_2\,(\mathcal{P}^t)_{2\,1}}{\pi_2}, \frac{\pi_2\,(\mathcal{P}^t)_{2\,2}}{\pi_2}, \frac{\pi_2\,(\mathcal{P}^t)_{2\,3}}{\pi_2}\,\Big) \,=$$

$$(\,(\mathcal{P}^t)_{21}\,,\, (\mathcal{P}^t)_{22}\,,\, (\mathcal{P}^t)_{23}\,) =$$

$$(0, 1, 0) \begin{pmatrix} (\mathcal{P}^t)_{11} & (\mathcal{P}^t)_{12} & (\mathcal{P}^t)_{13} \\ (\mathcal{P}^t)_{21} & (\mathcal{P}^t)_{22} & (\mathcal{P}^t)_{23} \\ (\mathcal{P}^t)_{31} & (\mathcal{P}^t)_{32} & (\mathcal{P}^t)_{33} \end{pmatrix}.$$

Seja S enumerável ou finito qualquer. Lembre que para todo $i, j \in S$ e $t \geq 0$ vale

$$(\mathcal{P}^t)_{ij} = P(X_t = j \,|\, X_0 = i).$$

Ainda, fixado π condição inicial, e X_t o processo estocástico associado a $\mathcal{P}^t$, $t \in \mathbb{R}^+$, se $i, j \in S = \{1, 2, ..., d\}$

$$(\, P(X_t = 1 \,|\, X_0 = i), P(X_t = 2 \,|\, X_0 = i), ...,$$

$$P(X_t = d \,|\, X_0 = i)) \,=$$

$$(\,(\mathcal{P}^t)_{i\,1}, (\mathcal{P}^t)_{i\,2}, ..., (\mathcal{P}^t)_{i\,d}\,) \,=\, e_i\, \mathcal{P}^t.$$

onde e_i é o vetor em Σ que tem todos os elementos nulos menos o i-ésimo que é igual a 1.

$\diamondsuit$

No caso do π geral, pode-se mostrar que a P definida da forma descrita acima, a partir da família $\mathcal{P}^t$ e do vetor π, define, de fato, um Processo Estocástico de Markov.

Vamos mostrar esta propriedade no caso particular

$$P(X_t = j | X_0 = i_0, X_{t_1} = i_1, X_{t_2} = i_2) = P(X_t = j | X_{t_2} = i_2),$$

$\forall t \geq 0$, $\forall j, i_0, i_1, i_2 \in S$, toda vez que $0 < t_1 < t_2 < t$.

De fato,

$$P(X_t = j | X_0 = i_0, X_{t_1} = i_1, X_{t_2} = i_2) =$$

$$= \frac{P(X_t = j, X_0 = i_0, X_{t_1} = i_1, X_{t_2} = i_2)}{P(X_0 = i_0, X_{t_1} = i_1, X_{t_2} = i_2)} =$$

$$= \frac{\pi_{i_0} \mathcal{P}^{t_1}_{i_0 i_1} \mathcal{P}^{t_2 - t_1}_{i_1 i_2} \mathcal{P}^{t - t_2}_{i_2 j}}{\pi_{i_0} \mathcal{P}^{t_1}_{i_0 i_1} \mathcal{P}^{t_2 - t_1}_{i_1 i_2}} = \mathcal{P}^{t-t_2}_{i_2 j} = \frac{\sum_{s \in S} \pi_s \mathcal{P}^{t_2}_{s\, i_2} \mathcal{P}^{t - t_2}_{i_2 j}}{\sum_{s \in S} \pi_s \mathcal{P}^{t_2}_{s\, i_2}} =$$

$$= \frac{\sum_{s \in S} P(X_0 = s, X_{t_2} = i_2, X_t = j)}{\sum_{s \in S} P(X_0 = s, X_{t_2} = i_2)} =$$

$$= \frac{P(X_0 \in S, X_{t_2} = i_2, X_t = j)}{P(X_0 \in S, X_{t_2} = i_2)} = \frac{P(X_{t_2} = i_2, X_t = j)}{P(X_{t_2} = i_2)} =$$

$$= P(X_t = j \,|\, X_{t_2} = i_2).$$

O caso geral pode ser demonstrado de maneira semelhante. Fica assim provada a afirmação acima.

Considere $S = \{1, 2, 3, ..., d\}$ e seja a distribuição de probabilidade inicial o vetor de probabilidade π^0. Denotamos por

$$\pi^t = (P(X_t) = 1, P(X_t) = 2, ..., P(X_t) = d) \,=\, (\pi^0_1, \pi^0_2, ..., \pi^0_d)\, \mathcal{P}^t = \pi^0\, \mathcal{P}^t$$

para todo $t \geq 0$.

Um dos objetivos do que segue é entender a evolução dinâmica de π^t, $t \geq 0$. De forma semelhante ao caso do tempo discreto podemos perguntar se vai haver algum limite de π^t quando $t \to \infty$?

Definição 4.2. *Fixada uma família $\mathcal{P}^t$ satisfazendo 1), 2) e 3) acima, uma Cadeia de Markov é o conjunto de todos os Processos Estocásticos obtidos a partir de qualquer vetor de probabilidade π.*

Uma pergunta natural é se existem famílias de matrizes estocásticas $\mathcal{P}^t$ indexadas por t com tais propriedades? Vamos analisar tal questão a seguir.

Definição 4.3. *Uma matriz $\mathcal{L}$ com entradas reais da forma $\#S$ por $\#S$ tal que*

1) $L_{ii} \leq 0$, para todo $i \in S$,
2) $L_{ij} \geq 0$, para todo $i \neq j$, $i, j \in S$,
3) $\sum_{j \in S} L_{ij} = 0$, para todo $i \in S$,
é dita uma matriz do tipo **linha soma zero**.

Exemplo 4.2. No caso em que $S = \{1, 2, 3, 4\}$, um possível exemplo seria

$$A = \begin{pmatrix} -2 & 1 & 0 & 1 \\ 1 & -2 & 1 & 0 \\ 0 & 0 & -2 & 2 \\ 1 & 0 & 1 & -2 \end{pmatrix}.$$

$\diamondsuit$

Seja B_n uma sequência de matrizes da forma $\#S$ por $\#S$, $n \in \mathbb{N}$, então dizemos que a matriz A é o limite da sequencia B_n,

$$\lim_{n \to \infty} B_n = A,$$

se, para cada $i, j \in S$, vale a propriedade que a entrada $(B_n)_{ij}$ da matriz B_n satisfaz

$$\lim_{n\to\infty} (B_n)_{ij} = A_{ij}.$$

Note que o conjunto das matrizes n por n é isomorfo ao conjunto $\mathbb{R}^{n^2}$. Por exemplo,

$$M(2,2) = \{ \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \mid a_{11}, a_{12}, a_{21}, a_{22} \in \mathbb{R} \},$$

tem a mesma dimensão de $\mathbb{R}^4$, ou seja, é isomorfo a este espaço.

Considere $G : M(2,2) \to \mathbb{R}^4$, tal que

$$G(\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}) = (a_{11}, a_{12}, a_{21}, a_{22}).$$

G é claramente uma bijeção linear. Sendo assim, o sentido de convergência acima descrito é o mesmo que o leitor está familiarizado de convergência em $\mathbb{R}^k$. Só que $k = n^2$.

Vamos ser mais explícitos, diremos, por exemplo, que a sequência de matrizes

$$\begin{pmatrix} b_{11}^n & b_{12}^n \\ b_{21}^n & b_{22}^n \end{pmatrix} \text{ converge à matriz } \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \text{ quando } n \to \infty,$$

se o vetor

$$b_n = (b_{11}^n, b_{12}^n, b_{21}^n, b_{22}^n)$$

em $\mathbb{R}^4$ converge ao vetor

$$a = (a_{11}, a_{12}, a_{21}, a_{22}).$$

Por exemplo,

$$\lim_{n\to\infty} \begin{pmatrix} 1 + \frac{1}{n} & (\frac{1}{2})^n \\ 3 - \frac{1}{2^n} & \cos(\frac{1}{n}) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix}.$$

Formalmente, para qualquer ϵ existe $N > 0$ tal que, para todo $n > N$, vale que

$$||b^n - a|| = ||(b_{11}^n, b_{12}^n, b_{21}^n, b_{22}^n) - (a_{11}, a_{12}, a_{21}, a_{22})|| \leq \epsilon.$$

Acima $||(x_1, x_2, x_3, x_4)|| = \sqrt{x_1^2 + x_2^2 + x_3^2 + x_4^2}$, é denominada de norma euclidiana do vetor (x_1, x_2, x_3, x_4) em $\mathbb{R}^4$. No caso em que o vetor b^n converge ao vetor a, então para todo i, j fixados vale que o número real b_{ij}^n converge ao número real a_{ij}. As propriedades análogas para matrizes do tipo n por n seguem deste fato usando uma identificação similar a dada por G acima. O caso em $\mathbb{R}^{n^2}$ é em tudo semelhante ao que foi descrito acima para $\mathbb{R}^4$.

Dada uma sequência de matrizes A_n, dizemos que

$$\sum_{n=0}^{\infty} A_n = A,$$

se

$$\lim_{n\to\infty} \sum_{k=0}^{n} A_k = A.$$

Definição 4.4. *Considere uma matriz A da forma $\#S$ por $\#S$, então e^A é matriz da forma $\#S$ por $\#S$ dada por*

$$e^A = I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \frac{1}{4!}A^4 + ... + \frac{1}{n!}A^n + ... = \sum_{n=0}^{\infty} \frac{1}{n!}\, A^n.$$

e^A é chamada de exponencial da matriz A.

Exemplo 4.3. Seja $a \in \mathbb{R}$. Dada a matriz A da forma 2 por 2

$$A = \begin{pmatrix} 0 & 0 \\ a & 0 \end{pmatrix},$$

então

$$e^A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ a & 0 \end{pmatrix} + \frac{1}{2!} \begin{pmatrix} 0 & 0 \\ a & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ a & 0 \end{pmatrix} +$$

$$\frac{1}{3!}\begin{pmatrix}0 & 0\\ a & 0\end{pmatrix}\begin{pmatrix}0 & 0\\ a & 0\end{pmatrix}\begin{pmatrix}0 & 0\\ a & 0\end{pmatrix}+ ... =$$

$$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}+\begin{pmatrix}0 & 0\\ a & 0\end{pmatrix}+\frac{1}{2!}\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}+\frac{1}{3!}\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}+ ... =$$

$$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}+\begin{pmatrix}0 & 0\\ a & 0\end{pmatrix}=\begin{pmatrix}1 & 0\\ a & 1\end{pmatrix}$$

♢

Exemplo 4.4. Dada a matriz 2 por 2

$$A=\begin{pmatrix}b & 0\\ 0 & c\end{pmatrix},$$

então é fácil ver que

$$e^A=\begin{pmatrix}e^b & 0\\ 0 & e^c\end{pmatrix},$$

onde $a, b \in \mathbb{R}$.

♢

Se S for finito de cardinalidade n, dada a matriz A da forma n por n, sempre existe e^A (ver [DL]). Se S for infinito, é necessário considerar alguma norma que faça o papel da norma euclidiana em $\mathbb{R}^n$. Desta forma poderemos dar sentido à convergencia de matrizes. Isto pode ser feito e, neste caso, B_n converge a B se para cada par $i, j \in S$, cada elemento $(B_n)_{ij}$ converge a B_{ij} quando n vai à infinito. A seguir vamos considerar apenas as matrizes A tais que existe e^A.

Considere duas matrizes A e B (com $\#S$ finita) da forma $A = \sum_k^\infty A_k$ e $B = \sum_j^\infty B_j$, pode-se mostrar (ver [DL]) que seu produto é dado por

$$AB = C = \lim_{r\to\infty} \sum_{k+j\leq r} A_j B_k.$$

Dadas duas matrizes A e B, nem sempre vale que $e^{A+B} = e^A\, e^B$. O resultado seguinte nos dá uma condição suficiente para que isto ocorra.

Lema 4.1. *Note que dadas A e B do tipo $\#S$ por $\#S$, se $AB = BA$, então $e^{A+B} = e^B\, e^A = e^A\, e^B$.*

Demonstração: De fato, usando o fato que $AB = BA$,

$$e^A\, e^B = \lim_{n\to\infty} \sum_{i=0}^{n} \frac{1}{i!} A^i \cdot \lim_{n\to\infty} \sum_{j=0}^{m} \frac{1}{j!} B^j =$$

$$= \lim_{l\to\infty} \sum_{k+j\leq l} \frac{1}{j!k!} A^k\, B^j = e^{A+B}.$$

□

Referimos o leitor para [DL] para uma demonstração mais cuidadosa do resultado acima.

Note que segue diretamente da definição de exponencial de matriz que se B for a matriz com todas as entradas nulas, então $e^B = I$.

Por definição, para cada t real, a matriz $e^{t\,A}$ é a expressão

$$e^{t\,A} = I + t\,A + \frac{1}{2!} t^2\, A^2 + \frac{1}{3!} t^3\, A^3 + \frac{1}{4!} t^4\, A^4 + ... + \frac{1}{n!} t^n\, A^n + ... = \sum_{n=0}^{\infty} \frac{1}{n!}\, t^n\, A^n.$$

Note que $e^{0\,A} = I$, para qualquer A. Ou seja, $(e^{0\,A})_{ij} = \delta_{ij}$, onde δ_{ij} é a delta de Kronecker.

Segue do Lema acima que dado t e A, então e^{tA} é inversível e $(e^{t\,A})^{-1} = e^{-t\,A}$.

De fato, $(t\,A)\,(-t\,A) = (-\,t\,A)\,(t\,A)$ e assim

$$e^{t\,A}\, e^{-t\,A} = e^{t\,A\,-\,t\,A} = e^0 = I = e^{-t\,A}\, e^{t\,A}.$$

Dada uma matriz $\mathcal{L}$ tipo linha soma zero, denote $\mathcal{P}^t = e^{t\,\mathcal{L}}$, onde $t \geq 0$. Vamos mostrar que tal $\mathcal{P}^t$ define um semigrupo, e este será denominado de semigrupo gerado por $\mathcal{L}$. A matriz $\mathcal{L}$ é denominada de gerador infinitesimal do semigrupo.

Fixado $\mathcal{L}$ tipo linha soma zero qualquer, vamos mostrar que de fato tal $\mathcal{P}^t = e^{t\,\mathcal{L}}$ satisfaz as propriedades 1), 2) e 3) acima (mencionadas antes do Exemplo 4.1). Deste modo obteremos uma classe grande de exemplos de cadeias de Markov com tempo contínuo.

Desta forma se $S = \{1, 2, 3, ..., d\}$ e π^0 o vetor de probabilidade teremos que

$$\pi^t = (P(X_t) = 1, P(X_t) = 2, ..., P(X_t) = d) =$$
$$(\pi^0_1, \pi^0_2, ..., \pi^0_d)\,\mathcal{P}^t = (\pi^0_1, \pi^0_2, ..., \pi^0_d)\,e^{t\,\mathcal{L}} = \pi^0\,e^{t\,\mathcal{L}}.$$

para todo $t \geq 0$.

Note que a solução do sistema linear $x' = x\,\mathcal{L}$ com condição inicial $x(0) = \pi^0$ é $x(t) = \pi^0\,e^{t\,\mathcal{L}}$, $t \in \mathbb{R}$ (ver por exemplo [DL]).

Vamos mostrar a seguir um exemplo interessante e importante do cálculo do semigrupo $\mathcal{P}^t$ a partir do gerador infinitesimal.

Exemplo 4.5. Seja $\mathcal{L}$ a matriz dois por dois tipo linha soma zero dada por

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

onde

$$-a_{11} = a_{12} = \mu \geq 0 \quad \text{e} \quad -a_{22} = a_{21} = \lambda \geq 0.$$

Esta matriz

$$L = \begin{pmatrix} -\mu & \mu \\ \lambda & -\lambda \end{pmatrix},$$

descreve o caso geral de uma matriz tipo linha soma zero no caso $\#S = 2$.

Note que

$$L^2 = L\,L = -(\mu + \lambda)\,L.$$

De fato,

$$L\,L = \begin{pmatrix} -\mu & \mu \\ \lambda & -\lambda \end{pmatrix} \begin{pmatrix} -\mu & \mu \\ \lambda & -\lambda \end{pmatrix} = \begin{pmatrix} \mu^2 + \mu\,\lambda & -\mu^2 - \mu\,\lambda \\ -\lambda\,\mu - \lambda^2 & \mu\lambda + \lambda^2 \end{pmatrix} =$$

$$-(\mu+\lambda) \begin{pmatrix} -\mu & \mu \\ \lambda & -\lambda \end{pmatrix} = -(\mu+\lambda)\,L.$$

Por indução é fácil ver que

$$L^n = (-1)^{n-1}\,(\mu+\lambda)^{n-1} \begin{pmatrix} -\mu & \mu \\ \lambda & -\lambda \end{pmatrix} = (-1)^{n-1}\,(\mu+\lambda)^{n-1}\,L.$$

Para cada t fixo, podemos fazer o mesmo raciocínio para $\mathcal{L}$ tal que

$$-a_{11} = a_{12} = t\,\mu \quad \text{e} \quad -a_{22} = a_{21} = t\,\lambda,$$

obtendo L^n com expressão semelhante a acima.

Portanto, para t fixo, obtemos a matriz

$$\mathcal{P}^t = e^{t\,L} = I - \frac{1}{\mu+\lambda}\,(e^{-t\,(\mu+\lambda)} - 1)\,L =$$

$$= \frac{1}{\mu+\lambda} \begin{pmatrix} \lambda + \mu\,e^{-t\,(\mu+\lambda)} & \mu - \mu\,e^{-t\,(\mu+\lambda)} \\ \lambda - \lambda\,e^{-t\,(\mu+\lambda)} & \mu + \lambda\,e^{-t\,(\mu+\lambda)} \end{pmatrix}.$$

Note que para todo t fixo, tal matriz $\mathcal{P}^t$ é estocástica.

$\diamondsuit$

Vamos agora demonstrar que no caso geral, uma matriz $\mathcal{L}$ tipo linha soma zero determina através de $\mathcal{P}^t = e^{t\,L}, t \in \mathbb{R}^+$, um semigrupo de matrizes.

Fixada a matriz $\mathcal{L}$, dados $t, u \in \mathbb{R}^+$, temos que

$$(t\,A)\,(u\,A) = t\,u\,A\,A = u\,t\,A\,A = (u\,A)\,(t\,A).$$

Logo, pelo lema acima

$$\mathcal{P}^{t+u} = e^{(t+u)\,L} = e^{t\,L}\,e^{u\,L} = \mathcal{P}^t\,\mathcal{P}^u.$$

Já vimos acima que $e^{0\,L} = I$

Falta então a propriedade 2) requerida para $\mathcal{P}^t$ ser semigrupo. Aqui necessitamos assumir que $\mathcal{L}$ seja tipo linha soma zero.

Seja B_t uma família de matrizes da forma $\#S = m \in \mathbb{N}$ por $\#S = m$, indexadas por $t \in \mathbb{R}^+$. Seja s fixo em $\mathbb{R}^+$, então dizemos que a matriz A é o limite da sequência B_t quando t tende a s,

$$\lim_{t\to s} B_t = A,$$

se para cada $i, j \in S$ vale que a entrada $(B_t)_{ij}$ da matriz B_t satisfaz

$$\lim_{t\to s} (B_t)_{ij} = A_{ij}.$$

A convergencia com $t \to s$ é em tudo similar ao que foi descrito no caso de matrizes B_n convergindo a A quando n vai a infinito. A diferença, é claro, é que uma vai a infinito e a outra a s.

Por exemplo, se s está fixado e

$$B_t = \begin{pmatrix} \sin(t) + t^2 & t+1 \\ \\ t^2 + t & -e^t \end{pmatrix},$$

então

$$\lim_{t\to s} B_t = \begin{pmatrix} \sin(s) + s^2 & s+1 \\ \\ s^2 + s & -e^s \end{pmatrix}.$$

Teorema 4.1. *Dada uma matriz finita A então*

$$\lim_{t\to 0} \frac{e^{t\,A} - I}{t} = A.$$

Sendo assim, para $i, j \in S$, $i = j$ vale que

$$A_{ii} = \lim_{t \to 0} \frac{(e^{t\,A})_{ii} - 1}{t} = \lim_{t \to 0} \frac{(e^{t\,A})_{ii} - (e^{0\,A})_{ii}}{t} = (\frac{d\,(e^{t\,A})_{ii}}{dt})_{t=0}.$$

Ainda, para $i, j \in S$, $i \neq j$ vale que

$$A_{ij} = \lim_{t \to 0} \frac{(e^{t\,A})_{ij} - 0}{t} = \lim_{t \to 0} \frac{(e^{t\,A})_{ij} - (e^{0\,A})_{ij}}{t} = (\frac{d\,(e^{t\,A})_{ij}}{dt})_{t=0}.$$

Dito de outra forma: para qualquer vetor x vale que

$$\lim_{t \to 0} v\, \frac{e^{t\,A} - I}{t} = \lim_{t \to 0} \frac{v\, e^{t\,A} - v}{t} \;=\; v\,A$$

Demonstração: Ora,

$$\lim_{t \to 0} \frac{e^{tA} - I}{t} =$$

$$\lim_{t \to 0} \frac{t\,A \,+\, \frac{1}{2!}\,t^2\,A^2 + \frac{1}{3!}\,t^3\,A^3 + ... + \frac{1}{n!}\,t^n\,A^n + ...}{t} = A.$$

Na afirmação acima temos um limite em t e uma soma infinita em n, sendo assim é preciso ser cauteloso. O leitor que desejar obter uma prova rigorosa do fato acima pode encontrá-la em [DL] Seção 2.1.

Finalmente,

$$A_{ij} = \lim_{t \to 0} \frac{(e^{t\,A})_{ij} - \delta_{ij}}{t},$$

onde δ_{ij} é a delta de Kronecker, segue da expressão acima quando analisamos a entrada ij.

□

Teorema 4.2. *Seja A uma matriz finita. Então e^{tA} é uma matriz estocástica para todo $t \in \mathbb{R}$, se e só se, A é uma matriz tipo linha soma zero.*

Demonstração: Note que (ver [DL])

$$\lim_{t\to 0} \frac{e^{tA} - I - tA}{t} =$$

$$= \lim_{t\to 0} \frac{\frac{1}{2!}\, t^2\, A^2 + \frac{1}{3!}\, t^3\, A^3 + \frac{1}{4!}\, t^4\, A^4 + ... + \frac{1}{n!}\, t^n\, A^n + ...}{t} = 0.$$

Acima 0 significa, naturalmente, a matriz com todas as entradas nulas.

Suponha que A seja tipo linha soma zero. Devemos mostrar que $(e^{t\,A})_{ij} \geq 0$ para todo i, j, e todo $t \geq 0$. Primeiro vamos mostrar que vale para t pequeno.

Ora, seja $i \neq j$, ambos em S, então $A_{ij} \geq 0$, e assim

$$0 = \lim_{t\to 0} \frac{(e^{tA})_{ij} - I_{ij} - tA_{ij}}{t} = \lim_{t\to 0} \frac{(e^{t\,A})_{ij} - (e^{0\,A})_{ij}}{t} - A_{ij}.$$

Desta forma, para $i \neq j$ fixo vale que

$$\frac{d(e^{t\,A})_{ij}}{dt}|_{t=0} \geq 0,$$

e, ainda $(e^{0\,A})_{i,j} = 0$, logo temos, para t pequeno, $(e^{t\,A})_{ij} \geq 0$.

Ainda,

$$\lim_{t\to 0}(e^{tA} - I) = \lim_{t\to 0} \left(t\,A + \frac{1}{2!}\, t^2\, A^2 + \frac{1}{3!}\, t^3\, A^3 + ... + \frac{1}{n!}\, t^n\, A^n + ... \right) = 0.$$

Sendo assim, para $i \in S$ fixo, temos que para t pequeno vale $(e^{t\,A})_{ii} \geq 0$, pois $(e^{t\,A})_{ii} \to 1$.

Concluímos assim que existe $\delta > 0$ tal que para $t < \delta$, todas as entradas de $e^{t\,A}$ são positivas.

Seja agora um $u \in \mathbb{R}$ qualquer fixo e $n \in \mathbb{N}$. Ora, pelo Lema 4.1 temos que

$$e^{u\,A} = \underbrace{e^{(\frac{u}{n})\,A}\, e^{(\frac{u}{n})\,A} ...\, e^{(\frac{u}{n})\,A}}_{n\ vezes} = (\, e^{(\frac{u}{n})\,A}\,)^n.$$

Tomando n grande, temos que $\frac{u}{n} < \delta$, e assim o produto acima envolve apenas matrizes com coeficientes não negativos. Logo, $e^{u\,A}$ é uma matriz com entradas não negativas para todo $u \in \mathbb{R}^+$.

É fácil ver que se e^{uA} tem todas as entradas não negativas para todo $u \geq 0$, então para todo $i \neq j$ vale que $A_{ij} \geq 0$. De fato,

$$\frac{d(e^{t\,A})_{ij}}{dt} = A_{ij},$$

assim, se $A_{ij} < 0$, teríamos $(e^{t\,A})_{ij} < 0$, para t pequeno pois, $(e^{0\,A})_{ij} = I_{ij} = 0$.

Da mesma forma, se e^{uA} tem todas as entradas não negativas para todo $u \geq 0$, e a soma de cada linha é 1, então, todas as entradas são menores que 1.

Note que para todo i vale que $\frac{d(e^{t\,A})_{ii}}{dt} = A_{ii}$, Desta forma, $A_{ii} \leq 0$. Isto porque $e^{0\,A}_{ii} = I_{ii} = 1$ e $(e^{uA})_{ii} \leq 1$.

Vamos agora usar o fato que cada linha de A soma zero para mostrar que a soma de cada linha de $e^{t\,A}$ é igual a 1.

Afirmamos que se cada linha de A soma zero então A^n, para $n > 0$, também vale tal propriedade. A afirmação não vale para $n = 0$.

Isto segue do fato que B é do tipo linha soma zero, se e só se, dado o vetor coluna p, que tem todas as entradas 1, então $B\,p = q$, onde q é o vetor coluna que tem todas as entradas 0, ou seja $q = 0$. Deste modo $A^2\,p = A\,(A\,p) = A\,q = A\,0 = 0$, e assim por diante...

Seja $(A^n)_{ij}$ a entrada ij da matriz A^n.

Segue do que foi dito acima que, se A é do tipo linha soma zero, então para $i \in S$

$$\sum_{j\in S} (e^{t\,A})_{ij} = 1 + \sum_{n=1}^{\infty} \frac{t^n}{n!} \sum_{j \in S} (A^n)_{ij} = 1.$$

Logo, $e^{t\,A}$ é estocástica, para qualquer $t \geq 0$ fixado.

Reciprocamente, se $e^{t\,A}$ é estocástica, então para $i \in S$ fixo,

$$0 = (\,\frac{d}{dt}\,1\,)_{t=0} = \Big(\,\frac{d}{dt} \sum_{j\in S} (e^{t\,A})_{ij}\,\Big)_{t=0} = \sum_{j \in S} A_{ij}.$$

Sendo assim a soma em cada linha de A é zero.

□

O resultado acima quando S é infinito, sob certas hipótese naturais, também é válido, mas a demonstração é mais delicada e não será feita aqui. Por exemplo, na última passagem da demonstração acima usamos o fato que

$$\Big(\frac{d}{dt} \sum_{j \in S} (e^{t\,A})_{ij} \Big)_{t=0} = \sum_{j \in S} A_{ij}.$$

Quando S é infinito, em certos casos, não se pode passar a derivada para dentro do somatório.

Em resumo, se $\mathcal{L}$ é tipo linha soma zero, então o semigrupo $\mathcal{P}^t = e^{t\,\mathcal{L}}$ é estocástico e desta forma, a partir de um vetor de probabilidade π (sobre S), podemos definir uma probabilidade P no espaço Ω de caminhos ω em $S^{\mathbb{R}^+}$ contínuos a direita (e com limite a esquerda).

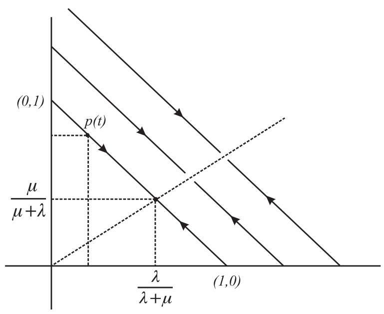


Figura 4.2: Distintas soluções no plano da E.D.O. linear associada $p' = \mathcal{L}(p)$. Quando $t \to \infty$, então, $p(t)$ converge a um determinado ponto na reta pontilhada. Esta reta passa por $\overline{p}$ tal que $\overline{p}\,\mathcal{L} = 0$.

Note que a matriz $\mathcal{L}$, denominada de gerador infinitesimal, determina as probabilidades de transição de i para j em tempo t através de

$$(\mathcal{P}^t)_{ij} = (e^{t\,\mathcal{L}})_{ij}.$$

Denotamos por $\Sigma = \{(p_1, p_2, ..., p_n) \,|\, p_j \geq 0, j = 1, 2, ..., n,$ e ainda, $p_1 + p_2 + ... + p_n = 1\}$.

Dada a equação diferencial $p'(t) = p(t)\,\mathcal{L}$, $t \geq 0$, e fixada uma condição inicial $p(0) = \pi_0 \in \Sigma$, se obtém uma solução $p(t)$ (ver [DL] ou o apêndice 4.5 ao fim deste capítulo), e ela é tal que $p_j(t)$, $j = 1, 2, .., n \in S$, descreve a evolução ao longo do tempo de $P(X_t = j) = p_j(t)$. Isto segue do fato que esta solução $p(t)$ satisfaz $p(t) = \pi_0\, e^{t\,\mathcal{L}}$.

Está é a propriedade fundamental desta parte da teoria.

Por exemplo, seja $S = \{1, 2\}$. Então

$$(P(X_t = 1), P(X_t = 2)\,) = \,(\pi_1, \pi_2)\,\mathcal{P}^t \,=\, (\pi_1, \pi_2)\, e^{t\,\mathcal{L}}.$$

Exemplo 4.6. Suponha que o Processo X_t tomando valores em $S = \{1, 2\}$, tenha como gerador infinitesimal a matriz

$$\mathcal{L} \,=\, \begin{pmatrix} -\mu & \mu \\ \lambda & -\lambda \end{pmatrix},$$

onde λ e μ são positivos.

Conforme calculamos antes, para t fixo, obtemos a matriz

$$\mathcal{P}^t = e^{t\,\mathcal{L}} = \frac{1}{\mu+\lambda} \begin{pmatrix} \lambda + \mu\, e^{-t\,(\mu+\lambda)} & \mu - \mu\, e^{-t\,(\mu+\lambda)} \\ \lambda - \lambda\, e^{-t\,(\mu+\lambda)} & \mu + \lambda\, e^{-t\,(\mu+\lambda)} \end{pmatrix}.$$

Logo, se $p_1 = P(X_0 = 1)$ e $p_2 = P(X_0 = 2)$, temos que

$$(P(X_t = 1), P(X_t = 2)) \,=\, (p_1, p_2)\, \frac{1}{\mu+\lambda} \begin{pmatrix} \lambda + \mu\, e^{-t\,(\mu+\lambda)} & \mu - \mu\, e^{-t\,(\mu+\lambda)} \\ \lambda - \lambda\, e^{-t\,(\mu+\lambda)} & \mu + \lambda\, e^{-t\,(\mu+\lambda)} \end{pmatrix}.$$

Ainda,

$$\begin{pmatrix} \mathcal{P}^t_{1,1} & \mathcal{P}^t_{1,2} \\ \mathcal{P}^t_{2,1} & \mathcal{P}^t_{2,2} \end{pmatrix} \,=\, \begin{pmatrix} P(X_t = 1\,|\,X_0 = 1) & P(X_t = 2\,|\,X_0 = 1) \\ P(X_t = 1\,|\,X_0 = 2) & P(X_t = 2\,|\,X_0 = 2) \end{pmatrix} =$$

$$= \frac{1}{\mu+\lambda}\begin{pmatrix} \lambda+\mu\, e^{-t\,(\mu+\lambda)} & \mu-\mu\, e^{-t\,(\mu+\lambda)} \\ \lambda-\lambda\, e^{-t\,(\mu+\lambda)} & \mu+\lambda\, e^{-t\,(\mu+\lambda)} \end{pmatrix}.$$

Observe que quando $t\to\infty$ temos que

$$\mathcal{P}^t \to \begin{pmatrix} \frac{\lambda}{\lambda+\mu} & \frac{\mu}{\lambda+\mu} \\ \frac{\lambda}{\lambda+\mu} & \frac{\mu}{\lambda+\mu} \end{pmatrix}.$$

Note, neste caso, que para qualquer vetor inicial $p=(p_1,p_2)$ de probabilidade vale que

$$\lim_{t\to\infty} p\,\mathcal{P}^t = (\frac{\lambda}{\lambda+\mu}, \frac{\mu}{\lambda+\mu}),$$

conforme Figura 4.2.

♢

No caso de S infinito, muitos exemplos também podem ser obtidos em que propriedades análogas as que foram descritas acima são válidas. Basta que para $\mathcal{L}$ (uma matriz infinito por infinito) tipo linha soma zero, se consiga provar que está bem definido $e^{t\,\mathcal{L}}$, para todo $t\geq 0$.

Em vários exemplos vamos proceder de maneira inversa. Iremos supor que valham certas propriedades para o Processo Estocástico definido por certa probabilidade P sobre $\Omega\subset S^{\mathbb{R}^+}$ (definida anteriormente). Deduziremos a partir desta informação qual é a matriz $\mathcal{L}$. Este procedimento será utilizado no caso do Processo de Poisson.

Em muitos exemplos aplicados o que se tem na verdade é apenas uma informação de probabilidades de transição $(\mathcal{P}^t)_{ij}$ próxima a $t=0$. Assim pode-se determinar (de alguma forma) a matriz $\mathcal{L}$. Ou seja, $\mathcal{P}^t$, na totalidade dos valores t, surge apenas após se descobrir o $\mathcal{L}$, e só então determinamos explicitamente $\mathcal{P}^t = e^{t\,\mathcal{L}}$.

Vamos formalizar o que estamos afirmando acima: considere novamente o caso geral de Processo Estocástico Markoviano: seja $(\hat{\Omega},\hat{\mathcal{A}},\hat{P})$ espaço de probabilidade, $(S, \mathcal{G})$ espaço mensurável e uma família de variáveis aleatórias X_t indexadas por um parâmetro $t \in \mathbb{R}^+$, onde cada $X_t : (\hat{\Omega}, \hat{\mathcal{A}}, \hat{P}) \to (S, \mathcal{G})$ é mensurável com S enumerável contido em $\mathbb{Z}$.

Suponha que para cada $s \geq 0$ fixado,

$$P(X_{t+s} = j \mid X_s = i),$$

seja independente de $s \geq 0$ e que valha a condição

$$P(X_t = j | X_0 = i_0, X_{t_1} = i_1, \ldots, X_{t_n} = i_n) = P(X_t = j | X_{t_n} = i_n),$$

$$\forall\, t \geq 0,\, \forall j, i_0, i_1, \ldots, i_n \in S,$$

toda vez que $0 < t_1 < t_2 < t_3 < ... < t_n < t$.

Então, defina
$$(\mathcal{P}^t)_{ij} = P(X_t = j \mid X_0 = i).$$

Fixe $t_1 < t_2 < t_3$.

Usando a regra de Bayes, para i_1 e i_3 fixos, podemos condicionar em t_2 e obtemos

$$(\mathcal{P}^{t_3 - t_1})_{i_1\, i_3} = P(X_{t_3} = i_3 | X_{t_1} = i_1) = \frac{P(X_{t_3} = i_3, X_{t_1} = i_1)}{P(X_{t_1} = i_1)} =$$

$$= \frac{\sum_{i_2 \in S} P(X_{t_3} = i_3, X_{t_2} = i_2, X_{t_1} = i_1)}{P(X_{t_1} = i_1)} =$$

$$= \sum_{i_2 \in S} \frac{P(X_{t_3} = i_3, X_{t_2} = i_2, X_{t_1} = i_1)}{P(X_{t_2} = i_2, X_{t_1} = i_1)} \frac{P(X_{t_2} = i_2, X_{t_1} = i_1)}{P(X_{t_1} = i_1)} =$$

$$= \sum_{i_2 \in S} P(X_{t_3} = i_3 \mid X_{t_2} = i_2, X_{t_1} = i_1)\, P(X_{t_2} = i_2 \mid X_{t_1} = i_1) =$$

$$= \sum_{i_2 \in S} P(X_{t_3} = i_3 \mid X_{t_2} = i_2)\, P(X_{t_2} = i_2 \mid X_{t_1} = i_1) =$$

$$= \sum_{i_2 \in S} (\mathcal{P}^{t_2-t_1})_{i_1\, i_2}\, (\mathcal{P}^{t_3-t_2})_{i_2\, i_3}.$$

Definição 4.5. *A identidade*

$$(\mathcal{P}^{t_3-t_1})_{i_1\, i_3} = \sum_{i_2 \in S} (\mathcal{P}^{t_2-t_1})_{i_1\, i_2}\, (\mathcal{P}^{t_3-t_2})_{i_2\, i_3},$$

para $t_1 < t_2 < t_3$ e i_1, i_3 fixos é chamada de Equação de Chapman-Kolmogorov.

Em sua versão mais simples: para qualquer $i, j \in S$, $t, s > 0$

$$P_{i,j}^{t+s} = \sum_{k \in S} P_{i,k}^{s} P_{k,j}^{t}.$$

Neste caso se diz que $\mathcal{P}^t, t \geq 0$ possui a propriedade de semigrupo.

A expressão da equação de Chapman-Kolmogorov também pode ser escrita para $t_1 < t_2 < t_3$ como

$$P(X_{t_3} = i_3 \,|\, X_{t_1} = i_1) \;=\; \sum_{i_2 \in S} P(X_{t_2} = i_2 \,|\, X_{t_1} = i_1)\, P(X_{t_3} = i_3 \,|\, X_{t_2} = i_2),$$

ou

$$P_{i_1,i_3}^{t_3-t_1} = \sum_{i_2 \in S} P_{i_1,i_2}^{t_2-t_1}\;\, P_{i_2,i_3}^{t_3-t_2}.$$

Ou seja, a passagem de i_1 em tempo t_1, para i_3 em tempo t_3, pode ser condicionada a um tempo intermediário t_2, variando todos os possíveis valores $s \in S$.

Note que (pela expressão de produto de matrizes)

$$(\mathcal{P}^{t_3-t_1})_{i_1\, i_3} = \sum_{i_2 \in S} (\mathcal{P}^{t_2-t_1})_{i_1\, i_2}\, (\mathcal{P}^{t_3-t_2})_{i_2\, i_3} = (\mathcal{P}^{t_2-t_1}\, \mathcal{P}^{t_3-t_2})_{i_1\, i_3},$$

onde a úlima expressão é o elemento $i_1\, i_3$ da matriz obtida como produto de $\mathcal{P}^{t_2-t_1}\, \mathcal{P}^{t_3-t_2}$

Para $t_1 < t_2 < t_3$ fixos, variando i_1, i_3 em S, obtemos a partir da igualdade acima para todo i_1, i_3, a identidade entre matrizes

$$\mathcal{P}^{t_3-t_1} = \mathcal{P}^{t_2-t_1}\,\mathcal{P}^{t_3-t_2}.$$

Em outras palavras, para $t, s \geq 0$

$$\mathcal{P}^{t+s} = \mathcal{P}^t\,\mathcal{P}^s,$$

ou seja, que vale que $\mathcal{P}^t, t \in \mathbb{R}$, é um semigrupo a tempo contínuo. Por definição as matrizes $\mathcal{P}^t$ são estocásticas para cada t fixo.

Sendo assim, a equação de Chapman-Kolmogorov traduz a propriedade de semigrupo associado ao processo obtido através de P e X_t, $t \geq 0$.

Nesta formulação não falamos em $\mathcal{L}$. Algumas vezes, em problemas do mundo real aparece de maneira natural o $\mathcal{P}^t$. Outras vezes, o que aparece de maneira natural é o $\mathcal{L}$. Exemplos serão apresentados em breve.

Note que por indução também vale o seguinte: dados $t_1, t_2, ..., t_n$ maiores que zero, então

$$\mathcal{P}^{t_1+t_2+...+t_n} = \mathcal{P}^{t_1}\,\mathcal{P}^{t_2}\,...\,\mathcal{P}^{t_n}.$$

Denote por $P^{s,t} = e^{(t-s)\,A}$, $t > s$, a matriz que tem na entrada i, j a probabilidade de $X_t = j$ dado que $X_s = i$, $i, j \in S$.

Obtemos assim entradas $P_{ij}(s,t)$.

Observe que se A é uma matriz n por n então a solução da equação diferencial matricial $X'(t) = AX(t)$, com a condição incial $X(0) = I$, é $X(t) = e^{t\,A}$. Mesma coisa para a equação diferencial matricial $X'(t) = X(t)\,A$, com a condição inicial $X(0) = I$. Disso vão seguir as equações forward e backward que descreveremos a seguir.

Se o gerador infinitesimal $\mathcal{L}$ tiver entradas $A_{j,k}$, $j, k \in S$, então valem as seguintes equações (quando S é enumerável):

a) Equação diferencial forward (para frente) de Chapmann-Kolomogorov: para s, t

$$\frac{\partial P_{ij}}{\partial t}(s,t) = \sum_k P_{ik}(s,t) A_{kj},$$

b) Equação diferencial backward (para trás) de Chapmann-Kolomogorov: para s, t

$$\frac{\partial P_{ij}}{\partial s}(s,t) = -\sum_k A_{ik} P_{kj}(s,t).$$

Estas equações podem ser deduzidas diretamente da equação de Chapman-Kolmogorov por diferenciação. De fato, para $\mathcal{P}^t$ da forma $\mathcal{P}^t = e^{t\,\mathcal{L}}$,

a) a Equação diferencial forward (para frente) de Chapman-Kolomogorov é

$$\frac{d}{dt} e^{t\,\mathcal{L}} = \mathcal{P}' = \mathcal{P}\;\mathcal{L}\,,$$

e

b) a Equação diferencial backward (para trás) de Chapman-Kolomogorov é

$$\frac{d}{dt} e^{t\,\mathcal{L}} = \mathcal{P}' = \mathcal{L}\,\mathcal{P}\,.$$

Dado $\mathcal{P}^t$, podemos nos perguntar: existe um $\mathcal{L}$ natural associado ao problema? Ou seja, será que existe $\mathcal{L}$ tal que $\mathcal{P}^t$, é da forma $\mathcal{P}^t = e^{t\,\mathcal{L}}$? A resposta como veremos (em muitos casos) é sim, e desta forma a) e b) acima são válidos.

Com as mesmas hipóteses anteriores vamos supor agora que seja verdadeira ainda a *hipótese de diferenciabilidade em* $t = 0$, ou seja, que existam os limites abaixo: para todo $i \neq j \in S$ fixados

$$\lim_{t\to 0} \frac{(\mathcal{P}^t)_{ij}}{t} := L_{ij},$$

e para todo $i \in S$

$$\lim_{t\to 0} \frac{(\mathcal{P}^t)_{ii} - 1}{t} := L_{ii},$$

onde

$$(\mathcal{P}^t)_{ij},$$

descreve a probabilidade de transição de i para j em tempo t pequeno.

Vamos denotar por L_{ij} o primeiro limite acima e L_{ii} o segundo limite acima.

De maneira compacta, as duas expressões acima significam que

$$\lim_{t\to 0} \frac{\mathcal{P}^t - I}{t} = \mathcal{L}. \quad (*)$$

Ou seja, para qualquer π_0 vale

$$\lim_{t\to 0} \frac{\pi_0 \mathcal{P}^t - \pi_0}{t} = \pi_0\, \mathcal{L}. \quad (*)$$

Denominamos a matriz $\mathcal{L}$ com ordenadas $L_{i,j}$, $i, j \in S$, de gerador infinitesimal do processo.

Note que $L_{ij} \geq 0$ para $i \neq j$ pois $(\mathcal{P}^t)_{ij} \geq 0$.

Por outro lado, $L_{ii} \leq 0$ para $i \in S$ pois $(\mathcal{P}^t)_{ii} - 1 \leq 0$.

Desta forma, dado o processo estocástico Markoviano X_t, $t \geq 0$, tomando valores em $S = \{1, 2, ..., d\}$, definimos a familia de matrizes estocásticas $\mathcal{P}^t, t \geq 0$, satisfazendo

$$(\mathcal{P}^t)_{ij} = P(X_t = j \,|\, X_0 = i)$$

e a partir de (*) descobrimos seu gerador infinitesimal $\mathcal{L}$.

Lembre que $P_{i,j}(t) = P(X_{t+u} = j \,|\, X_u = i)$, $t, u \geq 0$, independe de u.

Ainda, a matriz $\mathcal{L}$ assim obtida é tipo linha soma zero, pois para i fixo

$$\sum_{j\in S} L_{ij} = \sum_{j\in S, j\neq i} L_{ij} + L_{ii} =$$

$$= \lim_{t\to 0} \sum_{j\in S, j\neq i} \frac{(\mathcal{P}^t)_{ij}}{t} + \lim_{t\to 0} \frac{(\mathcal{P}^t)_{ii} - 1}{t} =$$

$$= \lim_{t\to 0} \frac{\sum_{j\in S}(\mathcal{P}^t)_{ij} - 1}{t} = \lim_{t\to 0} \frac{0}{t} = 0.$$

Alertamos ao leitor que no caso em que S é infinito a dedução acima pode não ser válida por envolver somas infinitas e limites em t. Em muitas situações em que S é infinito, mesmo assim, a afirmação é verdadeira.

Exemplo: No caso em que $S = \{1,2,3,4\}$, e

$$\mathcal{L} = \begin{pmatrix} -2 & 2 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -3 & 2 \\ 0 & 0 & 2 & -2 \end{pmatrix},$$

o limite (*) acima deve ser interpretado da seguinte forma:

a)$P_{1,2}(t) = 2t + o(t)$

b)$P_{1,1}(t) = 1 - 2t + o(t)$

c)$P_{1,3}(t) = o(t)$

d)$P_{2,3}(t) = t + o(t)$

e)$P_{2,1}(t) = t + o(t)$

f)$P_{3,4}(t) = 2t + o(t)$

g)$P_{3,2}(t) = 1t + o(t)$

h)$P_{3,3}(t) = 1 - 3\,t + o(t)$

i)$P_{4,3}(t) = 2t + o(t)$,

onde $\lim_{t\to 0} \frac{o(t)}{t} = 0$.

Sendo assim, o modelo afirma que estando no tempo 0 em 3 a probabilidade de saltar para 4 é duas vezes maior do que saltar para 2 quando o tempo t é pequeno. Da mesma forma, pela propriedade de homogeniedade que assumimos temos que estando no tempo s em 3 a probabilidade de saltar no futuro para 4 é duas vezes maior que saltar para 2 quando o tempo t é próximo de s.

No caso S finito, ou $S = \mathbb{N} = \{1,2,3,...,n,...\}$ (infinito), usaremos a notação

$$\frac{d\,\mathcal{P}^s}{ds} = \begin{pmatrix} \frac{d\,(\mathcal{P}^s)_{11}}{ds} & \frac{d\,(\mathcal{P}^s)_{12}}{ds} & \frac{d\,(\mathcal{P}^s)_{13}}{ds} & \cdots \\ \frac{d\,(\mathcal{P}^s)_{21}}{ds} & \frac{d\,(\mathcal{P}^s)_{22}}{ds} & \frac{d\,(\mathcal{P}^s)_{23}}{ds} & \cdots \\ \frac{d\,(\mathcal{P}^s)_{31}}{ds} & \frac{d\,(\mathcal{P}^s)_{32}}{ds} & \frac{d\,(\mathcal{P}^s)_{33}}{ds} & \cdots \\ \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdots \\ \frac{d\,(\mathcal{P}^s)_{n\,1}}{ds} & \frac{d\,(\mathcal{P}^s)_{n\,2}}{ds} & \frac{d\,(\mathcal{P}^s)_{n\,3}}{ds} & \cdots \\ \cdot & \cdot & \cdot & \cdots \end{pmatrix}.$$

Acima, o elemento $i\,j$ da matriz $\mathbb{N}$ por $\mathbb{N}$ é $\frac{d\,(\mathcal{P}^s)_{i\,j}}{ds}$.

No caso em que $S = \mathbb{Z}$, temos que

$$\frac{d\,\mathcal{P}^s}{ds} = \begin{pmatrix} \cdots & \cdot & \cdot & \cdot & \cdots \\ \cdots & \cdot & \cdot & \cdot & \cdots \\ \cdots & \cdot & \cdot & \cdot & \cdots \\ \cdots & \frac{d\,(\mathcal{P}^s)_{-1\,-1}}{ds} & \frac{d\,(\mathcal{P}^s)_{-1\,0}}{ds} & \frac{d\,(\mathcal{P}^s)_{-1\,1}}{ds} & \cdots \\ \cdots & \frac{d\,(\mathcal{P}^s)_{0\,-1}}{ds} & \frac{d\,(\mathcal{P}^s)_{0\,0}}{ds} & \frac{d\,(\mathcal{P})_{0\,1}}{ds} & \cdots \\ \cdots & \frac{d\,(\mathcal{P}^s)_{1\,-1}}{ds} & \frac{d\,(\mathcal{P}^s)_{1\,0}}{ds} & \frac{d\,(\mathcal{P}^s)_{1\,1}}{ds} & \cdots \\ \cdots & \cdot & \cdot & \cdot & \cdots \\ \cdots & \cdot & \cdot & \cdot & \cdots \\ \cdots & \cdot & \cdot & \cdot & \cdots \end{pmatrix}.$$

Acima, o elemento $i\,j$ da matriz $\mathbb{Z}$ por $\mathbb{Z}$ é $\frac{d\,(\mathcal{P}^s)_{i\,j}}{ds}$.

O que descrevemos acima, de maneira sintética, é que

$$\frac{d\mathcal{P}^s}{ds}|_{s=0} = \mathcal{L}.$$

Dito de outra forma, para cada i, j fixos vale

$$\frac{d\,(\mathcal{P}^s)_{i,j}}{ds}\,|_{s=0} \,=\, \frac{d\,P(X_s = j \,|\, X_0 = i)}{d\,s}\,|_{s=0} \,=\, L_{i,j}.$$

Exemplo 4.7. Considere o semigrupo

$$\mathcal{P}^t = \frac{1}{\mu+\lambda}\begin{pmatrix} \lambda + \mu\, e^{-t\,(\mu+\lambda)} & \mu - \mu\, e^{-t\,(\mu+\lambda)} \\ \lambda - \lambda\, e^{-t\,(\mu+\lambda)} & \mu + \lambda\, e^{-t\,(\mu+\lambda)} \end{pmatrix}.$$

Derivando em cada entrada da matriz obtemos que

$$\frac{d\,\mathcal{P}^t}{dt}\Big|_{t=0} = \begin{pmatrix} -\mu & \mu \\ \lambda & -\lambda \end{pmatrix}.$$

$\diamondsuit$

Gostaríamos de concluir, a partir das hipóteses acima, que para cada $t \geq 0$ a matriz $\mathcal{P}^t$ que tem como entrada ij o valor $(\mathcal{P}^t)_{ij}$ satisfaz $\mathcal{P}^t = e^{t\mathcal{L}}$. Para que isto ocorra necessitamos, de fato, apenas a hipótese de diferenciabilidade em $t = 0$.

Teorema 4.3. *Se $\mathcal{P}^t$ é uma familia de matrizes da forma n por n, indexadas por $t \geq 0$, que satisfaz a propriedade de semigrupo, com $\mathcal{P}^0 = I$, e ainda $\lim_{s\to 0}\frac{\mathcal{P}^s - I}{s} = \mathcal{L}$, então*

$$\mathcal{P}^t = e^{t\,\mathcal{L}},$$

para todo $t \geq 0$.

Demonstração: De fato, note primeiro que estamos assumindo acima que

$$\lim_{s\to 0}\frac{\mathcal{P}^s - I}{s} = \mathcal{L}. \qquad (*)$$

Como $\mathcal{P}^s, s \geq 0$ é um semigrupo, para cada t fixo,

$$\lim_{s\to 0}\frac{\mathcal{P}^{s+t} - \mathcal{P}^t}{s} = \lim_{s\to 0}\frac{\mathcal{P}^s - \mathcal{P}^0}{s}\,\mathcal{P}^t = \lim_{s\to 0}\frac{\mathcal{P}^s - I}{s}\,\mathcal{P}^t = \mathcal{L}\,\mathcal{P}^t. \qquad (**)$$

O produto na última expressão à direita é produto de matrizes.

Logo, concluímos que a existência de derivada (à direita) em $t = 0$ (*), ou seja,

$$\Big(\frac{d\mathcal{P}^t}{dt}\Big)_{t=0} = \mathcal{L}$$

implica (pela propriedade de semigrupo) a existência de derivada (à direita) em qualquer ponto $t \geq 0$ (**), ou seja, a existência de

$$\frac{d\mathcal{P}^t}{dt}.$$

Considere a curva (no espaço das matrizes $\mathbb{R}^{n^2}$ com a norma descrita anteriormente) $\mathcal{P}^t$ com $t \geq 0$. Segue do que foi dito acima, que para qualquer $t \geq 0$, vale

$$\frac{d\mathcal{P}^t}{dt} = \mathcal{P}^t\,\mathcal{L}.$$

Ou seja, o semigrupo $\mathcal{P}^t = X(t) \in \mathbb{R}^{n^2}$ satisfaz a equação diferencial

$$X'(t) = X(t)\,\mathcal{L}.$$

Ainda $X(0) = I$.

Da Teoria das Equações Diferenciais Ordinárias (ver [DL] ou Apêndice 4.5) sabe-se que

$$\mathcal{P}^t = e^{t\,\mathcal{L}}, t \geq 0.$$

$\square$

A conclusão é que se a família de matrizes $\mathcal{P}^t$, $t \geq 0$, da forma n por n, oriundas de um Processo Estocástico Markoviano com $\#S = n$, satisfaz a hipótese de diferenciabilidade em $t = 0$, então existe $\mathcal{L}$ tal que para todo $t \geq 0$

$$\mathcal{P}^t = e^{t\,\mathcal{L}}.$$

A evolução temporal do processo depende basicamente das propriedades da matriz $\mathcal{L}$, denominada de gerador infinitesimal do Processo Estocástico Markoviano.

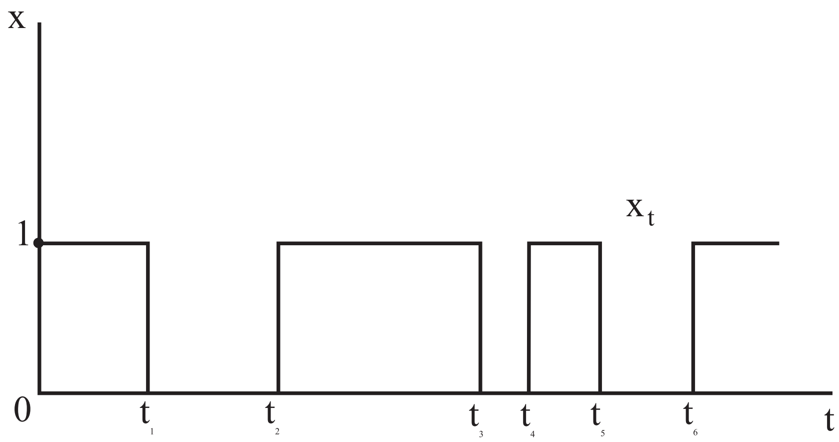


Figura 4.3: Caminho amostral $x(t) = x_t$ em que o conjunto de estados é $S = \{0, 1\}$

Sendo assim, suponha que $\pi^0 = (\pi_1^0, \pi_2^0, ..., \pi_d^0)$ denota a probabilidade inicial do Processo Markoviano. Ainda, denote

$$\pi_t = (P(X_t = 1), P(X_t = 2), ..., P(X_t = d)),$$

a probabilidade no tempo t. A evolução temporal de π_t é dada por

$$\pi_t = (P(X_t = 1), P(X_t = 2), ..., P(X_t = d)) =$$

$$(\pi_1^0, \pi_2^0, ..., \pi_d^0)\,\mathcal{P}^t = \pi^0\, e^{t\,\mathcal{L}} \tag{4.1}$$

para todo $t \geq 0$.

π_t satisfaz a equaão diferencial $\frac{d}{dt}\pi_t = \pi_t\,\mathcal{L}$.

Deste modo, nosso procedimento inicial de definir uma probabilidade P em $S^{\mathbb{R}^+}$ a partir de um semigrupo $e^{t\,\mathcal{L}}, t \geq 0$, onde $\mathcal{L}$ é matriz do tipo linha soma zero, é bastante natural.

$\mathcal{L}$ é denominado de gerador infinitesimal do Processo Markoviano a tempo contínuo.

A Figura 4.4 ilustra o fato que $(\mathcal{P}^0)_{ii} = 1$ e $(\mathcal{P}^0)_{ij} = 0$. Ainda,

$$\frac{d}{dt}\,(\mathcal{P}^t_{ij})_{t=0} \,=\, \lambda_{ij} \,=\, L_{ij}.$$

A medida que $t > 0$ cresce, a função $P_{ij}(t) \,=\, (\mathcal{P}^t)_{ij}$ cresce, quando $i \neq j$. A medida que $t > 0$ cresce, a função $P_{ii}(t) \,=\, (\mathcal{P}^t)_{ii}$ decresce, quando $i \in S$.

Lembre que a evolução de π_t, $t \geq 0$, é descrita por (4.1).

Definição 4.6. *Fixada a matriz tipo linha soma zero $\mathcal{L}$, dizemos que o vetor de probabilidade π é um vetor estacionário para o semigrupo $\mathcal{P}^t = e^{t\,\mathcal{L}}, t \in \mathbb{R}^+$, se para todo $t \in \mathbb{R}^+$ vale que*

$$\pi\,\mathcal{P}^t = \pi\,e^{t\,\mathcal{L}} = \pi.$$

Teorema 4.4. *Seja S finito, $\mathcal{L}$ matriz tipo linha soma zero fixada, e $\mathcal{P}^t = e^{t\,\mathcal{L}}$, então π é vetor estacionário para $\mathcal{P}^t$, se e só se, $\pi\,\mathcal{L} = 0$.*

Demonstração: Ora, se $\pi\,\mathcal{L} = 0$, então a solução de $x'(t) = x(t)\,\mathcal{L}$, com a condição inicial $x(0) = \pi$ é $x(t) = \pi$ para todo t real. Mas sabemos que tal $x(t)$ satisfaz

$$x(t) = \pi\,e^{t\,\mathcal{L}},$$

e assim segue que π é vetor estacionário (pela unicidade da solução da EDO).

Reciprocamente, se $x(t) = \pi\,e^{t\,\mathcal{L}} = \pi$, para todo $t \geq 0$, então, a solução do sistema de equações diferenciais lineares $x'(t) = x(t)\,\mathcal{L}$ com a condição inicial $x(0) = \pi$ é constante, logo

$$0 = x'(t) = \pi\,\mathcal{L}.$$

□

Exemplo 4.8. (ver [CD]) Para a matriz geral da forma linha soma zero do tipo dois por dois

$$\mathcal{L} = \begin{pmatrix} -\mu & \mu \\ \lambda & -\lambda \end{pmatrix} = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix},$$

com $\lambda, \mu \geq 0$, temos que

$$(\frac{\lambda}{\lambda+\mu}, \frac{\mu}{\lambda+\mu})\, \mathcal{L} \,=\, (0,0).$$

Logo, o vetor $(\pi_1, \pi_2) = \pi = (\frac{\lambda}{\lambda+\mu}, \frac{\mu}{\lambda+\mu})$ é estacionário e assim,

$$\pi = (\frac{\lambda}{\lambda+\mu}, \frac{\mu}{\lambda+\mu}) \,= (\frac{\lambda}{\lambda+\mu}, \frac{\mu}{\lambda+\mu})\; e^{t\,\mathcal{L}} \,=\, \pi\; e^{t\,\mathcal{L}},$$

para todo $t \geq 0$.

Podemos pensar que tal matriz descreve uma Cadeia Markoviana com tempo contínuo tomando valores em $S = \{0,1\}$. Neste caso, os caminhos amostrais são sempre do tipo descritos na figura 4.3. As probabilidades de transição seriam dadas por

$$e^{t\,\mathcal{L}} \,=\, \begin{pmatrix} p_{00}(t) & p_{01}(t) \\ p_{10}(t) & p_{11}(t) \end{pmatrix}.$$

A evoluão ao longo do tempo t da matriz $e^{t\,\mathcal{L}}$ é de fundamental importância na análise do Processo. A evolução das entradas $p_{00}(t)$ e $p_{11}(t)$ é descrita na figura 4.4.

$\diamondsuit$

Dizemos que o processo a tempo contínuo X_t, $t \geq 0$, é estacionário se para qualquer t e qualquer n, qualquer escolha $t_1 < t_2 < ... < t_n$, vale

$$P(\,X_0 = a_0, X_{t_1} = a_1, X_{t_2} = a_2, X_{t_3} = a_3, ..., X_{t_n} = a_n\,) =$$

$$= P(\, X_t = a_0, X_{t+t_1} = a_1, X_{t+t_2} = a_2, X_{t+t_3} = a_3, ..., X_{t+t_n} = a_n\,).$$

É fácil ver que a escolha de um vetor de probabilidade inicial π estacionário torna o Processo de Markov X_t assim obtido um Processo Estacionário.

Desta forma, considerando o exemplo acima de processo X_t, $t \in \mathbb{R}^+$, com $S = \{1, 2\}$, associado ao semigrupo com gerador infinitesimal $\mathcal{L}$, temos que se considerarmos o vetor π estacionário, então vale

$$(P(X_t = 1), P(X_t = 2)) \,=\, (\pi_1, \pi_2)\, \mathcal{P}^t \,=$$

$$=\, (\pi_1, \pi_2)\, e^{t\,\mathcal{L}} = (\pi_1, \pi_2) = (P(X_0 = 1), P(X_0 = 2)),$$

para todo $t \geq 0$.

Ainda, $P(\, X_0 = 1, X_{3.1} = 1, X_{7.3} = 2) = P(\, X_{1.2} = 1, X_{4.3} = 1, X_{8.5} = 2)$.

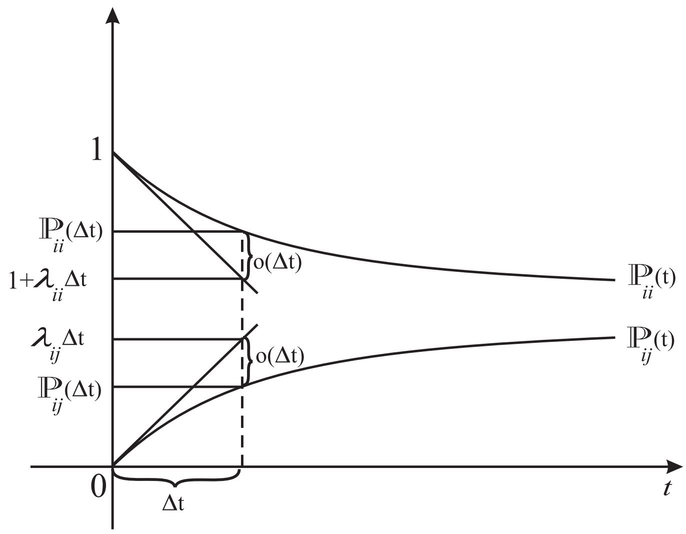


Figura 4.4: A evolução ao longo do tempo das entradas $p_{00}(t)$ (ou, $p_{11}(t)$) e $p_{10}(t)$ (ou, $p_{01}(t)$) da matriz $e^{t\,\mathcal{L}}$.

Exemplo 4.9. No caso em que $S = \{1, 2, 3, 4\}$, e

$$\mathcal{L} = \begin{pmatrix} -2 & 1 & 0 & 1 \\ 1 & -2 & 1 & 0 \\ 0 & 0 & -2 & 2 \\ 1 & 0 & 1 & -2 \end{pmatrix},$$

podemos obter o $\pi \in \Sigma$, tal que $\pi\,\mathcal{L} = 0$, da forma a ser descrita abaixo.

Seja $x = (1, x_2, x_3, x_4) \in \mathbb{R}^4$, desejamos inicialmente resolver $x\,L = 0$, ou seja resolver o sistema linear

$$\begin{aligned} -2 + x_2 + x_4 &= 0, \\ 1 - 2x_2 &= 0, \\ x_2 - 2x_3 + x_4 &= 0. \end{aligned}$$

Este sistema tem três equações e três incógnitas. A quarta equação é redundante (pois a quarta coluna é combinação linear das outras já que cada linha soma zero).

Da segunda equação obtemos que $x_2 = \frac{1}{2}$. Substituindo $x_2 = \frac{1}{2}$ na primeira obtemos $x_4 = \frac{3}{2}$. Finalmente, da última equação obtemos $x_3 = 1$. Sendo assim,

$$(1\ ,\ \frac{1}{2}\ ,\ 1\ ,\ \frac{3}{2})\ L = 0.$$

Como $1 + \frac{1}{2} + 1 + \frac{3}{2} = 4$, temos que o único $\pi \in \Sigma$ tal que $\pi\,L = 0$ é

$$\pi = (\frac{1}{4}\,,\ \frac{1}{8}\ ,\ \frac{1}{4}\ ,\ \frac{3}{8}).$$

Desta forma, para todo t

$$(\frac{1}{4}\ ,\ \frac{1}{8}\ ,\ \frac{1}{4}\ ,\ \frac{3}{8})\, e^{t\,\mathcal{L}} = (\frac{1}{4}\ ,\ \frac{1}{8}\ ,\ \frac{1}{4}\ ,\ \frac{3}{8})$$

e tal π define a probabilidade inicial que torna o Processo de Markov X_t associado ao semigrupo $\mathcal{P}^t = e^{t\,\mathcal{L}}$ um Processo Estacionário.

$\diamondsuit$

A demonstração dos próximos três resultados pode ser evitada numa primeira leitura. É importante, no entanto, o bom entendimento do que afirmam os três teoremas a seguir.

Teorema 4.5. *Quando S é finito, dada $\mathcal{L}$ tipo linha soma zero existe um vetor de probabilidade π estacionário.*

Demonstração: Para cada $t = 1/n \geq 0$ fixo, $n \in \mathbb{N}$, a matriz $\mathcal{P}^{1/n}$ é estocástica, logo, existe um vetor q_n tal que

$$q_n \, \mathcal{P}^{1/n} = q_n.$$

Assim, para qualquer n

$$\frac{q_n \, (\, \mathcal{P}^{1/n} \, - I \,)}{1/n} = 0.$$

O conjunto Σ é compacto, desta forma existe um elemento q que é limite da subsequencia q_n com n natural.

Ora, como

$$\lim_{t \to 0} \frac{\mathcal{P}^t \, - I}{t} = \mathcal{L}$$

uniformemente, então

$$0 = \lim_{n \to \infty} \frac{q_n \, (\, \mathcal{P}^{1/n} \, - I \,)}{1/n} = \lim_{n \to \infty} \frac{q \, (\, \mathcal{P}^{1/n} \, - I \,)}{1/n} \; = \; q \, \mathcal{L}.$$

Sendo assim, q satisfaz $q \, \mathcal{L} = 0$ e portanto a partir do último teorema q é um vetor estacionário para o semigrupo $e^{t \, \mathcal{L}}$, $t \geq 0$.

$\square$

Teorema 4.6. *Seja $\mathcal{L}$ matriz k por k tipo linha soma zero, denote por $\mathbb{I}$ a matriz k por k, que tem todas as entradas iguais a 1, então se $\mathbb{I} - \mathcal{L}$ for inversível, temos que*

$$\pi \,=\, (1\,1\,...1\,1)(\mathbb{I} - \mathcal{L})^{-1}$$

satisfaz

$$\pi\,\mathcal{L} = 0.$$

Demonstração: Sabemos que existe π tal que $\pi\,\mathcal{L} = 0$. Note que

$$\pi\,(\mathbb{I} - \mathcal{L}) = (1\,1\,...1\,1) - (0\,0\,...0\,0) = (1\,1\,...1\,1).$$

Aplicando em ambos os lados da igualdade acima (do lado direito) a matriz $(\mathbb{I} - \mathcal{L})^{-1}$ obtemos o resultado desejado.

□

Teorema 4.7. *Seja S finito com cardinalidade d. Dada a matriz $\mathcal{L}$ tipo linha soma zero, suponha que exista apenas um elemento $\pi \in \Sigma$ tal que $\pi\,\mathcal{L} \,=\, 0$. Assuma também que $e^{t\,\mathcal{L}}$ seja regular, para todo $t \geq 0$. Então para qualquer elemento $p \in \Sigma$, vale que*

$$\lim_{t\to\infty} p\,e^{t\,\mathcal{L}} \,=\, \pi.$$

Demonstração: Seja t real positivo fixo.

Considere o espaço

$$V = \{v \in \mathbb{R}^d \,|\; 0 =< v, (1,1,1,...,1) >= \,v_1\, +\, v_2 + \,...\, + v_d\,\}.$$

O espaço V tem dimensão $d-1$. O espaço gerado pelo vetor $(1,1,...,1)$ e o espaço vetorial V geram o $\mathbb{R}^d$.

Vamos mostrar agora que dado $v \in V$, temos que $v\,\mathcal{L} \,\in\, V$.

Como a matriz $\mathcal{L}$ é tipo linha soma zero temos que vale

$$\mathcal{L} \begin{pmatrix} 1 \\ 1 \\ . \\ . \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ . \\ . \\ 0 \end{pmatrix},$$

logo,

$$(1\ 1\ 1 \ldots 1)\ \mathcal{L}^* = (0\ 0\ 0 \ldots 0),$$

onde $\mathcal{L}^*$ é a matriz adjunta de $\mathcal{L}$ (ver [DL] ou [Li1]).

Sendo assim,

$$< v\,\mathcal{L}\,,\,(1,1,...,1) > = < v\,,\,(1,1,...,1)\,\mathcal{L}^* > = < v\,,\,(0,0,...,0) > = 0.$$

Assim, $v\,\mathcal{L} \in V$, se $v \in V$.

Desta forma, também vale que $v\,(\,t^n\ \mathcal{L}^n\,) \in V$, para todo n, e finalmente que

$$v\,(\sum_{n=0}^{\infty} \frac{1}{n!}\,(t\,\mathcal{L})^n\,) = v\,e^{t\,\mathcal{L}} \in V,$$

se $v \in V$.

Acima usamos o fato que como V é um conjunto fechado, toda sequencia convergente de elementos de $V \subset \mathbb{R}^d$, converge a um elemento de V.

Afirmamos que todo autovalor generalizado de $\mathcal{L}$, outro que 1, tem parte real negativa.

Considere a Transformação linear $T : V \to V$, induzida por $e^{\mathcal{L}}$, isto ï¿$\frac{1}{2}$, $T(v) = v\,e^{\mathcal{L}}$. Os autovalores de T são tais que seus autovalores tem norma menor ou igual a 1, pois $e^{\mathcal{L}}$ é estocástica (ver Teorema 2.7). Ainda, como π é único que satisfaz $\pi\mathcal{L} = 0$, então pelo Teorema 2.7, os outros autovalores de T são distintos de 1.

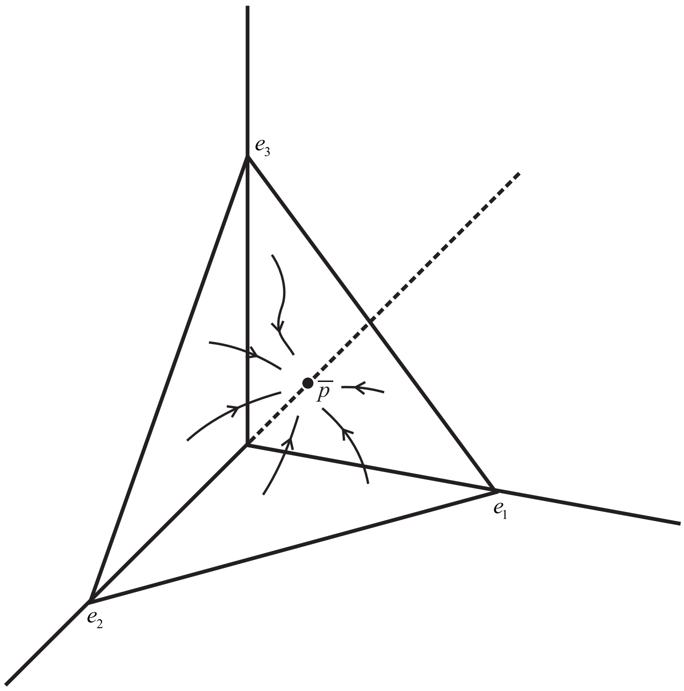


Figura 4.5: As distintas soluções $p(t)$ no caso em que $\Sigma \subset \mathbb{R}^3$. Neste caso, $\overline{p}\ \mathcal{L} = 0$.

Denote por c, onde $c < 1$ o maior destes autovalores. Desta forma, $|T(v)| < c\,|v|$, para todo v em V.

Seja $u \in \mathbb{C}^{d-1}$ autovetor generalizado associado a $\lambda \in \mathbb{C}$, autovalor de $\mathcal{L}_V$ (ou seja, $\mathcal{L}$ restrito a V). Então vale que $u\,(\,t^n\ \mathcal{L}^n\,) = (t\,\lambda)^n\,u$, e assim

$$u\,e^{t\,\mathcal{L}} = e^{t\,\lambda}u.$$

Sendo assim, como $e^{\lambda} \leq c < 1$, então, dados $v_1, v_2 \in V$, vale que

$$|(v_1\, -\, v_2)\,e^{t\,\mathcal{L}}|\, \leq\, c^t\,|v_1\, -\, v_2|.$$

Note que π não está em V pois $< \pi\,,\,(1,1,...,1) >= 1$. Seja agora, $x_1, x_2 \in$

Σ, e escreva $x_1 = v_1 + c_1\,\pi$, e $x_2 = v_2 + c_2\,\pi$, onde $c_1, c_2 \in \mathbb{R}$.

Ora,

$$1 \;=< \, x_1, (1,1,...,1) >=$$

$$=< v_1, (1,1,...,1) > \; + c_1 \; < \pi, (1,1,...,1) >= \; 0 \; + \; c_1.$$

Logo $c_1 = 1$. Aplicando o mesmo a x_2 obtemos que $c_2 = 1$.

Temos então que

$$x_1\, e^{t\,\mathcal{L}} \; - \; x_2\, e^{t\,\mathcal{L}} = (v_1 + \; \pi)\, e^{t\,\mathcal{L}} - (v_2 + \; \pi)\, e^{t\,\mathcal{L}} =$$

$$= v_1\, e^{t\,\mathcal{L}} \; - \; v_2\, e^{t\,\mathcal{L}}.$$

Sendo assim,

$$|\, x_1\, e^{t\,\mathcal{L}} \; - \; x_2\, e^{t\,\mathcal{L}}\,| \; < \; c^t\, |\, v_1 \; - \; v_2\,| \; =$$

$$= c^t\, |(v_1 + \pi) \; - \; (v_2 + \pi)\,| \leq c^t\, |\, x_1 \; - \; x_2\,|.$$

Portanto, para $p \in \Sigma$ qualquer e t real vale que

$$|\, \pi \; - \; p\, e^{t\,\mathcal{L}}\,| \; = \; |\, \pi\, e^{t\,\mathcal{L}} \; - \; p\, e^{t\,\mathcal{L}}\,| \; \leq \; c^t\, |\, \pi \; - \; p\,|\,.$$

Desta forma,

$$\lim_{t \to \infty} p\, e^{t\,\mathcal{L}} \; = \; \pi.$$

A velocidade de convergência a π em t é da ordem de c^t, com $c < 1$.

Logo, esta convergencia é bem rápida.

$\square$

O Exemplo 4.6 acima mostra um caso particular do último teorema, numa situação em que vale para qualquer p inicial fixado

$$\lim_{t \to \infty} p\, e^{t\,\mathcal{L}} \; = \; (\frac{\lambda}{\lambda + \mu}, \frac{\mu}{\lambda + \mu}) \; = \; \pi.$$

Ilustrando o que afirma o teorema acima no caso $S = \{1, 2, 3\}$, na Figura 4.5 exibimos várias curvas $x(t) = p\, e^{t\,\mathcal{L}}$, $t \in \mathbb{R}^+$, descritas por distintas condições iniciais p, e sua convergencia ao vetor invariante $\bar{p}$, quando t vai a infinito.

Teorema 4.8. *Suponha que S tem cardinalidade d e π é o único elemento em Σ tal que $\pi\,\mathcal{L} = 0$, onde $\mathcal{L}$ é matriz d por d tipo linha soma zero. Considere o semigrupo $\mathcal{P}^t$ gerado por $\mathcal{L}$ e o correspondente Processo Estocástico Markoviano X_t, $t \in \mathbb{R}^+$. Dados i e j em S, temos que*

$$\lim_{t\to\infty} P(X_t = j \,|\, X_0 = i) \,=\, \pi_j,$$

onde $\pi = (\pi_1, \pi_2, ..., \pi_d)$.

Demonstração: O semigrupo $\mathcal{P}^t$ tem gerador infinitesimal $\mathcal{L}$, logo

$$\begin{gathered}(\,P(X_t = 1|X_0 = i), P(X_t = 2|X_0 = i), ...,\\ P(X_t = d|X_0 = i)\,) \,=\\ = (\,(\mathcal{P}^t)_{i\,1}, (\mathcal{P}^t)_{i\,2}, ..., (\mathcal{P}^t)_{i\,d}\,) \,=\\ = e_i\,\mathcal{P}^t = \, e_i\,e^{t\,\mathcal{L}}.\end{gathered}$$

Pelo último teorema temos que

$$\lim_{t\to\infty} e_i\,e^{t\,\mathcal{L}} \,=\, \pi \,=\, (\pi_1, \pi_2, ..., \pi_d).$$

□

É fácil ver a partir do que foi mostrado acima que se considerarmos um vetor inicial de probabilidade qualquer p_0, o semigrupo $\mathcal{P}^t$ gerado por $\mathcal{L}$ e o correspondente Processo Estocástico Markoviano X_t, $t \in \mathbb{R}^+$, então, dado j em S, temos que

$$\lim_{t\to\infty} P(X_t = j) \,=\, \pi_j,$$

onde $\pi = (\pi_1, \pi_2, ..., \pi_d)$ é o vetor estacionário.

Exemplo 4.10. Como vimos antes, no caso em que $S = \{1, 2, 3, 4\}$, e

$$\mathcal{L} = \begin{pmatrix} -2 & 1 & 0 & 1\\ 1 & -2 & 1 & 0\\ 0 & 0 & -2 & 2\\ 1 & 0 & 1 & -2 \end{pmatrix},$$

o vetor $\pi = (\frac{1}{4}\,, \frac{1}{8}\,, \frac{1}{4}\,, \frac{3}{8}) \in \Sigma$ é tal que $\pi\,\mathcal{L} \,=\, 0$. Isto ï¿½$\frac{1}{2}$,

$$(\frac{1}{4}\,, \frac{1}{8}\,, \frac{1}{4}\,, \frac{3}{8})$$

é estacionário para $e^{t\,\mathcal{L}}$, $t \in \mathbb{R}^+$.

O Processo de Markov X_t associado ao semigrupo $\mathcal{P}^t = e^{t\,\mathcal{L}}$ e a probabilidade inicial $\pi = (\frac{1}{4}\,, \frac{1}{8}\,, \frac{1}{4}\,, \frac{3}{8}) \in \Sigma$ é estacionário. Apenas para esta escolha de probabilidade inicial o Processo de Markov associado a $\mathcal{L}$ será estacionário.

Considere o Processo de Markov X_t associado ao semigrupo $\mathcal{P}^t = e^{t\,\mathcal{L}}$ e uma probabilidade inicial qualquer $p \in \Sigma$.

Neste caso, então vale que

$$\lim_{t\to\infty} P(X_t = 3 \,|\, X_0 = 2) \,=\, \frac{1}{4}.$$

Ainda,

$$\lim_{t\to\infty} P(X_t = 3) \,=\, \lim_{t\to\infty} \sum_{j=1}^{4} P(X_t = 3 \,|\, X_0 = j)\, p_j \,= \sum_{j=1}^{4} \frac{1}{4}\; p_j \,=\, \frac{1}{4}.$$

$\diamondsuit$

Algumas palavras de cautela ao leitor. No desenvolvimento acima deixamos alguns pontos conceituais não totalmente bem resolvidos. Destacamos que em Processos Estocásticos a tempo contínuo, as probabilidades finito dimensionais não caracterizam de maneira única o Processo. Com isto queremos dizer que dois Processos Estocásticos X_t e Y_t (duas probabilidades P_X e P_Y sobre $S^{\mathbb{R}^+}$) poderiam ter as mesmas distribuições finito-dimensionais e mesmo assim existir um conjunto K tal que $P_X(K) = 0$ e $P_Y(K) = 1$.

Sob certas hipóteses, resultados similares ao do caso tempo discreto podem ser obtidos de qualquer forma (propriedade forte de Markov, ergodicidade, etc.) mas a complexidade matemática da análise destas questões é bem maior e está fora do escopo do presente texto. Na Seção 4.4 faremos um apanhado resumido de alguns resultados básicos da teoria geral das Cadeias de Markov a tempo contínuo.

4.2 O Processo de Poisson

Um exemplo muito importante de Processo Estocástico a tempo contínuo é o de Poisson. Seja $S = \mathbb{N} = \{0, 1, 2, ..., n, ...\}$, vamos voltar a considerar o Processo de Poisson de parâmetro λ. Afirmamos que este processo pode ser descrito como $\mathcal{P}^t = e^{t\mathcal{L}}$, onde a matriz $\mathcal{L}$ tipo linha soma zero é

$$\mathcal{L} = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & 0 & ..... & ... & .. & ... \\ 0 & -\lambda & \lambda & 0 & 0 & 0 & ..... & ... & .. & ... \\ 0 & 0 & -\lambda & \lambda & 0 & 0 & ..... & ... & .. & ... \\ 0 & 0 & 0 & -\lambda & \lambda & 0 & ..... & ... & .. & ... \\ . & . & . & . & . & . & . & .... & .. & ... \\ 0 & 0 & 0 & ...... & 0 & 0 & -\lambda & \lambda & 0 & ... \\ . & . & . & . & . & . & . & .... & .. & ... \end{pmatrix},$$

Nesta matriz temos sempre $-\lambda$ na diagonal e na entrada logo à direita da diagonal um λ. As outras entradas são iguais a zero.

A matriz $\mathcal{L}$ associada ao Processo de Poisson, não possui vetor de probabilidade π, tal que $\pi\,\mathcal{L} = 0$.

Lembre que assumimos no Processo de Poisson que $P(X_0 = 0) = 1$, logo estamos considerando o vetor inicial

$$\pi = (1, 0, 0, 0, 0, ...).$$

Como vimos antes vale para todo $j \geq 1$ e todo t

$$P(\,X_{t+h} - X_t = 1\,|\,X_t = j\,) = \lambda h + o(h),$$

e

$$P(\,X_{t+h} - X_t = 0\,|\,X_t = j\,) = 1 - \lambda h + o(h),$$

Considere a equação diferencial $x'(t) = x(t)\,\mathcal{L}$, onde $\mathcal{L}$ é a matriz acima, com a condição inicial $x(0) = (1, 0, 0, 0, 0, ...)$, e sua solução

$$x(t) = (x_0(t), x_1(t), x_2(t), x_3(t), ..., x_n(t), ...).$$

A solução $x(t)$, neste caso, nos dá uma informação importante para o Processo de Poisson (conforme descrito no começo desta seção em que assumimos que $P(X_0 = 0) = 1$):

$$(P(X_t\; =\; 0),\; P(X_t\; =\; 1),\; P(X_t\; =\; 2), ...) = x(t)\; =$$

$$(1, 0, 0, 0, ..., 0, ...)\; e^{t\,\mathcal{L}} =\; (1, 0, 0, 0, ..., 0, ...)\; \mathcal{P}^t.$$

Note que segue direto da expressão de $\mathcal{L}$ que a equação $x' = x\,\mathcal{L}$, é equivalente a relação

$$x_k^{'}(t) = -\lambda x_k(t)\; +\; \lambda\, x_{k-1}(t),$$

para qualquer $k \geq 1$, e

$$x_0^{'}(t) = -\lambda x_0(t). \tag{4.2}$$

Teríamos assim, equivalentemente, a relação

$$\frac{d}{dt}\, P(X_t = k) = -\lambda\, P(X_t = k)\; +\; \lambda\, P(X_t = k-1),$$

para qualquer $k \geq 1$, e

$$\frac{d}{dt}\, P(X_t = 0) = -\lambda\, P(X_t = 0).$$

Vamos calcular $e^{t\,\mathcal{L}}$ neste caso. Note que $t\,\mathcal{L} = t\,\mathcal{L}_1 + t\mathcal{L}_2$, onde

$$
t\,\mathcal{L}_1 = \begin{pmatrix}
-\lambda\,t & 0 & 0 & 0 & 0 & 0 & ..... & ... & .. & ... \\
0 & -\lambda\,t & 0 & 0 & 0 & 0 & ..... & ... & .. & ... \\
0 & 0 & -\lambda\,t & 0 & 0 & 0 & ..... & ... & .. & ... \\
0 & 0 & 0 & -\lambda\,t & 0 & 0 & ..... & ... & .. & ... \\
. & . & . & . & . & . & . & .... & .. & ... \\
0 & 0 & 0 & ...... & 0 & 0 & -\lambda\,t & 0 & 0 & ... \\
. & . & . & . & . & . & . & .... & .. & ...
\end{pmatrix},
$$

e

$$
t\,\mathcal{L}_2 = \begin{pmatrix}
0 & \lambda\,t & 0 & 0 & 0 & 0 & ..... & ... & .. & ... \\
0 & 0 & \lambda\,t & 0 & 0 & 0 & ..... & ... & .. & ... \\
0 & 0 & 0 & \lambda\,t & 0 & 0 & ..... & ... & .. & ... \\
0 & 0 & 0 & 0 & \lambda\,t & 0 & ..... & ... & .. & ... \\
. & . & . & . & . & . & . & .... & .. & ... \\
0 & 0 & 0 & ...... & 0 & 0 & 0 & \lambda\,t & 0 & ... \\
. & . & . & . & . & . & . & .... & .. & ...
\end{pmatrix}.
$$

Note que $(t\,\mathcal{L}_1)\,(t\,\mathcal{L}_2) \,=\, (t\,\mathcal{L}_2)\,(t\,\mathcal{L}_1)$.

Neste caso, também se pode mostrar que $e^{t\,\mathcal{L}} = e^{t\,(\mathcal{L}_1+\mathcal{L}_2)} = e^{t\,\mathcal{L}_1}\,e^{t\,\mathcal{L}_2}$.

Ora, $e^{t\,\mathcal{L}_1}$ é matriz diagonal que tem $e^{-t\,\lambda}$ em cada elemento da diagonal.

Ainda,

$$
(t\mathcal{L}_2)\,(t\mathcal{L}_2) = \begin{pmatrix}
0 & 0 & (\lambda\,t)^2 & 0 & 0 & 0 & 0 & . & 0 & ... \\
0 & 0 & 0 & (\lambda\,t)^2 & 0 & 0 & 0 & . & 0 & ... \\
0 & 0 & 0 & 0 & (\lambda\,t)^2 & 0 & 0 & . & 0 & ... \\
0 & 0 & 0 & 0 & 0 & (\lambda\,t)^2 & 0 & ... & 0 & ... \\
. & . & . & . & . & . & . & .... & .. & ... \\
0 & 0 & 0 & 0 & ...... & 0 & 0 & . & (\lambda\,t)^2 & ... \\
. & . & . & . & . & . & . & .... & .. &
\end{pmatrix},
$$

e

$$(t\,\mathcal{L}_2)^3 = \begin{pmatrix} 0 & 0 & 0 & (\lambda\, t)^3 & 0 & 0 & 0 & 0 & . & ... \\ 0 & 0 & 0 & 0 & (\lambda\, t)^3 & 0 & 0 & 0 & . & ... \\ 0 & 0 & 0 & 0 & 0 & (\lambda\, t)^3 & 0 & 0 & . & ... \\ 0 & 0 & 0 & 0 & 0 & 0 & (\lambda\, t)^3 & 0 & . & ... \\ . & . & . & . & . & . & . & .... & .. & \\ 0 & 0 & 0 & 0 & 0 & ...... & 0 & . & 0 & (\lambda\, t)^3\, .. \\ . & . & . & . & . & . & . & .. & .. & .. \end{pmatrix},$$

e assim por diante.

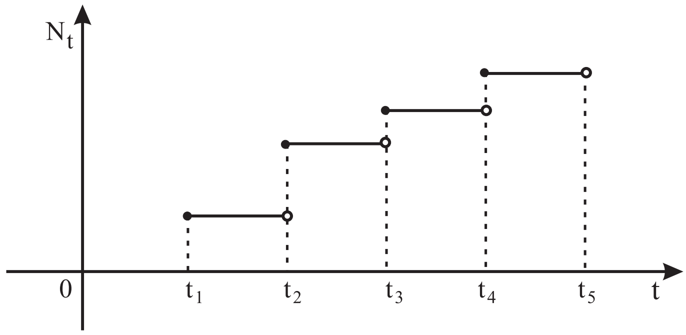


Figura 4.6: Trajetória amostral típica $w(t)$ do Processo de Poisson

Logo,

$$e^{t\,\mathcal{L}_2} = \begin{pmatrix} 1 & (\lambda\, t) & \frac{(\lambda\, t)^2}{2!} & \frac{(\lambda\, t)^3}{3!} & \frac{(\lambda\, t)^4}{4!} & \frac{(\lambda\, t)^5}{5!} & ... & ... & ... & ... \\ 0 & 1 & (\lambda\, t) & \frac{(\lambda\, t)^2}{2!} & \frac{(\lambda\, t)^3}{3!} & \frac{(\lambda\, t)^4}{4!} & ... & ... & ... & ... \\ 0 & 0 & 1 & (\lambda\, t) & \frac{(\lambda\, t)^2}{2!} & \frac{(\lambda\, t)^3}{3!} & . & ... & ... & ... \\ 0 & 0 & 0 & 1 & (\lambda\, t) & \frac{(\lambda\, t)^2}{2!} & ... & ... & ... & ... \\ . & . & . & . & . & . & . & .... & .. & ... \\ 0 & 0 & 0 & 0 & ...... & 0 & . & 1 & (\lambda\, t) & ... \\ . & . & . & . & . & . & . & ... & .. & ... \end{pmatrix}.$$

Sendo assim, obtivemos explicitamente $e^{t\,\mathcal{L}} = e^{t\,\mathcal{L}_1}\, e^{t\,\mathcal{L}_2} = e^{-\lambda\, t}\, e^{t\,\mathcal{L}_2}$ usando a expressão acima.

Desta forma, considerando $x(t) = (x_1(t), x_2(t), x_3(t), ...)$, onde $x_k(t) = P(X_t = k)$, obteremos explicitamente

$$x(t) \;=\; (1, 0, 0, 0, ..., 0, ...)\; e^{t\,\mathcal{L}}.$$

Portanto,

$$P(X_t = k) = x_k(t) = \frac{(\lambda\, t)^k}{k!}\, e^{-\lambda\, t}.$$

Existem outra formas alternativas de se apresentar tal processo. A razão porque tal processo tem este gerador infinitesimal $\mathcal{L}$ será fornecida a seguir.

Como sabemos, o Processo de Poisson X_t é um processo estocástico a parâmetro contínuo: (X_t), $t \in [0, +\infty)$, $S = \mathbb{N} = \{0, 1, 2, 3, 4, ..., n, ...\}$

Note que faz sentido também considerar um processo com uma condição inicial $\pi_0 = (\pi_0, \pi_1, \pi_2, ..., \pi_n, ...)$, tal que cada $\pi_n \geq 0$, e $\sum_{n=0}^{\infty} \pi_n = 1$ (em vez de $\pi_0 = (1, 0, 0, 0, ..., 0, ...)$ como acima). Mas não vamos considerar tal situaão aqui.

Note que $P(X_{9.43} = 7\,|\, X_{3.4} = 11) = 0$. Com probabilidade 1 os caminhos amostrais $w(t)$ do Processo de Poisson são constantes em intervalos e monótonos crescentes. Uma descrição geométrica de um caminho típico é apresentada na figura 4.6.

Vamos partir de algumas premissas naturais e concluir que devemos modelá-lo através de uma cadeia de Markov com tempo contínuo e com gerador infinitesimal $\mathcal{L}$ como descrito acima.

Assim, devemos deduzir do modelo que vale

$$\frac{d}{dt}\, P(X_t = k) = -\lambda\, P(X_t = k) \,+\, \lambda\, P(X_t = k - 1),$$

para qualquer $k \geq 1$, e

$$\frac{d}{dt}\, P(X_t = 0) = -\lambda\, P(X_t = 0).$$

Descreveremos o processo de Poisson através de uma aplicaão bem definida: chamadas telefônicas recebidas numa central telefônica.

Definimos $X_t =$ número de chamadas recebidas até o instante t

Suponhamos que uma pessoa queira construir um modelo probabilístico para este processo. A partir de observações feitas, ela chega às seguintes conclusões com relação às chamadas realizadas nessas linhas.

1. A distribuição do número de chamadas recebidas durante qualquer intervalo de tempo dado parece depender somente da duração do intervalo de tempo. Quanto maior o intervalo de tempo maior tende a ser o número de chamadas.

2. As chamadas parecem chegar independentemente, isto é, um excesso ou declínio acidental do número de chamadas em algum intervalo de tempo dado não parece exercer nenhum efeito sobre o número de chamadas ocorridas durante qualquer outro intervalo de tempo.

3. A probabilidade de duas ou mais chamadas chegarem durante um intervalo de tempo pequeno é muito pequena quando comparada à probabilidade de uma única chamada.

Cada resultado do experimento "contar o número de chamadas recebidas até os tempos $t_1, t_2, t_3, ..., t_n, ...$", gera um caminho tipo

$$\omega = (w_t)_{t\in\mathbb{R}^+}$$

e $w_t \in S = \mathbb{N}$ para todo t. Logo, $\omega \in \Omega \subset \mathbb{N}^{\mathbb{R}}$. Temos que w_t é constante em intervalos de tempo. Mais exatamente, em cada intervalo $[t_i, t_{i+1})$, a função (o caminho amostral) w_t é constante.

Para $s, t \in \mathbb{R}^+, k \in \mathbb{N}$ consideraremos o evento (conjunto) $A^k_{s,\,t} = \{\omega$ tal que ocorreram exatamente $\mathbf{k}$ chamadas no intervalo $(s, s+t]\} \subset \Omega$.

Em outras palavras $A^k_{s,\,t} = \{w \in \Omega ; w(s+t) - w(s) = k\}$, $s, t \geq 0$ *e* $k = 0, 1, 2, ...$

A σ-álgebra $\mathcal{A}$ que devemos considerar sobre Ω deve conter todos os eventos (conjuntos) $A^k_{s,\,t}$.

Vamos a seguir descrever mais explicitamente $P(A^k_{s,\,t})$.

Com base nas observaães que foram feitas são formuladas as seguintes hipóteses:

Hipótese 1: (Incrementos Estacionários)

A probabilidade de ocorrencia de **k** chamadas no intervalo (s,s+t] depende somente de **t** e não de **s**.

$P(A^k_{s,\,t}) = P(A^k_{s',\,t}) \stackrel{\text{def}}{=} P_k(t) \quad \forall s \neq s'$ em particular, $P(A^k_{s,\,t}) = P(A^k_{0,\,t}) = P_k(t)$. Assim, $P_k(t)$ descreve a probabilidade de exatamente k saltos no intervalo $[0, t]$.

Hipótese 2: (Incrementos Independentes)

O número de chamadas durante intervalos disjuntos de tempo são independentes, isto é,

$A^k_{s,\,t}$ e $A^j_{u,\,v}$ são independentes para toda a escolha de k e j se $(s, s+t] \cap (u, u+v] = \emptyset$

Portanto,

$P(A^k_{s,\,t} \cap A^j_{u,\,v}) = P(A^k_{s,\,t}).P(A^j_{u,\,v}) = P_k(t).P_j(v)$, sempre que $(s, s+t] \cap (u, u+v] = \emptyset$. Supomos sem perda de generalidade que $s+t < u$.

Em particular,

$P(A^k_{s,\,t} \cap A^{n-k}_{u,\,v}) = P(A^k_{s,\,t}).P(A^{n-k}_{u,\,v}) = P_k(t).P_{n-k}(v), \forall\, n\; e\; \forall\, k \leq n.$

Uma forma compacta de descrever esta hipótese é dizer que o conjunto $A^k_{s,\,t}$ é independente da sigma álgebra gerada por $X_r - X_z$, $r, z \in (u, u+v]$, onde $r > z$.

Hipótese 3: (As chamadas chegam sozinhas e não simultaneamente)

A probabilidade de ocorrerem duas ou mais chamadas durante um intervalo de tempo pequeno é muito pequena quando comparada à de ocorrer uma única chamada. Observe que:

P(ocorrencia de duas ou mais chamadas)= $P(X_t \geq 2) = \sum_{k=2}^{\infty} P_k(t) = 1 - P_0(t) - P_1(t)$

Assim, a **hipótese 3** diz que

$$\lim_{t\to 0} \frac{1 - P_0(t) - P_1(t)}{P_1(t)} = \lim_{t\to 0} \frac{1 - P_0(t) - P_1(t)}{1 - P_0(t)} = 0,$$

ou seja,

$$\lim_{t\to 0} \frac{P_1(t)}{1 - P_0(t)} = 1.$$

Queremos mostrar que

A) $P_0(t) = e^{\lambda t}$, para todo $t \geq 0$,

B) $P_k'(t) = -\lambda\, P_k(t) + \lambda\, P_{k-1}(t)$ para todo $t \geq 0$

A) O evento "nenhuma chamada no intervalo $(0, t]$" é o mesmo que os eventos "nenhuma chamada nos intervalos $\left(0, \frac{t}{n}\right], \left(\frac{t}{n}, \frac{2t}{n}\right], \left(\frac{2t}{n}, \frac{3t}{n}\right], \ldots \left(\frac{(n-1)t}{n}, \frac{nt}{n}\right]$", onde $n \in \mathbb{N}$ está fixo. Desta forma

$A^0_{0,t} = A^0_{0,\frac{t}{n}} \cap A^0_{\frac{t}{n},\frac{t}{n}} \ldots \cap A^0_{\frac{(n-1)t}{n},\frac{t}{n}}$

Como os intervalos são disjuntos, os eventos são independentes (**hipótese 2**) e então para $t \in \mathbb{R}^+$ e $n \in \mathbb{N}$:

$$P_0(t) = P(A^0_{0,t}) = \prod_{i=0}^{n-1} P\left(A^0_{\frac{it}{n},\frac{t}{n}}\right) = \left[P\left(A^0_{0,\frac{t}{n}}\right)\right]^n = (P_0(\frac{t}{n}))^n.$$

Logo, tomando $t = n$, obtemos

$$P_0(n) = (P_0(1))^n.$$

Vamos estender a afirmação para n feita acima primeiramente para números racionais da forma m/n.

Para $m \in \mathbb{N}$ e $t \in \mathbb{R}^+$ fixados temos $P_0(mt) = (P_0(t))^m$ e assim

$$P_0(\frac{m}{n}t) = (P_0(\frac{t}{n}))^m = \left[(P_0(t))^{\frac{1}{n}}\right]^m = (P_0(t))^{\frac{m}{n}}.$$

Acima usamos o fato que $P_0(t) = (P_0(\frac{t}{n}))^n$.

Portanto, tomando $t = 1$ se $r = \frac{m}{n} \in \mathbb{Q}$,

$$P_0(r) = (P_0(1))^r.$$

Queremos agora estender este resultado para qualquer real $t > 0$.

Observe que $P_0(t)$ é uma função decrescente pois

$t \leq s \Rightarrow A^0_{0,s} \subset A^0_{0,t} \Rightarrow P_0(t) \geq P_0(s)$

Seja $t > 0$ fixo e sejam $r_1, r_2 \in \mathbb{Q}$ tais que $r_1 \leq t \leq r_2$.

Então, $(P_0(1))^{r_1} = P_0(r_1) \geq P_0(t) \geq P_0(r_2) = (P_0(1))^{r_2}$.

Se $r_1 \uparrow t$ e $r_2 \downarrow t$ temos que $(P_0(1))^{r_1}_0 \downarrow (P_0(1))^t$ e $(P_0(1))^{r_2} \uparrow (P_0(1))^t$

Portanto,

$$P_0(t) = (P_0(1))^t, \ \forall\, t > 0.$$

Queremos agora encontrar $P_0(t)$ explicitamente.

Afirmamos que $0 < P_0(1) < 1$.

De fato, note que

$P_0(1) = 1 \Rightarrow P_0(t) = 1, \forall\, t > 0 \Rightarrow$ probabilidade 1 de não ocorrer nenhuma chamada em $(0, t]$.
(O que não é de interesse algum)

$P_0(1) = 0 \Rightarrow P_0(t) = 0, \forall\, t > 0 \Rightarrow$ probabilidade 1 de ocorrer pelo menos uma chamada em$(0, t]$,para todo $t > 0$.

Vamos analisar com cuidado este segundo caso. se $P_0(1) = 0$, teriam que ocorrer pelo menos duas chamadas em $(0,t]$, com probabilidade 1, pois a ocorrência de pelo menos uma chamada em $(0, \frac{t}{2}]$ e de pelo menos uma chamada em $(\frac{t}{2}, t]$ (este evento também seria de probabilidade um, pela **hipótese 1**), implica a ocorrência de pelo menos duas chamadas em $(0,t]$.

Em consequencia disto, teríamos $1 - P_0(t) = 1$ e $1 - P_0(t) - P_1(t) = 1$, para todo $t > 0$ o que contradiz a **hipótese 3**.

Lembre que $P_0(1) = P(\{w \,|\, w_t = 0 \text{ em } [0,1]\})$.

Logo, $0 < P_0(1) < 1$ e definimos $\lambda = -\log P_0(1)$.

Portanto,

$P_0(t) = [P_0(1)]^t = [e^{-\lambda}]^t = e^{-\lambda t}\,, \forall\, t > 0$.

Observe que $P_0(0) = 1$. De fato,

$P_0(0) = P\,(A^0_{0,0}) = P\,(w \in \Omega; w(0) - w(0) = 0) = P\,(\Omega) = 1$.

Sendo assim,

$P_0(t) = e^{-\lambda t}\,, \forall\, t \geq 0$

$P_0(0) = 1\,;\; P_0'(t) = -\lambda e^{-\lambda t} = -\lambda P_0(t), \forall\, t > 0$

B) Mostraremos agora que

$$P_k(s+t) = \sum_{i=0}^{k} P_i(s)\, P_{k-i}(t).$$

Ou seja, uma equação tipo Chapman-Kolmogorov.

Sejam $k \geq 1$, $s \geq 0$ e $t > 0$. O evento “ocorrem k chamadas em $(0, s+t]$” pode ser escrito como

$$A^k_{0,s+t} = (A^0_{0,s} \cap A^k_{s,t}) \cup (A^1_{0,s} \cap A^{k-1}_{s,t}) \cup (A^2_{0,s} \cap A^{k-2}_{s,t}) \cup ... \cup (A^k_{0,s} \cap A^0_{s,t}).$$

Observe que os intervalos $(0,s]$ e $(s, s+t]$ são disjuntos o que garante a independencia dos eventos $A^i_{0,s}$ e $A^{k-i}_{s,t}, 0 \leq i \leq k$, pela **hipótese 2** de incrementos independentes.

Alï¿½m disso, os eventos $A^i_{0,s} \cap A^{k-i}_{s,t}$ são disjuntos em i, isto ï¿½,

$$(A^i_{0,s} \cap A^{k-i}_{s,t}) \cap (A^j_{0,s} \cap A^{k-j}_{s,t}) = \emptyset,$$

se $i \neq j$.

Portanto,

$$P_k(s+t) = P(A^k_{0,s+t}) = \sum_{i=0}^{k} P(A^i_{0,s}) P(A^{k-i}_{s,t}) = \sum_{i=0}^{k} P_i(s) P_{k-i}(t).$$

Sendo assim obtivemos a equação de Chapman-Kolmogorov desejada.

Vamos agora obter um sistema de equações diferenciais para os $P_k(t)$.

Ora,

$$P_k(s+t) = \sum_{i=0}^{k} P_i(s) P_{k-i}(t) =$$

$$= \sum_{i=0}^{k-2} P_i(s) P_{k-i}(t) + P_{k-1}(s) P_1(t) + P_k(s) P_0(t) =$$

$$= \sum_{i=0}^{k-2} P_i(s) P_{k-i}(t) + P_{k-1}(s) P_1(t) + e^{-\lambda t} P_k(s).$$

Agora, queremos determinar a derivada à direita e à esquerda de $P_k(s)$. Sabemos que a derivada à direita de P_k em s é dada por

$$\lim_{t \to 0^+} \frac{P_k(s+t) - P_k(s)}{t} = P'_k(s).$$

Observe que como $e^{-\lambda\, t} = 1 - \lambda\, t + \frac{1}{2!} \lambda^2\, t^2 - ...$, então

1)

$$\lim_{t \to 0} \frac{1 - P_0(t)}{t} = \lim_{t \to 0} \frac{1 - e^{-\lambda t}}{t} = \lambda.$$

2)

$$\lim_{t \to 0} \frac{P_1(t)}{t} = \lim_{t \to 0} \left(\frac{P_1(t)}{1 - P_0(t)} . \frac{1 - P_0(t)}{t} \right) =$$

$$= \lim_{t \to 0} \frac{P_1(t)}{1 - P_0(t)} . \lim_{t \to 0} \frac{1 - P_0(t)}{t} = 1.\lambda = \lambda.$$

3)

$$\lim_{t \to 0} \frac{1 - P_0(t) - P_1(t)}{t} = \lim_{t \to 0} \left(\frac{1 - P_0(t) - P_1(t)}{1 - P_0(t)} . \frac{1 - P_0(t)}{t} \right) = 0.\lambda = 0.$$

Então,

$$P_k'(s) = \lim_{t \to 0} \frac{P_k(s+t) - P_k(s)}{t} =$$

$$= \lim_{t \to 0} \frac{1}{t} \left[\sum_{i=0}^{k-2} P_i(s) P_{k-i}(t) + P_{k-1}(s) P_1(t) + P_k(s)(e^{-\lambda t} - 1) \right] =$$

$$= \lim_{t \to 0} \frac{1}{t} \sum_{i=0}^{k-2} P_i(s) P_{k-i}(t) +$$

$$+ \lim_{t \to 0} \frac{1}{t} P_{k-1}(s) P_1(t) + \lim_{t \to 0} P_k(s) \frac{e^{-\lambda t} - 1}{t} =$$

$$= (\lim_{t \to 0} \frac{1}{t} \sum_{i=0}^{k-2} P_i(s)\, P_{k-i}(t)) \; + \; \lambda P_{k-1}(s) - \lambda P_k(s).$$

Ora,

$$0 \leq \frac{1}{t} \sum_{i=0}^{k-2} P_i(s) P_{k-i}(t) \leq \frac{1}{t} \sum_{i=0}^{k-2} P_{k-i}(t) \stackrel{j=k-i}{=}$$

$$= \frac{1}{t} \sum_{j=2}^{k} P_j(t) = \frac{1}{t} (P_2(t) + P_3(t) + ... + P_k(t)) =$$

$$= \frac{1}{t} (1 - P_0(t) - P_1(t)) \underset{t \to 0}{\longrightarrow} 0.$$

Logo, para todo k

$$P_k'(s) = \lambda P_{k-1}(s) - \lambda P_k(s).$$

Desta forma, $P_k(t)$ satisfaz um sistema de equações diferenciais que é equivalente a $x' = x\mathcal{L}$, onde a matriz $\mathcal{L}$ foi descrita acima.

Uma maneira de resolver diretamente as equações acima (sem passar por exponencial de matriz) é observar que

$$P_k'(t) = \lambda P_{k-1}(t) - \lambda P_k(t),$$

pode ser escrito como

$$P_k'(t) \, + \, \lambda P_k(t) = \lambda P_{k-1}(t),$$

para $k \geq 1$.

Assim, resolvendo a primeira equação ($k = 0$),

$$P_0^{'}(t) \, = \lambda P_0(t),$$

com a condição inicial $P_0(0) = 1$, obtemos

$$P_0\,(t) = e^{\lambda\, t}.$$

Na notação da seção 4.5, estamos analisando acima a equação diferencial

$$x'(t) = \lambda\, x(t)$$

com a condição inicial

$$x(0) = 1,$$

cuja solução é $x(t) = e^{\lambda\, t}$.

A seguir, consideramos $k = 1$

$$P_1^{'}(s) \, + \, \lambda P_1(s) = \lambda P_0(s) = \lambda\, e^{\lambda\, s}.$$

Da expressão anterior (4.2), se obtém a primeira solução, que seria

$$P_1(s) = \lambda\, s e^{-\,\lambda\, s},$$

para todo $s \geq 0$.

Vamos a seguir obter indutivamente as outras soluções $P_k(s)$, $s \geq 0$, do sistema de equações diferenciais descrito no começo desta seção.

Consideramos então $k = 2$, logo

$$P_2'(s) + \lambda P_2(s) = \lambda P_1(s) = \lambda^2 s e^{-\lambda s}.$$

Aplicando o procedimento descrito na seão 4.5 (para a equação linear não homogenea no exemplo 4.14) se obtém que

$$P_2(s) = \frac{(\lambda s)^2}{2!} e^{-\lambda s}.$$

E assim, indutivamente, se obtém que

$$P(X_s = k) = P_k(s) = \frac{(\lambda s)^k}{k!} e^{-\lambda s}.$$

De qualquer forma, em resumo, o Processo de Poisson tem a matriz $\mathcal{L}$ como gerador infinitesimal.

Já havíamos mostrado antes (via exponencial) que, neste caso,

$$P(X_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

Desta forma, para cada t fixo, a função característica de X_t é dada por

$$E[e^{i w X_t}] = e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k e^{i w k}}{k!} = e^{\lambda t (e^{i w} - 1)}.$$

Sendo assim, $E(X_t) = \lambda t$ e $Var(X_t) = \lambda t$.

Exercício: Mostre que a derivada à esquerda de $P_k(s)$ é dada pela mesma expressão acima, utilizando a equação

$$P_k(s) = \sum_{i=0}^{k} P_i(s - t) P_{k-i}(t).$$

Ou seja, neste caso mostre que também vale o sistema de equações diferenciais

$$(*)\begin{cases} P_k'(t) = \lambda P_{k-1}(t) - \lambda P_k(t), \quad k \geq 1 \\ P_0'(t) = -\lambda P_0(t) \end{cases}$$

sujeito às condições iniciais

$$P_k\,(0) = 0\,, k \geq 1 \text{ e } P_0\,(0) = 1.$$

Vamos definir agora os tempos de salto dos caminhos $\omega \in \Omega$. Considere um Processo de Poisson de parâmetro λ.

Cada caminho $\omega = (X_t(\omega))_{t\in\mathbb{R}^+} = (w_t)_{t\in\mathbb{R}^+}$, é constante em conjuntos da forma $[a, b)$ e ainda monótono não decrescente em t. Fixado ω, os tempos de saltos de ω são aqueles t em que

$$\lim_{h\to t,\, h\geq t} w_h = \lim_{h\to t,\, h\leq t} w_h + 1.$$

Definição 4.7. *Seja $T_0(\omega) = 0$, e*

$$T_1(\omega) = \inf\{t \mid X_t(\omega) = 1\},$$

e de forma mais geral,

$$T_n(\omega) = \inf\{t \mid X_t = n\}.$$

Note que $T_n(\omega) < T_{n+1}(\omega)$ para qualquer n. Dizemos que $T_n(\omega)$ é o tempo de n-ésimo salto do caminho ω.

Para cada n, temos que T_n é uma variável aleatória sobre Ω, ou seja, $T_n : \Omega \to \mathbb{R}^+$ é uma função $\mathcal{F}$ mensurável.

Vamos denotar por Y_n a variável aleatória

$$Y_n\,(\omega) = T_n\,(\omega) - T_{n-1}\,(\omega).$$

A função mensurável Y_n mede o tempo entre saltos sucessivos.

Teorema 4.9. *As variáveis aleatórias Y_n, $n \in \mathbb{N}$, são independentes entre si e cada uma delas tem distribuição exponencial com parâmetro λ.*

Demonstração: Primeiro note que $P(Y_1 > t) = P(X_t = 0) = e^{-t\lambda}$.

Logo, Y_1 tem distribuição exponencial.

Ora, sejam t_1 e t fixos

$$P(Y_2 > t \,|\, Y_1 = t_1) \;=\; P(X_{t_1+t} - X_{t_1} = 0 \,|\, Y_1 = t_1) =$$

$$P(X_{t_1+t} - X_{t_1} = 0 \,|\, \{X_{t_1} - X_0 = 1\} \cap_{s<t_1} \{X_s - X_0 = 0\}\,).$$

Ora, a hipótese 2 acima diz que o conjunto $A^0_{t_1,\,t}$ é independente da sigma álgebra gerada por $X_s - X_z$, $s, z \in [0, t_1]$, onde $s > z$.

Logo,

$$P(Y_2 > t \,|\, Y_1 = t_1) \;=\; P(X_{t_1+t} - X_{t_1} = 0) = P(X_t - X_0 = 0) = e^{-t\lambda}.$$

Assim, Y_1 é independente de Y_0 e tem distribuição exponencial.

Na próxima etapa consideramos

$$P\,(\,Y_3 > t \,|\, Y_1 = t_1,\, Y_2 = t_2)$$

e procedemos de maneira similar.

O resultado segue por indução.

$\square$

4.3 Processos de Nascimento e Morte

Uma outra forma de introduzir Cadeias de Markov a tempo contínuo é a seguinte: Suponha X_t, S=$\mathbb{N}$, que define uma probabilidade P sobre a σ-álgebra $\mathcal{F}$ sobre Ω.

Denote $P_{i,j}(t) = P(X(t+s) = j \,|\, X(s) = i)$ para todo $i, j \in \{0, 1, 2, 3, 4, ...\}$. A expressão acima é independente de s pois assumimos que o processo é homogêneo no tempo.

Considere uma sequencia de números reais positivos γ_k, $k \in \mathbb{N}$.

Suponha que o processo, por alguma razão natural, é tal que para todo k, a probabilidade $P_{k,k+1}(h)$ é aproximadamente linear em h com taxa γ_k, para tempos h muito pequenos, ou seja, $P_{k,k+1}(h) \sim \gamma_k\, h$.

Vamos formalizar isto de uma forma adequada.

Seja uma função $r(h)$, dizemos que r tem ordem de h se

$$\lim_{h\to 0} \frac{r(h)}{h} = 0.$$

Em geral vamos pedir só convergencia pela direita de zero, ou seja no limite acima basta considerar $h \geq 0$.

Ou seja, não só $r(h)$ vai a zero quando $h \to 0$, mas até dividido por h, também vai a zero. Diremos neste caso que $r(h)$ é um o(h).

Ainda, $r(h)$ só precisa estar definida para h num intervalo $[0, \epsilon)$, onde $\epsilon > 0$ pode ser pequeno.

Por exemplo, seja f uma função real tal que que $f'(2) = 5$, então, se chamamos de $r(h)$, o erro na aproximação da função por sua aproximação linear, ou seja,

$$f(2+h) - f(2) - 5\,h \,= r(h),$$

então,

$$\lim_{h\to 0} \frac{r(h)}{h} = 0.$$

Logo, tal $r(h)$ é um o(h).

Diremos que $P_{k,k+1}(h) \sim \gamma_k\, h$, se

$$r(h) = P_{k,k+1}(h) \,-\, \gamma_k\, h, \;\; h \geq 0,$$

é um o(h).

Finalmente, descreveremos este fato da seguinte forma

$$P_{k,k+1}(h) \,=\, \gamma_k\, h \,+\, \mathrm{o}\,(h).$$

Vamos assumir no presente exemplo que, da mesma forma como no processo de Poisson, uma vez que ocorre k, com o decorrer do tempo, ou k se mantém ou passa a $k+1$. A seguir, atingido este valor, ele se mantém ou passa a $k+2$, e assim por diante...

Por exemplo, X_t poderia estar descrevendo o número de habitantes numa certa cidade. A cada momento poderia nascer mais uma pessoa. Assumimos também que não haveria mortes no modelo nem abandono da cidade por nenhuma pessoa.

Desta forma, vamos assumir que $P_{k,k}(h) + P_{k,k+1}(h) \,=\, 1 + \mathrm{o}(h)$. Note que, pela homogeneidade no tempo segue que para todo $s \geq 0$ também vale $P_{k,k}(h+s) + P_{k,k+1}(h+s) = 1 + \mathrm{o}(h)$.

Resumindo as hipóteses acima, dizemos que os $P_{i,j}(t)$ satisfazem

a)

$$P_{k,k+1}(h) \,=\, \lambda_k\, h \,+\, \mathrm{o}(h), \qquad \forall\, k \in \mathbb{N},$$

b)

$$P_{k,k}(h) \,=\, 1 \,-\, \lambda_k\, h \,+\, \mathrm{o}(h), \qquad \forall\, k \in \mathbb{N},$$

c)

$$P_{i,j}(h) \,=\, \mathrm{o}(h), \qquad \forall\, i, j \in \mathbb{N},\, j \neq i,\, j \neq i+1.$$

Note que é sempre verdade que $P_{k,k}(0) = 1$, para todo k, e ainda $P_{i,j}(0) = 0$ para $i \neq j$.

Neste caso será válido que

a)

$$\lim_{h \to 0} \frac{P_{k,k+1}(h) \,-\, 0}{h} \,=\, \frac{d\, P_{k,k+1}(t)}{dt}\Big|_{t=0} = \lambda_k, \qquad \forall\, k \in \mathbb{N}.$$

b)

$$\lim_{h\to 0} \frac{P_{k,k}(h) \, - \, 1}{h} \; = \; \frac{d\, P_{k,k}(t)}{dt}|_{t=0} \, = - \, \lambda_k, \qquad \forall\, k \in \mathbb{N}.$$

c)

$$\lim_{h\to 0} \frac{P_{i,j}(h) \, - \, 0}{h} \; = \; \frac{d\, P_{i,j}(t)}{dt}|_{t=0} = 0, \qquad \forall\, i, j \in \mathbb{N},\, j \neq i,\, j \neq i+1.$$

Desta forma, o gerador infinitesimal do semigrupo $\mathcal{P}^t$ associado ao Processo de Markov X_t, $t \in \mathbb{R}^+$, seria a matriz $\mathcal{L}$ dada por

$$\mathcal{L} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & 0 & ..... & ... & .. & ... \\ 0 & -\lambda_1 & \lambda_1 & 0 & 0 & 0 & ..... & ... & .. & ... \\ 0 & 0 & -\lambda_2 & \lambda_2 & 0 & 0 & ..... & ... & .. & ... \\ 0 & 0 & 0 & -\lambda_3 & \lambda_3 & 0 & ..... & ... & .. & ... \\ . & . & . & . & . & . & . & .... & .. & ... \\ 0 & 0 & 0 & ...... & 0 & 0 & -\lambda_k & \lambda_k & 0 & ... \\ . & . & . & . & . & . & . & .... & .. & ... \end{pmatrix}.$$

Um processo com estas características é denominado de Processo de Nascimento.

Esta classe de Processos Estocásticos inclui como caso particular o Processo de Poisson. Poderíamos considerar, por exemplo, o caso de uma central que recebe telefonemas. Este modelo contempla a situação em que a probabilidade de novas futuras chamadas depende de quantas chamadas, no caso k, já foram recebidas até o presente momento.

Como ilustração vamos descrever o Processo de Yule. Assuma que numa certa população cada membro individual tem uma probabilidade $\alpha\, h +$ o(h) de produzir um novo membro.

É natural então supor que para todo $t \geq 0$ e $k \geq 0$ vale

$$P(X_{t+h} - X_t \;= 1 \,|\, X_t = k) = k\, \alpha\, h + \; \text{o}\,(h).$$

Ou seja,

$$P_{k,k+1}(h) \,=\, k\,\alpha\,h + \;\mathrm{o}\,(h).$$

Neste caso, $\gamma_k \,=\, \alpha\,k$, $k \geq 0$.

Note que $P(X_t \,=\, 0\,|\,X_0 \,=\, 0) \,=\, 1$. Ou seja, se a população era zero, ninguém vai nascer.

Vamos supor primeiro que a população inicial era igual a 1, ou seja, $P(X_0 = 1) = 1$.

Note que neste modelo foi natural descrever o modelo pelas probabilidades de transição para h pequeno. Para calcular $P(X_t \,=\, k) \,=\, x_k(t)$, devemos resolver o sistema

$$x'\,(t) \,=\, x(t)\,\mathcal{L},$$

onde

$$x(t) = (x_0(t), x_1(t), x_2(t), x_3(t), ..., x_k(t), ...),$$

associado a matriz infinitesimal $\mathcal{L}$, ou seja, para todo $k \geq 0$

$$x_k^{'}(t) \,=\, -\,\alpha\,(\,k\,x_k(t) - (k-1)\,x_{k-1}(t)\,),$$

com $x(0) = (0, 1, 0, ...)$.

Note que $x_0(t) = 0$ para todo t.

Logo, $x_1'(t) = -\alpha\,x_1(t)$, e assim $x_1(t) = e^{-\alpha\,t}$.

A segunda equação torna-se

$$x_2^{'}(t) \,=\, -\,\alpha\,(\,2\,x_2(t) - e^{-\alpha\,t}\,) \,=\, -\,\alpha\;2\,x_2(t) + \,\alpha\,e^{-\alpha\,t}\,.$$

Temos ainda a condição inicial $x_2(0) = 0$.

Esta equação é linear de primeira ordem não homogênea (cuja solução geral aparece no exemplo 4.14 da seção 4.5). A solução $x_2(t)$ pode ser facilmente calculada, e se obtém

$$x_2(t) = e^{-\alpha t}\,(1 - e^{-\alpha\,t}).$$

A seguir se deve considerar

$$x_3'(t) = -\alpha\,(3\,x_3(t) - 2\,x_2(t)) = -\alpha\,3\,x_3(t) + 2\,\alpha\,x_2(t),$$

onde $x_2(t)$ já é conhecida.

De novo temos que $x_3(t)$ satisfaz uma equação linear de primeira ordem não homogênea, e assim pode ser explicitamente obtida. Procedendo de maneira indutiva se mostra que para $k \geq 1$, vale que

$$P(X_t = k) = x_k\,(t) = e^{-\alpha t}\,(1 - e^{-\alpha\,t})^{k-1}.$$

Vamos calcular, para t fixo, a função geradora de momentos

$$f_t\,(s) = f(s) = \sum_{k=1}^{\infty} P(X_t = k)\,s^k = s\,e^{-\alpha t} \sum_{k=1}^{\infty} [\,(1 - e^{-\alpha\,t})\,s]^{k-1} =$$

$$= \frac{s\,e^{-\alpha\,t}}{1 - (1 - e^{-\alpha\,t})\,s}.$$

Consideraremos agora o processo em que $P(X_0 = N) = 1$, ou seja, iremos supor que no tempo $t = 0$, a população é igual a N.

Supondo ainda que não exista interação entre os membros da população, e que no tocante a geração de descendentes, eles hajam de forma independente, podemos supor que esta população evolui como a soma de N processos independentes tipo Yule em que $N = 1$.

Sendo assim, podemos facilmente calcular a nova função geradora. De fato, se

$$P_{N,k}(t) = P(X_t(w) = k \,|\, X(0) = N),$$

e denotamos por

$$f_{N,t}\,(s) = f_N(s) = \sum_{k=N}^{\infty} P_{N,k}(t)\,s^k,$$

então, em função da independência (ver também Seção 3.4) temos que

$$f_N(s) = f_1(s)^N = f(s)^N = \left(\frac{s\, e^{-\alpha t}}{1 - (1 - e^{-\alpha t})\, s}\right)^N =$$

$$= \sum_{k=N}^{\infty} C_{k-N}^{k-1}\, (e^{-t\alpha})^N\, (1 - e^{-t\alpha})^{k-N}\, s^k,$$

onde C_{k-N}^{k-1} denota combinação de $(k-N)$, de $(k-1)$ a $(k-1)$.

Obtemos assim que para cada valor t e k

$$P(X_t(w) = k \,|\, X(0) = N) = C_{k-N}^{k-1}\, (e^{-t\alpha})^N\, (1 - e^{-t\alpha})^{k-N},$$

onde $k \geq N$.

Para um t fixo se quiséssemos calcular a variância bastaria derivar a função $f_{N,t}\,(s)$ no ponto um, etc... Ou seja, neste caso conseguimos calcular as informações mais importantes do modelo de maneira explícita.

Vamos agora considerar processos de nascimento e morte. Suponha fixadas as sequencias $\lambda_k \geq 0$ e $\beta_k \geq 0$, onde $k \in \mathbb{N}$. Assumimos que $\beta_0 = 0$.

Considere um Processo Markoviano X_t, $t \in \mathbb{N}$, sobre o conjunto de estados $S = \mathbb{N} = \{0, 1, 2, 3, ..., k, ...\}$, e denote $P_{i,j}(t) = P(X(t+s) = j \,|\, X(s) = i)$ para todo $i, j \in \{0, 1, 2, 3, 4, ...\}$

Assuma que os $P_{i,j}(t)$ satisfazem

a)

$$P_{k,k+1}(h) = \lambda_k\, h + \mathrm{o}(h), \qquad \forall\, k \in \mathbb{N},$$

b)

$$P_{k,k-1}(h) = \beta_k\, h + \mathrm{o}(h), \qquad \forall\, k \in \mathbb{N},$$

c)

$$P_{k,k}(h) = 1 - (\lambda_k + \beta_k)\, h + \mathrm{o}(h), \qquad \forall\, k \in \mathbb{N},$$

c)

$$P_{i,j}(h) = \mathrm{o}(h), \qquad \forall\, i, j \in \mathbb{N},\; j \neq i,\; j \neq i-1,\; j \neq i+1.$$

Como sempre, $P_{k,k}(0) = 1$ e $P_{i,j}(0) = 0$ se $i \neq j$.

Lembre que $\beta_0 = 0$.

Ou seja, no presente modelo podemos pensar que X_t descreve o número de habitantes de uma população em que existe uma probabilidade $\lambda_k\, h$, de nascimento, para h pequeno, e uma probabilidade $\beta_k\, h$, de morte, para h pequeno, dado que a população no tempo 0 é k.

O sistema de equações associado seria: para todo $t \geq 0$,

a)

$$\frac{d\, P_{0,k}(t)}{dt} \;=\; -\,\lambda_0\, P_{0,k}(t) \;+\; \lambda_0\, P_{1,k}(t), \qquad \forall\, k \in \mathbb{N}.$$

b)

$$\frac{d\, P_{i,k}(t)}{dt} \;= \beta_k\, P_{i-1,k}\,(t) \;-\; (\lambda_k + \beta_k)\, P_{i,k}(t) + \lambda_k\, P_{i+1,k}(t), \;\; \forall\, k \geq 1.$$

Calculando a derivada acima em $t = 0$ podemos obter o gerador infinitesimal do semigrupo $\mathcal{P}^t$ associado ao Processo de Markov X_t, $t \in \mathbb{R}^+$, através da matriz $\mathcal{L}$ dada por

$$\mathcal{L} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & 0 & ... \\ \beta_1 & -(\lambda_1+\beta_1) & \lambda_1 & 0 & 0 & 0 & .... \\ 0 & \beta_2 & -(\lambda_2+\beta_2) & \lambda_2 & 0 & 0 & .... \\ 0 & 0 & \beta_3 & -(\lambda_3+\beta_3) & \lambda_3 & 0 & ... \\ . & . & . & . & . & . & ... \\ . & . & . & . & . & . & ... \end{pmatrix}.$$

De fato, os elementos da matriz acima são obtidos como

$$a_{i,k} = \frac{d\, P_{i,k}(t)}{dt}|_{t=0} \;= \beta_k\, P_{i-1,k}\,(0) \;-\; (\lambda_k + \beta_k)\, P_{i,k}(0) + \lambda_k\, P_{i+1,k}(0), \;\; \forall k \geq 1.$$

O resultado segue de que $P_{i-1,k}(0) = 0$, a menos que $k = i-1$ e $P_{i+1,k}(0) = 0$, a menos que $k = i+1$, etc...

Um processo com estas características é denominado de Processo de Nascimento e Morte. Ele generaliza o Processo de Nascimento.

O sistema de equações acima traduz

$$\frac{d\,\mathcal{P}^t}{dt} = \mathcal{L}\,\mathcal{P}^t.$$

onde $(\mathcal{P}^t)_{i,j} = (P(X(t) = j \,|\, X(0) = i)$.

Este sistema é o sistema de equações diferenciais para frente de Chapman-Kolmogorov deste processo.

De maneira equivalente, os $(P(X(t) = j \,|\, X(0) = i)$, poderiam também ser obtidos resolvendo

$$\frac{d\,\mathcal{P}^t}{dt} = \mathcal{P}^t\,\mathcal{L},$$

Este é o sistema de equações diferenciais para trás de Chapman-Kolmogorov associado ao processo em consideração.

É fácil ver que no presente caso, as equações para trás se traduziriam como

a')

$$\frac{d\,P_{i,0}(t)}{dt} = -\,\lambda_0\,P_{i,0}(t) + \beta_1\,P_{i,1}(t), \quad \forall k \in \mathbb{N}.$$

b')

$$\frac{d\,P_{i,j}(t)}{dt} = \lambda_{j-1}\,P_{i,j-1}\,(t) - (\lambda_j + \beta_j)\,P_{i,j}(t) + \beta_{j+1}\,P_{i,j+1}(t), \;\; \forall j \geq 1.$$

Vamos analisar um exemplo particular interessante de um processo que possui as seguintes características: seja $\lambda > 0$, $\beta > 0$ e $\alpha > 0$, então usando a notação acima

$$\lambda_k = k\,\lambda + \alpha, \quad \text{e} \quad \beta_k = k\,\beta.$$

Ou seja, assuma que os $P_{i,j}(t)$ satisfazem

a)

$$P_{k,k+1}(h) = (\lambda\,k + \alpha)\,h + \mathrm{o}(h), \quad \forall\,k \in \mathbb{N},$$

b)
$$P_{k,k-1}(h) \,=\, \beta\, k\, h \,+\, \mathrm{o}(h), \qquad \forall\, k \in \mathbb{N},$$

c)
$$P_{k,k}(h) \,=\, 1 - \,\, (\lambda\, k \,+\, \alpha \,+\, \beta k)\, h \,+\, \mathrm{o}(h), \qquad \forall\, k \in \mathbb{N},$$

c)

$$P_{i,j}(h) \,=\, \mathrm{o}(h), \qquad \forall\, i, j \in \mathbb{N},\; j \neq i,\; j \neq i-1,\; j \neq i+1.$$

Este modelo descreveria uma situação semelhante ao Processo de Yule em que cada indivíduo poderia gerar um descendente (nos daria assim um termo $P_{k,k+1}(h) \sim \lambda\, k\, h$) mas consideramos agora a possibilidade de uma taxa fixa de imigração α linear em h, o que nos dá na verdade, a expressão $P_{k,k+1}(h) \sim (\lambda\, k \,+\, \alpha)\, h$.

Neste modelo assumimos ainda que cada indivíduo tem uma probabilidade de morrer com uma taxa β linear em h, e assim aparece o termo $P_{k,k-1}(h) \sim \beta\, k\, h$.

A maneira de calcular as probabilidades $P_{i,j}(h)$ é mais difícil que nos casos anteriores. Vamos calcular, para cada t o valor esperado da variável X_t de uma forma interessante.

Considere i fixo e o processo X_t, $t \geq 0$, tal que $P(X(0) = i) = 1$. Denote para cada valor $t \geq 0$,

$$E(X(t)) = g_i(t) = g(t) = \sum_{j=1}^{\infty} j\, P_{i,j}(t).$$

Vamos usar a seguir as equações diferenciais para trás, dadas por, a') e b') acima.

Ora,
$$g'(t) = \sum_{j=1}^{\infty} j \;\; \frac{d\, P_{i,j}(t)}{dt} =$$

$$= \sum_{j=1}^{\infty} j \left[\left(\lambda (j-1) + \alpha \right) P_{i,j-1}(t) - (\lambda j + \alpha + \beta j) P_{i,j}(t) + \beta (j+1) P_{i,j+1}(t) \right] =$$

$$= (\lambda - \beta)\, g(t) + \alpha.$$

A razão pela preferencia da equação de Kolmogorov para trás é que neste caso, como pode ser visto acima, aparecem termos $P_{i,j-1}$, $P_{i,j}\, P_{i,j+1}$ sempre com i à esquerda.

A solução $g(t)$ da equação diferencial acima, quando $\lambda \neq \beta$ é

$$g(t) = \frac{\alpha}{\lambda - \beta} (e^{t(\lambda-\beta)} - 1) + i\, e^{t(\lambda-\beta)}.$$

Ainda, $g(t)$ (solução) é tal que, quando $\lambda = \beta$, então

$$g(t) = \alpha\, t + i.$$

Note que se $(\lambda - \beta) > 0$, ou seja, maior probabilidade de nascimento do que de morte, por indivíduo, então

$$\lim_{t \to \infty} E[X(t)] = \lim_{t \to \infty} g(t) = +\infty.$$

Se $(\lambda - \beta) < 0$, ou seja, maior probabilidade de morte do que de nascimento, por indivíduo, então

$$\lim_{t \to \infty} E[X(t)] = \lim_{t \to \infty} g(t) = \frac{\alpha}{\beta - \lambda}.$$

Vamos calcular a função geradora para X_t, $t \in \mathbb{R}^+$ do exemplo acima no caso em que $\alpha = 0$, ou seja, sem imigração.

Assuma que i está fixo, $P(X(0) = i) = 1$, e desejamos calcular $p_j(t) = P(X(t) = j)$, para todo $j \in \mathbb{N}$.

Desta forma, $p_j(0) = 0$, para $j \neq i$, e $p_i(0) = 1$.

Seja $p(t) = (p_0(t), p_1(t), p_2(t), ...)$.

A equação dos $p_j(t)$ é dada a partir da equação $p'(t) = p(t)\,\mathcal{L}$ e onde $p(0)$ é o vetor de probabilidade que é zero em todas as coordenadas menos na coordenada i.

A partir disto obtemos que $p(t)$ satisfaz

$$p_0'(t) = \beta p_1(t),$$

e para $j \geq 1$

$$p_j^{'}(t) \,=\, (\,\lambda\,(j-1)+\alpha\,)\;p_{j-1}\,(t)\,-\,(\lambda\,+\,\beta\,)\,j\,p_j(t)\,+\,\beta\,(j+1)\,p_{j+1}(t).$$

Considere

$$G(s,t) = \sum_{j=0}^{\infty} s^j\, p_j(t).$$

Sendo assim, utilizando a expressão para $p_j^{'}(t)$ descrita acima

$$\frac{\partial G(s,t)}{\partial t} = \sum_{j=0}^{\infty} s^j\, p_j^{'}(t) =$$

$$= \lambda\, s^2 \sum_{j=0}^{\infty} (j-1) s^{j-2} p_{j-1}(t) \,-\, (\lambda+\beta)\, s \sum_{j=0}^{\infty} j\, s^{j-1}\, p_j(t) +$$

$$+\beta \sum_{j=0}^{\infty} (j+1) s^j p_{j+1}(t).$$

Note que

$$\frac{\partial G(s,t)}{\partial s} = \sum_{j=0}^{\infty} (j+1) s^j p_{j+1}(t).$$

Logo, a equação acima pode ser descrita como

$$\frac{\partial G(s,t)}{\partial t} \,=\, \lambda\, s^2\, \frac{\partial G(s,t)}{\partial s} \,-\, (\lambda+\beta)\, s\, \frac{\partial G(s,t)}{\partial s} \,+\, \beta\, \frac{\partial G(s,t)}{\partial s},$$

ou seja,

$$\frac{\partial G(s,t)}{\partial t} = (s-1)\,(\lambda\, s - \beta)\,\frac{\partial G(s,t)}{\partial s},$$

com a condição de fronteira $G(s,0) = s^i$ (pois, $p_j(0) = 0$, para $j \neq i$, e $p_i(0) = 1$)

Esta equação diferencial parcial pode ser facilmente calculada (via o método descrito no fim da Seção 4.5) e se obtém que

$$G(s,t) = \Big(\frac{\lambda\, t\,(1-s)\, + s}{\lambda\, t\,(1-s)\, + 1}\Big)^t, \quad \text{se } \beta = \lambda,$$

$$G(s,t) = \Big(\frac{\beta\,(1-s)\, - (\beta - \lambda\, s)\, e^{-t\,(\lambda-\beta)}}{\lambda\,(1-s)\, - (\beta - \lambda\, s)\, e^{-t\,(\lambda-\beta)}}\Big)^t, \quad \text{se } \beta \neq \lambda.$$

Desta forma podemos obter explicitamente todos os $p_j(t) = P(X(t) = j)$ a partir de $G(s,t)$ como descrito na Seção 4.5.

Note que $G(0,t) = p_0(t)$, que no caso $\beta = \lambda$ obtemos

$$p_0(t) = \Big(\frac{\lambda\, t}{\lambda\, t\, + 1}\Big)^t.$$

Desta forma, quando $t \to \infty$, temos que

$$\lim_{t\to\infty} p_0\,(t) = 1,$$

o que nos diz, neste caso que a probabilidade de que a população se extingua vai a 1 quando o tempo vai a ∞.

A probabilidade de extinção nos outros casos (em que $\beta \neq \alpha$) podem ser analisados de forma semelhante.

Em alguns casos as contas envolvidas na solução dos sistemas de equações diferenciais de Kolomogorov são longas, mas muitas vezes encontrar o vetor de probabilidade invariante π não é tão difícil.

É importante destacar que não existem vetores π tal que $\pi\, L = 0$, no caso do processo de Poisson ou seus assemelhados (com os $\lambda_k \geq 0$), mas quando consideramos processos de nascimento e morte (com $\lambda_k \geq 0$ e $\beta_k \geq 0$), algumas vezes este problema pode ser resolvido.

Heuristicamente falando, a razão desta dicotomia é que quando só existe nascimento, não existe um fator que impeça a “massa” da probabilidade ir embora para $+\infty$. Os β_k, por sua vez, contrabalanceiam esta tendencia reestabelecendo uma maior probabilidade de que X_t assuma valores menores.

Seja $\pi = (\pi_0, \pi_1, \pi_2, \pi_3, ...)$, então $\pi\, \mathcal{L} = 0$ significa

$$(0, 0, 0, ...) =$$

$$= (\pi_0\, \pi_1\, \pi_2 \, ...) \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & 0 & ... \\ \beta_1 & -(\lambda_1 + \beta_1) & \lambda_1 & 0 & 0 & 0 & .... \\ 0 & \beta_2 & -(\lambda_2 + \beta_2) & \lambda_2 & 0 & 0 & .... \\ 0 & 0 & \beta_3 & -(\lambda_3 + \beta_3) & \lambda_3 & 0 & ... \\ . & . & . & . & . & . & ... \\ . & . & . & . & . & . & ... \end{pmatrix},$$

logo, a primeira linha nos dá

$$-\lambda_0\, \pi_0 + \beta_1\, \pi_1 \;=\; 0.$$

A segunda, nos dá,

$$\lambda_0\, \pi_0 - (\lambda_1 + \beta_1)\, \pi_1 + \beta_2\, \pi_2 \;=\; 0,$$

e assim por diante, a k-ésima linha nos dá

$$\lambda_{k-1}\, \pi_{k-1} - (\lambda_k + \beta_k)\, \pi_k + \beta_{k+1}\, \pi_{k+1} = 0\,.$$

Da equação da primeira linha resulta

$$\pi_1 = \frac{\lambda_0}{\beta_1}\, \pi_0.$$

Da segunda equação temos que

$$\beta_2\,\pi_2 \,=\, -\lambda_0\,\pi_0 + \lambda_1\,\pi_1 + \beta_1\pi_1 \,=$$

$$-\lambda_0\,\pi_0 + \lambda_1\,\frac{\lambda_0}{\beta_1}\,\pi_0 + \beta_1\frac{\lambda_0}{\beta_1}\,\pi_0 \,=$$

$$-\lambda_0\,\pi_0 + \,\frac{\lambda_0\,\lambda_1}{\beta_1}\,\pi_0 + \lambda_0\,\pi_0 \,=\, \frac{\lambda_0\,\lambda_1}{\beta_1}\,\pi_0.$$

Portanto,

$$\pi_2 \,=\, \frac{\lambda_0\,\lambda_1}{\beta_1\beta_2}\,\pi_0.$$

Procedendo de maneira indutiva, sempre substituindo a expressão obtida numa certa etapa na seguinte, obtemos que para todo $k \geq 1$ vale

$$\pi_k \,=\, \frac{\lambda_0\,\lambda_1\,...\,\lambda_{k-1}}{\beta_1\beta_2\,...\,\beta_k}\,\pi_0.$$

O ponto agora é que π deve ser um vetor de probabilidade e assim é natural escolher π_0 tal que

$$1 \,=\, \sum_{k=0}^{\infty}\,\pi_k \,=\, \pi_0\,(1 + \sum_{k=1}^{\infty}\,\frac{\lambda_0\,\lambda_1\,...\,\lambda_{k-1}}{\beta_1\beta_2\,...\,\beta_k})\,.$$

Logo, bastaria tomar

$$\pi_0 \,=\, \frac{1}{1 + \sum_{k=1}^{\infty}\,\frac{\lambda_0\,\lambda_1\,...\,\lambda_{k-1}}{\beta_1\beta_2\,...\,\beta_k}}.$$

O procedimento faz completo sentido contanto que

$$\sum_{k=1}^{\infty}\,\frac{\lambda_0\,\lambda_1\,...\,\lambda_{k-1}}{\beta_1\beta_2\,...\,\beta_k}\,<\,\infty.$$

Se a probabilidade cumulativa dada pela sequencia de λ_j for maior do que aquela dos β_j então esta série diverge.

No caso do somatório acima convergir então existe vetor π estacionário para o processo cujo gerador infinitesimal é $\mathcal{L}$.

Um exemplo em que isto ocorre é quando existe $\lambda > 0$ e $\beta > 0$ tal que todos os $\lambda_k = \lambda$ e todos os $\beta_k = \beta$, e ainda $\beta > \lambda$.

Neste caso,

$$1 + \sum_{k=1}^{\infty} \frac{\lambda_0\,\lambda_1 \ldots \lambda_{k-1}}{\beta_1 \beta_2 \ldots \beta_k} = 1 + \sum_{k=1}^{\infty} \frac{\lambda^k}{\beta^k} < \infty,$$

pois uma progressão geométrica de razão $\frac{\lambda}{\beta}$ converge.

4.4 Estados Recorrentes e Cadeias Irredutíveis

Nesta seção vamos descrever brevemente, sem apresentar as provas dos teoremas, algumas das propriedades básicas de Cadeias de Markov X_t, $t \in \mathbb{R}^+$, tomando valores em S. Vamos supor abaixo que S é finito, embora muitos resultados sejam verdadeiros para S enumerável. Algumas das demonstrações dos teoremas abaixo são semelhantes ao do caso com tempo discreto, outras, no entanto, são bem mais complexas.

Como sempre Ω é o subconjunto de $S^{\mathbb{R}^+}$ que contém apenas os caminhos contínuos a direita, etc... Vamos supor que $\mathcal{P}_{i,j}^t$ indica a probabilidade de passar de i a j em tempo t, e que este semigrupo descreva o Processo Estocástico.

Uma otima referencia no assunto Cadeias de Markov a tempo continuo sao as notas [Lal1].

Definição 4.8. *Dizemos que $i \in S$ conduz a $j \in S$, se existe $t > 0$ tal que $\mathcal{P}_{i,j}^t > 0$. Neste caso usamos a notação $i \to j$. Dizemos que "i e j se comunicam", se $i \to j$ e $j \to i$ (os tempos t de passagem podem ser diferentes). Denotamos tal fato por $i \leftrightarrow j$.*

Teorema 4.10. *Seja X_t Cadeia de Markov a tempo contínuo sobre S. Então,*

a) Se $i \to j$ e $j \to k$, então $i \to k$;

b) A relação $i \leftrightarrow j$ é uma relação de equivalência e divide o espaço S em classes de equivalência.

Os conceitos e resultados para tais tipos de cadeias X_t a tempo contínuo são basicamente os mesmos que o do caso com tempo discreto. No entanto, não faz sentido o conceito de periodicidade ou aperiodicidade.

Definição 4.9. *Dizemos que a cadeia X_t, $t \in \mathbb{R}^+$, é irredutível se só existe uma classe de equivalencia.*

Note que se o gerador infinitesimal L de um processo X_t for tal que a matriz $\mathcal{L}$ possa ser escrita como

$$\mathcal{L} = \begin{pmatrix} \mathcal{L}_1 & 0 \\ 0 & \mathcal{L}_2 \end{pmatrix}.$$

Então X_t, que é descrito pelo semigrupo $\mathcal{P}^t = e^{t\mathcal{L}}$, não é irredutível.

Definição 4.10. *Um estado $i \in S$ é dito recorrente se*

$$P(\sup\{t \,|\, X_t = i\} \, = \infty \;|\; X_0 = i) = 1.$$

Definição 4.11. *Um estado $i \in S$ é dito transiente se*

$$P(\sup\{t \,|\, X_t = i\} \, < \infty \;|\; X_0 = i) = 1.$$

Definição 4.12. *Fixado i e $\omega = (w_t)_{t\in\mathbb{R}^+}$, tal que $w_0 = i$, dizemos que o tempo de salto do estado i, para o caminho ω, é o número real*

$$T_0^i(\omega) \, = \, \inf\{s > 0 \,|\, X_s(\omega) \, \neq \, i\}.$$

Como os caminhos ω em Ω que consideramos são sempre contínuos a direita, temos que $T_0(\omega) > 0$ (sob a hipótese de $w_0 = i$).

Definição 4.13. *Fixado i e $\omega = (w_t)_{t\in\mathbb{R}^+}$, tal que $w_0 = i$, dizemos que o tempo de primeiro retorno ao estado i, para o caminho ω, é o número real*

$$T_1^i(\omega) \,=\, \inf\{t > T_0^i\,(\omega)\,|\,X_t\,(\omega) = i\}.$$

O valor $T_1^i(\omega)$ poderia ser infinito mas é sempre positivo.

Definição 4.14. *Fixado i, a probabilidade de retorno em tempo finito a i, começando em i, é o valor*

$$\eta_i \,=\, P(T_1 \,< \infty\,|\,X_0 = i).$$

Teorema 4.11. *Seja X_t Cadeia de Markov a tempo contínuo sobre S. Então,*

a) i é recorrente, se e só se, $\eta_i = 1$;

b) se i é recorrente e $j \leftrightarrow i$, então, j é recorrente.

Definição 4.15. *Um estado $i \in S$ é dito positivo recorrente se,*

$$E[\,T_1^i\;\,|\,X_0 = i] < \infty.$$

Teorema 4.12. *Suponha que a Cadeia de Markov X_t seja irredutível e que todo estado i seja recorrente positivo, então, para qualquer $i, j \in S$, vale que*

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t \mathcal{P}^s_{i,j}\;ds \,=\, \frac{1}{E[T_1^j\,|\,X_0 = i]}.$$

Neste caso, se denotamos por $\pi_i = \frac{1}{E[T_1^j\,|\,X_0=i]}$, $i \in S$, então

$$\pi = (\pi_1, \pi_2, ..., \pi_{\#S}),$$

é um vetor de probabilidade invariante para o semigrupo $\mathcal{P}^t$ e é único vetor invariante.

Para concluir, algumas considerações de natureza geral.

Se todas as entradas da matriz $\mathcal{L}$, da forma n por n, forem não nulas (assim, estritamente negativas na diagonal e estritamente positivas fora da diagonal), então o estado estacionário π, ou seja, tal que satisfaz $\pi\,\mathcal{L} = 0$, é único [N].

Destacamos o fato que em Ω pode ser definida uma métrica d que torna Ω completo e separável (ver [EK] Seção 3.5). Tal d faz com que todos os cilindros sejam conjuntos abertos. Ω no entanto não é compacto mesmo quando S é finito. A métrica d é chamada de métrica de Skhorohod.

A σ-álgebra natural para considerar o Processo de Markov X_t a tempo contínuo é a σ-álgebra de Borel $\mathcal{B}$ sobre Ω, ou seja, a σ-álgebra gerada pelos abertos obtidos por tal métrica d.

Para cada $s \in \mathbb{R}^+ \cup \{0\}$ podemos definir a transformação $\mathcal{B}$-mensurável $\Theta_s : \Omega \to \Omega$ dada por $\Theta_s(w_t) = w_{t+s}$. Para s fixo, Θ_s muda a contagem do “tempo do relógio” do processo. De outra forma, cada caminho amostral $\omega = (w_t)_{t\in T} \in \Omega$ é levado por Θ_s num outro caminho ω_s que começa a contar tempo 0 no tempo que era s para ω. Esta família de tranformações Θ_s é denominada de shift em $\mathbb{R}^+$. Ela desempenha um papel análogo ao shift do caso de tempo discreto que foi considerado anteriormente.

Uma probabilidade P sobre Ω é dita invariante se, para qualquer $s \geq 0$, vale que $P(\,\Theta_s^{-1}(A)\,) = P(A)$, para qualquer $A \in \mathcal{B}$. Da mesma forma como antes, o Processo de Markov X_t é estacionário, se e só se, a P associada é invariante. Pode-se falar em Processo Ergódico, Teorema de Birkhof, etc. Se todas as entradas da matriz $\mathcal{L}$, da forma n por n, forem não nulas, então o processo estacionário associado a tempo contínuo será ergódico (ver [N]). Estes tópicos estão acima do escopo do presente texto (ver [S], [EK], [N]).

Uma otima referencia para as questoes ergodicas sao as notas [Lal1].

4.5 Apêndice - Breve Introdução às Equações Diferenciais

Vamos inicialmente tentar descrever da maneira mais simples possível o que é uma equação diferencial. Para um estudo mais profundo do assunto recomendamos ao leitor o texto [DL].

Uma equação diferencial é um expressão que envolve a derivada $x'(t)$ de uma função $x(t)$, e o nosso objetivo é tentar descobrir quem é tal função $x(t)$. Por exemplo, considere a expressão

$$x^{'}(t) \,=\, 5\,x(t).$$

Perguntamos: quem é a função $x(t)$ (onde $x:\mathbb{R}\to\mathbb{R}$) que satisfaz a equação acima para todo t real? A resposta é bem simples: seja $x(t)=e^{5t}$, então para "todo" valor real t,

$$x^{'}(t) \,=\, 5\,e^{5t} \,=\, 5\,x(t).$$

Portanto, a resposta que buscávamos é $x(t)=e^{5t}$. Note que desejamos que a equação seja verdadeira para todos os valores de t e não apenas para um valor específico de t.

Exercício: Mostre que $x(t) = e^{-7t}$ satisfaz a equação diferencial $x^{'}(t) = -7\,x(t)$. Neste caso, quem é $f(x)$?

Definição 4.16. *Mais geralmente, uma equação diferencial é uma expressão da forma*

$$x'(t) = f(x(t)),$$

onde $f:\mathbb{R}\to\mathbb{R}$ está fixada. A incógnita é a função $x(t)$.

No caso $x'(t) = 5\,x(t)$, temos que $f(x) = 5\,x$.

No caso da equação $x'(t) = \cos(x(t)) + x(t)^2$, temos que $f(x) = \cos(x) + x^2$.

Exercício: Mostre que $x(t) = 3\,e^{7t}$ satisfaz a equação diferencial $x^{'}(t) = 7x(t)$.

Para determinar $x(t)$ de maneira única é necessário fixar a condição inicial, isto é, estabelecer que $x'(t_0) = x_0$.

Exemplo 4.11. Vamos dar um exemplo: para qualquer constante k, é fácil ver que $x(t) = k\,e^{7\,t}$, satisfaz

$$x'(t) \;= 7\,x(t).$$

Diremos que $x(t) \;=\; ke^{7\,t}$ é a solução geral de $x'(t) \;=\; 7\,x(t)$.

Neste caso, a função $x(t)$ está determinada a menos da constante k.

Estabelecendo que $x(0) = 10$, então

$$10 \;=\; x(0) \;=\; k\,e^{7\,0} = k.$$

Logo, $k = 10$ e assim a função $x(t) = 10\,e^{7\,t}$, fica determinada de maneira única.

$\diamondsuit$

Diremos que tal $x(t) \;=\; 10\,e^{7\,t}$, é uma solução particular de $x'(t) \;=\; 7\,x(t)$.

Exemplo 4.12. Como vimos para qualquer constante k, temos que $x(t) = k\,e^{7\,t}$, satisfaz

$$x'(t) \;= 7\,x(t).$$

Assumindo que $x(1,2) = 3$, temos que

$$3 \;=\; x(1,2) \;=\; k\,e^{7\,\times\,1,2}.$$

Deste modo, $k = \frac{3}{e^{8,4}}$ e assim a função $x(t) = \frac{3}{e^{8,4}}\,e^{7\,t}$, fica determinada de maneira única.

$\diamondsuit$

Exemplo 4.13. Considere fixada uma constante a, então a solução geral de

$$x'(t) = a\,x(t)$$

é

$$x(t) = e^{a\,t}\,k.$$

Fixada a condição inicial, $x(0) = x_0 \in \mathbb{R}$, a solução $x(t)$ é única e dada por

$$x(t) = e^{a\,t}\,x_0.$$

$\diamondsuit$

Exercício: Mostre que $x(t) = t + s$, onde s é constante, satisfaz a equação diferencial $x^{'}(t) = 1$, com a condição inicial $x(0) = s$.

Exemplo 4.14. Para a e b fixos, seja $x(t) = c\,e^{b\,t} - \frac{a}{b}$, onde c denota uma constante. Então, tal $x(t)$ é a solução geral da equação $x'(t) = b\,x(t) + a$. Isto porque para tal $x(t)$

$$x'(t) = c\,b\,e^{b\,t} = b\,(ce^{b\,t} - \frac{a}{b}) + a = b\,x(t) + a.$$

Neste caso $f(x) = b\,x + a$, onde a é constante.

Em resumo, a solução geral $x(t)$ de

$$x'(t) - b\,x(t) = a$$

é $x(t) = c\,e^{b\,t} - \frac{a}{b}$.

$\diamondsuit$

Exemplo 4.15. Vamos considerar agora uma caso bem distinto dos anteriores. Suponha que é dada uma função $a(t)$, perguntamos: quem é a função $x(t)$ tal que

$$x'(t) = a(t)\,x(t)?$$

A resposta é fácil, é $x(t) = k\,e^{\int_0^t a(s)\,ds}$. De fato, pela regra da cadeia sabemos que

$$\frac{d}{dt}\,e^{b(t)} = b'(t)\,e^{b(t)}.$$

Logo,

$$\frac{d}{dt}\,x(t) = \frac{d}{dt}\,k\,e^{\int_0^t a(s)\,ds} = a(t)\,k\,e^{\int_0^t a(s)\,ds} = a(t)\,x(t).$$

♢

Exemplo 4.16. A partir do que vimos acima tomando $a(t) = t^3$, podemos considerar a equação $x'(t) = t^3\,x(t)$. A solução de $x(t)$ de $x'(t) = t^3\,x(t)$ é

$$x(t) = k\,e^{\int_0^t s^3\,ds} = k\,e^{t^4/4}.$$

Se considerarmos a condição inicial $x(0) = 2$ teremos que

$$2 = x(0) = k\,e^{0^4/4} = k.$$

Assim, fica determinado $k = 2$ e finalmente obtemos de maneira única a incógnita $x(t)$ que é $2\,e^{t^4/4}$.

♢

Motivados pelos exemplos acima introduzimos a seguinte generalização:

Definição 4.17. *Generalizando a definição inicial, seja $f(x,t)$, onde $f : \mathbb{R}^2 \to \mathbb{R}$. Denominamos de equação diferencial a expressão*

$$x'(t) = f(x(t), t).$$

A incógnita é a função $x(t)$.

O último exemplo ilustra tal situação. Neste caso, $f(x,t) = a(t)\,x = t^3\,x$.

Vamos analisar agora o caso

$$x'(t) + a\,x(t) = b(t)$$

(que corresponde ao caso da equação linear de primeira ordem não homogênea) com a condição inicial

$$x(t_0) = x_0.$$

Acima a é constante e $b(t)$ uma função de t.

Pode-se mostrar que a solução $x(t)$ da equação linear de primeira ordem não homogênea

$$x'(t) + a\,x(t) = b(t),$$

é

$$x(t) = e^{-a\,(t-t_0)}\,(\int_{t_0}^{t} e^{a\,(r\,-\,t_0)}\,b(r)\,dr\; +\; x_0\,),$$

onde k é constante.

Exemplo 4.17. Para encontrar $x(t)$ tal que

$$x'\,(t)\; -\; x(t) =\; 1,$$

e

$$x(0) = s$$

podemos utilizar a expressão acima e obter

$$x(t) = e^{-a\,(t-t_0)}\,(\int_{t_0}^{t} e^{a\,(r\,-\,t_0)}\,b(r)\,dr\; +\; x_0\,)\; = e^{\,t}\,(\int_{0}^{t} e^{-1\,r}\,1\,dr\; +\; s\,)\; =$$

$$= e^{\,t}\,(\,(-e^{-\,t}\; +\; 1)\; +\; s\,)\; = -1\; + e^{t}\,(1+s).$$

Acima $a = -1$, $t_0 = 0$, $x_0 = s$, e $b(t)$ é a função constante igual a 1.

◊

Exemplo 4.18. Para encontrar $x(t)$ tal que

$$x'(t) + a\,x(t) = a\,e^{-a\,t},$$

e

$$x(0) = x_0$$

podemos utilizar a expressão acima e obter

$$x(t) = e^{-a\,t}\,(\int_0^t e^{a\,r}\,a\,e^{-a\,r}\,dr + x_0) = e^{-a\,t}\,(a\,t + x_0).$$

$\diamondsuit$

Vamos considerar agora um caso mais geral e complexo de equações diferenciais. Desejamos obter agora $x_1(t)$ e $x_2(t)$ que satisfazem certo tipo de equações que envolvem $x_1'(t)$ e $x_2'(t)$.

Definição 4.18. *Um sistema linear de equações diferenciais em* $\mathbb{R}^2$ *é uma expressão do tipo*

$$\begin{cases} \frac{d\,x_1}{dt} = & a\,x_1(t) + b\,x_2(t) \\ \\ \frac{d\,x_2}{dt} = & c\,x_1(t) + d\,x_2(t) \end{cases},$$

onde a,b,c,d são constantes reais

As incógnitas no caso são as funções $x_1(t)$ *e* $x_2(t)$. *É útil as vezes considerar o par* $(x_1(t), x_2(t))$ *evoluindo com o tempo* t *sobre o plano* $\mathbb{R}^2$.

$x(t) = (x_1(t), x_2(t))$ *é denominada de solução do sistema linear de equações diferenciais.*

$x_1'(t)$ descreve a taxa de variação de $x_1(t)$ no tempo t e $x_2'(t)$ descreve a taxa de variação de $x_2(t)$ no tempo t.

Exemplo 4.19. Por exemplo se $a = 2$, $b = 0$, $c = 0$ e $d = -5$ obtemos o sistema

$$\begin{cases} \frac{d\,x_1}{dt} = & 2\,x_1(t) + 0\,x_2(t) = 2\,x_1(t) \\ \\ \frac{d\,x_2}{dt} = & 0\,x_1(t) - 5\,x_2(t) = -5\,x_2(t) \end{cases} .$$

Neste caso, podemos resolver individualmente primeiro a equação em $x_1(t)$ e depois em $x_2(t)$.

Obtemos $x_1(t) = x_0\,e^{2\,t}$ e $x_2(t) = y_0 e^{-5\,t}$.

Deste modo, $(x_1(t), x_2(t)) = (6\,e^{2\,t}, 7\,e^{-5\,t})$ é uma curva solução evoluindo com o tempo t em $\mathbb{R}^2$. Por exemplo, para $t = 9$ a solução vai estar em $(x_1(9), x_2(9)) = (6\,e^{2\times 9}, 7\,e^{-5\times 9}) = (6\,e^{18}, 7\,e^{-45})$.

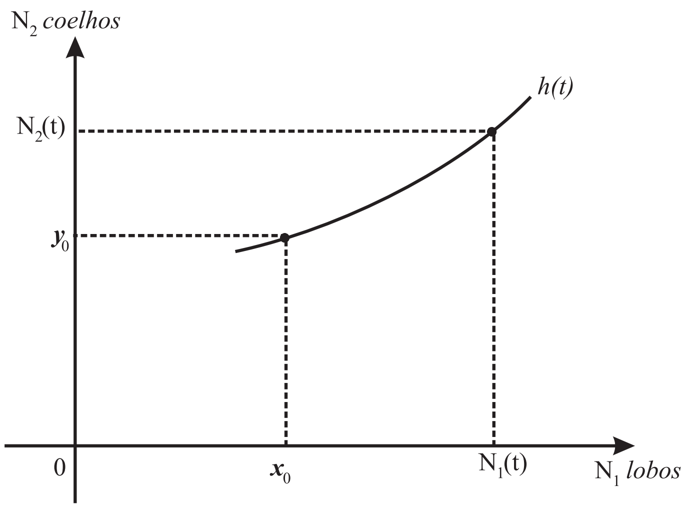


Figura 4.7: Evoluão temporal da população de coelhos e lobos. Esta é descrita pela solução de um sistema de equações diferenciais.

Faça você mesmo as contas e comprove que tal $x_1(t)$ e $x_2(t)$ satisfazem a equação

$$\begin{cases} \frac{dx_1}{dt} = 2\,x_1(t) \\ \\ \frac{dx_2}{dt} = -\,5\,x_2(t) \end{cases}$$

♢

No caso do exemplo acima foi muito simples encontrar a solução $(x_1(t), x_2(t))$, pois a evolução de $x_1(t)$ não sofre interferência da evolução de $x_2(t)$ e vice versa. O caso mais interessante é quando existe interferencia. Isto acontece quando $b \neq 0$ e $c \neq 0$. Esta análise será o objeto do que segue.

Pode-se mostrar que para $a, b, c, d, x_0, y_0 \in \mathbb{R}$ fixos, só existe uma única função $(x_1(t), x_2(t))$ com $t \in \mathbb{R}$, tal que satisfaz ao mesmo tempo

$$x_1(0) = x_0,\ \ x_2(0) = y_0,$$

e a equação

$$\begin{cases} \frac{dx}{dt} = & a\ x(t) + b\ y(t) \\ \\ \frac{dy}{dt} = & c\ x(t) + d\ y(t) \end{cases}.$$

O problema, é claro, é como obter $(x(t), y(t))$ a partir dos dados acima.

Podemos expressar o sistema de equações diferenciais acima da seguinte forma: seja

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Denote $z(t) = (x(t), y(t))$, então, o sistema acima equivale a

$$z'(t) = M\,z(t).$$

O $z(t)$ solução é denominado de solução da equação diferencial linear obtida a partir de M. Note que $z(t) = (x(t), y(t))$ também é solução do sistema linear associado.

A matriz M está agindo no vetor $z(t)$ direita.

A análise do sistema

$$z'(t) = z(t)\, M\,,$$

ou seja, a matriz M está agindo agora no vetor $z(t)$ à esquerda, é em tudo semelhante ao outro caso.

A situação que nos interessa é exatamente o sistema

$$p'(t) \;=\; p(t)\, \mathcal{L},$$

no caso em que $\mathcal{L}$ é uma matriz linha soma zero.

Exemplo 4.20. Considere o sistema linear

$$\begin{cases} x' = & 7x - 4y \\ y' = & -9x + 7y \end{cases}.$$

Vamos assumir que a condição inicial é $x(0) = 4$ e $y(0) = -1$.

Pergunta: quem são $x(t)$ e $y(t)$?

A solução é dada por

$$\begin{cases} x(t) = & \frac{7}{3}e^{13t} + \frac{5}{3}e^{t} \\ y(t) = & -\frac{7}{2}e^{13t} + \frac{5}{2}e^{t} \end{cases}, \text{ que naturalmente satisfaz } \begin{cases} x(0) = & 4 \\ y(0) = & -1 \end{cases}.$$

De fato,

$$x'(t) = \frac{7}{3}\, 13\, e^{13t} \,+\, \frac{5}{3}\, e^{t} = \frac{91}{3}\, e^{13t} \,+\, \frac{5}{3}\, e^{t} =$$

$$= (7\,\frac{7}{3} + 4\,\frac{7}{2})\, e^{13t} \,+\, (7\,\frac{5}{3} - 4\frac{5}{2})\, e^{t} =$$

$$= 7\,(\frac{7}{3}\, e^{13t} \,+\, \frac{5}{3}\, e^{t}) \,-\, 4\,(-\frac{7}{2}\, e^{13t} + \frac{5}{2}\, e^{t}) = 7\, x(t) - 4\, y(t).$$

Para mostrar que

$$y(t) = -\frac{7}{2}e^{13t} + \frac{5}{2}e^{t}$$

satisfaz

$$y'(t) = -9\,x'(t) + 7\,y(t),$$

devemos seguir um procedimento semelhante ao que foi feito acima.

Note que, neste caso,

$$\lim_{t\to+\infty} x(t) = \lim_{t\to+\infty} (\frac{7}{3}e^{13t} + \frac{5}{3}e^{t}) = +\infty$$

e

$$\lim_{t\to+\infty} y(t) = \lim_{t\to+\infty} (-\frac{7}{2}e^{13t} + \frac{5}{2}e^{t}) = -\infty.$$

A análise se $x(t)$ satisfaz

$$\lim_{t\to+\infty} x(t) = 0,$$

ou,

$$\lim_{t\to+\infty} x(t) = \infty,$$

ou,

$$\lim_{t\to+\infty} x(t) = a \in \mathbb{R},$$

ou, se não vale nenhuma dessas propriedades, é de grande importância na teoria. Naturalmente, a mesma pergunta se pode fazer para $y(t)$.

Em alguns casos interessantes, distintos do presente exemplo, ocorre que

$$\lim_{t\to+\infty} (x(t), y(t)) = (a,b) \in \mathbb{R}^2.$$

Neste caso, estamos obtendo uma informação importante sobre o comportamento assintótico (quando o tempo t fica arbitrariamente grande) da solução $(x(t), y(t))$.

$\diamondsuit$

Exercício: Mostre que $(x(t), y(t)) = (2e^{7t}, e^{7t})$ satisfaz $(x(0), y(0)) = (2, 1)$ e a equação

$$\begin{cases} \frac{dx}{dt} = & x(t) + 12\ y(t) \\ \\ \frac{dy}{dt} = & 3\ x(t) + 1\ y(t) \end{cases}.$$

Exercício: Mostre que $(x(t), y(t)) = (-2e^{-5t}, e^{-5t})$ satisfaz $(x(0), y(0)) = (-2, 1)$ e a equação

$$\begin{cases} \frac{dx}{dt} = & x(t) + 12\ y(t) \\ \\ \frac{dy}{dt} = & 3\ x(t) + 1\ y(t) \end{cases}.$$

Vamos dar agora um exemplo aplicado a uma situação real.

Exemplo 4.21. Suponha que estamos analisando uma população de N_1 lobos e N_2 coelhos e nosso objetivo é analisar como estas duas populações se propagam numa disputa natural. Assim, podemos supor (simplificando o modelo) que um sistema linear de equações diferenciais dado por

$$\begin{cases} N_1'(t) = \frac{dN_1}{dt} = & aN_1(t) + bN_2(t) \\ \\ N_2'(t) = \frac{dN_2}{dt} = & cN_1(t) + dN_2(t) \end{cases}$$

onde aqui $b > 0$, $c \leq 0$, $a > 0$ e $d > 0$, descreve a evolução do sistema.

$N_1'(t)$ descreve a taxa de variação da população $N_1(t)$ e $N_2'(t)$ descreve a taxa de variação da populacão $N_2(t)$.

Neste caso, estamos supondo que a taxa de variação do número de lobos $N_1'(t)$ no decorrer do tempo é uma combinação linear do número de coelhos e o número de lobos. $N_1(t)$ vai crescer ($b > 0$) quanto mais coelhos existirem. Os lobos (o predador) tem alimento. O coeficiente $a > 0$ da conta do crescimento populacional de lobos por cruzamento dentro da espécie.

A taxa de variação do número de coelhos $N_2'(t)$ no decorrer do tempo é combinação linear do número de lobos e do número de coelhos. $N_2(t)$ vai decrescer muito ($c \leq 0$) se o número de lobos for grande. Os coelhos (a presa) são comidos pelos lobos. O coeficiente $d > 0$ da conta do crescimento populacional de coelhos por cruzamento.

Os valores a, b, c, d podem ser estimados em função de dados reais.

O que desejamos determinar é quem é $(N_1(t), N_2(t))$.

A Figura 4.7 descreve no plano $\mathbb{R}^2$ a evolução temporal de $(N_1(t), N_2(t))$ para distintos valores de t.

Na situação em que $a = 0,1$, $b = 0 = c$, e $d = 0,3$ com $N_1(0) = x_0$ e $N_2(0) = y_0$, temos que

$$\frac{dN_1}{dt} = 0,1\, N_1(t)$$

e

$$\frac{dN_2}{dt} = 0,3\, N_2(t),$$

ou seja,

$$N_1(t) = x_0 e^{0,1\,t}$$

$$N_2(t) = y_0 e^{0,3\,t}.$$

Se $x_0 = 100$ lobos e $y_0 = 110$ coelhos, então, $h(t) = (100\, e^{0,1t}, 110\, e^{0,3t})$.

Voltemos ao caso geral. Vamos supor interferencia entre as populações N_1 e N_2 da seguinte forma.

$$\begin{cases} \frac{dN_1}{dt} = & a\, N_1(t) + b\, N_2(t) = N_1'(t) \\ \\ \frac{dN_2}{dt} = & c\, N_1(t) + d\, N_2(t) = N_2'(t) \end{cases}$$

onde $N_1(0) = x_0$ e $N_2(0) = y_0$.

Gostaríamos de saber quem é a curva solução $(N_1(t), N_2(t))$ no plano.

Pode-se mostrar que se a matriz

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

tem autovalores reais distintos λ_1 e λ_2 então a solução do sistema linear de equações diferenciais mencionado acima é da forma

$$N_1(t) = A_1 e^{\lambda_1 t} + A_2 e^{\lambda_2 t}$$

$$N_2(t) = B_1 e^{\lambda_1 t} + B_2 e^{\lambda_2 t}$$

onde A_1, A_2, B_1, B_2 são constantes reais (a serem determinadas). Estas constantes devem ser escolhidas de forma que $N_1(t)$ e $N_2(t)$ satisfaçam as condições iniciais $N_1(0) = x_0$ e $N_2(0) = y_0$.

Se ocorresse (no presente exemplo não é o caso) de existir o limite

$$\lim_{t \to +\infty} (N_1(t), N_2(t)) = (150, 450) \in \mathbb{R}^2,$$

então é porque a população de lobos, a longo prazo, iria para o total de 150 e a de coelhos para o total de 450.

O exemplo que estamos descrevendo é ilustrativo, de fato, um pouco simplista demais (para capturar a complexidade do problema real). Acreditamos, que, de qualquer forma, ele auxilia no entedimento do interesse prático da teoria dos sistemas lineares de equações diferenciais. Um modelo mais realista existe, mas isto requer a introdução do estudo de sistemas não lineares de equações diferenciais [DL]. Não vamos tratar deste caso aqui.

$\diamondsuit$

Vamos utilizar o que foi explicado acima para calcular a solução em um exemplo numérico.

Exemplo 4.22. Suponha que $\begin{cases} N_1' = & 3N_1 - 2N_2 \\ N_2' = & 2N_1 - 2N_2 \end{cases}$.

Neste exemplo N_1 e N_2 não representam de fato populações pois a equação não se enquadra no modelo populacional descrito acima; no presente caso $a = 3$, $b = -2$; $c = 2$; $d = -2$. Note que não vale $c < 0$. De qualquer modo, o exemplo é instrutivo, e vamos calcular a solução $(N_1(t), N_2(t))$ explicitamente. Ainda, assumimos que $N_1(0) = 100$, $N_2(0) = 300$.

Determinemos inicialmente os valores de $\lambda_1, \lambda_2, A_1, A_2, B_1$ e B_2.

Queremos resolver o sistema $\det(\lambda I - M) = 0$, onde $\lambda \in \mathbb{R}$, e

$$M = \begin{pmatrix} 3 & -2 \\ 2 & -2 \end{pmatrix},$$

isto é,

$$\det \begin{pmatrix} \lambda - 3 & 2 \\ -2 & \lambda + 2 \end{pmatrix} = 0 \Leftrightarrow (\lambda - 3)(\lambda + 2) + 4 = 0 \Leftrightarrow \lambda^2 - \lambda - 6 + 4 = 0$$

$$\Leftrightarrow \lambda^2 - \lambda - 2 = 0,$$

e desta forma,

$$\lambda = \frac{1 \overset{+}{-} \sqrt{1 - 4(-2)}}{2}.$$

Portanto, obtemos, $\lambda_1 = 2$ e $\lambda_2 = -1$. Logo,

$$N_1(t) = A_1 e^{2t} + A_2 e^{-t},$$

$$N_2(t) = B_1 e^{2t} + B_2 e^{-t},$$

onde, $N_1(0) = 100$ e $N_2(0) = 300$.

Note que M não satisfaz a propriedade de ser linha soma zero.

Devemos agora determinar A_1, A_2, B_1, e B_2.

Sabemos que

$$100 = A_1 + A_2,$$

$$300 = B_1 + B_2,$$

considerando os valores acima em $t = 0$.

Pelas equações que envolvem a derivada de N_1 temos que vale

$$2\,A_1\,e^{2\,t} - A_2\,e^{-t} = {N_1}'\,(t) = 3\,N_1(t) - 2N_2(t) =$$

$$= 3\,(A_1 e^{2t} + A_2 e^{-t}) - 2\,(B_1 e^{2t} + B_2 e^{-t}).$$

Coletando os termos que multiplicam respectivamente $e^{2\,t}$ e e^{-t}, obtemos da expressão acima

$$2\,A_1 \;=\; 3\,A_1 - 2\,B_1,$$

e

$$-\,A_2 \;=\; 3\,A_2 - 2\,B_2.$$

Obtemos então o conjunto de equações

$$A_1 + A_2 = 100,$$

$$B_1 + B_2 = 300,$$

e

$$A_1(-1) = -2B_1,$$

$$A_2(-4) = -2B_2.$$

Este último par significa resolver

$$A_1 = 2B_1,$$

$$2A_2 = B_2,$$

o que equivale a

$$B_1 = \frac{A_1}{2},$$
$$B_2 = 2A_2.$$

Obtemos assim, olhando apenas as equações que envolvem A_1, A_2, que vale

$$\begin{cases} A_1 + A_2 = & 100 \\ \frac{A_1}{2} + 2A_2 = & 300 \end{cases} \Leftrightarrow$$

$$\begin{cases} A_1 = & 100 - A_2 \\ 50 - \frac{A_2}{2} & +2A_2 = 300 \end{cases} \Leftrightarrow$$

$$\begin{cases} A_1 = 100 & -A_2 \\ \frac{3}{2}A_2 = & 250 \end{cases} \Leftrightarrow$$

$$\boxed{\begin{array}{ll} A_1 = & -\frac{200}{3} \\ \\ A_2 = & \frac{500}{3} \end{array}}$$

Como $B_1 = \frac{A_1}{2}$ e $B_2 = 2A_2$, obtemos, finalmente, a partir de

$$A_1 = \frac{-200}{3}, \ \ A_2 = \frac{500}{3},$$

que

$$B_1 = \frac{-100}{3}, \ \ B_2 = \frac{1000}{3}.$$

Portanto, a solução que buscávamos $(N_1(t), N_2(t))$ é dada por

$$\boxed{\begin{array}{ll} N_1(t) = & -\frac{200}{3}e^{2t} + \frac{500}{3}e^{-t} \\ \\ N_2(t) = & -\frac{100}{3}e^{2t} + \frac{1000}{3}e^{-t} \end{array}}$$

Supondo que todos os coeficientes $A_1, A_2, B_1, B_2, \lambda_1, \lambda_2$ são diferentes de zero, é natural perguntar sobre os limites, quando $t \to \infty$, de $N_1(t)$ e $N_2(t)$.

Note que $\lim_{t\to+\infty} N_1(t) = \lim_{t\to+\infty} A_1 e^{\lambda_1 t} + A_2 e^{\lambda_2 t} = \lim_{t\to+\infty} A_1 e^{2t} + A_2 e^{-t} =$

$$A_1 \lim_{t\to+\infty} e^{2t} = -\infty.$$

Ainda, $\lim_{t\to+\infty} N_2(t) = \lim_{t\to+\infty} B_1 e^{\lambda_1 t} + B_2 e^{\lambda_2 t} = \lim_{t\to+\infty} B_1 e^{2t} + B_2 e^{-t} =$

$$B_1 \lim_{t\to+\infty} e^{2t} = -\infty.$$

♢

Observaão: Note que $(x(t), y(t))$ constante e igual a $(0,0)$, para todo t real, sempre satisfaz

$$\begin{cases} \frac{dx}{dt} = & a\,x(t) + b\,y(t) \\ \\ \frac{dy}{dt} = & c\,x(t) + d\,y(t) \end{cases},$$

e, naturalmente, também satisfaz a condição inicial $(x(0), y(0)) = (0,0)$.

Exercício: Determine a solução $(x(t), y(t))$ satisfazendo a condição inicial $(x(0), y(0)) = (2,3)$ do sistema linear de equações diferenciais

$$\begin{cases} \frac{dx}{dt} = & x(t) + 12\ y(t) \\ \\ \frac{dy}{dt} = & 3\ x(t) + 1\ y(t) \end{cases}.$$

Sugestão: Calcule primeiro as soluções em λ do polinômio do segundo grau

$$\det(\lambda\, I - M) = 0,$$

onde

$$M = \begin{pmatrix} 1 & 12 \\ 3 & 1 \end{pmatrix}.$$

Estes valores são $\lambda = 7$ e $\lambda = -5$. A seguir proceda como acima resolvendo um sistema linear e usando o fato que $(x(0), y(0)) = (2, 3)$.

A resposta é $(x(t), y(t)) = (4\,e^{7\,t} - 2\,e^{-5\,t}, 2\,e^{7\,t} + e^{-5\,t})$.

Exercício: Determine a solução geral $(x(t), y(t))$ satisfazendo a condição inicial $(x(0), y(0)) = (x_0, y_0)$ do sistema de equações diferenciais

$$\begin{cases} \frac{dx}{dt} = & x(t) + 12\; y(t) \\ \\ \frac{dy}{dt} = & 3\; x(t) + 1\; y(t) \end{cases}.$$

Definição 4.19. *O caso mais geral de equação diferencial em* $\mathbb{R}^2$ *seria*

$$\begin{cases} \frac{dx}{dt} &= f(x(t), y(t)) \\ \frac{dy}{dt} &= g(x(t), y(t)) \end{cases},$$

onde $f : \mathbb{R}^2 \to \mathbb{R}$ *e* $g : \mathbb{R}^2 \to \mathbb{R}$.

A incógnita é o caminho $(x(t), y(t))$ que satisfaz tal equação.

Quando

$$f(x, y) = a\,x + b\,y$$

e

$$g(x, y) = c\,x + d\,y$$

obtemos um sistema linear de equações diferenciais (como anteriormente descrito).

No caso em que $f(x, y)$ e $g(x, y)$ não são lineares se obtem maior flexibilidade para descrever interferências complexas entre duas populações.

O Teorema de existencia e unicidade de equações diferenciais (ver por exemplo [DL]) assegura que, dada uma condição inicial $(x(0), y(0)) = (c, d)$, então, a solução $(x(t), y(t))$ de $(x'(t), y'(t)) = (\,f(x(t), y(t)), g(x(t), y(t))\,)$, é única, se f e g forem de classe C^1.

Tais equações são denominadas de não lineares. Elas não vão ocorrer na teoria desenvolvida no presente volume que cobre questões menos complexas da Teoria do Processos Estocásticos. Com o fim de dar um breve panorama mais amplo sobre o assunto vamos sucintamente fazer algumas considerações simples, antes de voltar a analisar sistemas lineares na sua formulação mais geral.

Note que no caso geral, se (a, b) é tal que $f(a, b) = 0$ e $g(a, b) = 0$, então , se a curva $(x(t), y(t))$, $t \in \mathbb{R}$, for constante e igual a (a, b) então $(x(t), y(t))$, $t \in \mathbb{R}$ é solução ou seja satisfaz a equação acima. Dizemos neste caso que $(x(t), y(t))$ é uma **solução em equilíbrio** (fica parada sobre (a, b)). Em muitos problemas reais, o que se deseja é saber quem é (a, b), tal que as duas populações $x(t)$ e $y(t)$ ficam em **equilíbrio** nestes valores. Para tanto, basta descobrir os (a, b) tal que $(f(a, b), g(a, b)) = (0, 0)$. Neste caso, a solução $(x(t), y(t)$ fica parada na posição (a, b), independente da variação de t.

Isto porque se $x(t)$ e $y(t)$ são constantes, então para todo t real

$$(x'(t), y'(t)) = (0, 0) = (f(a, b), g(a, b)) = (f(x(t), y(t)), g(x(t), y(t)).$$

Considere uma condição inicial (x_0, y_0) (distinta de (a, b)) e então $(x(t), y(t))$ vai denotar a solução de

$$\begin{aligned} x'(t) &= f(x(t), y(t), \\ y'(t) &= g(x(t), y(t), \end{aligned}$$

que satisfaz a condição inicial $(x(0), y(0)) = (x_0, y_0)$.

Em muitos problemas reais, vale ainda que para grandes conjuntos de condições iniciais (x_0, y_0), existe o limite

$$\lim_{t \to +\infty} (x(t), y(t)) = (a, b) \in \mathbb{R}^2.$$

Estamos assumindo acima, é claro, que (a, b) é ponto de equilíbrio.

Se $x(t) = N_1(t)$ e $y(t) = N_2(t)$ descrevem duas populações, então no caso da ocorrência do que descrevemos acima, se poderia dizer, que independente das populações iniciais (x_0, y_0), se observaria no modelo, que as populações N_1 e N_2, a longo prazo, se aproximariam respectivamente dos valores a e b.

Neste caso, o ponto de equilíbrio (a, b) seria dito estável.

Comportamento semelhante ocorre em (certos) Processos de Markov em tempo contínuo com dois estados, em que (a, b) denota o estado de equilíbrio $\pi \in \mathbb{R}^2$. Deve-se, supor no entanto, que as condições iniciais (x_0, y_0) estejam no conjunto dos elementos que satisfazem $x_0 + y_0 = 1$.

O leitor deve encarar as considerações acima como uma discussão heurística envolvendo assuntos completamente distintos.

Vamos voltar agora aos sistemas lineares e analisar a questão de maneira bastante geral.

Definição 4.20. *Seja uma matriz real*

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & ... & a_{1n} \\ a_{21} & a_{22} & a_{23} & ... & a_{2n} \\ . & . & . & ... & . \\ a_{n1} & a_{n2} & a_{n3} & ... & a_{nn} \end{pmatrix},$$

do tipo n por n.

Um sistema linear de equações diferenciais em $\mathbb{R}^n$ é um sistema do tipo

$$\begin{cases} \frac{d\,x_1}{dt} = a_{11}\,x_1(t) + a_{12}\,x_2(t) + a_{13}\,x_3(t) + ... + a_{1n}\,x_n(t) \\ \frac{d\,x_2}{dt} = a_{21}\,x_1(t) + a_{22}\,x_2(t) + a_{23}\,x_3(t) + ... + a_{2n}\,x_n(t) \\ \frac{d\,x_3}{dt} = a_{31}\,x_1(t) + a_{32}\,x_2(t) + a_{33}\,x_3(t) + ... + a_{3n}\,x_n(t) \\ \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \frac{d\,x_n}{dt} = a_{n1}\,x_1(t) + a_{n2}\,x_2(t) + a_{n3}\,x_3(t) + ... + a_{nn}\,x_n(t) \end{cases}.$$

A incógnita é a função $x(t) = (x_1(t), x_2(t), ..., x_n(t))$.

A equação acima pode ser escrita da seguinte forma:

$$\begin{pmatrix} x_1'(t) \\ x_2'(t) \\ . \\ x_n'(t) \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & ... & a_{1n} \\ a_{21} & a_{22} & a_{23} & ... & a_{2n} \\ . & . & . & ... & . \\ a_{n1} & a_{n2} & a_{n3} & ... & a_{nn} \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \\ . \\ x_n(t) \end{pmatrix} = A \begin{pmatrix} x_1(t) \\ x_2(t) \\ . \\ x_n(t) \end{pmatrix}.$$

A solução $x(t) = (x_1(t), x_2(t), ..., x_n(t))$ do sistema acima sempre existe. Fixada uma condição inicial $x(0) = (x_0^1, x_0^2, ..., x_0^n) \in \mathbb{R}^n$, a função $x(t)$ solução do sistema acima é única.

Vamos descrever brevemente como, dado a matriz A e a condição inicial $x(0) = (x_0^1, x_0^2, ..., x_0^n) \in \mathbb{R}^n$, se encontra a solução $x(t) = (x_1(t), x_2(t), ..., x_n(t))$.

Para a matriz A fixada e t real fixado, denote por

$$e^{t\,A} = I + t\,A + \frac{1}{2!}(t\,A)^2 + \frac{1}{3!}(t\,A)^3 + \frac{1}{4!}(t\,A)^4 + ... + \frac{1}{n!}(t\,A)^n + ... = \sum_{n=0}^{\infty} \frac{1}{n!}\,(t\,A)^n.$$

Acima $(t\,A)^n$ significa que multiplicamos a matriz $t\,A$ por si mesma n vezes.

A expressão acima é conhecida como exponencial de uma matriz (no caso da matriz $t\,A$).

Podemos também escrever

$$e^{t\,A} = I + t\,A + \frac{1}{2!}\,t^2\,A^2 + \frac{1}{3!}\,t^3\,A^3 + \frac{1}{4!}\,t^4\,A^4 + ... + \frac{1}{n!}\,t^n\,A^n + ...$$

Fixado A e variando t obtemos uma familia $e^{t\,A}$ a um parâmetro t de matrizes da forma n por n.

Seja

$$x(t) = e^{t\,A} \begin{pmatrix} x_0^1 \\ x_0^2 \\ . \\ . \\ x_0^n \end{pmatrix}.$$

Pode-se mostrar (ver [DL]) que fixada a matriz A e a condição inicial $x(0) = (x_0^1, x_0^2, ..., x_0^n) \in \mathbb{R}^n$, a solução $x(t) = (x_1(t), x_2(t), ..., x_n(t))$ é dada pela expressão acima. Desta forma a expressão $e^{t\,A}$ desempenha um papel fundamental.

A expressão acima

$$x(t) \,=\, e^{t\,A} \begin{pmatrix} x_0^1 \\ x_0^2 \\ . \\ . \\ x_0^n \end{pmatrix},$$

resolve o caso geral de qualquer sistema linear de equações diferenciais $x' = A\,x$.

Exemplo 4.23. Podemos, por exemplo, calcular para qualquer t real a exponencial $e^{t\,A}$ de uma matriz diagonal

$$A = \begin{pmatrix} a_1 & 0 & 0 \\ 0 & a_2 & 0 \\ 0 & 0 & a_3 \end{pmatrix}.$$

É fácil ver que neste caso

$$e^{t\,A} = \begin{pmatrix} e^{t\,a_1} & 0 & 0 \\ 0 & e^{t\,a_2} & 0 \\ 0 & 0 & e^{t\,a_3} \end{pmatrix}.$$

$\diamondsuit$

Exemplo 4.24. Pode-se mostrar (ver [DL]) que

$$e^{t \begin{pmatrix} a & b \\ -b & a \end{pmatrix}} = e^{t\,a} \begin{pmatrix} \cos(b\,t) & \operatorname{sen}(b\,t) \\ -\operatorname{sen}(b\,t) & \cos(b\,t) \end{pmatrix}.$$

$\diamondsuit$

Exemplo 4.25. Note que no caso unidimensional, fixada a matriz a da forma 1 por 1 (um número real) e a condição inicial x_0, a solução de $x'(t) = a\,x(t)$, é dada por

$$x(t) = e^{t\,a}\,x_0 = [\sum_{n=0}^{\infty} \frac{1}{n!}\,(t\,a)^n\,]\,x_0.$$

$\diamondsuit$

Exemplo 4.26. Fixada a matriz

$$\begin{pmatrix} 2,3 & 4,5 & -1 \\ 1,2 & 5 & 3 \\ -10,3 & 7 & 8 \end{pmatrix}$$

e a condição inicial $x(0) = (1,0,0)$, temos que a solução $x(t)$ de

$$x'(t) = A\,x(t)$$

é dada por

$$x(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{pmatrix} = e^{t\,A}\,\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} =$$

$$[\sum_{n=0}^{\infty} \frac{1}{n!}\,\begin{pmatrix} 2,3\,t & 4,5\,t & -t \\ 1,2\,t & 5\,t & 3\,t \\ -10,3\,t & 7\,t & 8\,t \end{pmatrix}^n\,]\,\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} =$$

$$[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 2.3\,t & 4.5\,t & -t \\ 1.2\,t & 5\,t & 3\,t \\ -10.3\,t & 7\,t & 8\,t \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 2.3\,t & 4.5\,t & -t \\ 1.2\,t & 5\,t & 3\,t \\ -10.3\,t & 7\,t & 8\,t \end{pmatrix}^2 + ...]\,\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$\diamondsuit$

De maneira análoga, a partir de uma matriz A, podemos considerar o sistema de equações diferenciais

$$\begin{pmatrix} x_1'(t) & x_2'(t) & ... & x_n'(t) \end{pmatrix} =$$

$$\begin{pmatrix} x_1(t) & x_2(t) & ... & x_n(t) \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} & ... & a_{1n} \\ a_{21} & a_{22} & a_{23} & ... & a_{2n} \\ . & . & . & ... & . \\ a_{n1} & a_{n2} & a_{n3} & ... & a_{nn} \end{pmatrix} =$$

$$\begin{pmatrix} x_1(t) & x_2(t) & ... & x_n(t) \end{pmatrix} A.$$

Fixada uma condição inicial $x(0) = (x_0^1, x_0^2, ..., x_0^n) \in \mathbb{R}^n$, desejamos saber quem é a solução $x(t) = (x_1(t), x_2(t), ..., x_n(t))$ do sistema acima. A resposta é

$$x(t) \,=\, \begin{pmatrix} x_0^1 & x_0^2 & ... & x_0^n \end{pmatrix} e^{t\,A}\,.$$

Os resultados correspondentes para a ação da matriz A agindo à direita ou à esquerda são os mesmos. No caso da Teoria dos Processos Estocásticos (matrizes $A = \mathcal{L}$ tipo linha soma zero) é mais natural considerar a ação à esquerda.

Exemplo 4.27. Fixada a matriz

$$\begin{pmatrix} 2,3 & 4,5 & -1 \\ 1,2 & 5 & 3 \\ -10,3 & 7 & 8 \end{pmatrix}$$

e a condição inicial $x(0) = (1, 0, 0)$, temos que a solução $x(t)$ de

$$x'(t) \,=\, x(t)\,A$$

é dada por

$$x(t) \,=\, \begin{pmatrix} x_1(t) & x_2(t) & x_3(t) \end{pmatrix} \,=\, \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} e^{t\,A} \,=$$

$$\begin{pmatrix} 1 & 0 & 0 \end{pmatrix} [\sum_{n=0}^{\infty} \frac{1}{n!}\, t^n \begin{pmatrix} 2,3 & 4,5 & -1 \\ 1,2 & 5 & 3 \\ -10,3 & 7 & 8 \end{pmatrix}^n].$$

$\diamondsuit$

Finalmente vamos analisar a equação

$$X'(t)\, A = X(t), \;\; X(0) = I,$$

com t real, onde $X(t)$ é matriz da forma n por n que depende de t e A é matriz fixa da forma n por n.

Acima $X'(t)$, para cada t fixo, é a matriz em que tomamos a derivada de cada um de suas entradas. Por exemplo, se

$$X(t) = \begin{pmatrix} x_{11}(t) & x_{12}(t) & x_{13}(t) \\ x_{21}(t) & x_{22}(t) & x_{23}(t) \\ x_{31}(t) & x_{32}(t) & x_{33}(t) \end{pmatrix}$$

então

$$X'(t) = \begin{pmatrix} {x_{11}}'(t) & {x_{12}}'(t) & {x_{13}}'(t) \\ {x_{21}}'(t) & {x_{22}}'(t) & {x_{23}}'(t) \\ {x_{31}}'(t) & {x_{32}}'(t) & {x_{33}}'(t) \end{pmatrix}.$$

Dado A desejamos encontrar quem é $X(t)$.

A resposta é fácil, é a família de matrizes indexada por t dada por

$$X(t) = e^{t\,A},$$

onde

$$e^{t\,A} = I + t\,A + \frac{1}{2!}(t\,A)^2 + \frac{1}{3!}(t\,A)^3 + \frac{1}{4!}(t\,A)^4 + ... + \frac{1}{n!}(t\,A)^n + ... = \sum_{n=0}^{\infty} \frac{1}{n!}\,(t\,A)^n.$$

Exemplo 4.28. A razão para a última afirmação é bem simples. Vamos mostrar este fato no caso em que A é da forma dois por dois. O caso geral é semelhante. Ora,

$$X'(t)\,A = X(t),$$

com

$$X(t) = \begin{pmatrix} x_{11}(t) & x_{12}(t) \\ x_{21}(t) & x_{22}(t) \end{pmatrix},$$

e

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

significa

$$\begin{pmatrix} x'_{11}(t) & x'_{12}(t) \\ x'_{21}(t) & x'_{22}(t) \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} x_{11}(t) & x_{12}(t) \\ x_{21}(t) & x_{22}(t) \end{pmatrix}.$$

A igualdade acima pode ser decomposta em duas como

$$\begin{pmatrix} x'_{11}(t) & x'_{12}(t) \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} x_{11}(t) & x_{12}(t) \end{pmatrix},$$

e

$$\begin{pmatrix} x'_{21}(t) & x'_{22}(t) \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} x_{21}(t) & x_{22}(t) \end{pmatrix}.$$

Por outro lado, a igualdade $X(0) = I$ significa

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} x_{11}(0) & x_{12}(0) \\ x_{21}(0) & x_{22}(0) \end{pmatrix}.$$

Esta, por sua vez, pode ser decomposta em duas

$$\begin{pmatrix} 1 & 0 \end{pmatrix} = \begin{pmatrix} x_{11}(0) & x_{12}(0) \end{pmatrix},$$

e

$$\begin{pmatrix} 0 & 1 \end{pmatrix} = \begin{pmatrix} x_{21}(0) & x_{22}(0) \end{pmatrix}.$$

Recaímos assim nos dois pares de equações

$$\begin{pmatrix} x'_{11}(t) & x'_{12}(t) \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} x_{11}(t) & x_{12}(t) \end{pmatrix}, \text{ e } \begin{pmatrix} 1 & 0 \end{pmatrix} = \begin{pmatrix} x_{11}(0) & x_{12}(0) \end{pmatrix},$$

e, ainda

$$\begin{pmatrix} x'_{21}(t) & x'_{22}(t) \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} x_{21}(t) & x_{22}(t) \end{pmatrix}, \text{ e } \begin{pmatrix} 0 & 1 \end{pmatrix} = \begin{pmatrix} x_{21}(0) & x_{22}(0) \end{pmatrix}.$$

Vamos denominar de $x(t) = \begin{pmatrix} x_{11}(t) & x_{12}(t) \end{pmatrix}$, a solução do primeiro problema e $y(t) = \begin{pmatrix} x_{21}(t) & x_{22}(t) \end{pmatrix}$, a do segundo.

Sabemos que a solução do primeiro problema é

$$x(t) = (1 \;\; 0)\; e^{t\,A}$$

e do segundo

$$y(t) = (0 \;\; 1)\; e^{t\,A}.$$

Logo,

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \; e^{t\,A} \; = \; e^{t\,A}.$$

Sendo assim $X(t)$, solução de $X'(t) = X(t)\,A$ e $X(0) = I$, satisfaz $X(t) = e^{t\,A}$.

$\diamondsuit$

Vamos agora analisar um outro tipo de problema: considere

$$P : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R},$$

da forma $P(s,t)$ tal que satisfaça a equação

$$\frac{\partial P}{\partial t} + (s + a) \frac{\partial P}{\partial s} = 1,$$

com a condição de fronteira $P(s,0) = h(s)$, onde a é uma constante e $h(s)$ úma função real fixada.

A equação acima é diferencial parcial: ela envolve derivação parcial, diferentemente dos casos anteriormente considerados acima nesta seção.

Um dos procedimentos mais usuais em equações diferenciais parciais é tentar fazer a equação recair em um problema em que existe separação de variáveis.

Por exemplo, o ideal seria que $P(s,t)$ fosse da forma

$$P(s,t) = g(t) + h(s).$$

Para simplificar mais a questão, da forma

$$P(s,t) = t + h(s).$$

Desta forma teríamos

$$\frac{\partial P}{\partial t} = 1.$$

Reciprocamente, se $P(s,t)$ satisfaz a equação acima, então por integração

$$\int \frac{\partial P\,(s,t)}{\partial t}\, dt = \int 1\, dt = t + z(s).$$

Como em geral tal não acontece, ou seja, não ocorre $\frac{\partial P}{\partial t} = 1$, então nosso objetivo será fazer uma mudança de coordenadas para P de tal forma que isto ocorra.

Seja $B(v,\tau) = (s(v,\tau), t(v,\tau))$, onde $B : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R} \times \mathbb{R}^+$.

v e τ serão as novas coordenadas que substituirão s e t.

Nas novas coordenadas denotamos

$$\tilde{P}(v,\tau) = P(s(v,\tau), t(v,\tau)).$$

Gostaríamos que B fosse tal que

$$\frac{\partial \tilde{P}(v,\tau)}{\partial \tau} = 1.$$

É fundamental supor abaixo que $P(s,t)$, embora desconhecida, satisfaz a equação desejada. Descobrir a $B(v,\tau)$ conveniente será nosso objetivo.

Vamos ver qual a equação satisfeita por $\tilde{P}(v,\tau)$.

Ora, pela regra da cadeia

$$\frac{\partial \tilde{P}}{\partial \tau} \,=\, \frac{\partial P}{\partial s}\,\frac{\partial s}{\partial \tau} + \frac{\partial P}{\partial t}\,\frac{\partial t}{\partial \tau}.$$

Ora, se ocorrer

$$\frac{\partial s}{\partial \tau} \,=\, (s+a),$$

e

$$\frac{\partial t}{\partial \tau} \,=\, 1,$$

a equação acima se torna

$$\frac{\partial \tilde{P}}{\partial \tau} \,=\, \frac{\partial P}{\partial s}\,(s+a) + \frac{\partial P}{\partial t} \,=\, 1.$$

Desta forma, obteremos o resultado desejado, ou seja $\frac{\partial \tilde{P}}{\partial \tau} \,=\, 1$, e será fácil determinar $P(v,\tau)$.

Vamos então determinar $s(v,\tau)$ e $t(v,\tau)$. As equações que necessitamos resolver são ordinárias. Para tanto necessitamos de condições iniciais.

Para

$$\frac{\partial t}{\partial \tau} \,=\, 1,$$

a condição natural (para cada v fixo) é $t(v,\tau)=0$ em $\tau=0$.

Logo,

$$t(v,\tau)=\tau.$$

Ainda, para a equação ordinária (dependendo de um parâmetro v)

$$\frac{\partial s}{\partial \tau} - s = a,$$

a condição inicial natural (para cada τ fixo) é $s(v,\tau) = v$,

Logo,

$$s(v,\tau) = (v+a)\, e^{\tau} - a,$$

conforme vimos no exemplo 4.14 no começo desta seção. O resultado segue de considerar para cada v fixo, a expressão obtida no exemplo 4.14, para a variável s.

Desta forma determinamos $B(v,\tau)$.

Para determinar $\tilde{P}(v,\tau)$, usamos a equação ordinária

$$\frac{\tilde{P}(v,\tau)}{\partial \tau} = 1,$$

com a condição inicial $P(v,0) = h(v)$.

Logo,

$$\tilde{P}(v,\tau) = \tau + h(v).$$

Finalmente, vamos recuperar $P(s,t)$.

Invertendo B, obtemos $\tau = t$ e

$$v = e^{-\tau}\,(s+a) - a = e^{-t}\,(s+a) - a.$$

Logo,

$$P(s,t) = \tilde{P}(v(s,t),\tau(s,t)) = t + h(\, e^{-t}\,(s+a) - a\,).$$

Por exemplo, se a condição de fronteira for $h(s) = s^i$, então

$$P(s,t) = t + (\, e^{-t}\,(s+a) - a\,)^i.$$

O método descrito acima pode ser aplicado em diversos tipos de equações diferenciais parciais. Em essencia, ele se vale de equações diferenciais ordinárias para resolver o problema. É conhecido como o método das características.

4.6 Apêndice - Distribuição Geométrica e Exponencial

Definição 4.21. *Dada uma variável aleatória $X : \Omega \to \mathbb{R}^+$, dizemos que ela tem distribuição exponencial de parâmetro α, se*

$$P(X \le x) = 1 - e^{-\alpha x}, \ x \geq 0,$$

e $P(X \le x)$ é o valor 0 se $x < 0$.

Segundo nossa notaão $X \in \epsilon(\lambda)$.

Nosso objetivo nesta seção é descrever o fenômeno da perda de memória associado a Processo de Poisson X_t, $t \in \mathbb{R}^+$, com parâmetro α. Considere a variável X que descreve o tempo de primeiro salto de um certo caminho amostral $\omega \in S^{\mathbb{R}^+}$. Ora, sabemos que X, que toma valores em $\mathbb{R}^+$, é governado pela distribuição exponencial de parâmetro α.

Inicialmente, como motivação, vamos considerar uma versão em que X toma valores discretos deste fenômeno, o que será obtido através da análise da distribuição geométrica.

Considere uma moeda que tem probabilidade p de sair cara e tem probabilidade $q = (1-p)$ de sair coroa.

Vamos jogar a moeda sucessivamente várias vezes e a variável aleatória X vai descrever a probabilidade do tempo em que acontece pela primeira vez o evento sair coroa.

Desta forma X toma valores em $\{1, 2, 3, ...\}$ e

$$P(X = n) = p^{n-1} q, \quad \forall n \geq 1.$$

Definição 4.22. *Dizemos, neste caso, que X tem distribuição geométrica.*

Segundo nossa notação $X \in \mathcal{G}(p)$, $0 \leq p \leq 1$

Vamos descrever o que entendemos por falta de memória da distribuição: vamos calcular

$$P(X = n+m|\, X > n) = \frac{P(X = n+m\,,\, X > n)}{P(X > n)} =$$

$$= \frac{P(X = n+m)}{P(X > n)} = \frac{p^{n+m-1}\, q}{\sum_{j=n+1}^{\infty} p^{j-1}\, q} =$$

$$= \frac{p^{n+m-1}}{\frac{p^n}{1-p}} = p^{m-1} q = P(X = m).$$

Portanto, condicionando em $[X > n]$ (isto é, ainda não saiu coroa no tempo n), o número de tentativas restantes até que pela primeira vez apareça coroa no tempo $m+n$, tem função de distribuição de probabilidade igual a $P(X = m)$ (ou seja, é como supor que o processo tivesse começado a contar as tentativas considerando o tempo m como se fosse o tempo 0).

Esta propriedade é o que entendemos de falta de memória. Num certo sentido, a cada momento, o processo esquece o que aconteceu antes.

A recíproca é verdadeira no seguinte sentido: Seja X uma variável aleatória discreta tomando valores em $\{1, 2, 3, ...\}$, e denote $p_n = P(X = n)$. Assuma a propriedade da falta de memória. Vamos mostrar que então X tem distribuição geométrica. De fato, por hipótese,

$$P(X = i) = P(X = n+i\,|\, X > n) = \frac{P(X = n+i)}{P(X > n)} =$$

$$= \frac{P(X = n+i)}{\sum_{j \geq n+1} P(X = j)} = \frac{p_{\,n+i}}{\sum_{j \geq n+1} p_j}.$$

Desta forma, $p_i = \frac{p_{\,n+i}}{\sum_{j \geq n+1} p_j}$.

Sendo assim, considerando acima $n = 1$, temos

$$p_i = \frac{p_{\,i+1}}{\sum_{j \geq 2} p_j} \Rightarrow \sum_{j \geq 2} p_j = \frac{p_{i+1}}{p_i} = 1 - p_1.$$

Denote $p = (1 - p_1)$.

Logo,

$$p_{i+1} = p\, p_i \Rightarrow p_i = p\, p_{i-1} \Rightarrow \;... \Rightarrow p_i = p^{i-1}\, p_1.$$

Denote $q = p_1 = 1 - p$, e obtemos assim que X possui distribuição geométrica, pois $P(X = i) = p_i = p^{i-1}\, q$.

Vamos agora considerar o caso em que a variável aleatória X toma valores sobre os reais.

Suponha que X tenha distribuição exponencial com parâmetro α, ou seja,

$$f_X(x) = \begin{cases} \alpha e^{-\alpha x}, & x \geq 0 \\ 0, & c.c. \end{cases} \quad , \quad P(X \leq x) = F_X(x) = \begin{cases} 1 - e^{-\alpha x}, & x \geq 0 \\ 0, & c.c. \end{cases}$$

A distribuição exponencial ocorre como:

- densidade de variáveis aleatórias que representam tempos de vida de determinados equipamentos.
- intervalo de tempo entre eventos que ocorrem em sequencia aleatória, como encomendas a um atacadista, chamadas telefônicas em um centro telefônico, demandas de serviço de computador, etc...

Vamos analisar se ocorre aqui também o fenômeno de falta de memória da distribuição da variável X.

Seja $s > 0$ fixo e G a função de distribuição condicional de $X' = X - s$ dado que $X > s$. Assim, para $x \geq 0$, temos:

$$G(x) = P(X' \leq x | X > s) = P(X - s \leq x | X > s) = P(X \leq x + s | X > s) =$$

$$= \frac{P(X \leq x + s, X > s)}{P(X > s)} = \frac{P(s < X \leq x + s)}{P(X > s)} = \frac{1 - e^{-\alpha(x+s)} - 1 + e^{-\alpha s}}{e^{-\alpha s}} =$$

$$= \frac{-e^{-\alpha x} e^{-\alpha s} + e^{-\alpha s}}{e^{-\alpha s}} = 1 - e^{-\alpha x}.$$

Logo, $G(x) = F_X(x)$. Isto significa, no caso dos exemplos de aplicações acima, que a distribuição da vida restante do equipamento não depende do tempo em que o mesmo estava em funcionamento.

Reciprocamente, a densidade exponencial é a única que tem esta propriedade. De fato, suponha que X seja uma v.a. não negativa contínua para qual a função de distribuição condicional de $X' = X - s$, dado que $X > s$, é igual a função de distribuição inicial $F_X(x)$.

Sendo assim,

$$G(x) = \frac{\int_s^{s+x} f(t)dt}{\int_s^{+\infty} f(t)dt} = F(x) \Rightarrow \frac{F(x+s) - F(s)}{1 - F(s)} = F(x).$$

Ora $F(0) = 0$, pois a variável assume apenas valores não-negativos, logo

$$\frac{F(x+s) - F(s)}{x} = [1 - F(s)] \left[\frac{F(x) - F(0)}{x} \right].$$

Vamos fazer aqui a hipótese que F é diferenciável (resultados mais gerais, sem supor que valha tal propriedade são verdadeiros, ver [KT1])

Fazendo $x \to 0^+$ temos que

$$F'(s) = [1 - F(s)]F'(0)$$

Seja $U(s) = 1 - F(s)$ e denote por $\alpha = F'(0)$. Entï¿½o,

$$U'(s) = -F'(s) = -U(s)\ \alpha.$$

Sabe-se, da Teoria das Equações Diferenciais, que a solução $U(s)$ da equação acima é

$$U(s) = ce^{-\alpha s}.$$

Como $U(0) = 1 - F(0) = 1 - 0 = 1$ concluímos que $c = 1$, e assim, $U(s) = e^{-\alpha s} = 1 - F(s) = 1 - F_X(s)$, para todo $s > 0$.

Logo, X tem distribuição exponencial com parâmetro α.

Finalmente, voltando ao Processo de Poisson, se X descreve o tempo do primeiro salto, então, tal variável X tem distribuição exponencial.

Condicionando em que ainda não houve um salto no tempo s fixo, $X' = X - s$, vai descrever o tempo do primeiro salto após o tempo s. Acima, mostramos que a variável descrita por X perde memória.

4.7 Exercícios

1. Seja $(X_t)_{t\geq 0}$ um processo de Poisson com parâmetro contínuo. Usando a fórmula
$$P_{ij}(t) = \begin{cases} \exp\{-\lambda t\}\frac{(\lambda t)^{j-i}}{(j-i)!}\,, \text{ se } j \geq i \\ 0\,, \;\; c.c. \end{cases}$$

   a. Mostre que $P_{ij}(\Delta t) = o(\Delta t)$, sempre que $|j-i| > 1$.

   b. Mostre que $\sum_{k=i}^{j} P_{ik}(t)P_{kj}(s) = P_{ij}(t+s)$, sempre que $i < j$, $t \geq 0$, $s \geq 0$.

2. Seja $(X_t)_{t\geq 0}$ uma Cadeia de Markov com espaço de estados $\{0, 1\}$ e probabilidades de transição
$$P_{ij}(t) = \frac{1 + (-1)^{i+j}e^{-2t}}{2}, \text{ para } i, j = 0, 1.$$

   Mostre que

   a. $0 \leq P_{ij}(t) \leq 1$.

   b. $\sum_j P_{ij}(t) = 1$.

   c.
$$P_{ij}(0) = \begin{cases} 1\,, \text{ se } i = j \\ 0\,, \text{ se } i \neq j \end{cases}.$$

d. $P_{ij}(t+\tau) = \sum_k P_{ik}(t)P_{kj}(\tau)$.

Exercício 3:

3. Determine a matriz $\mathcal{L}$ para o processo do Exercício 2 acima e mostre que:

   a. $P'_{ij}(t) = \frac{dP_{ij}(t)}{dt} = \sum_k P_{ik}(t)\lambda_{kj}$.
   b. $\frac{dP_{ij}(\tau)}{d\tau} = \sum_k \lambda_{ik}P_{kj}(\tau)$.

4. Dois satélites de comunicação são colocados em órbita. O tempo de vida de um satélite é exponencial com média $\frac{1}{\mu}$. Se há falha, efetua-se uma substituição. O tempo necessário para preparo e envio de uma substituição é exponencial com média $\frac{1}{\lambda}$. Seja $X_t =$ o número de satélites em órbita no tempo t. Suponha que $(X_t)_{t\geq 0}$ seja um processo markoviano.

   a. Obtenha o gerador infinitesimal do processo.

   b. Estabeleça as equações de Kolmogorov.

   c. Resolva o sistema em (b) para encontrar as probabilidades de transição $P_{ij}(t)$.

5. Seja X_t um processo de Poisson com parâmetro λ. Determine para t_1 e t_2 fixados, com $t_2 > t_1$, a covariância associada a estes dois tempos, ou seja, calcule
$$E(\,[X_{t_1} - E(X_{t_1})\,]\,[\,X_{t_2} - E(X_{t_2})\,]\,).$$

6. Seja X_t um processo de Poisson com parâmetro λ que descreve o número de chamadas telefônicas recebidas por uma central telefônica. Qual a probabilidade de que nenhuma chamada seja recebida entre o tempo t_1 e t_2, com $t_1 < t_2$? Qual a probabilidade de que nenhuma chamada seja recebida entre o tempo t_1 e t_2, e duas sejam recebidas entre o tempo t_3 e t_4, onde $t_1 < t_2 < t_3 < t_4$?

5

Revisão de Teoria da Medida e Propriedades Gerais de Processos

5.1 Introdução

Vamos iniciar esta seção mencionando os resultados de Teoria da Medida que necessitaremos para formalizar rigorosamente a Teoria dos Processos Estocásticos. Após esta parte preliminar, iremos nos dedicar a utilizar tais resultados no contexto específico que necessitamos no texto.

Ótimas referências gerais sobre Teoria da Medida que complementam o que apresentamos aqui são [Ba] e [Fe]. As notas por S. Lalley in [Lal] abrangem parte do material do começo desta seção.

Em particular, o exemplo 5.8 vai descrever, de forma matematicamente, precisa o que é um processo Markoviano tomando valores num conjunto de estados S obtido a partir de uma matriz de transição $\mathcal{P}$ e uma probabilidade inicial π.

Definição 5.1. *Uma família $\mathcal{A} \subset \mathfrak{p}\ (X)$ de subconjuntos de um conjunto fixo X é denominada de uma σ-álgebra sobre X se:*

a) X está em $\mathcal{A}$,

b) se A pertence a $\mathcal{A}$, então seu complemento $X - A$ também está em $\mathcal{A}$, *c) Se $(A_n)_{n\in\mathbb{N}}$ é uma sequencia enumerável de conjuntos em $\mathcal{A}$, então sua união $\cup_{n\in\mathbb{N}} A_n$ também está em $\mathcal{A}$.*

Definição 5.2. *Os conjuntos A em $\mathcal{A}$ são denominados de conjuntos $\mathcal{A}$-mensuráveis ou, abreviadamente, mensuráveis.*

No nosso contexto o conjunto X será em geral o espaço amostral Ω.

O conjunto $\mathfrak{p}(X)$ das partes de X é sempre uma σ-álgebra. O conjunto dos conjuntos $\{\emptyset, X\}$ também é uma σ-álgebra. Estes dois exemplos são os casos extremos das possibilidades. Em geral vamos nos interessar em coleções de conjuntos que são casos intermediários. No entanto, toda vez que o conjunto X for enumerável, a única σ-álgebra que vamos considerar é o conjunto das partes de X, ou seja, $\mathfrak{p}(X)$.

O conjunto X que prioritariamente estaremos interessados aqui é $X = S^T$, onde $T = \mathbb{N}$ e S é o conjunto de estados. A análise do conjunto $X = S^{\mathbb{R}^+}$ também é de grande importância, mas sua abordagem é muito mais delicada (ver [EK]) e não vamos aqui aprofundar muito tal questão.

Observação: Na observação feita após o exemplo 5.10 ficará bem claro o sentido e a pertinência das exigências requeridas na definição de sigma-algebra.

Definição 5.3. *Dada uma família $\mathcal{A}_\theta \subset \mathfrak{p}(X)$ de σ-álgebras sobre o mesmo conjunto X indexadas por θ, denominamos $\cap_\theta \mathcal{A}_\theta$ a interseção de todas as σ-álgebras $\mathcal{A}_\theta$. Neste caso, se chamarmos de $\mathcal{F} = \cap_\theta \mathcal{A}_\theta$, então um subconjunto A de X está em $\mathcal{F}$, se e só se, $A \in \mathcal{A}_\theta$ para todo θ.*

Teorema 5.1. *Dado $\mathcal{A}_\theta \subset \mathfrak{p}(X)$, uma família de σ-álgebras sobre o conjunto X indexadas por θ, então $\mathcal{F} = \cap_\theta \mathcal{A}_\theta$ é uma σ-álgebra.*

Demonstração: Ora, os conjuntos $\emptyset$ e X estão, cada um deles, em todos os $\mathcal{A}_\theta$, logo, $\emptyset, X \in \mathcal{F}$, portanto a) na definição 5.1 é verdadeiro.

Se $A \in \mathcal{F}$, então para todo θ fixo, temos que $A \in \mathcal{A}_\theta$. Logo, como $\mathcal{A}_\theta$ é uma σ-álgebra, então $A^c = X - A \in \mathcal{A}_\theta$, para todo θ, assim $A^c \in \mathcal{F}$. Isto mostra b).

Para provar c), considere uma sequencia de conjuntos $A_n \in \mathcal{F}$ com $n \in \mathbb{N}$, então para cada n, temos que $A_n \in \mathcal{A}_\theta$, para todo θ. Logo, como $\mathcal{A}_\theta$ é uma σ-álgebra, $\cup_n A_n \in \mathcal{A}_\theta$, para todo θ, e assim segue que $\cup_n A_n \in \mathcal{F}$.

$\square$

Definição 5.4. *Seja $\mathcal{C} \subset \mathfrak{p}(X)$, uma coleção de subconjuntos de X (não necessariamente uma σ-álgebra). Seja $\mathcal{A}_\theta$ a coleção de todas as σ-álgebras que contém a coleção de conjuntos $\mathcal{C}$. A σ-álgebra $\mathcal{F}$ gerada por $\mathcal{C}$ é a interseção de todas as σ-álgebras $\mathcal{A}_\theta$ contendo a coleção $\mathcal{C}$. Denotamos por $\sigma(\mathcal{C})$ esta interseção.*

Como $\mathcal{C}$ está contido em $\mathfrak{p}(X)$, sempre existe uma σ-álgebra contendo $\mathcal{C}$. Sendo assim, o conceito acima sempre faz sentido.

Esta σ-álgebra $\sigma(\mathcal{C})$ é a menor σ-álgebra que contém o conjunto $\mathcal{C}$. Em outras palavras, se $\mathcal{G}$ é uma σ-álgebra e contém $\mathcal{C}$, então $\sigma(\mathcal{C}) \subset \mathcal{G}$.

Definição 5.5. *Seja $X = \mathbb{R}^n$, então a σ-álgebra de Borel é aquela gerada pelas bolas abertas de $\mathbb{R}^n$, isto é, a σ-álgebra gerada pelos conjuntos da forma $B(x,r) = \{y \in \mathbb{R}^n \mid \|x-y\| < r\}$, onde $x \in \mathbb{R}^n$ e $r > 0$. Existem subconjuntos $A \subset \mathbb{R}^n$ que não estão na σ-álgebra de Borel de $\mathbb{R}^n$ (ver [Fe] ou [Ba]). Ou seja, a σ-álgebra de Borel não é o conjunto $\mathfrak{p}(\mathbb{R}^n)$.*

Usaremos a notação $\mathcal{R}^n$ para a σ-álgebra de Borel em $\mathbb{R}^n$.

A única σ-álgebra que iremos considerar sobre $\mathbb{R}$ ou sobre $\mathbb{R}^n$ é a σ-álgebra de Borel.

No caso de $X = \mathbb{R}$, pode-se mostrar que a σ-álgebra de Borel também pode ser obtida como a σ-álgebra gerada por todos os intervalos da reta da forma (a, ∞), com a real (ver [F], [Fe]).

Alternativamente, pode-se mostrar também que a σ-álgebra de Borel sobre $\mathbb{R}$ pode ser obtida como a σ-álgebra gerada por todos os intervalos da reta da forma (a, b), com a, b reais e $a < b$ (ver [Fe]).

Ainda, pode-se mostrar também que a σ-álgebra de Borel sobre $\mathbb{R}^n$ pode ser obtida como a σ-álgebra gerada por todos os produtos de intervalos da forma $(a_1, b_1) \times (a_2, b_2) \times ... \times (a_n, b_n)$, com $a_i < b_i$ reais, e $i = 1, 2, ..., n$ (ver [Fe]).

Seja X um espaço métrico com métrica d. A bola aberta de centro $x \in X$ e raio ϵ é o conjunto $B(x, \epsilon) = \{y \in x \,|\, d(x, y) < \epsilon\}$.

Um subconjunto $A \subset X$ é dito um conjunto aberto se para qualquer $x \in A$ existe ϵ tal que $B(x, \epsilon) \subset A$.

Neste caso a σ-álgebra de Borel é aquela gerada pelos conjuntos abertos de X.

Ainda, dado um subconjunto fixo $X \subset \mathbb{R}^n$, a única σ-álgebra que iremos considerar sobre tal conjunto X é a σ-álgebra

$$\mathcal{F} = \{A \cap X \,|\, A \text{ é um elemento da } \sigma\text{-álgebra de Borel de } \mathbb{R}^n\}.$$

Esta σ-álgebra é denominada de σ-álgebra induzida em X pela σ-álgebra de Borel.

Dado um subconjunto B de $\mathbb{R}^n$, usaremos também a notação $\mathcal{R}^n$ para a σ-álgebra de Borel induzida em B pela σ-álgebra de Borel em $\mathbb{R}^n$.

Desta forma, $\mathcal{R}$ denotará tanto a σ-álgebra de Borel em $\mathbb{R}$ como a σ-álgebra de Borel induzida em $[0, 1]$. Acreditamos que isto não causará confusão ao leitor.

Considere um Processo Markoviano X_t, $t \in \mathbb{N}$, tomando valores em $S = \{1, 2, ..., d\}$. Seja $\mathcal{C}$ o conjunto de todos os cilindros de todos os possíveis tamanhos (finitos) $n \in \mathbb{N}$. O principal objeto de estudo deste texto é a σ-álgebra $\sigma(\mathcal{C})$.

Definição 5.6. *Uma lei μ que associa números não negativos em $\mathbb{R} \cup \{+\infty\}$ aos subconjuntos B em uma σ-álgebra $\mathcal{F}$ (isto é, $\mu(B) \geq 0$) é chamada de uma medida se satisfaz a propriedade:*

a) $\mu(\emptyset) = 0$,

b) Para qualquer coleção E_n, $n \in \mathbb{N}$, *de conjuntos em* $\mathcal{F}$ *tais que* $E_n \cap E_m = \emptyset$ *com* $m \neq n$, *então vale que* $\mu(\cup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \mu(E_n)$.

Se $\mu(X) = 1$ *dizemos que* μ *é uma probabilidade.*

Uma **medida com sinal** é uma lei μ que associa números em $\mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$ (aos subconjuntos B em uma σ-álgebra $\mathcal{F}$) que satisfaz as propriedades a) e b) (mas não se exige que $\mu(B) \geq 0$). O conjunto das medidas com sinal é um espaço vetorial para a estrutura natural de soma e multiplicação por escalar. Note que o conjunto das medidas não é um espaço vetorial.

A partir de agora vamos considerar prioritariamente medidas μ que são probabilidades.

Exemplo 5.1. Se $X = \{a_1, a_2, ..., a_n\}$ é finito e $\mathcal{F}$ é a σ-álgebra das partes de X, então escolhendo números $p_j \geq 0$, $j \in \{1, 2, ..., n\}$ tais que $\sum_{j=1}^{n} p_n = 1$, podemos definir $\mu(B)$ como a soma dos p_j tais que $a_j \in B$. É fácil ver que tal μ é uma probabilidade sobre a σ-álgebra das partes de X.

$\diamondsuit$

Exemplo 5.2. Seja X um conjunto qualquer e seja $\mathcal{F} = \mathfrak{p}(X)$. Fixado um ponto x de X, podemos definir μ de forma que $\mu(B) = 1$ se $x \in B$ e $\mu(B) = 0$ se x não está em B. É fácil ver que tal μ define uma probabilidade sobre X. Denominamos tal probabilidade de delta-Dirac em x e denotamos $\mu = \delta_x$.

$\diamondsuit$

Observamos que se μ é medida e $A \subset B$, então $\mu(A) \leq \mu(B)$. Isto porque

$$\mu(B) = \mu(\,(B \cap A^c) \,\cup\, A\,) = \mu(B \cap A^c) \,+\, \mu(A) \geq \mu(A).$$

Considere um Processo Markoviano X_t, $t \in \mathbb{N}$, tomando valores em $S = \{1, 2, ..., d\}$. Seja $\mathcal{C}$ o conjunto de todos os cilindros de todos os possíveis tamanhos (finitos) $n \in \mathbb{N}$. O principal objeto de estudo deste texto é a probabilidade P que se obtém de forma natural em $\sigma(\mathcal{C})$.

Definição 5.7. *Fixada uma medida (ou probabilidade) μ sobre uma σ-álgebra $\mathcal{A}$ em um certo conjunto X, dizemos que uma determinada propriedade é válida μ-quase toda parte, se existe um conjunto $\mathcal{A}$-mensurável K, com $\mu(K) = 0$, tal que a propriedade é válida para todos os pontos x em $X - K$.*

O principal objetivo da Teoria das Probabilidades é fazer afirmações que são válidas P-quase toda parte para uma probabilidade P.

Definição 5.8. *Denominamos de uma álgebra sobre X a uma coleção $\mathcal{F} \subset \mathfrak{p}(X)$ de subconjuntos de X, tal que $\mathcal{F}$ satisfaz as seguintes propriedades:*

a) Se B e A estão em $\mathcal{F}$, então $B - A$ também pertence a $\mathcal{F}$,

b) $X \in \mathcal{F}$.

c) Se $(A_n)_{n\in\{1,2,...,r\}}$ é uma coleção de conjuntos de $\mathcal{F}$, então a união finita $\cup_{n\in\{1,2,...,r\}} A_n$ também é um elemento de $\mathcal{F}$.

Qualquer σ-álgebra é uma álgebra.

Exemplo 5.3. Um exemplo de álgebra sobre $X = [0, 1] \subset \mathbb{R}$ é a a coleção $\mathcal{F}$ das uniões finitas

$$\cup_{n\in\{1,2,...,r\}} I_n,$$

onde $r \in \mathbb{N}$, e I_n são intervalos contidos em $[0, 1] \subset \mathbb{R}$. Os intervalos I_n podem ser da forma (a_n, b_n) , $[a_n, b_n)$, $(a_n, b_n]$ ou $[a_n, b_n]$ (onde $a_n < b_n$ são números reais).

$\diamondsuit$

Exemplo 5.4. Um exemplo que terá grande importância no que segue é o seguinte: considere S um conjunto enumerável e $X = S^{\mathbb{N}}$. Quando $X = S^{\mathbb{N}}$, consideramos na introdução os conjuntos cilindros C (com finitas restrições temporais), ou seja os conjuntos da forma

$$C = \{ w \in \Omega = S^{\mathbb{N}} \text{ tais que } X_{t_1}(w) = a_1,\, X_{t_2}(w) = a_2, ..., \, X_{t_n}(w) = a_n\} =$$

$$= \{w \in \Omega \,|\, w = (w_1, w_2, ..., w_n, w_{n+1}, w_{n+2}, ...), w_{t_i} = a_i, i \in \{1, 2, ..., n\}\},$$

onde $n \in \mathbb{N}$, $t_1 < t_2 < ... < t_n \in \mathbb{N}$, $a_1, a_2, ..., a_n \in S$.

Acima já estamos fazendo a simplificação canônica (descrita na seção 1) de modo que X_t é tal que $X_t(w) = w_t$.

Quando $t_i = i$, $i \in \{1, 2, ..., n\}$, então se denota

$$C = \overline{a_1, a_2, ..., a_n}.$$

Vamos considerar a seguir $\mathcal{C}$ o conjunto de todos os cilindros C de todos os tamanhos n.

Estaremos interessados na álgebra da uniões finitas de cilindros e na sigma álgebra gerada por tal coleção. Isto vai nos permitir analisar pr exemplo processo estocásticos X_n, $n \in \mathbb{N}$, com valores em $S = \{1, 2, ..., k\}$.

$\diamondsuit$

Vamos estender um pouco mais o conceito de cilindro.

Definição 5.9. *Seja k finito fixo, um cilindro de ordem k é um subconjunto C de $S^{\mathbb{N}}$ tal que exista $B \subset S^k$ tal que*

$$C = \{\omega \in S^{\mathbb{N}} \,|\, (X_1(\omega), X_2(\omega), ..., X_k(\omega)) \in B\} =$$

$$\{\omega = (w_1, w_2, .., w_k, ...) \in S^{\mathbb{N}} \,|\, (w_1, w_2, ..., w_k) \in B\}.$$

Denotamos por $\mathcal{F}_k$ a sigma algebra gerada pelos cilindros de ordem k.

Por exemplo se $S = \{1, 2\}$, $\Omega = \{1, 2\}^{\mathbb{N}}$, os cilindros de ordem 3 são

$$\overline{1,1,1},\ \overline{1,1,2},\ \overline{1,2,1},\ \overline{1,2,2},\ \overline{2,1,1},\ \overline{2,1,2},\ \overline{2,2,1},\ \overline{2,2,2}.$$

A sigma algebra $\mathcal{F}_3$ é a gerada por tais cilindros. Ela descreve a "informação"que obtemos a partir de X_1, X_2, X_3. Ou seja, se 1 e 2 estão associados a cara e coroa, a sigma-algebra $\mathcal{F}_3$ é a informação do jogo definido por lançar a moeda três vezes.

A sigma-algebra $\mathcal{F}_3$ é a coleção de todas as uniões finitas de cilindros da forma $\overline{a_1, a_2, a_3}$, $a_1, a_2, a_3 \in \{1, 2\}$.

Da mesma forma, a sigma-algebra $\mathcal{F}_k$ sobre $\Omega = S^{\mathbb{N}}$, é a coleção de todas as uniões finitas de cilindros da forma $\overline{a_1, a_2, a_3, ..., a_k}$, $a_1, a_2, a_3, ..., a_k \in S$.

Note que em geral (quando a sigma algebra não é gerada por uma coleção finita) não é verdade que a sigma algebra gerada seja constituida apenas por uniões finitas (ou, mesmo enumeráveis) dos geradores.

O conceito de cilindro anteriormente considerado está contemplado na versão acima, por exemplo,

$$C = \{X_1(\omega) = a_1, X_3(\omega) = a_3\} = \{X_1(\omega) = a_1, X_2(\omega) \in S, X_3(\omega) = a_3\},$$

onde $a_1, a_3 \in S$. Ou seja, neste caso o cilindro tem rank 3 (depende de três coordenadas) e B pode ser tomado como $B = \{a_1\} \times S \times \{a_2\}$, e assim

$$C = \{X_1(\omega) = a_1, X_3(\omega) = a_3\} =$$

$$\{(X_1(\omega), X_2(\omega), X_3(\omega)) \in B\}.$$

Naturalmente, existem algumas ambiguidades, por exemplo, podemos tomar $B' = \{a_1\} \times S \times \{a_2\} \times S$ e obteríamos da mesma forma

$$C = \{X_1(\omega) = a_1, X_3(\omega) = a_3\} =$$

$$\{(X_1(\omega), X_2(\omega), X_3(\omega), X_4(\omega)) \in B'\}.$$

Neste caso poderíamos também dizer que C tem rank 4.

Vamos denotar no que segue por $\mathcal{C}$, o conjunto de todos cilindros C com todos os ranks possíveis, ou seja, $k \in \mathbb{N}$, conforme a última definição. Note que o conjunto $X = S^{\mathbb{N}}$ está em $\mathcal{C}$, bastando tomar para isto k qualquer e $B = S^k$ na notação acima. Se $B = \emptyset$ então o resultante cilindro $C = \emptyset$.

Sejam C_1 e C_2 dois cilindros de rank k, dados respectivamente por

$$C_1 = \{\omega \in S^{\mathbb{N}} \,|\, (X_1(\omega), X_2(\omega), ..., X_k(\omega)) \in B\}$$

e

$$C_2 = \{\omega \in S^{\mathbb{N}} \,|\, (X_1(\omega), X_2(\omega), ..., X_k(\omega)) \in B'\},$$

onde $B, B' \subset S^k$.

É fácil ver que

$$C_1 \cup C_2 = \{\omega \in S^{\mathbb{N}} \,|\, (X_1(\omega), X_2(\omega), ..., X_k(\omega)) \in B \cup B'\}.$$

Da mesma forma,

$$C_1 - C_2 = \{\omega \in S^{\mathbb{N}} \,|\, (X_1(\omega), X_2(\omega), ..., X_k(\omega)) \in B - B'\}.$$

Estamos muito próximos de mostrar que $\mathcal{C}$ é de fato uma álgebra. Considere agora C_1 um cilindro de rank k e C_2 um cilindro de rank m. Sem perda de generalidade suponha que $k \geq m$

Podemos considerar que C_2 também é de rank k, basta substituir o correspondente B' por $B'' = B' \times S^{k-m}$.

Aplicando o que foi dito acima concluímos que, neste caso, também vale que $C_1 \cup C_2$ e também $C_1 - C_2$ são cilindros.

Podemos passar por indução a somas do tipo $C_1 \cup C_2 \cup C_3$ e assim por diante. Logo, as uniões finitas de cilindros são cilindros. Ficam assim comprovados os itens a), b) e c) da definição 5.8 de álgebra para a coleção dos cilindros $\mathcal{C}$.

Vamos denominar de $\mathcal{C}$ *a álgebra dos cilindros de* $S^{\mathbb{N}}$. A principal σ-álgebra $\mathcal{F}$ sobre $X = S^{\mathbb{N}}$ que iremos considerar aqui é $\mathcal{F} = \sigma(\mathcal{C})$.

Para evitar tecnicalidades desnecessárias, neste texto, vamos formalizar matematicamente apenas o caso $X = S^{\mathbb{N}}$ ou $X = S^{\mathbb{Z}}$, onde S é finito ou enumerável (ver [EK] para a análise do caso $X = S^{\mathbb{R}^+}$).

Definição 5.10. *Uma lei $\tilde{\mu}$ que associa a cada subconjunto B em uma álgebra $\mathcal{F}$ (sobre um conjunto X) um número não negativo em $\mathbb{R}$, (isto é, $\tilde{\mu}(B) \geq 0$) é chamada de uma lei aditiva se satisfaz a propriedade:*

para qualquer coleção B_n, $n \in \{1, 2, 3, ..., k\}$, com $k \in \mathbb{N}$, de conjuntos em $\mathcal{F}$ tais que $B_n \cap B_m = \emptyset$ com $m \neq n$, sabemos que $\cup_{n=1}^k B_n \in \mathcal{F}$, e assumimos que vale $\tilde{\mu}(\cup_{n=1}^k B_n) = \sum_{n=1}^k \tilde{\mu}(B_n)$.

Definição 5.11. *Uma lei $\tilde{\mu}$ que associa a cada subconjunto B em uma álgebra $\mathcal{F}$ (sobre um conjunto X) um número não negativo em $\mathbb{R}$, (isto é, $\tilde{\mu}(B) \geq 0$) é chamada de uma lei σ-aditiva se satisfaz a propriedade:*

para qualquer coleção B_n, $n \in \mathbb{N}$, de conjuntos em $\mathcal{F}$ tais que $B_n \cap B_m = \emptyset$ com $m \neq n$, se $\cup_{n=1}^\infty B_n \in \mathcal{F}$, então vale que $\tilde{\mu}(\cup_{n=1}^\infty B_n) = \sum_{n=1}^\infty \tilde{\mu}(B_n)$.

Note que, em geral, se E_n está na álgebra $\mathcal{F}$, nem sempre $\cup_{n=1}^\infty E_n$ está em $\mathcal{F}$. Lembre que $\mathcal{F}$ é fechado apenas para uniões finitas de elementos E_n. Acima, afirmamos que **se** $\cup_{n=1}^\infty E_n \in \mathcal{F}$, então vale algo.

Observamos que, em princípio, nem toda a lei σ-aditiva $\tilde{\mu}$, tal que $\tilde{\mu}(X) = 1$, é uma probabilidade, pois a primeira está definida sobre álgebras e a segunda sobre σ-álgebras.

Note que na definição 5.11 acima, se não existir elemento na álgebra $\mathcal{F}$ que possa ser escrito como união infinita de elementos da álgebra, então está satisfeita a hipótese por vacuidade. Ou seja, vale a lei σ-aditiva.

É usual a denominação de propriedade aditiva a relativa a validade de

$$\mu(\cup_{n=1}^k B_n) = \sum_{n=1}^k \mu(B_n),$$

e de propriedade σ-aditiva aquela relativa a validade de

$$\mu(\cup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} \mu(B_n),$$

ou seja, a relativa a soma de infinitos elementos.

Exemplo 5.5. O exemplo mais simples de lei aditiva $\tilde{\mu}$ é sobre a álgebra das uniões finitas dos intervalos I contidos em $X = [0,1] \subset \mathbb{R}$, onde

$$\tilde{\mu}([a,b]) = b - a = \text{ comprimento de } [a,b]$$

e onde $\tilde{\mu}(\cup_{n\in\{1,2,\ldots,r\}} I_n) = \sum_{n=1}^{r}$ " comprimento de (I_n)", $r \in \mathbb{N}$, quando os intervalos I_n acima são disjuntos. Referimos o leitor a [Fe] para a prova de tal afirmação.

Note que $(0, \frac{1}{2}] = \cup_{n\geq 2}(\frac{1}{n+1}, \frac{1}{n}]$.

$\diamondsuit$

Exemplo 5.6. Um exemplo de lei aditiva $\tilde{\mu}$ sobre a álgebra das uniões finitas dos intervalos finitos I contidos em $X = \mathbb{R}$ é a seguinte: defina

$$\tilde{\mu}([a,b]) = b - a = \text{ comprimento de } [a,b],$$

e $\tilde{\mu}(\cup_{n\in\{1,2,\ldots,r\}} I_n) = \sum_{n=1}^{r}$ " comprimento de (I_n)", $r \in \mathbb{N}$, quando os intervalos I_n acima são disjuntos. Referimos o leitor a [Fe] para a prova de tal afirmação.

$\diamondsuit$

Definição 5.12. *Dizemos que uma lei aditiva (ou σ-aditiva) $\tilde{\mu}$ sobre uma álgebra $\mathcal{F}$ é σ-finita se existe uma coleção enumerável de conjuntos mensuráveis $E_n \in \mathcal{F}, n \in \mathbb{N}$, tal que $X = \cup_{n\in\mathbb{N}} E_n$ e $\tilde{\mu}(E_n) < \infty$.*

O típico exemplo é a reta real (munida da algebra das uniões finitas de intervalos) que pode ser expressa como união de intervalos da forma $E_n = [-n, n]$, $n \in \mathbb{N}$, em que cada E_n é tal que $\tilde{\mu}(E_n) = 2\,n < \infty$.

No caso da medida do espaço X ser igual a um (o caso de interesse aqui) então vale a σ-finitude, é claro.

Teorema 5.2 (**Teorema da Extensão de Caratheodori-Kolmogorov**). *Dada uma lei σ-aditiva $\tilde{\mu}$ em X sobre uma álgebra σ-finita $\mathcal{F} \subset \mathfrak{p}(X)$, tal que $\tilde{\mu}(X) = 1$, existe um e apenas um meio de definir uma medida μ sobre $\mathcal{A} = \sigma(\mathcal{F})$, a σ- álgebra gerada por $\mathcal{F}$, tal que $\mu(B)$ e $\tilde{\mu}(B)$ coincidem sobre todo $B \in \mathcal{F}$. Se $\tilde{\mu}(X) = 1$ então μ é uma probabilidade.*

(Ver [Fe] para a demonstração).

Chamamos a probabilidade μ de extensão da pré-probabilidade $\tilde{\mu}$ sobre a σ-álgebra $\mathcal{A}$ gerada pela álgebra $\mathcal{F}$.

O teorema análogo considerando uma lei σ-aditiva σ-finita e resultando assim em uma única medida com sua extensão também é verdadeiro.

Em resumo, o teorema acima permite definir uma probabilidade μ agindo em subconjuntos “complicados" do espaço X a partir da informação dada por uma lei σ-aditiva $\tilde{\mu}$ agindo em subconjuntos de X com “natureza mais simples". Esta extensão é única no sentido acima.

A Observação 1 apresentada na introdução do livro e o exemplo analisado antes do Teorema 3.9, mostram a necessidade e a importância de se calcular a probabilidade de conjuntos que não são cilindros. Note que esta informação (sobre conjuntos não cilindros) depende, de qualquer forma, da "probabilidade" que provêm dos cilindros.

O teorema acima é extremamente geral e a única restrição na sua aplicação (a um dado exemplo específico) aparece ao se tentar mostrar a propriedade da σ-aditividade na álgebra, o que muitas vezes não é tão fácil. Felizmente, nos exemplos em que estamos prioritariamente interessados aqui, ou seja $S^{\mathbb{N}}$, este problema pode ser contornado como veremos em breve.

Exemplo 5.7. A medida μ usualmente considerada sobre a σ-álgebra de Borel em $\mathbb{R}$ (a σ-álgebra gerada pelos subintervalos da forma (a,b) de $\mathbb{R}$) é a extensão da pré-medida $\tilde{\mu}$ que é definida como $\tilde{\mu}(\,[a,b]\,) = b-a =$ comprimento de $(a,b) \subset \mathbb{R}$. Nós chamamos tal medida de Lebesgue sobre $\mathbb{R}$ e denotamos ela por dx ou $d\lambda$ e guardamos a expressão $\lambda(A)$ para o conjunto de Borel A contido em $\mathbb{R}$.

O conjunto com um único elemento $\{p\}$, onde p é um ponto em $\mathbb{R}$, é um conjunto da σ-álgebra de Borel (mencionada no último parágrafo), ou seja, $\{p\}$ é um conjunto mensurável Borel. Isto porque,

$$\{p\} = \cap_{n\in\mathbb{N}}(p-1/n, p+1/n).$$

Note que a medida de Lebesgue λ do conjunto $\{p\}$ neste caso é zero. Desta forma $\lambda(\,\{p\}\,)=0$.

Isto porque: $\lambda(\,(p-1/n, p+1/n)\,) = 2/n$, para todo $n\in\mathbb{N}$.

$\diamondsuit$

Note que nem sempre a medida de um conjunto da forma $\{p\}$ é nula (para uma probabilidade em geral).

Segue então da definição de medida (ou probabilidade) que qualquer conjunto A com um número enumerável de elementos tem medida de Lebesgue zero.

Segue também do fato que $\lambda(\{a\})=0, \lambda(\{b\})=0$ que vale

$$\lambda(\,(a,b)\,) = \lambda(\,[a,b]\,).$$

A σ-álgebra de Borel $\mathcal{B}$ sobre $\mathbb{R}$, denotada por $\mathcal{R}$, não é igual a coleção das partes de $\mathbb{R}$ (ver [Fe], Cor. 2.5.1).

Exemplo 5.8. Associado a um processo estocástico X_n, $n \in \mathbb{N}$, tomando valores em S, podemos considerar a sigma algebra $\mathcal{F}$ sobre $\Omega = S^{\mathbb{N}}$, gerada pela união de todos os cilindros de tamanho k (conforme definimos antes). Esta sigma algebra é a mais importante que vamos considerar.

Considere fixada uma certa lei P (obtida de maneira natural, como nos exemplos descritos na seção 1 para processos independentes ou markovianos) que associa números reais (não negativos) a cilindros em $S^{\mathbb{N}}$ (de todos os tamanhos). Desempenhará um papel muito importante aqui esta lei P que inicialmente sabemos ser apenas aditiva sobre *a álgebra* $\mathcal{C}$ *das uniões finitas de cilindros de* $S^{\mathbb{N}}$ (de todos os tamanhos). Tal lei será denominada de *lei aditiva* P *agindo sobre a álgebra* $\mathcal{C}$ *dos cilindros de* $S^{\mathbb{N}}$. Um fato fundamental será mostrar (ver Teorema 5.5) que, de fato, vale a σ-aditividade da lei P em $\mathcal{C}$, ou seja, se $\cup_{n=1}^{\infty} E_n \in \mathcal{C}$, então vale que $\tilde{\mu}(\cup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \tilde{\mu}(E_n)$.

Assim, pelo Teorema de Caratheodori-Kolmogorov obteremos uma probabilidade sobre $\mathcal{F}$. **Por exemplo, uma matriz estocástica $\mathcal{P}$ e uma probabilidade inicial π permitem definir a probabilidade dos cilindros $\overline{a_1, a_2, ..., a_k}$, para todo k. Desta forma o Teorema de Caratheodori-Kolomogorov nos permite falar da probabilidade de conjuntos quaisquer na sigma algebra $\mathcal{F}$ gerada pelos cilindros. Esta sigma-algebra será chamada de sigma-algebra de Borel sobre $S^{\mathbb{N}}$.**

Como veremos em breve a necessidade de se poder atribuir probabilidade a conjuntos que não são cilindros será de importância fundamental. Isto ficará claro quando enunciarmos o Teorema Ergódico.

Considere, por exemplo, $S = \{1, 2, ..., d\}$, uma matriz estocástica $\mathcal{P}$ do tipo d por d e um vetor de probabilidade inicial $\pi \in \mathbb{R}^d$. Ficam assim definidas as probabilidades dos cilindros $\overline{a_1, a_2, ..., a_k}$ de tamanho k e por extensão uma probabilidade P sobre a sigma algebra $\mathcal{F}$ gerada pelos cilindros. Denotamos por X_n, $n \in \mathbb{N}$, $X_n : (\{1, 2, ..., d\}^{\mathbb{N}}, \mathcal{F}, P) \to$

$(S, \mathfrak{p}(S))$ **o processo estocástico Markoviano associado. Mais precisamente** $X_n(w) = X_n(w_1, w_2, w_3, ..., w_n, ...) = w_n$.

$\diamondsuit$

Lembre que o shift σ age em $\Omega = S^{\mathbb{N}}$ da seguinte forma $\sigma(x_1, x_2, x_3, ...) = (x_2, x_3, ...)$.

Dizemos que uma probabilidade μ sobre Ω é **invariante** para o shift $\sigma : \Omega \to \Omega$ se para qualquer conjunto A Borel mensurável vale que $\mu(A) = \mu(\sigma^{-1}(A))$.

Mais geralmente uma probabilidade μ sobre X é **invariante** para $T : (\Omega, \mathcal{F}) \to (\Omega, \mathcal{F})$ se para qualquer conjunto Borel mensurável $A \in \mathcal{F}$ vale que $\mu(A) = \mu(T^{-1}(A))$.

Para que μ seja invariante para o shift basta mostrar que para conjuntos A que são uniões finitas de cilindros (a algebra que gera a sigma-algebra) vale que $\mu(A) = \mu(T^{-1}(A))$. Isto segue do fato que a probabilidade μ e a probabilidade ν tal que $\nu(A) = \mu(T^{-1}(A))$ coincidem na algebra dos conjuntos A que são uniões finitas de cilindros.

Definição 5.13. *Como vimos antes, dado um subconjunto H de X e uma σ-álgebra $\mathcal{F}$ em X, podemos associar uma outra σ-álgebra $\mathcal{G} = \mathcal{F}_H$ induzida em H, como aquela obtida tomando a interseção de todos os conjuntos da σ-álgebra $\mathcal{F}$ com H. Isto é, $\mathcal{F}_H = \{A \cap H \,:\, A \in \mathcal{F}\}$.*

Definição 5.14. *Se H é também $\mathcal{F}$-mensurável e fixamos uma medida (ou probabilidade) μ definida em $\mathcal{F}$, podemos considerar a medida induzida em $\mathcal{G}$, denotada por $\nu = \mu_H$ definida por $\nu(B) = \mu(B)$ para $B = H \cap A \in \mathcal{G}$, onde $A \in \mathcal{F}$. É fácil ver que tal ν é de fato uma medida. Se $\mu(H) < \infty$, então $\tilde{\nu}$ definida por*

$$\tilde{\nu}(B) = \frac{\mu(B)}{\mu(H)},$$

para $B \in \mathcal{G}$, define uma probabilidade sobre $\mathcal{G}$. Esta probabilidade será denominada de probabilidade induzida por μ sobre H.

Quando consideramos a σ-álgebra de Borel em $[0,1]$ e a medida de Lebesgue induzida em $[0,1]$, estamos nos referindo ao que descrevemos acima.

Definição 5.15. *Vamos denotar por $(X, \mathcal{F})$ um conjunto geral X munido de uma σ-álgebra $\mathcal{F}$ fixada. Este par será denominado de um Espaço Mensurável.*

Definição 5.16. *Vamos denotar por $(X, \mathcal{F}, P)$ um conjunto geral X munido de uma σ-álgebra $\mathcal{F}$ fixada e também uma probabilidade P fixada. Chamaremos esta tripla de Espaço de Probabilidade.*

Definição 5.17. *Considere um conjunto X equipado com uma σ-álgebra $\mathcal{A}$ e Y outro conjunto equipado com uma σ-álgebra $\mathcal{G}$. Uma função $\phi : X \to Y$ é dita mensurável se $\phi^{-1}(A) \in \mathcal{A}$ para todo $A \in \mathcal{G}$.*

É usual a notação $\phi : (X, \mathcal{A}) \to (Y, \mathcal{G})$ para descrever o fato acima.

É fácil ver que composição de funções mensuráveis é mensurável.

Quando $Y = \mathbb{R}$ (e assim, $\mathcal{G}$ é a σ-álgebra de Borel, gerada pelos intervalos), se considerarmos funções mensuráveis $\phi_1 : (X, \mathcal{A}) \to (Y, \mathcal{G})$ e $\phi_2 : (X, \mathcal{A}) \to (Y, \mathcal{G})$, então $\phi_1 + \phi_2$, $\phi_1\,\phi_2$, e o produto cartesiano $\phi_1 \times \phi_2$ são também funções mensuráveis ([Fe] e [B]).

Por $g = \phi_1 \times \phi_2$ deve se entender a função $g(x) = (\phi_1(x), \phi_2(x)) \in \mathbb{R}^2$, $g : X \to \mathbb{R}^2$.

Se $\phi_1, \phi_2, ..., \phi_k$ são mensuráveis então $\phi_1 + \phi_2 + ... + \phi_k$ é mensurável bem como o produto cartesianoo $\phi_1 \times \phi_2 \times ... \times \phi_k : X \to \mathbb{R}^k$.

Quando a função $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{R})$ é mensurável, então é usual chamar X de **Variável Aleatória**.

Quando consideramos uma transformação mensurável T definida sobre o espaço mensurável $(X, \mathcal{A})$ tomando valores neste mesmo espaço, isto é, $T : (X, \mathcal{A}) \to (X, \mathcal{A})$ chamamos tal T de endomorfismo sobre $(X, \mathcal{A})$.

É fácil ver que se $f : X \to Y$ é uma função qualquer, então dado uma σ-álgebra $\mathcal{G}$ sobre o conjunto Y, temos que a coleção de conjuntos da forma $f^{-1}(A) \subset X$, onde $A \in \mathcal{G}$, define uma σ-álgebra $\mathcal{A}$ sobre X. Nós chamamos a σ-álgebra $\mathcal{A}$ de σ-álgebra induzida em X por f e $\mathcal{G}$. Neste caso, nós denotamos $\mathcal{A} = f^{-1}\mathcal{G}$. Desta forma se pode encontrar facilmente exemplos de transformações $f : (X, \mathcal{A}) \to (Y, \mathcal{G})$.

Dada uma coleção Z de subconjuntos de Y, considere $\sigma(Z)$ a menor σ-álgebra contendo Z.

Observação: Considere $S = \{1, 2\}$, $\Omega = \{1, 2\}^{\mathbb{N}}$. Então, os cilindros de ordem 3 são

$$\overline{1,1,1},\ \overline{1,1,2},\ \overline{1,2,1},\ \overline{1,2,2},\ \overline{2,1,1},\ \overline{2,1,2},\ \overline{2,2,1},\ \overline{2,2,2}.$$

A sigma-algebra $\mathcal{F}_3$ é o conjunto das uniões finitas de cilindros de tamanho 3. Afirmamos que uma função $F : (\{1, 2\}^{\mathbb{N}}, \mathcal{F}_3) \to (\mathbb{R}, \mathcal{R})$ é mensurável, se e só se, F é constante em cilindros de tamanho 3. De fato, por exemplo, suponha que F assuma dois valores distintos α e β sobre o cilindro $\overline{a_1, a_2, a_3}$. Seja γ um número entre α e β. Então $F^{-1}(\,(\gamma, \infty)\,) \cap \overline{a_1, a_2, a_3}$, define um subconjunto J não vazio e estritamente contido em $\overline{a_1, a_2, a_3}$. O conjunto J claramente não está na sigma-algebra $\mathcal{F}_3$.

A afirmação análoga é válida para $\mathcal{F}_k$. Ou seja, $F : (\{1, 2\}^{\mathbb{N}}, \mathcal{F}_k) \to (\mathbb{R}, \mathcal{R})$ é mensurável, se e só se, F é constante em cilindros de tamanho k.

Teorema 5.3. *Se $f : X \to Y$ é uma função, então $f^{-1}(\sigma(Z)) = \sigma(f^{-1}(Z))$, onde $f^{-1}(Z)$ é a coleção de conjuntos $f^{-1}(B)$, onde $B \in Z$.*

Demonstração: É fácil ver que se $Z_1 \subset Z_2$ então $\sigma(Z_1) \subset \sigma(Z_2)$. Logo, como $f^{-1}(Z) \subset f^{-1}(\sigma(Z))$, então $\sigma(f^{-1}(Z)) \subset \sigma(f^{-1}(\sigma(Z)))$. A coleção de conjuntos $f^{-1}(\sigma(Z))$ é uma σ-álgebra, portanto

$$\sigma(f^{-1}(Z)) \subset f^{-1}(\sigma(Z)).$$

Considere a coleção

$$\mathcal{D} = \{A \subset Y; f^{-1}(A) \in \sigma(f^{-1}(Z))\}.$$

É fácil ver que $\mathcal{D}$ é uma σ-álgebra e contém Z. Logo, contém $\sigma(Z)$. A conclusão é que $f^{-1}(A) \in \sigma(f^{-1}(Z))$ para todo conjunto $A \in \sigma(Z)$ e isto significa que

$$f^{-1}(\sigma(Z)) \subset \sigma(f^{-1}(Z)).$$

□

Teorema 5.4. *Considere um conjunto X equipado com uma σ-álgebra $\mathcal{A}$ e Y outro conjunto equipado com uma σ-álgebra $\mathcal{G}$. Suponha que o conjunto $\mathcal{E}$ gera a σ-álgebra $\mathcal{G}$, isto é, $\mathcal{G} = \sigma(\mathcal{E})$. Então, $f : X \to Y$ é mensurável, se e só se, $f^{-1}(B) \in \mathcal{A}$ para todo conjunto $B \in \mathcal{E}$.*

Demonstração: Se f é mensurável então é claro que $f^{-1}(B) \in \mathcal{A}$ para todo $B \in \mathcal{E}$.

Suponha agora que, $f^{-1}(B) \in \mathcal{A}$ para todo conjunto $B \in \mathcal{E}$.

Então, $\sigma(f^{-1}(\mathcal{E})) \subset \mathcal{A}$. Como $f^{-1}(\sigma(\mathcal{E})) = \sigma(f^{-1}(\mathcal{E}))$, então, $f^{-1}\mathcal{G} \subset \mathcal{A}$.

Acima usamos o Teorema 5.3.

□

Exercício: Mostre que a transformação $T(x) = 2x \pmod 1$ sobre o conjunto $[0, 1)$, com a σ-álgebra de Borel $\mathcal{B}$, é mensurável,

$$T : ([0, 1), \mathcal{B}) \to ([0, 1), \mathcal{B}).$$

(Use o fato que imagem inversa de um intervalo $I \subset [0,1]$, ou seja, o conjunto $T^{-1}(I)$, é a união de dois intervalos e os intervalos geram a σ-álgebra de Borel).

Exemplo 5.9. Suponha que X esteja munido de uma métrica d. Considere a álgebra dos abertos $\mathcal{C}$ e seja $\mathcal{A} = \sigma(\mathcal{C})$. Tal σ-álgebra é denominada de Borel sobre (X,d). Considere uma função contínua $f : X \to \mathbb{R}$. Como vale que a imagem inversa $f^{-1}(A)$ de um intervalo $A \in \mathcal{R}$ é um aberto em X, temos pela proposição acima que qualquer função contínua é mensurável $f(X,\mathcal{A}) \to (\mathbb{R},\mathcal{R})$.

Da mesma forma, dados X_1 com um distância d_1 e X_2 com um distância d_2, se $f : X_1 \to X_2$ for contínua também vale que $f : (X_1,\mathcal{A}_1) \to (X_2,\mathcal{A}_2)$ é mensurável, onde $\mathcal{A}_1$ é a σ-álgebra de Borel de X_1 e $\mathcal{A}_2$ é a σ-álgebra de Borel de X_2.

O espaço de Bernoulli $\Omega = S^{\mathbb{N}}$ pode ser equipado com um distância natural $d_\theta : \Omega \times \Omega \to \mathbb{R}$ da seguinte forma: para um valor fixado θ com $0 < \theta < 1$, definimos para $x = (x_1, x_2, ..., x_n, ...)$ e $y = (y_1, y_2, ..., y_n, ...)$ o valor $d_\theta(x,y) = \theta^N$, onde N é o menor número natural tal que $x_i \neq y_i$, para $1 \leq i \leq N-1$, e $x_N = y_N$, se x é diferente de y. Quando x é igual a y então declare que $d(x,y) = 0$.

Neste caso, temos que $d(x,y) < \theta^k$, se e só se,

$$(x_1, x_2, ..., x_k) = (y_1, y_2, ..., y_k).$$

A sigma algebra sobre $S^{\mathbb{N}}$ gerada pelos cilindros é igual a sigma algebra gerada pelos abertos com a métrica D acima definida.

Se S é finito, então Ω é compacto.

O espaço de Bernoulli $\Omega = S^{\mathbb{Z}}$ também pode ser equipado com um distância natural $d_\theta : \Omega \times \Omega \to \mathbb{R}$ da seguinte forma: para um valor fixado θ com $0 < \theta < 1$, definimos para $x = (..., x_{-2}, x_{-1} \,|\, x_1, x_2, ..., x_n, ...)$ e $y = (..., y_{-2}, y_{-1} \,|\, y_1, y_2, ..., y_n, ...)$ o valor $d_\theta(x,y) = \theta^N$, onde N é o menor número natural tal que $x_i \neq y_i$ para $i \in \{-N+1, ..., -1, 1, 2, ..., N-1\}$ e $x_N = y_N$, ou $x_{-N} = y_{-N}$.

Se S é finito, então Ω é compacto.

$\diamondsuit$

Exemplo 5.10. Seja $\theta = 0.3$, $x = (1, 1, 0, 1, 0, 0, 1, ...)$ e $\epsilon = 0,0081 = (0,3)^4$, então $B(x, \epsilon)$ (a bola aberta de centro x e raio ϵ) é igual a $\{X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 1\}$.

É fácil ver que todo cilindro C em $\mathcal{C}$ é um conjunto aberto. Por exemplo, se C tem rank k, então, se $x \in C$, temos que a bola aberta de centro x e raio θ^k está contida em C.

Como todo cilindro é um aberto, pode-se mostrar que a σ-álgebra de Borel associada a tal métrica coincide com a σ-álgebra gerada pelos cilindros.

Cada cilindro também é fechado, ou seja o complementar de um aberto.

$\diamondsuit$

Quando S é finito, o espaço $S^{\mathbb{N}}$, com a métrica d acima definida é um espaço métrico compacto.

Exemplo 5.11. Exemplo Importante: Considere a transformação shift σ agindo em $\{1, 2, .., d\}^{\mathbb{N}}$ dada por

$$\sigma(x_1, x_2, ..., x_n, ..) = (x_2, x_3, ..., x_n, ..),$$

que claramente não é bijetiva.

Uma pergunta natural é se $\sigma : \{1, 2, .., d\}^{\mathbb{N}} \to \{1, 2, ..., d\}^{\mathbb{N}}$ é mensurável quando consideramos a sigma-álgebra de Borel.

Observe que $\sigma^{-1}(\overline{a_1, a_2, .., a_n}) = \cup_{j=1}^{d} \overline{j, a_1, a_2, .., a_n}$.

Segue que a imagem inversa de cilindros (os geradores da sigma álgebra de Borel) é uma união finita de cilindros (que está neste anel), que por sua vez gera a sigma-álgebra de Borel. Assim, σ é mensurável.

Dizemos que uma probabilidade μ sobre $\{1, 2, ..., d\}^{\mathbb{N}}$ é invariante para σ se para qualquer Boreleano A vale que $\mu(\sigma^{-1}(A)) = \mu(A)$. Como veremos em futura seção esta propriedade está associada a processos estocásticos X_n, $n \in \mathbb{N}$, que são estacionários.

Suponha que a propriedade acima seja válida quando A é um cilindro. Afirmamos que isto basta para afirmar que μ é invariante por σ. De fato, considere a lei sigma-aditiva $\tilde{\mu}$ tal que $\tilde{\mu}(A) = \mu(\sigma^{-1}(A))$, para todo conjunto de Borel A.

Em particular

$$\overline{a_1, a_2, .., a_n} \to \tilde{\mu}(\overline{a_1, a_2, .., a_n}) = \mu(\sigma^{-1}(\overline{a_1, a_2, .., a_n}))$$

define uma lei no anel dos geradores da sigma-álgebra de Borel.

Estamos assumindo que $\mu(\sigma^{-1}(\overline{a_1, a_2, .., a_n})) = \mu(\overline{a_1, a_2, .., a_n})$.

A extenção de $\tilde{\mu}$ definida no anel a sigma-algebra de Borel é unica se $\tilde{\mu}$ é finitamente aditiva no anel dos geradores da sigma-álgebra de Borel. Note que se $A_1, A_2, ..., A_n$ são cilindros disjuntos, então a família $B_j = \sigma^{-1}(A_j)$, $j = 1, 2, ..., n$, também é disjunta e

$$\tilde{\mu}(\cup_j A_j) = \sum_j \tilde{\mu}(A_j) = \sum_j \mu(\sigma^{-1}(A_j)) = \sum_j \tilde{\mu}(A_j).$$

Assim, $\tilde{\mu}$ no anel se estende de forma única sigma-álgebra de Borel. Mas como $\tilde{\mu}(\overline{a_1, a_2, .., a_n}) = \mu(\overline{a_1, a_2, .., a_n})$ nos geradores, então $\mu = \tilde{\mu}$ na sigma-álgebra de Borel.

Logo para todo Boreleano A vale que $\mu(A) = \tilde{\mu}(A) = \mu(\sigma^{-1}(A))$. Assim, μ é σ-invariante.

$\diamondsuit$

Observação: Considere $S = \{1,2\}$, $\Omega = \{1,2\}^{\mathbb{N}}$. Então é mesurável a função $I_{\overline{1,1,2}}$ que é a função indicador do cilindro $\overline{1,1,2}$. Mais geralmente, se $S = \{1,2,...,d\}$ e $\Omega = \{1,2,...,d\}^{\mathbb{N}}$, a função indicador de qualquer cilindro é mensurável.

Ainda, note que dado um processo estocástico X_n, $n \in \mathbb{N}$, $\alpha \in \mathbb{R}$ e $\epsilon > 0$, o conjunto

$$\{w \,|\, |(X_1 + X_2 + ... + X_n)(w) - \alpha| < \epsilon\},$$

é mensurável em relação a sigma-algebra de Borel sobre $\Omega = S^{\mathbb{N}}$.

Em muitos casos podemos estar interessados em dizer que existe α tal que para P-quase todo $w \in \Omega$ vale que

$$\lim_{n\to\infty} (X_1 + X_2 + ... + X_n)(w) = \alpha.$$

Por exemplo no caso da moeda honesta é natural pensar que $\alpha = 1/2$.

Para fazer sentido a afirmação é necessário que seja mensurável o conjunto de tais w.

Observe então que é mensurável o conjunto

$$\cap_{k\in\mathbb{N}} \cup_{N\in\mathbb{N}} \cap_{n>N} \{w \,|\, |(X_1 + X_2 + ... + X_n)(w) - \alpha| < 1/k\}.$$

Fica bem claro neste momento que as exigencias feitas na definição de sigma algebra são suficientemente adequadas para a formalização das questões mais importantes que podemos estar interessados.

Definição 5.18. *Seja o espaço mensurável* $(\Omega, \mathcal{F})$. *Uma classe de conjuntos* $\mathcal{B} \subset \mathcal{F}$ *é denominada de um* **sistema-**π *se pra qualquer* $A, B \in \mathcal{B}$ *vale que* $A \cap B \in \mathcal{B}$.

Teorema 5.5. *Considere duas probabilidades* P_1 *e* P_2 *sobre* $(\Omega, \mathcal{F})$, *e uma classe de conjuntos* $\mathcal{B} \subset \mathcal{F}$ *que seja um sistema-*π. *Suponha que para qualquer elemento* $B \in \mathcal{B}$ *vale que*

$$P_1(B) = P_2(B),$$

então, para qualquer elemento $B \in \sigma(\mathcal{B})$ *vale que*

$$P_1(B) = P_2(B).$$

A demonstração deste teorema pode ser encontrada em [B] seção 3. O resultado em si é muito útil, mas o conhecimento da sua prova não é de fundamental importância para o que segue. Não vamos usar tal teorema nesta seção mas sim na próxima.

Quando consideramos um Processo Estocástico X_n, $n \in \mathbb{N}$, com valores num conjunto enumerável S, é natural, como vimos antes, que a correspondente lei P seja definida a partir das distribuições finito-dimensionais (ou dito, de outra forma, sobre os cilindros). Após algumas considerações e resultados que seguem, utilizando basicamente o Teorema de Caratheodori-Kolmogorov, tal P se estenderá a uma probabilidade sobre a σ-álgebra de Borel $\mathcal{F} = \sigma(\mathcal{C})$ gerada pelos cilindros $\mathcal{C}$ sobre $\Omega = S^{\mathbb{N}}$. Para simplificar, quando não ocorrer confusão, denotaremos também por P tal probabilidade.

Como se pode assegurar que uma dada lei aditiva numa algebra é na verdade sigma-aditiva? Isto é necessário para utilizar o teorema de extensão à sigma-algebra gerada.

Para nos auxiliar na análise da questão acima, necessitamos dos seguintes resultados:

Teorema 5.6. *Suponha que* $\tilde{\mu}$ *seja uma lei aditiva em uma álgebra* $\mathcal{F}$ *(sobre um conjunto* X*) e que vale a propriedade:*

para toda seqüência C_n, $n \in \mathbb{N}$.*, de conjuntos em* $\mathcal{F}$ *tal que* $C_{n+1} \subset C_n$ *para todo* n*, se* $\emptyset = \cap_{n\in\mathbb{N}} C_n$*, então* $\lim_{n\to\infty} \tilde{\mu}(C_n) = 0$.

Então $\tilde{\mu}$ *é* σ*-aditiva sobre a álgebra* $\mathcal{F}$.

Demonstração: Temos que mostrar que vale a σ-aditividade. Seja uma coleção B_n, $n \in \mathbb{N}$, de conjuntos em $\mathcal{F}$ tais que $B_n \cap B_m = \emptyset$ com $m \neq n$, e suponha que $\cup_{n=1}^{\infty} B_n = A \in \mathcal{F}$.

Seja, $C_n = A - \cup_{j<n} B_j$, então C_n está em $\mathcal{F}$ e C_n satisfaz a propriedade descrita acima (na hipótese do presente teorema).

Logo, $\tilde{\mu}(C_n) \to 0$, quando $n \to \infty$.

Note que para todo n fixo vale $A = C_n \cup (\cup_{j<n} B_j)$, logo

$$\tilde{\mu}(A) = \tilde{\mu}(C_n) + \sum_{j<n} \tilde{\mu}(B_j),$$

pois C_n é disjunto de $\cup_{j<n} B_j$.

Tomando limite, quando n vai a infinito, na expressão acima mostramos que

$$\tilde{\mu}(A) = \sum_{j\in\mathbb{N}} \tilde{\mu}(B_j).$$

□

Desejamos utilizar o resultado acima quando a álgebra $\mathcal{F}$ é a das uniões finitas de cilindros.

O resultado acima nos dá um critério muito útil que nos permite obter a propriedade da σ-aditividade sobre uma álgebra, requisito para se aplicar o Teorema de Caratheodori-Kolmogorov.

Note que $C_n = \underbrace{\overline{111...11}}_{n} - \{(11111111111111...)\}$, em $\{1,2\}^{\mathbb{N}}$, é tal que C_n decresce ao $\emptyset$, quando $n \to \infty$. Esta propriedade explica de maneira apropriada a questão brevemente discutida na seção inicial sobre o caso em que se lança sucessivamente uma moeda honesta e associamos 1 à cara e 2 à coroa. Sair apenas cara sucessivamente infinitas vezes é um evento que tem probabilidade zero.

Teorema 5.7. *Seja S finito, $\mathcal{C}$ a álgebra dos cilindros em $S^{\mathbb{N}}$ e ν uma lei finitamente aditiva sobre $\mathcal{C}$. Então ν é σ-aditiva sobre $\mathcal{C}$.*

Antes de demonstrar tal proposição necessitamos do seguinte lema.

Lema 5.1. *Seja S finito. Suponha que $A_n, n \in \mathbb{N}$, seja uma sequencia decrescente de uniões finitas de cilindros (não vazios) em $S^{\mathbb{N}}$, tais que $\cap_{n\in\mathbb{N}} A_n = A$. Então A é não vazio.*

Demonstração:

Lembre que cada cilindro é um conjunto compacto quando S é finito.

Este resultado pode ser obtido do seguinte resultado geral bem conhecido em espaços métricos: Uma sequencia decrescente de compactos não vazios é tal que a interceção infinita deles é um conjunto não vazio (ver [Li3]). De fato, $\Omega = S^{\mathbb{N}}$ é compacto e cada cilindro é um conjunto fechado (assim compacto). Ainda, a união finita de compactos é compacto. Logo, a união finita de cilindros é um conjunto compacto.

Para o leitor que não conhece este resultado vamos apresentar abaixo uma prova ao caso particular que tratamos aqui.

A demonstração não se altera se consideramos que $A_n, n \in \mathbb{N}$, é uma sequência decrescente de cilindros e não uma sequência decrescente de uniões finitas de cilindros

Vamos denotar, sem perda de generalidade, o cilindro A_n por

$$A_n \,=\, \{X_1 \in V_1^n, X_2 \in V_2^n, ..., X_{t_n} \in V_{t_n}^n\}$$

com $t_n \in \mathbb{N}$, e $V_j^n \subset S$.

Note que como por hipótese $A_{n+1} \subset A_n$, temos que para todo n fixo vale que $t_n \leq t_{n+1}$, e $V_j^{n+1} \subset V_j^n$.

Observe que (como cada cilindro é não vazio) t_n converge a infinito quando $n \to \infty$, senão o resultado seria trivial, ou seja $A \neq \emptyset$.

Escolha um elemento ω^n em cada A_n. Denotamos

$$\omega^n = (w_1^n, w_2^n, w_3^n, ..., w_m^n, ...)$$

para n fixo em $\mathbb{N}$. Logo, $w_j^n \in V_j^n$ para $j \leq t_n$

Considere agora a sequência $w_1^n, n \in \mathbb{N}$, de elementos em S. Como S é finito, existe um elemento s_1 tal que existem infinitos índices $n \in \mathbb{N}$ tais que $w_1^n = s_1$. Denote em ordem crescente estes tais índices n através de n_1^k, com $k \in \mathbb{N}$. Observe que por construção $n_1^k \geq k$.

Considere agora os elementos $\omega^{n_1^k} = (w_1^{n_1^k}, w_2^{n_1^k}, w_3^{n_1^k}, ...)$, $k \in \mathbb{N}$, e os correspondentes $w_2^{n_1^k}$ com $k \in \mathbb{N}$. Como S é finito existe um elemento s_2 tal que existem infinitos índices $k \in \mathbb{N}$ tais que $w_2^{n_1^k} = s_2$. Denote em ordem crescente estes índices n_1^k através de n_2^r, com $r \in \mathbb{N}$. Logo, $w_2^{n_2^r} = s_2$. Note que por construção $n_2^r \geq r$.

O procedimento agora se repete: considere agora os elementos $\omega^{n_2^r} = (w_1^{n_2^r}, w_2^{n_2^r}, w_3^{n_2^r}, ...)$, $r \in \mathbb{N}$, e os correspondentes $w_3^{n_2^r}$ com $r \in \mathbb{N}$. Como S é finito existe um elemento s_3 e uma sequência n_3^v com $v \in \mathbb{N}$, tal que $w_3^{n_3^v} = s_3$ e os n_3^v são obtidos do conjunto dos n_2^r.

Note que

$$X_1(\omega^{n_2^r}) = s_1, X_2(\omega^{n_2^r}) = s_2,$$

para todo $r \in \mathbb{N}$. Ainda,

$$X_1(\omega^{n_3^v}) = s_1, X_2(\omega^{n_3^v}) = s_2, X_3(\omega^{n_3^v}) = s_3,$$

para todo $v \in \mathbb{N}$.

Dito de outra forma

$$w_1^{n_2^r} = s_1, w_2^{n_2^r} = s_2,$$

para todo $r \in \mathbb{N}$. Ainda,

$$w_1^{n_3^v} = s_1, w_2^{n_3^v} = s_2, w_3^{n_3^v} = s_3$$

para todo $v \in \mathbb{N}$.

Em particular,

$$w_1^{n_3^3} = s_1, w_2^{n_3^3} = s_2, w_3^{n_3^3} = s_3.$$

Desta forma, procedendo de forma indutiva, obtemos uma seqüência de elementos $s_b \in S$ para todo $b \in \mathbb{N}$, e uma família de índices n_u^j com $j, u \in \mathbb{N}$, tal que $w_b^{n_u^j} = s_b$ para todo $b \leq u$, com $u \in \mathbb{N}$. Pela mesma razão como acima vale que $n_u^j \geq j$.

Considere agora $\omega = (s_1, s_2, s_3, ...)$. Afirmamos que $\omega \in \cap_{m\in\mathbb{N}} A_m$. De fato, fixando m vamos mostrar que ω está em A_m. Escolha u tal que $u > t_m$ e $u > m$. Note que como $n_u^u \geq u$, temos que $(s_1, s_2, ..., s_{t_m}) = (w_1^{n_u^u}, w_2^{n_u^u}, ..., w_{t_m}^{n_u^u})$. De fato, $\{n_u^j, j \in \mathbb{N}\}$ está contido no conjunto $\{n_1^k, k \in \mathbb{N}\}$, e no conjunto $\{n_2^r, r \in \mathbb{N}\}$, e assim por diante até $u - 1$.

Assim, como $V_j^{n_u^u} \subset V_j^m$, então

$$X_1(\omega) = s_1 \in V_1^m, X_2(\omega) = s_2 \in V_2^m, ..., X_{t_m}(\omega) = s_{t_m} \in V_{t_m}^m,$$

para todo $v \in \mathbb{N}$.

$\square$

Agora que demonstramos o lema vamos demonstrar o Teorema 5.7

Demonstração: A ideia é mostrar agora que dada uma seqüência C_n, $n \in \mathbb{N}$ de conjuntos cilindros em $\mathcal{C}$, a álgebra gerada pelos cilindros em $S^{\mathbb{N}}$, tal que $C_{n+1} \subset C_n$ para todo n, se $\emptyset = \cap_{n\in\mathbb{N}} C_n$, então $\lim_{n\to\infty} \tilde{\mu}(C_n) = 0$.

O último teorema seguirá então do penúltimo.

A demonstração é por contradição. Considere uma seqüência decrescente de conjuntos C_n (união finita disjunta de cilindros), tal que existe $a > 0$ satisfazendo $\tilde{\mu}(C_n) \geq a$. Logo, cada C_n é não vazio. Então, pelo Lema acima, $A = \cap_{n\in\mathbb{N}} C_n$ é não vazio.

$\square$

A partir do resultado acima, uma grande classe de probabilidades (σ-aditivas, naturalmente) ficam definidas na σ-álgebra gerada por $\mathcal{C}$ (os cilindros de $S^{\mathbb{N}}$), a partir de leis finitamente aditivas sobre $\mathcal{C}$. Entre estes se encontram os processos independentes e os Markovianos, por exemplo. Em geral não é difícil mostrar a aditividade finita sobre $\mathcal{C}$. Vamos mostrar alguns exemplos agora.

O primeiro exemplo que vamos considerar é a construção de processos independentes (e identicamente distribuídos).

Suponha S finito com n elementos (o caso de S infinito é em tudo similar) e $p_1, p_2, ..., p_n$, números reais tais que $0 \leq p_i \leq 1$ e que satisfazem

$$p_1 + p_2 + ... + p_n = 1.$$

Dado um cilindro C da forma

$$C = \{\omega \,|\, (X_1(\omega), X_2(\omega), ..., X_m(\omega)) = (a_1, a_2, ..., a_m)\},$$

onde $a_1, a_2, ..., a_m \in S$ estão fixados, definimos

$$P(C) = p_{a_1}\, p_{a_2} \, ... p_{a_m}.$$

O cilindro geral de rank m pode ser escrito como

$$C = \{\omega \,|\, (X_1(\omega), X_2(\omega), ..., X_m(\omega)) \in B\},$$

onde $B \subset S^m$.

Neste caso definimos

$$P(C) = \sum_{(a_1, a_2, ..., a_m) \in B} p_{a_1}\, p_{a_2} \, ... p_{a_m}.$$

Novamente, algumas ambiguidades podem aparecer, e isto, em princípio, poderia tornar esta P não bem definida. Afirmamos, no entanto, que a P está bem definida. De fato, suponha que C possa também ser definida por

$$C = \{\omega \,|\, (X_1(\omega), X_2(\omega), ..., X_k(\omega)) \in B'\},$$

onde $B' \subset S^k$.

Sem perda de generalidade suponha que $k > m$. Logo, $B' = B \times S^{k-m}$.

Logo,

$$P(C) = \sum_{(a_1,a_2,...,a_k)\in B'} p_{a_1}\, p_{a_2} \dots p_{a_k} =$$

$$\sum_{(a_1,a_2,...,a_m)\in B} p_{a_1}\, p_{a_2} \dots p_{a_m} \sum_{(a_{m+1},a_{m+2},...,a_k)\in S^{k-m}} p_{a_{m+1}}\, p_{a_{m+2}} \dots p_{a_k} =$$

$$\sum_{(a_1,a_2,...,a_m)\in B} p_{a_1}\, p_{a_2} \dots p_{a_m} \sum_{(a_{m+2},...,a_k)\in S^{k-m-1}} p_{a_{m+2}} \dots p_{a_k} =$$

$$\sum_{(a_1,a_2,...,a_m)\in B} p_{a_1}\, p_{a_2} \dots p_{a_m}.$$

Assim, P está bem definida. Afirmamos que é aditiva.

De fato, dados dois cilindros C_1 e C_2, $C_1 \cap C_2 = \emptyset$, podemos supor sem perda de generalidade que eles tem o mesmo rank m e são dados respectivamente por

$$C_1 = \{\omega \,|\, (X_1(\omega), X_2(\omega), ..., X_m(\omega)) \in B\},$$

$$C_2 = \{\omega \,|\, (X_1(\omega), X_2(\omega), ..., X_m(\omega)) \in B'\},$$

onde $B, B' \subset S^m$.

Logo,

$$P(C_1 \cup C_2) = \sum_{(a_1,a_2,...,a_m)\in B\cup B'} p_{a_1}\, p_{a_2} \dots p_{a_m} =$$

$$\sum_{(a_1,a_2,...,a_m)\in B} p_{a_1}\, p_{a_2} \dots p_{a_m} + \sum_{(a_1,a_2,...,a_m)\in B'} p_{a_1}\, p_{a_2} \dots p_{a_m} =$$

$$P(C_1) + P(C_2).$$

Por indução vale a aditividade finita na álgebra $\mathcal{C}$. Podemos agora usar os Teoremas 5.5, 5.6 e 5.7 para concluir que a lei aditiva P acima definida tem uma extensão única à σ-álgebra $\sigma(\mathcal{C})$.

Esta probabilidade P sobre a σ-álgebra $\sigma(\mathcal{C})$ definida sobre subconjuntos de $S^{\mathbb{N}}$ é o que entendemos por Processo Estocástico Independente obtido a partir de $p_1, p_2, ..., p_n$ como acima. Algumas vezes é denominado de Probabilidade de Bernoulli associada a $p_1, p_2, ..., p_n$.

Neste caso, temos as seguintes versões da mesma informação

$$P(\{\omega \,|\, (X_1(\omega), X_2(\omega), ..., X_m(\omega)) = (a_1, a_2, ..., a_m)\}) =$$

$$P(X_1(\omega) = a_1, X_2(\omega) = a_2, ..., X_m(\omega) = a_m) =$$

$$P(\{a_1\} \times \{a_2\} \times ... \times \{a_m\} \times S^{\mathbb{N}}) =$$

$$p_{a_1}\, p_{a_2} \,...p_{a_m}.$$

Fica assim definido sobre $S^{\mathbb{N}}$ o conceito de Processo Independente e identicamente distribuído.

O conceito de Processo Independente (mas não identicamente distribuído) pode ser entendido da seguinte forma.

Seja S finito com n elementos (o caso de S infinito é similar) e para cada $t \in \mathbb{N}$ fixo, considere $p_1^t, p_2^t, ..., p_n^t$, números reais tais que, $0 \leq p_i^t \leq 1$, e que satisfazem

$$p_1^t + p_2^t + ... + p_n^t = 1.$$

Dado um cilindro C da forma

$$C = \{\omega \,|\, (X_1(\omega), X_2(\omega), ..., X_m(\omega)) = (a_1, a_2, ..., a_m)\},$$

onde $a_1, a_2, ..., a_m \in S$ estão fixados, definimos

$$P(C) = p^1_{a_1}\, p^2_{a_2} \,...p^m_{a_m}.$$

O cilindro geral de rank m pode ser escrito como

$$C = \{\omega \,|\, (X_1(\omega), X_2(\omega), ..., X_m(\omega)) \in B\},$$

onde $B \subset S^m$.

Neste caso definimos

$$P(C) = \sum_{(a_1,a_2,...,a_m)\in B} p^1_{a_1}\, p^2_{a_2}\, ...p^m_{a_m}.$$

Tal P está bem definida e pode ser estendida à σ-álgebra $\sigma(\mathcal{C})$ gerada pelos cilindros de maneira única.

O caso anterior é um caso particular deste.

Note que em ambos os casos

$$P(X_{t_1}(\omega) = a_1, X_{t_2}(\omega) = a_2, ..., X_{t_m}(\omega) = a_m) =$$

$$P(X_{t_1}(\omega) = a_1)\, P(X_{t_2}(\omega) = a_2)...P(X_{t_m}(\omega) = a_m).$$

No caso de processo independente e identicamente distribuído temos ainda que

$$P(X_{t_1}(\omega) = a_1, X_{t_2}(\omega) = a_2, ..., X_{t_m}(\omega) = a_m) =$$

$$P(X_1(\omega) = a_1)\, P(X_1(\omega) = a_2)...P(X_1(\omega) = a_m).$$

Note que o processo independente e identicamente distribuído é estacionário, mas o caso apenas independente não, pois, por exemplo, podemos ter que $P(X_1 = a_1) \neq P(X_2 = a_1)$.

No caso de Processos de Markov podemos proceder de forma semelhante. Considere um vetor de probabilidade sobre S dado por $\pi = (\pi_i)_{i\in S}$ e uma matriz estocástica $\mathcal{P} = (P_{ij})_{i,j\in S}$.

Sobre a álgebra dos cilindros, defina primeiro P para $C \in \mathcal{C}$ da forma

$$C = \{X_1(\omega) = a_1, X_2(\omega) = a_2, ..., X_m(\omega) = a_m\},\ a_i \in S,\ i \in \{1, 2, 3, ..., m\},$$

como

$$P(C) = \pi_{a_1}\, P_{a_1a_2}\, P_{a_2a_3}...P_{a_{m-2}a_{m-1}}\, P_{a_{m-1}a_m}.$$

Finalmente, para o cilindro geral de rank m

$$C = \{\omega \,|\, (X_1(\omega), X_2(\omega), ..., X_m(\omega)) \in B\},$$

onde $B \subset S^m$, definimos

$$P(C) = \sum_{(a_1, a_2, ..., a_m) \in B} \pi_{a_1} \, P_{a_1 a_2} \, P_{a_2 a_3} ... P_{a_{m-2} a_{m-1}} \, P_{a_{m-1} a_m}.$$

Dados dois cilindros que não se interceptam C_1 e C_2 definimos $P(C_1 \cup C_2)$ como $P(C_1) + P(C_2)$.

Pode-se mostrar via Teorema 5.7, da mesma forma como no caso anterior, que tal lei aditiva P (sobre o conjunto das uniões finitas de cilindros disjuntos) está bem definida e que é σ-aditiva sobre a álgebra $\mathcal{C}$. Sendo assim podemos estender de maneira única P a uma probabilidade sobre a σ-álgebra gerada pelos cilindros.

Este procedimento descreve os Processos Estocásticos Markovianos X_n, $n \in \mathbb{N}$, tomando valores em $S = \{1, 2, ..., d\}$. Em resumo, dado uma matriz estocástica $\mathcal{P}$, com entradas $P_{i\,j}$, e um vetor inicial de probabilidade π (que não precisa ser invariante para $\mathcal{P}$) obtemos uma probabilidade na sigma algebra de Borel em $\{1, 2, ..., d\}^{\mathbb{N}}$, tal que

$$P(X_1(\omega) = a_1, X_2(\omega) = a_2, ..., X_m(\omega) = a_m\,) =$$

$$\pi_{a_1} \, P_{a_1 a_2} \, P_{a_2 a_3} ... P_{a_{m-2} a_{m-1}} \, P_{a_{m-1} a_m}.$$

O processo X_n, $n \in \mathbb{N}$, é um processo estacionário se $\pi \mathcal{P} = \pi$.

Lembre que o shift σ agindo em $\{1, 2, ..., d\}^{\mathbb{N}}$ não é bijetivo.

Por definição o shift σ age em $\{1, 2, ..., d\}^{\mathbb{Z}}$ da seguinte forma: se

$$w = (...x_{-3}, x_{-2}, x_{-1} \,|\, x_1, x_2, x_3, ...)$$

então

$$\sigma(w) = (...x_{-3}, x_{-2}, x_{-1}, x_1 \,|\, x_2, x_3, ...).$$

Esta aplicação shift agindo em $\{1,2,...,d\}^{\mathbb{Z}}$ é por sua vez bijetiva.

Como exemplo observamos que

$$\sigma(X_{-2}(\omega)=a_{-2},X_{-1}(\omega)=a_{-1},X_1(\omega)=a_1,X_2(\omega)=a_2,X_3(\omega)=a_3)=$$

$$(X_{-3}(\omega)=a_{-2},X_{-2}(\omega)=a_{-1},X_{-1}(\omega)=a_1,X_1(\omega)=a_2,X_2(\omega)=a_3).$$

Podemos definir um Processo Estocástico Markoviano X_n, $n\in\mathbb{Z}$, tomando valores em $S=\{1,2,...,d\}$ da seguinte forma: seja $\mathcal{P}$ uma matriz estocástica, com entradas $P_{i\,j}$, e π tal que $\pi\,\mathcal{P}=\pi$, então

$$P(X_{-n}(\omega)=a_{-n},X_{-n+1}(\omega)=a_{-n+1},...,X_1(\omega)=a_1,...,X_m(\omega)=a_m\,)=$$

$$\pi_{a_{-n}}\,P_{a_{-n}a_{-n+1}}...\,P_{a_1a_2}...P_{a_{m-2}a_{m-1}}\,P_{a_{m-1}a_m}.$$

Esta lei aditiva P pode ser extendida a uma probabilidade sobre a sigma algebra gerada pelos cilindros em $S^{\mathbb{Z}}$.

O shift σ age em $S^{\mathbb{Z}}$ da seguinte forma

$$\sigma(...,x_{-2},x_{-1}\,|\,x_1,x_2,..)=(...,x_{-2},x_{-1},x_1\,|\,x_2,x_3..).$$

O processo X_n, $n\in\mathbb{Z}$, assim obtido é estacionário (a medida P sobre $S^{\mathbb{Z}}$ é invariante pelo shift σ em $S^{\mathbb{Z}}$).

Neste caso para se ter uma probabilidade P em $S^{\mathbb{Z}}$ não se pode tomar um π qualquer (sem satisfazer $\pi\mathcal{P}=\pi$). De fato, necessitamos a aditividade finita nos cilindros. Observe que se $S=\{1,2\}$, então

$$P(X_1=1)\neq P(X_0=1,X_1=1)+P(X_0=2,X_1=1)$$

a não ser que $\pi\mathcal{P}=\pi$

Definição 5.19. *Dado dois conjuntos A e B contidos em X, a diferença simétrica de A e B, denotada por $A\Delta B$, é o conjunto*

$$A\cup B-A\cap B=A\Delta B.$$

Podemos imaginar que a afirmação *A é quase igual à B, em termos da medida μ*, deveria significar que $\mu(A\Delta B)$ é um número pequeno.

Por outro lado, podemos imaginar que a afirmação *A é igual à B, em termos da medida μ*, deveria significar que $\mu(A\Delta B) = 0$.

Como exemplo, note que se λ é a medida de Lebesgue em $\mathbb{R}$, então vale $\lambda((a,b)\Delta[a,b]) = \lambda(\{a,b\}) = 0$.

Duas propriedades importantes de uma medida (ou probabilidade) μ, definida sobre a σ-álgebra $\mathcal{A}$, obtida como extensão de uma pré-medida $\tilde{\mu}$ sobre a álgebra $\mathcal{F}$ (de acordo com o Teorema da Extensão de Caratheodori-Kolmogorov) são descritas abaixo.

Um conjunto A da σ-álgebra $\mathcal{A}$ gerada pela coleção de conjuntos $\mathcal{F}$, ou seja, $A \in \mathcal{A} = \sigma(\mathcal{F})$, não é necessariamente uma união enumerável de conjuntos em $\mathcal{F}$, mas existem versões aproximadas desta ideia:

Teorema 5.8. *Considere uma medida σ-finita μ sobre uma σ-álgebra $\mathcal{A}$ que foi obtida como extensão de uma lei σ-aditiva $\tilde{\mu}$ sobre a álgebra $\mathcal{F}$. Seja A mensurável com respeito à $\sigma(\mathcal{F}) = \mathcal{A}$ e suponha que $\mu(A)$ é finito. Seja $\epsilon > 0$, então existe C em $\mathcal{F}$ tal que $\mu(A\Delta C) \leq \epsilon$.*

Esta é uma maneira de aproximar em medida elementos de $\sigma(\mathcal{F})$ por elementos de $\mathcal{F}$.

Seja $\Omega = \{1,2\}^{\mathbb{N}}$ e a probabilidade resultante em Ω associada a um processo independente dado por p_1, p_2 tal que $p_1 + p_2 = 1$. Então

$$A = 1 \times \{1,2\} \times 1 \times \{1,2\} \times 1 \times \{1,2\} \times ...$$

pode ser aproximado por cilindros da forma

$$C_n = \underbrace{1 \times \{1,2\} \times 1 \times \{1,2\} \times 1 \times ... \times \{1,2\}}_{n},$$

no sentido de que dado um ϵ existe um n tal que $\mu(A\Delta C_n) \leq \epsilon$.

Teorema 5.9. *Considere uma medida σ-finita μ sobre uma σ-álgebra $\mathcal{A}$ que foi obtida como extensão de uma lei σ-aditiva $\tilde{\mu}$ sobre a álgebra $\mathcal{F}$. Seja A mensurável com respeito a $\sigma(\mathcal{F}) = \mathcal{A}$. Dado $\epsilon > 0$, então existe uma coleção enumerável de conjuntos $C_i \in \mathcal{F}$, tal que*

$$\sum_{i=1}^{\infty} \mu(C_i) = \sum_{i=1}^{\infty} \tilde{\mu}(C_i) \leq \mu(A) + \epsilon,$$

e

$$A \subset \cup_{i=1}^{\infty} C_i.$$

Assim, se $\mu(A) = 0$, dado $\epsilon > 0$, existe uma coleção enumerável de conjuntos $C_i \in \mathcal{F}$, tal que

$$\sum_{i=1}^{\infty} \mu(C_i) = \sum_{i=1}^{\infty} \tilde{\mu}(C_i) \leq \epsilon,$$

e

$$A \subset \cup_{i=1}^{\infty} C_i.$$

Esta é uma maneira de aproximar (por cima, pois $A \subset \cup_{i=1}^{\infty} C_i$) em medida elementos de $\sigma(\mathcal{F})$ por uma união enumerável de elementos de $\mathcal{F}$.

Os dois teoremas acima permitem obter informação sobre $\mu(A)$ a partir da informação dada pela lei σ-aditiva $\tilde{\mu}$ sobre conjuntos em $\mathcal{F}$.

Vamos agora definir a integral de uma função mensurável $f : (X, \mathcal{A}, \mu) \to (\mathbb{R}, \mathcal{B})$, onde $\mathcal{B}$ denota a σ-álgebra de Borel sobre $\mathbb{R}$ e μ é uma medida sobre $\mathcal{A}$, σ-álgebra de subconjuntos de X. Gostaríamos de dar sentido à expressão $\int f(x) d\mu(x)$. Vamos assim generalizar a conhecida integral $\int f(x) dx$ definida para funções $f : \mathbb{R} \to \mathbb{R}$.

Primeiro vamos definir a integral para funções $f : X \to \mathbb{R}$ de uma certa forma bem simples. Por definição, dado um subconjunto A de X, a função indicador de A é a função $I_A : X \to \mathbb{R}$ tal que, $I_A(x) = 0$, se x está em $X - A$, e, $I_A(x) = 1$, se x está em A.

Se A é um conjunto $\mathcal{A}$ mensurável, então I_A é uma função $\mathcal{A}$-mensurável.

Inicialmente, vamos definir $\int f(x)d\mu(x)$ quando $f = I_A$ e A é $\mathcal{A}-$mensurável

Por definição,

$$\int I_A(x)d\mu(x) = \mu(A).$$

Considere uma coleção finita de conjuntos A_j em $\mathcal{A}$, $j \in \{1, 2, ..., n\}$, e uma sequência a_j de números reais, $j \in \{1, 2, ..., n\}$.

Quando f é da forma

$$f = \sum_{j=1}^{n} a_j I_{A_j},$$

então nós definimos que

$$\int f(x)d\mu(x) = \sum_{j=1}^{n} a_j \mu(A_j).$$

As funções da forma $\sum_{j=1}^{n} a_j I_{A_j}$ são chamadas de funções simples.

Seja $\Omega = \{1, 2\}^{\mathbb{N}}$ e a probabilidade resultante em Ω associada a um processo independente dado por p_1, p_2 tal que $p_1 + p_2 = 1$. Então $f = 2\, I_{\overline{1}} + 3,5\, I_{\overline{12}}$ é simples e sua integral é $\int\, f dP = 2p_1 + 3.5\, p_1 p_2$.

Seja

$$A = 1 \times \{1, 2\} \times 1 \times \{1, 2\} \times 1 \times \{1, 2\} \times ...$$

Então $f = 2\, I_A + 1.4\, I_{\overline{12}}$ é simples e sua integral é $\int\, f dP = 2P(A) + 1.4\, p_1 p_2$.

Pode ser mostrado (ver [Fe]) que qualquer função mensurável não negativa $f(x)$ é limite pontual de uma seqüência monótona crescente de funções simples f_i. Isto é, existem $f_i, i \in \mathbb{N}$ tal que para todo $x \in X$, temos que $0 \leq f_i(x) \leq f_{i+1}(x)$ e $\lim_{i\to\infty} f_i(x) = f(x)$.

Por definição

$$\int f(x)d\mu(x) = \lim_{i\to\infty} \int f_i(x)d\mu(x).$$

Pode-se mostrar que o valor acima obtido independe da escolha da seqüência f_i de funções simples (poderia haver mais de uma seqüência).

O valor limite acima poderia ser igual a ∞; no caso em que o valor $\int |\, f(x)\,|\, d\mu(x)$ é finito dizemos que a função f é μ-integrável e denotamos tal fato por $f \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$.

Se μ é uma probabilidade e f é mensurável e limitada, então ela é μ-integrável (ver [Fe]).

Se $f : \Omega \to \mathbb{R}$ é mensurável e atinge valores positivos e negativos, então f pode ser expressa como $f = f_+ - f_-$ onde f_+ e f_- são ambas não-negativas. No caso em que, ambas, f_+ e f_- estão em $\mathcal{L}^1(\Omega, \mathcal{A}, \mu)$, então definimos $\int f(x) d\mu(x) = \int f_+(x) d\mu(x) - \int f_-(x) d\mu(x)$. Para denotar tal fato escrevemos $f \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$.

Fica assim descrito o conceito de integral de uma função f em relação à uma medida μ.

Chamamos $\int f(x) d\mu(x)$ de **integral de Lebesgue** de f em relação à μ. Na definição acima estamos assumindo que μ é uma medida.

Assim, dada $f : \Omega \to \mathbb{R}$ mensurável, definimos $\int f d\mu = \int_P f\, d\mu - \int_N f d\mu^-$, onde nos valemos da expressão anterior para integral de uma função segundo uma medida.

No caso em que $\int f^2(x) d\mu(x)$ é finito dizemos que f é de quadrado integrável e denotamos tal fato por $f \in \mathcal{L}^2(\Omega, \mathcal{A}, \mu)$.

Note que se $\Omega = S^{\mathbb{N}}$ e ϕ depende das n primeiras coordenadas, isto é,

$$\phi = \sum_{a_1, a_2, \ldots, a_n \in S} \alpha_{(1,2,\ldots,n)} I_{\overline{a_1, a_2, \ldots, a_n}},$$

onde $\alpha_{1,2,\ldots,n} \in \mathbb{R}$ são constantes fixadas, então $\int \phi\, dP$, será dada por

$$E(\phi) \;=\; \int \phi\, dP = \sum_{a_1, a_2, \ldots, a_n \in S} \alpha_{(1,2,\ldots,n)}\, P(\overline{a_1, a_2, \ldots, a_n}).$$

Dada uma variável aleatória integrável $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{R})$, é usual denotar

$$E[X] = \int X(\omega) dP(\omega),$$

e dizer que $E[X]$ é o valor esperado de X.

Um conceito extremamente importante na teoria é:

Definição 5.20. *A variância de $X \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ é o valor*

$$Var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2.$$

A variância de X mede a dispersão dos valores $X(\omega)$ em torno da média $E[X]$. Quanto menor a variância, mais os valores $X(\omega)$ estão agrupados e concentrados em volta de $E[X]$.

Voltaremos a analisar tal conceito com mais cuidado em breve.

Definição 5.21. *Dados $(\Omega, \mathcal{A}, \mu)$ e $(\Omega, \mathcal{A}, \nu)$ duas medidas sobre a mesma σ-álgebra dizemos que μ é absolutamente contínua com respeito a ν, se para todo $A \in \mathcal{A}$ tal que $\nu(A) = 0$ então $\mu(A) = 0$. Denotamos tal fato por $\mu << \nu$.*

Exemplo 5.12. Considere uma função não negativa $\mathcal{A}$-mensurável $\phi(x)$ que seja ν-integrável.

Defina a medida μ sobre a σ-álgebra $\mathcal{A}$ por $\mu(A) = \int_A \phi(x) d\nu(x)$, para cada $A \in \mathcal{A}$. É fácil ver que μ está bem definida e de fato satisfaz as propriedades requeridas para ser uma medida. Não é dificil ver que tal μ é absolutamente contínua com relação a ν.

$\diamondsuit$

O próximo teorema afirma que a classe de exemplos acima descreve todos os casos possíveis.

Teorema 5.10 (Teorema de Radon-Nykodin). *Sejam* $(\Omega, \mathcal{A}, \nu)$ *e* $(\Omega, \mathcal{A}, \mu)$ *duas medidas na mesma* σ*-álgebra, onde assumimos que* ν *é* σ*-finita . Então* μ *é absolutamente contínua com respeito a* ν*, se e só se, existe uma função não negativa* $\phi : \Omega \to \mathbb{R}$*,* ν*-integrável, tal que, para todo* $A \in \mathcal{A}$*, vale*

$$\mu(A) = \int_A \phi(x) d\nu(x).$$

A função ϕ *é única* ν*-quase toda parte e denotamos* $\phi = \frac{d\mu}{d\nu}$.

O próximo teorema é de grande importância na teoria.

Teorema 5.11 (Teorema da Convergência Monótona). *Sejam* $X, (X_n)_{n\geqslant 1}$ *variáveis aleatórias em* $(\Omega, \mathcal{A}, \mathbb{P})$*. Se* $0 \leqslant X_n \uparrow X$*, isto é,* $X_n(w) \geqslant 0$ *e* $X_n(w) \uparrow X(w)$*, para todo* $w \in \Omega$*, então* $E(X_n) \uparrow E(X)$*.*

Demostração:: Pela propriedade da esperança, como $0 \leqslant X_n \uparrow X$ temos que $0 \leqslant E(X_n) \leqslant E(X)$ e $E(X_n) \uparrow$. Logo, $\lim\limits_{n\to\infty} E(X_n) \leqslant E(X)$. Portanto, basta provar que $\lim\limits_{n\to\infty} E(X_n) \geqslant E(X) - \varepsilon$, para todo $\varepsilon > 0$.

Para isso, vamos aproximar X por meio de uma variável aleatória discreta Y tal que $|X - Y| \leqslant \varepsilon$, onde $\varepsilon > 0$ está fixo.

Definimos o evento $B_n = [n\varepsilon < X \leqslant (n+1)\varepsilon], n \in \mathbb{N}$ e a variável aleatória $Y = \sum_{n=0}^{\infty} n\,\varepsilon\, I_{B_n}$. Note que os conjuntos B_n são disjuntos. Portanto, para cada $w \in \Omega$ temos que

$$Y(w) = \sum_{n=0}^{\infty} n\,\varepsilon\, I_{B_n}(w) = \begin{cases} n\varepsilon, & \text{se } w \in B_n \\ 0, & \text{se } w \notin \cup_{n\geq 0} B_n \end{cases} = \begin{cases} n\varepsilon, & \text{se } n\varepsilon < X(w) \leqslant (n+1)\varepsilon \\ 0, & \text{se } X(w) = 0. \end{cases}$$

Logo, $X - \varepsilon \leqslant Y \leqslant X$. De fato: $Y(w) \leqslant X(w)$ e ainda $X(w) \leqslant n\varepsilon + \varepsilon = Y(w) + \varepsilon$ o que implica em $X(w) - \varepsilon \leqslant Y(w)$.

Então, $EX - \varepsilon \leqslant E(Y) \leqslant E(X)$ (o que vale também no caso em que $EX = \infty$).

Vamos provar que $E(Y) \leqslant \lim_{n\to\infty} E(X_n)$, isto é, $\lim_{n\to\infty} E(X_n) \geqslant EY \geqslant EX - \varepsilon$, o que conclui a prova.

Para isso, considere $A_k = [X_k \geqslant Y]$. Observamos que $A_k \uparrow \Omega$. De fato: $X_k(w) \geqslant Y(w) \Rightarrow X_{k+1}(w) \geqslant X_k(w) \geqslant Y(w)$, pois $X_n \uparrow$. Portanto, $A_k \uparrow$. Mas a convergência de X_k para X implica que $X_k(w) \geqslant Y(w)$ para k suficientemente grande. Notemos que $Y(w) < X(w)$ a menos que $X(w) = 0$. Logo, $\Omega = \bigcup A_k = \lim A_k$. Portanto, $B_n \cap A_k \uparrow B_n \cap \Omega = B_n$, quando $k \to \infty$ (e n é fixo). Observamos que a variável aleatória YI_{A_k} é discreta e

$$Y(w)I_{A_k}(w) = \begin{cases} Y(w), & \text{se } w \in A_k \\ 0, & \text{se } w \notin A_k \end{cases} = \begin{cases} Y(w) \leqslant X_k(w), & \text{se } w \in A_k \\ 0 \leqslant X_k(w), & \text{se } w \notin A_k. \end{cases}$$

Logo, $0 \leqslant YI_{A_k} \leqslant X_k$ e $0 \leqslant E(YI_{A_k}) \leqslant E(X_k)$. Para calcular $E(YI_{A_k})$ é preciso notar que

$$Y(w)I_{A_k}(w) = \begin{cases} n\varepsilon, & \text{se } \ w \in B_n \cap A_k,\ n \in \mathbb{N} \\ 0, & \text{se } \ w \notin \bigcup\limits_{n\geqslant 0} (B_n \cap A_k). \end{cases}$$

Portanto,

$$E(X_k) \geqslant E(YI_{A_k}) = \sum_{n\geqslant 0} n\varepsilon P(B_n \cap A_k) \geqslant \sum_{n=0}^{m} n\varepsilon P(B_n \cap A_k), \forall\, m.$$

Mas $P(B_n \cap A_k) \uparrow P(B_n)$ quando $k \to +\infty$. Logo,

$$\lim_{k\to\infty} E(X_k) \geqslant \lim_{k\to\infty} \sum_{n=0}^{m} n\varepsilon P(B_n \cap A_k) = \sum_{n=0}^{m} n\varepsilon P(B_n), \text{ para todo } m.$$

Portanto,

$$\lim_{k\to\infty} E(X_k) \geqslant \sum_{n=0}^{\infty} n\varepsilon P(B_n) = EY \geqslant E(X) - \varepsilon, \text{para todo } \varepsilon > 0.$$

Concluímos que $E(X_n) \uparrow E(X)$.

□

Outro resultado importante é o seguinte:

Teorema 5.12 (**Teorema da Convergência Dominada**). *Sejam* $Y, X, (X_n)_{n\geqslant 1}$ *variáveis aleatórias em* $(\Omega, \mathcal{A}, P)$ *tais que* Y *é integrável,* $|X_n| \leqslant Y, \forall\, n$ *e* $X_n \to X$ *q.t.p. Então,* X *e* X_n *são integráveis e* $E(X_n) \to E(X)$.

A demonstração de tal afirmação pode ser encontrada em [Fe].

Um resultado interessante diz o seguinte: dadas duas probabilidades P_1 e P_2 na reta real, se para toda função contínua f vale que $\int f dP_1 = \int f dP_2$, então $P_1 = P_2$, no sentido de que $P_1(B) = P_2(B)$ para qualquer boreleano B. Isto segue do **Teorema de Riesz** (ver [Fe]) que vamos descrever abaixo.

Os resultados mencionados abaixo até o fim desta seção podem (devem ?) ser evitados numa primeira leitura. O leitor pode se dirigir diretamente a seção 5.2 sem maiores prejuízos ao entendimento do que seguirá.

Denotamos por $C_0(\Omega)$ o conjunto das funções contínuas sobre Ω, onde usamos a distância d_θ anteriomente descrita sobre $\Omega = \{1, 2, .., d\}^{\mathbb{N}}$.

Lembre que a sigma-algebra de Borel, neste caso, coincide com a sigma-algebra gerada pelos cilindros.

Considere a seguinte norma sobre as funções contínuas $f : \Omega \to \mathbb{R}$,

$$|f| = \sup_{w\in\Omega}\{|f(w)|\}.$$

Este valor é sempre finito pois Ω é compacto.

Dada uma função linear $L : C_0(\Omega) \to \mathbb{R}$, dizemos que ela é limitada se

$$\sup_{f \text{ tais que } |f| \leq 1} |L(f)| < \infty.$$

Dada uma função linear $L : C_0(\Omega) \to \mathbb{R}$, dizemos que uma função linear $L : C_0(\Omega) \to \mathbb{R}$ é positiva se $L(f) \geq 0$, toda vez que $f \geq 0$.

O Teorema de Riesz afirma o seguinte: seja um função linear $L : C_0(\Omega) \to \mathbb{R}$ tal que é limitada e positiva, e ainda que vale que $L(1) = 1$, então existe uma probabilidade P sobre a sigma-algebra de Borel de Ω tal que para qualquer f continua vale

$$L(f) = \int f \, d\, P.$$

Ainda, a lei que associa L a P é uma bijeção. Isto pode ser entendido da seguinte forma: uma probabilidade P é determinada de forma única apenas pela informação das integrais $\int f dP$ de todas as funções contínuas.

Dizemos que a sequencia de probabilidades P_n sobre o espaço mensurável $(\Omega, \mathcal{F})$ converge fracamente a probabilidade P, se para toda função contínua $f : \Omega \to \mathbb{R}$, vale que

$$\lim_{n\to\infty} \int f \, dP_n = \int f \, dP.$$

Note que isto não significa que para todo elemento $A \in \mathcal{F}$ vale que $P_n(A)$ converge a $P(A)$, quando $n \to \infty$.

Um teorema importante em probabilidade afirma que se $\Omega = \{1, 2, , ...d\}$, então dada uma sequencia de probabilidades P_n, sobre a sigma-algebra de Borel, sempre existe uma subsequencia P_{n_k}, $k \in \mathbb{N}$, e uma probabilidade P, tal que, P_{n_k} converge fracamente a P, quando $k \to \infty$, no sentido fraco.

O conjunto das probabilidade P sobre o conjunto $(\Omega, \mathcal{F})$ será denotado por $\mathcal{PW}$. A propriedade acima faz com que $\mathcal{PW}$ seja denominado sequencialmente compacto. Na verdade existe uma métrica d sobre $\mathcal{PW}$ de tal forma que $\lim_{n\to\infty} P_n = P$ no sentido fraco (como definido acima), se e só se , $\forall \epsilon$, existe N tal que para $n > N$, vale $d(P_n, P) < \epsilon$ (o sentido usual de convergência num espaço métrico).

Se μ é uma medida com sinal, também se pode considerar o conceito de $\int f d\mu$. Primeiro observamos que o Teorema de decomposição de Hahn-Jordan (ver seção 6.2 em [Fe]) afirma que dada uma medida sobre uma sigma-algebra $\mathcal{F}$, então existem dois conjuntos mensuráveis N e P tais que $N \cap P = \emptyset$,

$N \cup P = \Omega$, e que vale ainda para qualquer $A \subset P$ temos $\mu(A) \geq 0$, e para qualquer $A \subset N$ temos $\mu(A) \leq 0$. Desta maneira, se pode definir uma medida μ^- de forma que $\mu^-(A) = -\mu(A \cap N)$ para todo $A \in \mathcal{F}$.

O conjunto das medidas com sinal tais que $\mu(P) + \mu^-(N) < \infty$ será denotado por $\mathcal{SM}$. O conjunto $\mathcal{SM}$ é um espaço vetorial normado quando se considera a norma da variação total

$$|\mu|_{VT} = \mu(P) + \mu^-(N).$$

Um subconjunto G de um espaço vetorial com norma $|\ |$ é dito convexo se toda vez que x, $y \in G$, então a combinação convexa

$$\lambda x + (1-\lambda) y \in G, \quad \text{para todo} \ \ 0 \leq \lambda \leq 1.$$

É fácil ver que o conjunto $\mathcal{PW}$ é um conjunto convexo dentro do espaço vetorial das medidas com sinal $\mathcal{SM}$.

O Teorema de Schauder-Thychonov afirma que toda função contínua $F : G \to G$, onde G é convexo e sequencialmente compacto, possui um ponto fixo, isto é, existe $x \in G$ tal que $F(x) = x$.

Este teorema pode portanto ser aplicado quando $G = \mathcal{PW}$, no caso em que $\Omega = \{1, 2, .., d\}^{\mathbb{N}}$. A função contínua considerada acima é em relação a métrica d (compatível com a convergência fraca) que mencionamos antes, ou seja, é válida para $F : \mathcal{PW} \to \mathcal{PW}$.

Seja $T : (\Omega, \mathcal{A}) \to (\Omega, \mathcal{A})$ contínua, lembre que o conjunto das probabilidades P invariantes para T (isto é, tal que $P(T^{-1}(A)) = P(A), \forall A \in \mathcal{A}$) é denotado por $\mathcal{M}(T)$.

Este conjunto $\mathcal{M}(T)$ é não vazio por que a função $F : \mathcal{PW} \to \mathcal{PW}$ que leva P em $F(P) = Q$, tal que $Q(A) = P(T^{-1}(A)), \forall A \in \mathcal{A}$, é uma função contínua. Assim, quando $\Omega = \{1, 2, .., d\}^{\mathbb{N}}$, então pelo Teorema de Schauder-Thychonov existe um ponto fixo P. Tal P claramente pertence a $\mathcal{M}(T)$. Este resultado vale em particular para o shift, isto é, quando $T = \sigma$.

5.2 Propriedades Gerais de Processos Estocásticos

Lembre que dizemos que o processo X_t tomando valores em S (enumerável) com parâmetro $t \in T = \mathbb{N}$ é estacionário se para cada n e cada sequência $t_1 < t_2 < ... < t_n$, onde $t_i \in T$, $t > 0$ e para cada sequência de conjuntos $A_1, A_2, ..., A_n$, onde $A_i \subset S$ vale que

$$P(X_{t_1} \in A_1,\, X_{t_2} \in A_2, ...,\, X_{t_n} \in A_n) =$$

$$P(X_{t_1+t} \in A_1,\, X_{t_2+t} \in A_2, ..., X_{t_n+t} \in A_n).$$

Uma transformação

$$T : (S^{\mathbb{N}}, \mathcal{A}) \to (S^{\mathbb{N}}, \mathcal{A}),$$

onde $\mathcal{A}$ é a σ-álgebra gerada pelos cilindros, que possui grande importância é a seguinte:

$$T(w_1, w_2, w_3, w_4, ...) = (w_2, w_3, w_4, ...).$$

Esta transformação T é denominada de shift sobre $S^{\mathbb{N}}$ e muitas vezes denotada por σ.

A transformação T é mensurável pelo Teorema 5.9. De fato, note que se

$$C = \{X_1 = a_1, X_2 = a_2, X_3 = a_3, ..., X_k = a_k\},$$

então

$$T^{-1}(C) = \cup_{s\in S}\{X_1 = s, X_2 = a_1, X_3 = a_2, ..., X_k = a_{k-1}, X_{k+1} = a_k\}.$$

Outra forma de obter este resultado segue de mostrar que T é contínua, na verdade $d(T(\omega_1), T(\omega_2)) \leq \theta\ d(\omega_1, \omega_2)$.

Definição 5.22. *Considere um conjunto X com uma σ-álgebra $\mathcal{A}$ e uma probabilidade (ou medida) P definida sobre esta σ-álgebra e ainda um conjunto Y equipado com uma σ-álgebra $\mathcal{G}$ e uma probabilidade (ou medida) Q nesta σ-álgebra. Diremos que a função mensurável $f : X \to Y$ preserva probabilidade (ou medida) se $P(f^{-1}(A)) = Q(A)$, para todo $A \in \mathcal{G}$.*

Neste caso, usaremos a notação $f : (X, \mathcal{A}, P) \to (Y, \mathcal{G}, Q)$, para descrever tal fato

Estaremos interessados aqui principalmente no caso do endomorfismo shift acima definido

$$T : (X, \mathcal{A}) \to (X, \mathcal{A}),$$

onde $X = S^{\mathbb{N}}$.

Um Processo Estocástico com valores em S e com conjunto de valores temporais $\mathbb{N}$ é uma probabilidade P sobre a σ-álgebra gerada pelos cilindros de $S^{\mathbb{N}}$.

Anteriormente explicamos como o Teorema de extensão de Kolmogorov permite definir uma probabildade P sobre a sigma algebra gerada pelos cilindros no caso em que um processo Markoviano é obtido a partir de uma matriz estocástica $\mathcal{P}$ e uma probabilidade inicial π.

É natural perguntar quando P é invariante para $T : (S^{\mathbb{N}}, \mathcal{A}) \to (S^{\mathbb{N}}, \mathcal{A})$, onde T é o shift definido acima.

Teorema 5.13. *Seja X_n, $n \in \mathbb{N}$, um Processo Estocástico com valores em S e sua correspondente probabilidade P sobre $S^{\mathbb{N}}$.*

P é invariante para $T : (S^{\mathbb{N}}, \mathcal{A}) \to (S^{\mathbb{N}}, \mathcal{A})$, se e só se, X_n, $n \in \mathbb{N}$, é um Processo Estocástico Estacionário.

Demonstração: Considere um Processo Estocástico Estacionário, então vale que

$$P(\,\{X_1 = a_1, X_2 = a_2, X_3 = a_3, ..., X_k = a_k\}\,) =$$

$$P(\,\{X_{t+1} = a_1, X_{t+2} = a_2, X_{t+3} = a_3, ..., X_{t+k} = a_k\}\,).$$

Seja

$$C = \{X_1 = a_1, X_2 = a_2, X_3 = a_3, ..., X_k = a_k\},$$

então devemos analisar se

$$P(T^{-1}(C)) = P(\,\cup_{s\in S}\{X_1 = s, X_2 = a_1, X_3 = a_2, ..., X_k = a_{k-1}, X_{k+1} = a_k\}\,)$$

é igual a $P(C)$.

Ora, como

$$\cup_{s\in S}\{X_1 = s, X_2 = a_1, X_3 = a_2, ..., X_k = a_{k-1}, X_{k+1} = a_k\} =$$

$$\{X_2 = a_1, X_3 = a_2, ..., X_k = a_{k-1}, X_{k+1} = a_k\},$$

a afirmação de que tal P satisfaz $P(C) = P(T^{-1}(C))$ segue de

$$P(\,\{X_2 = a_1, X_3 = a_2, ..., X_k = a_{k-1}, X_{k+1} = a_k\}\,) =$$

$$P(\,\{X_1 = a_1, X_2 = a_2, ..., X_{k-1} = a_{k-1}, X_k = a_k\}\,).$$

A afirmação correspondente para o cilindro geral de rank k se demonstra de forma semelhante.

Sabemos que duas probabilidades que coincidem nos geradores de uma σ-álgebra coincidem na σ-álgebra pelo Teorema de Caratheodori-Kolmogorov.

Logo P é invariante para o shift T se o processo X_n é estacionário.

Vamos considerar a recíproca, suponha que o Processo Estocástico $X_n, n \in \mathbb{N}$, não seja estacionário, logo existe t tal que

$$P(\,\{X_{i_1} \in A_1, X_{i_2} \in A_2, X_{i_3} \in A_3, ..., X_{i_k} = a_k\}\,) \neq$$

$$P(\,\{X_{t+i_1} \in A_1, X_{t+i_2} \in A_2, X_{t+i_3} \in A_3, ..., X_{t+i_k} = a_k\}\,).$$

Note que por indução, se $P(T^{-1}(A)) = P(A)$ para todo $A \in \mathcal{A}$, então $P(T^{-n}(A)) = P(A)$, para qualquer $n \geq 0$ e para todo $A \in \mathcal{A}$.

Ora, se

$$C = \{X_{i_1} \in A_1, X_{i_2} \in A_2, X_{i_3} \in A_3, ..., X_{i_k} \in A_k\},$$

então $T^{-t}(C)$ é igual a

$$\cup_{s_1\in S, s_2\in S,\dots,s_t\in S}\{X_1 = s_1, ..., X_t = s_t, X_{t+i_1} \in A_1, ..., X_{t+i_k} \in A_k\} =$$

$$\{X_{t+i_1} \in A_1, ..., X_{t+i_k} \in A_k\}.$$

Logo,

$$P(T^{-t}(C)) = P(\{X_{t+i_1} \in A_1, ..., X_{t+i_k} \in A_k\}) \neq$$

$$P(\,\{X_{i_1} \in A_1, X_{i_2} \in A_2, X_{i_3} \in A_3, ..., X_{i_k} \in A_k\}\,) = P(C).$$

Concluímos que se o Processo Estocástico X_n não é estacionário, então a probabilidade P associada não é invariante para o shift T.

□

Nem sempre um processo Markoviano, como definido anteriormente, será estacionário. Como se vê na seção 1, é necessário e suficiente que esteja satisfeita a seguinte propriedade: o vetor de probabilidade (sobre S) inicial p deve ser tal que

$$p\,\mathcal{P} = p,$$

onde $\mathcal{P}$ é a matriz estocástica que define as transições no processo. Somente neste caso que a P sobre $S^{\mathbb{N}}$ será invariante para o shift T.

Dada $F:(X,\mathcal{A}) \to (Y,\mathcal{G})$ uma função mensurável e uma medida μ em $(X,\mathcal{A})$, podemos definir ν em $(Y,\mathcal{G})$ através de $\nu(A) = \mu(F^{-1})(A), \forall A \in \mathcal{G}$. Denotamos tal medida por $\nu = F^*(\mu)$ (que age sobre conjuntos de $\mathcal{G}$). Desta maneira, a partir de μ forçamos de forma trivial que a transformação F preserve medida, ou seja, $F : (X, \mathcal{A}, \mu) \to (Y, \mathcal{G}, \nu)$.

É usual a terminologia: ν é o **push-forward** da medida μ via F.

Um caso particular (de grande importância) da situação acima é o seguinte: seja X uma variável aleatória, ou seja, uma função $\mathcal{A}$ mensurável $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{R})$, onde $\mathcal{R}$ é a σ-álgebra de Borel sobre $\mathbb{R}$.

Definição 5.23. *Fixada uma probabilidade P sobre Ω, podemos definir $\nu(A) = P(X^{-1}(A)) = P(\{y \,|\, X(y) \in A\})$, $\forall A \in \mathcal{B}$. Fica assim definida uma probabilidade $\nu = \nu_X$ sobre a σ-álgebra de Borel $\mathcal{R}$ sobre $\mathbb{R}$. Tal ν_X é denominada de distribuição de probabilidade da variável aleatória X. Neste caso, $\nu_X = X^*(P)$ é uma probabilidade sobre $\mathbb{R}$.*

Dizemos que F_X é a distribuição da variável aleatória X se para qualquer $x \in \mathbb{R}$

$$F_X(x) = P(\{w \,|\, X(w) \leq x\}).$$

Se denotarmos a integral de Stieltjes (ver Cap. 4.2 em [Li2] ou [Ba2]) de uma função $g : \mathbb{R} \to \mathbb{R}$ por $\int g(x) d\, F_X(x)$, então, para qualquer g integrável

$$\int g(x)\, d\, F_X(x) = \int g(x)\, d\nu_X(x).$$

Se ν_X for absolutamente contínua em relação a medida de Lebesgue em $\mathcal{R}$ (nem sempre acontece), então pelo Teorema de Radon-Nykodin, existe uma função mensurável $f_X : (\mathbb{R}, \mathcal{R}) \to (\mathbb{R}, \mathcal{R})$, denominada densidade, que é não-negativa e que satisfaz para qualquer intervalo $[a, b]$

$$\int_a^b f_X(x)\, dx \;=\; \nu_X\,([a, b]).$$

Neste caso,

$$F_X(x) \;=\; \int_{-\infty}^x f_X(y)\, dy.$$

Note que para qualquer a e b, $a < b$, temos

$$F(b) - F(a) = \int_a^b f_X(x)\, dx.$$

Definição 5.24. *X tem distribuição exponencial de parâmetro λ, se para qualquer $x \geq 0$*

$$P(X < x) = F(x) = 1 - e^{-\lambda x},$$

e $P(X < x) = 0$ para $x < 0$.

Na nossa notação $X \sim \epsilon(\lambda)$.

Definição 5.25. *Dizemos que X tem distribuição Normal, ou Gaussiana, com média a e variância $\sigma^2 > 0$ se*

$$P(X < x) = F(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{x} e^{\frac{-(y-a)^2}{2\sigma^2}}\, dy.$$

Na nossa notação $X \sim \mathcal{N}(a, \sigma^2)$.

Exemplo 5.13. Considerando a X geral e a correspondente f_X como acima, note que

$$E[X^2 + 5\, e^X] = \int (x^2 + 5\, e^x)\, f_X(x)\, dx$$

(quando existir $E[X^2 + 5\, e^X]$, é claro).

$\diamondsuit$

Se $f : \mathbb{R} \to \mathbb{R}$ é densidade de X, isto é, se para qualquer intervalo (a, b) vale que

$$P(w : X(w) \in (a, b)) = \int_a^b f(x)\, dx.$$

então, esta propriedade se estende a qualquer conjunto Boreleano $B \in \mathcal{R}$:

$$P(w : X(w) \in B) = \int_B f(x)\, dx.$$

Dito de outra forma, o pushforward de X pela medida na reta definida por f é a probabilidade P. Para enfatizar a dependencia em X se denota tal f por f_X.

De fato, mais geralmente, para qualquer função mensurável e ν_X-integrável $g : \mathbb{R} \to \mathbb{R}$, vale que

$$E[g(X)] \,=\, \int g(x)\, f_X\,(x)\, dx.$$

A demonstração da afirmação acima para uma g qualquer segue o procedimento canônico em Teoria da Medida. Primeiro considere g da forma função indicador de um Boreleano A em $\mathcal{R}$, ou seja $g = I_A$. Neste caso,

$$E[g(X)] = P(\{\omega \,|\, X(\omega) \in A\}) = \nu_X(A) =$$

$$\int_A f_X(x)\, dx = \int I_A(x)\, f_X(x)\, dx \,=\, \int g(x)\, f_X(x) dx.$$

A seguir, se considera g da forma $g = \sum_{i=1}^k \alpha_i\, I_{A_i}$, onde $\alpha_i \in \mathbb{R}$ e A_i é elemento em $\mathcal{R}$. Finalmente, para uma função mesurável g qualquer se considera aproximações por funções simples, limites, etc... Deixamos a cargo do leitor completar a argumentação final até obter a afirmação correspondente para g qualquer ν_X-integrável.

No caso em que a variável X assume valores discretos, digamos sobre $G \subset \mathbb{Z}$ (G poderia ser finito ou não), denotamos por f_X, onde $f_X : G \to \mathbb{R}$, a função tal que $P(X = g) = f_X(g)$. Chamamos tal f_X, neste caso, de função massa de X.

Desta forma $E(X) = \int X(w)\, dP(w) = \int x\, f_X(x) dx$. No caso geral

$$\int g(X) dP = \int g(x) f_X(x) dx.$$

Ainda, $\int\, \cos(X)\, D\, P \,=\, \int\, \cos(X(w))\, dP(w) = \int \cos(x)\, f_X(x) dx.$

Sejam agora X e Y variáveis aleatórias sobre $(\Omega, \mathcal{A})$, ou seja,

$$X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{R}),$$

$$Y : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{R}).$$

Estamos interessados em analisar o par $(X, Y) : (\Omega, \mathcal{A}) \to (\mathbb{R}^2, \mathcal{R}^2)$

Definição 5.26. *Fixada uma probabilidade P sobre Ω, podemos definir*

$$\nu(A) = P(\,\{\omega \,|\, (X(w), Y(\omega)) \in A\}\,),$$

$\forall A \in \mathcal{R}^2$. Fica assim definida uma probabilidade $\nu = \nu_{X,Y}$ sobre a σ-álgebra de Borel $\mathcal{R}^2$ sobre $\mathbb{R}^2$. Tal $\nu_{X,Y}$ é denominada de distribuição de probabilidade conjunta do par ordenado de variáveis aleatórias (X, Y). Neste caso, $\nu_{X,Y}$ é uma probabilidade sobre $\mathbb{R}^2$.

Definição 5.27. *Da mesma forma como antes, se $\nu_{X,Y}$ for absolutamente contínua com respeito a medida de Lebesgue em $\mathcal{R}^2$, chamaremos de função de densidade conjunta do par (X, Y), a função $f_{X,Y}$ tal que $f_{X,Y} : (\mathbb{R}^2, \mathcal{R}^2) \to (\mathbb{R}, \mathcal{R})$, e*

$$\int_A f_{X,Y}(x, y) dx\, dy \;=\; \nu_{X,Y}(A),$$

para todo boreleano A sobre $\mathcal{R}^2$.

Definição 5.28. *Fixado $T : (\Omega, \mathcal{A}) \to (\Omega, \mathcal{A})$ mensurável, uma probabilidade P sobre $\mathcal{A}$ tal que $P(T^{-1}(A)) = P(A), \forall A \in \mathcal{A}$, é denominada de invariante. O conjunto de tais probabilidades é denotado por $\mathcal{M}(T)$.*

Vamos estar particularmente interessados no caso em que $\Omega = S^{\mathbb{N}}$ e $T = \sigma$.

Teorema 5.14. *Considere um conjunto Ω com uma σ-álgebra $\mathcal{A}$ e uma probabilidade P. Suponha que o conjunto $\mathcal{E} \subset \mathcal{P}(\Omega)$ gera a σ-álgebra $\mathcal{A}$ e é um sistema-π, isto é, $\sigma(\mathcal{E}) = \mathcal{A}$. Então, $T : \Omega \to \Omega$ mensurável preserva P, se e só se, $P(T^{-1}(B)) = P(B)$, para todo $B \in \mathcal{E}$.*

Demonstração: Suponha primeiro que $P(T^{-1}(B)) = P(B)$, para todo $B \in \mathcal{E}$. Defina a probabilidade $\nu = T^*(P)$, isto é, $\nu(A) = P(T^{-1}(A))$ para todo $A \in \mathcal{A}$. Pela definição de ν, segue que ν e P coincidem nos conjuntos da forma $B \in \mathcal{E}$. Segue da unicidade da extensão no Teorema de Caratheodori-Kolmogorov que para todo $A \in \mathcal{A}$, temos que

$$P(T^{-1}(A)) = \nu(A) = P(A).$$

A afirmação na outra direção é trivial.

$\square$

As duas últimas afirmações serão usadas extensivamente no texto.

Exemplo 5.14. Seja β uma constante real. A transformação mensurável $T(x) = x + \beta \pmod 1$ sobre o conjunto $\Omega = [0, 1)$ preserva a medida de Lebesgue $\lambda = P$ sobre a σ-álgebra de Borel. Para isto note que fixado um intervalo $(a, b) \subset [0, 1)$ (lembre que os intervalos geram a σ-álgebra de Borel), como a inclinação do gráfico de T é igual a 1, então $T^{-1}(a, b)$ será constituído por um ou dois intervalos, mas a soma do comprimento destes intervalos (no caso em que tivermos dois intervalos) é igual a $b - a$.

Portanto, pra todo intervalo (a, b), temos que $P(T^{-1}(a, b)) = (b - a) = P((a, b))$. Como os intervalos da forma (a, b) geram a σ-álgebra de Borel sobre $[0, 1]$, então T preserva a probabilidade P, onde $P((a, b)) = (b - a)$.

$\diamondsuit$

Exercício: Mostre que a transformação $T(x) = 2x \pmod 1$ sobre o conjunto $[0, 1)$ preserva a probabilidade $P = \lambda$ de Lebesgue sobre a σ-álgebra de Borel. Note que a pré-imagem de um intervalo (a, b) por T é a união de dois intervalos, cada um com metade do comprimento de (a, b).

Muitas vezes uma σ-álgebra pode ser encarada como um certo tipo de informação que dispomos. Por exemplo em $\{0, 1\}^{\mathbb{N}}$, podemos considerar a σ-álgebra $\mathcal{B}$ gerada pelos cilindros. Uma σ-álgebra $\mathcal{F}$ contida em $\mathcal{B}$ teria menos informações (certos conjuntos A em $\mathcal{B}$ não fazem sentido para quem olha a informação dada pelos conjuntos de $\mathcal{F}$).

Exemplo 5.15. Por exemplo, assuma que 0 está associado a cara e 1 a coroa. Quando jogamos a moeda três vezes, os conjuntos que ficam determinados por este evento seriam os conjuntos da forma

$$V_1 \times V_2 \times V_3 \times S^{\mathbb{N}},$$

onde $V_i \subset \{0,1\}$. O conjunto de todos os conjuntos da forma acima determina uma σ-álgebra $\mathcal{F}$ que traz informação apenas do que acontece quando se joga a moeda três vezes. Esta σ-álgebra $\mathcal{F}$ descreve o desconhecimento do que vai ocorrer após o tempo 3.

$\diamondsuit$

Pode se mostrar que uma probabilidade P definida nos boreleanos de $[0,1]$ é invariante para $T:[0,1]\to[0,1]$ contínua, se e só se, $\int f\,dP = \int (f\circ T)\,dP$ para toda f continua [PY]. Esta propriedade pode ser obtida a partir do Teorema da representação de Riesz.

Considere agora fixada um função mensurável $f:(\Omega,\mathcal{B})\to(\Omega',\mathcal{A})$, e a σ-álgebra $\mathcal{F}=f^{-1}(\mathcal{A})=\{f^{-1}(A)\,|\,A\in\mathcal{A}\}$. Considerando o conjunto Ω acima, tínhamos uma certa informação através dos conjuntos de $\mathcal{B}$. A σ-álgebra $\mathcal{F}$ acima, que está contida em $\mathcal{B}$, traduz um desconhecimento de qualquer coisa que não seja através da função f, ou seja, só conhecemos os conjuntos da forma $f^{-1}(A)$.

Dada uma função $g:(\Omega,\mathcal{B})\to(\mathbb{R},\mathcal{R})$, podemos perguntar se ela é $f^{-1}(\mathcal{A})$-mensurável.

Teorema 5.15. *A função $g:(\Omega,\mathcal{B})\to(\mathbb{R},\mathcal{R})$ é $f^{-1}(\mathcal{A})$ mensurável, se e só se, existe $h:(\Omega',\mathcal{A})\to(\mathbb{R},\mathcal{R})$ mensurável tal que*

$$g = h\circ f.$$

A demonstração deste fato pode ser encontrada em [Fe] prop. 3.22.

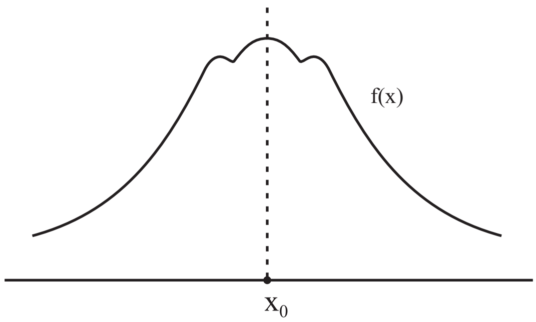


Figura 5.1: Densidade $f(x)$ com dados que possuem forte concentração em torno da média.

Definição 5.29. *Dada uma variável aleatória $X : \Omega \to \mathbb{R}$, sobre $(\Omega, \mathcal{F}, P)$, o número*

$$E(\,(X - E(X))^2\,) \,=$$

$$\int [\,(X(\omega) - E(X)) \;\times\; (X(\omega) - E(X))\,]\, dP(\omega) \,=$$

$$\int [\,(X(\omega) - \int X\, dP) \;\times\; (X(\omega) - \int X\, dP)\,]\, dP(\omega)\,,$$

é denominado de variância de X e é denotado por var(X) ou σ_X^2. Este número descreve como estão dispersos os valores $X(\omega)$ em torno de $E(X)$.

Imagine duas variáveis aleatórias X e Y tais que possuem o mesmo valor médio $E(X) = E(Y)$. Suponha que existe grande probabilidade de encontrar $X(\omega)$ tal que $|X(\omega) - E(X)| > 2$, e, por outro lado, não existe ω tal que $Y(\omega)$ satisfaz $|Y(\omega) - E(Y)| > 2$. Desta forma, se pode afirmar que os dados estão mais dispersos (em torno da média) para X do que para Y, neste caso, naturalmente, a variância de X é maior que a variância de Y.

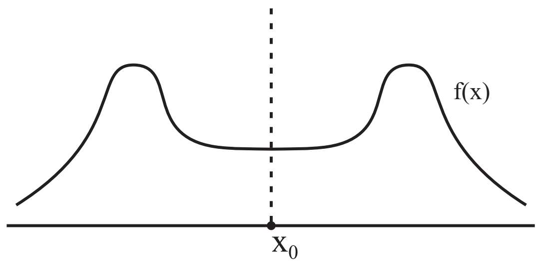


Figura 5.2: Densidade $f(x)$ com dados que possuem grande dispersão em torno da média. Alta probabilidade de encontrar valores bem distantes da média x_0.

Por exemplo, suponha que X descreve a altura média dos habitantes do país A e Y descreve a altura média dos habitantes do país B. Neste caso, por exemplo,

$$P(\, X(\omega) \in \, (1.60\,,\, 1.70)\,) =$$

$$= \frac{\text{número de pessoas do país } A \text{ que tem altura entre } 1.60 \text{ e } 1.70}{\text{população total do país } A}.$$

Suponhamos que o valor médio da altura dos dois países seja igual, ou seja, $E(X) = E(Y)$. Este número, em si, não captura a distribuição das alturas nas duas populações em torno da média. As variâncias de X e de Y vão descrever esta informação.

Por exemplo, suponha que X toma valores reais e seja descrita por uma densidade $f_X(x)$. Ou seja, $P(X \in (a,b)) = \int_a^b f_X(x)dx$, para todo intervalo (a,b). Vamos denotar o valor médio de X por x_0. Isto é, $\int x\, f_X(x)dx = x_0$.

Nas duas figuras 5.1 e 5.2 o leitor pode perceber em que situação ocorre o caso com pequena variância e o caso com grande variância.

Definição 5.30. *Dadas duas variáveis aleatórias* $X : \Omega \to \mathbb{R}$, $Y : \Omega \to \mathbb{R}$ *e fixado* $(\Omega, \mathcal{F}, P)$, *o número*

$$E(\,(X - E(X))\,(Y - E(Y))\,) \, =$$

$$\int [\,(X(\omega) - E(X)) \;\times\; (Y(\omega) - E(Y))\,]\, dP(\omega)\,,$$

$$\int [\,(X(\omega) - \int X\, dP) \;\times\; (Y(\omega) - \int Y\, dP)\,]\, dP(\omega)\,,$$

é denominado de covariância de X e Y e denotado por Cov(X,Y).

Este número descreve a existência de relação (maior ou menor) entre as variáveis X e Y. Dizemos que as variáveis não estão correlacionadas se $\text{Cov}(X,Y) = 0$

Por exemplo, suponha que X descreve a altura dos habitantes do país A e Y descreve a renda mensal da família do habitante do país A.

Em geral, se a renda familiar for maior, o indivíduo esteve sujeito a alimentação mais rica e a melhores condições de saúde durante sua infância, e assim, provavelmente terá altura maior do que aqueles oriundos de famílias de menor renda.

Note então que se $X(\omega) - E(X)$ for positivo (alguém com altura maior que a média) então $Y(\omega) - E(Y)$ deverá ser (provavelmente) positivo, logo $(X(\omega) - E(X))\,(Y(\omega) - E(Y)) > 0$. Por outro lado, se $X(\omega) - E(X)$ for negativo (alguém com altura menor que a média) então $Y(\omega) - E(Y)$ deverá ser (provavelmente) negativo também, logo $(X(\omega) - E(X))\,(Y(\omega) - E(Y)) > 0$. Desta forma, deveremos ter $\text{Cov}(X,Y)$ positiva, e tão maior quanto maior forem as diferenças $|X(\omega) - E(X)|$ e $|Y(\omega) - E(Y)|$ e suas respectivas probabilidades disto acontecer.

Dadas duas variáveis aleatórias quaisquer X e Y, se $\text{Cov}(X,Y)$ for um número real muito negativo, também existe indicação de que X influência muito a Y e vice-versa. Esta influencia pode se dar de forma inversa a anteriormente descrita: por exemplo, pode acontecer que $X(\omega) - E(X)$ negativo então $Y(\omega) - E(Y)$ positivo e $X(\omega) - E(X)$ positivo então $Y(\omega) - E(Y)$ negativo. Desta forma, $(X(\omega) - E(X))\,(Y(\omega) - E(Y)) < 0$. Neste caso as variáveis estão negativamente correlacionadas.

Por exemplo, considerando o universo das pessoas de mais de cinquenta anos; a idade está correlacionada com a capacidade de correr longas distâncias. A correlação é inversa (negativa portanto): em termos estatísticos, quanto maior a idade da pessoa, então, menor será a quantidade de quilômetros que ela é capaz de correr a uma velocidade fixada de, digamos, k quilômetros por hora.

Teorema 5.16. *Se X e Y são independentes tomando valores em um conjunto finito $S \subset \mathbb{R}$, então*

$$\int [\, X(\omega) \; \times \; Y(\omega)\,]\, dP(\omega) \; = \; E(\, X\, Y\,) = E(X)\, E(Y).$$

Demonstração: Ora,

$$E(X\,Y) = \sum_{x,y \in S} x\, y\, P(X = x, Y = y).$$

Ora, para cada par x, y, temos que

$$P(X = x,\, Y = y) = P(X = x)\, P(Y = y),$$

logo,

$$E(X\,Y) = \sum_{x,y \in S} x\, y\, P(X = x)\, P(Y = y) =$$

$$[\sum_{x \in S} x\, P(X = x)] \; \times \; [\sum_{y \in S} y\, P(Y = y)] = E(X)\, E(Y).$$

□

Teorema 5.17. *Se X e Y são independentes tomando valores em um conjunto finito $S \subset \mathbb{R}$, então $Cov(X, Y) = 0$.*

Demonstração: De fato,

$$E(\,(X - E(X))\,(Y - E(Y))\,) \; =$$

$$\int [\,(X(\omega) - E(X)) \;\times\; (Y(\omega) - E(Y))\,]\, dP(\omega) \;=$$

$$\int \,[\,X(\omega) \;\times\, Y(\omega)\,]\, dP(\omega) \;- E(X)\, E(Y) - E(X)\, E(Y) + E(X)\, E(Y) =$$

$$E(X)\, E(Y) - E(X)\, E(Y) - E(X)\, E(Y) + E(X)\, E(Y) = 0,$$

onde na última igualdade usamos o Teorema 5.16.

□

Segue do Teorema 5.17 que se X e Y forem independentes, elas não estão correlacionadas.

Definição 5.31. *Dados duas variáveis aleatórias $X : \Omega \to \mathbb{R}$, $Y : \Omega \to \mathbb{R}$ e fixado $(\Omega, \mathcal{F}, P)$, o número*

$$\frac{Cov(X,Y)}{\sigma_X\, \sigma_Y},$$

(onde σ_X^2 e σ_Y^2 são, respectivamente, as variâncias de X e Y) é denominado de coeficiente de correlação entre X e Y e denotado por $\rho_{X,Y}$.

Este número mede a covariância normalizada das duas variáveis X e Y. Isto é natural de se considerar, pois se X e Y tomam valores muito grandes, então $\mathrm{Cov}(X,Y)$ seria certamente um número grande. Quando tomamos o número $\rho_{X,Y}$, já expurgamos tal fato através da normalização obtida ao dividir por $\sigma_X\, \sigma_Y$. Assim, podemos comparar ρ_{X_1,Y_1} e ρ_{X_2,Y_2}, para pares de variáveis X_1, Y_1 e X_2, Y_2, de maneira absoluta.

Um valor $\rho_{X,Y}$ próximo de zero indica que X e Y não exercem muita influência um sobre o outro. Por outro lado, um valor $\rho_{X,Y}$ próximo de 1 indica que X influencia de forma intensa a Y e vice versa.

Por exemplo, suponha que X descreve uma pessoa genérica e sua probabilidade de ir a farmácia, e Y a probabilidade da pessoa estar com alguma moléstia não muito grave. Então X e Y estão correlacionadas: pessoas doentes tendem a ir a farmácia para poder comprar medicamentos. Neste caso, $\rho_{X,Y}$ deve ser grande e próximo de 1.

Um tópico de grande importância na Teoria dos Processos Estocásticos é a análise da questão: supondo que X_n, $n \in \mathbb{N}$, descreve um processo estocástico estacionário, para cada $n \in \mathbb{N}$, denote por c_n a correlação de X_0, X_n. Se o processo fosse independente então $c_n = 0$ para todo n. Se existisse $c, \lambda > 0$, tais que $c_n \leq c\,e^{-n\lambda}$, para todo n, dizemos que o processo X_n, $n \in \mathbb{N}$, possui decaimento exponencial de correlação. Esta propriedade corresponde a um processo X_n, que não é independente, mas que a evolução temporal de X_n é tal que a longo prazo se parece com ele. Ou seja, se n é grande, X_n quase não é influenciado por X_0. Neste caso se diz, de forma heurística, que o processo X_n começa a perder rapidamente memória com o decorrer do tempo.

Se existissem $0 < c, 0 < \beta < 1$, tais que $c_n \leq c\,\beta^n$, para todo n, dizemos que o processo X_n, $n \in \mathbb{N}$, possui decaimento polinomial (ou, hiperbólico) de correlação. Neste caso, se n é grande, X_n quase não é influenciado por X_0, mas a velocidade com que isto ocorre é bem mais lenta do que no caso de decaimento exponencial de correlação.

Muitas propriedade importantes, como o Teorema Central do Limite (em casos mais gerais de processos estacionários não independentes), podem ser obtidas a partir de propriedades oriundas da velocidade do decaimento de correlação [B] [S] [Du].

5.3 Processos Estocásticos Independentes

O Teorema Central do Limite para variáveis independentes foi obtido através da convergência de $\frac{S_n - n\mu}{\sqrt{n\,\sigma^2}}$, no sentido de distribuição, quando $n \to \infty$, e com $S_n = X_1 + X_2 + ... + X_n$. A Lei Fraca dos Grandes Números considera apenas a probabilidade de X_n, quando n vai a infinito. Desejamos analisar outros resultados para processos independentes que considerem convergência em P-quase todo ponto. Para isto, necessitaremos de vários resultados preliminares que serão abordados a seguir.

O leitor poderá encontrar uma versão mais abrangente e completa dos tópicos decritos aqui em [B].

Definição 5.32. *Uma sequência de conjuntos $B_1, B_2, ..., B_n \in \mathcal{F}$ são ditos independentes em relação à $(\Omega, \mathcal{F}, P)$ se*

$$P(B_1 \cap B_2 \cap ... \cap B_n) \,=\, P(B_1)\, P(B_2) \,...\, P(B_n).$$

Definição 5.33. *Uma sequência de classes de conjuntos $\mathcal{B}_1, \mathcal{B}_2, ..., \mathcal{B}_n \subset \mathcal{F}$ são ditos independentes em relação à $(\Omega, P, \mathcal{F})$ se*

$$P(B_1 \cap B_2 \cap ... \cap B_n) \,=\, P(B_1)\, P(B_2) \,...\, P(B_n),$$

para toda seqüência de conjuntos

$$B_1 \in \mathcal{B}_1, B_2 \in \mathcal{B}_2, ..., B_n \in \mathcal{B}_n.$$

Na definição acima as classes de conjuntos $\mathcal{B}_i$ **não são** necessariamente σ-álgebras.

Uma pequena sutileza, no que segue, iremos sempre considerar que Ω é um dos elementos de cada um dos $\mathcal{B}_i$, $i \in \{1, 2, ..., n\}$. Sendo assim, se por exemplo, temos três coleções de conjuntos independentes, $\mathcal{B}_1$, $\mathcal{B}_2$, $\mathcal{B}_3$, então para $B_1 \in \mathcal{B}_1$ e $B_3 \in \mathcal{B}_3$, vale que

$$P(B_1 \cap B_3) = P(B_1)\, P(B_3).$$

Teorema 5.18. *Suponha que as classes de conjuntos $\mathcal{B}_1, \mathcal{B}_2, ..., \mathcal{B}_n \subset \mathcal{F}$ são independentes para $(\Omega, P, \mathcal{F})$, que $\Omega \in \mathcal{B}_i$ para todo $i \in \{1, 2, ..., n\}$ e ainda que cada $\mathcal{B}_i$ seja um sistema-π, então*

$$\sigma(\mathcal{B}_1), \sigma(\mathcal{B}_2), ..., \sigma(\mathcal{B}_n)$$

são σ-álgebras independentes.

Demonstração: Podemos supor que cada B_i tem probabilidade positiva.

Fixados $B_2 \in \mathcal{B}_2, B_3 \in \mathcal{B}_3, ..., B_n \in \mathcal{B}_n$, a lei induzida P_1 sobre conjuntos $B \in \mathcal{F}$ por

$$P_1(B) = \frac{P(\, B \cap B_2 \cap B_3 ... \cap B_n\,)}{P(B_2)\, P(B_3) ... P(B_n)}$$

é uma probabilidade sobre $\mathcal{F}$.

Ainda, para conjuntos B em $\mathcal{B}_1$ temos a igualdade $P(B) = P_1(B)$. Logo, como assumimos que $\mathcal{B}_1$ é um sistema-π, só existe uma extensão de P_1 a $\sigma(\mathcal{B}_1)$ e esta deve coincidir com P. Logo, para todo $B_1 \in \sigma(\mathcal{B}_1)$ temos que

$$P(B_1) = \frac{P(\, B_1 \cap B_2 \cap B_3 ... \cap B_n\,)}{P(B_2)\, P(B_3) ... P(B_n)}.$$

Logo,

$$\sigma(\mathcal{B}_1), \mathcal{B}_2, \mathcal{B}_3, ..., \mathcal{B}_n,$$

são independentes.

Fixados $B_1 \in \sigma(\mathcal{B}_1), B_2 \in \mathcal{B}_2, ..., B_n \in \mathcal{B}_n$, a lei induzida P_2 sobre conjuntos $B \in \mathcal{F}$ por

$$P_2(B) = \frac{P(\, B_1 \cap B \cap B_3 ..., \cap B_n\,)}{P(B_1)\, P(B_3) ... P(B_n)}$$

é uma probabilidade sobre $\mathcal{F}$.

Ainda, para conjuntos B em $\mathcal{B}_2$ temos a igualdade $P(B) = P_2(B)$. Logo, como $\mathcal{B}_2$ é um sistema π, só existe uma extensão de P_2 a $\sigma(\mathcal{B}_2)$ e esta deve coincidir com P. Logo, para todo $B_2 \in \sigma(\mathcal{B}_2)$ temos que

$$P(B_2) = \frac{P(\, B_1 \cap B_2 \cap B_3 ..., \cap B_n\,)}{P(B_1)\, P(B_3) ... P(B_n)}.$$

Logo,

$$\sigma(\mathcal{B}_1), \sigma(\mathcal{B}_2), \mathcal{B}_3, ..., \mathcal{B}_n,$$

são independentes.

O resultado segue de aplicar o processo indutivo descrito acima.

□

Definição 5.34. *Seja uma coleção de índices* Θ. *Os conjuntos* $B_\theta \in \mathcal{F}$, *com* $\theta \in \Theta$, *são ditos independentes em relação à* $(\Omega, P, \mathcal{F})$, *se para qualquer escolha de finitos índices* $\theta_1, \theta_2, ..., \theta_n \in \Theta$, *vale*

$$P(B_{\theta_1} \cap B_{\theta_2} \cap ... \cap B_{\theta_n}) \,=\, P(B_{\theta_1})\, P(B_{\theta_2}) ... P(B_{\theta_n}).$$

O conjunto Θ acima não precisa ser enumerável.

Definição 5.35. *Seja uma coleção de índices* Θ. *As classes de conjuntos* $\mathcal{B}_\theta \subset \mathcal{F}$, *com* $\theta \in \Theta$, *são ditas independentes em relação à* $(\Omega, P, \mathcal{F})$, *se para qualquer escolha de finitos índices distintos* $\theta_1, \theta_2, ..., \theta_n \in \Theta$, *e de conjuntos* $B_{\theta_1} \in \mathcal{B}_{\theta_1}, B_{\theta_2} \in \mathcal{B}_{\theta_2}, ..., B_{\theta_n} \in \mathcal{B}_{\theta_n}$, *vale*

$$P(B_{\theta_1} \cap B_{\theta_2} \cap ... \cap B_{\theta_n}) \,=\, P(B_{\theta_1})\, P(B_{\theta_2}) ... P(B_{\theta_n}).$$

Dizemos que a **variável aleatória** $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{R})$**, é independente da** σ**-álgebra** $\mathcal{G}$, se a σ-álgebra $\mathcal{F} = \{X^{-1}(B)\,|\, B$ conjunto de Borel em $\mathcal{R}\}$ é independente da σ-álgebra $\mathcal{G}$.

Isto implica que se G é $\mathcal{G}$ mensurável, então

$$\int I_G\, X\, dP = \int I_G\, dP \int X\, dP.$$

De fato, $P(G \cap X^{-1}(B)) = P(G) \times P(X^{-1}(B))$. Assim,

$$\int I_G\, I_{X^{-1}(B)}\, dP = \int I_G\, dP \int I_{X^{-1}(B)}\, dP.$$

Desta forma se $Y_n = \sum_{j=1}^n a_j I_{X^{-1}(B_j)}$ temos que

$$\int I_G\, Y_n\, dP = \int I_G\, dP \int Y_n\, dP.$$

Tomando agora Y_n sequência monótona convergindo a X o resultado segue.

Teorema 5.19. *Seja Θ um conjunto de índices e $\mathcal{B}_\theta$, $\theta \in \Theta$, classes de conjuntos independentes, e suponha que cada $\mathcal{B}_\theta$ seja um sistema-π, então*

$$\sigma(\mathcal{B}_\theta),\ \theta \in \Theta,$$

são independentes.

Não apresentaremos a prova do resultado acima, pois segue raciocínio análogo aos casos anteriores.

Usaremos a notação $\sigma(\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_n, ...)$, onde $\mathcal{A}_i$, $i \in \mathbb{N}$, são classes de conjuntos em uma σ-álgebra $\mathcal{F}$, para denotar a σ-álgebra

$$\sigma\left(\mathcal{A}_1 \cup \mathcal{A}_2 \cup ... \cup \mathcal{A}_n \cup ...\right).$$

Fixada uma seqüência de classes $\mathcal{A}_i$, $i \in \mathbb{N}$, gostaríamos de considerar para um n fixo o conjunto $\sigma(\mathcal{A}_n, \mathcal{A}_{n+1}, \mathcal{A}_{n+2}, ...)$. Esta σ-álgebra só contém informação dos conjuntos em $\mathcal{A}_j$ para $j \geq n$.

Exemplo 5.16. Por exemplo, considere $S = \{1, 2\}$, $\mathcal{F}$ a σ-álgebra gerada pelos cilindros em $S^{\mathbb{N}}$, e

$$\mathcal{A}_i = \{\, \emptyset,\ S^{\mathbb{N}},\ \{\omega \,|\, X_i(w) = 1\}, \{\omega \,|\, X_i(w) = 2\}\,\}.$$

Associe 1 à cara e 2 à coroa. O conjunto $\mathcal{A}_i$ é uma σ-álgebra e descreve a informação obtida pelo que acontece apenas na i-ésima vez que jogamos a moeda.

Fixando $n = 4$, a σ-álgebra $\sigma(\mathcal{A}_4, \mathcal{A}_5, ..., \mathcal{A}_n, ...)$ descreve a informação total obtida ao jogar a moeda todas as vezes após o tempo $n = 3$. Esta σ-álgebra, por outro lado, não traz informação do que ocorre antes do tempo $n = 4$.

$\diamondsuit$

Definição 5.36. *Considere o espaço mensurável* $(\Omega, \mathcal{F})$ *e a seqüência de classes de conjuntos*

$$\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_n, ..., \subset \mathcal{F}.$$

A σ*-álgebra cauda é*

$$\mathcal{T} = \cap_{n=1}^{\infty} \sigma(\mathcal{A}_n, \mathcal{A}_{n+1}, ...).$$

Considere $\Omega = \{1,2\}^{\mathbb{N}}$, $\mathcal{F}$ a σ-álgebra gerada pelos cilindros (de todos os ranks) e $\mathcal{A}_n$ a σ-álgebra gerada pelos conjuntos $\{1,2\}^{n-1} \times \{1\} \times \{1,2\}^{\mathbb{N}}$ e $\{1,2\}^{n-1} \times \{2\} \times \{1,2\}^{\mathbb{N}}$.

O conjunto $\{1,2\} \times \{1,2\} \times \{1\}^{\mathbb{N}}$ está em $\sigma(\mathcal{A}_3, \mathcal{A}_4, ...)$

Exemplo 5.17. Seja uma seqüência A_i de elementos de $\mathcal{F}$ e tome $\mathcal{A}_i = \{A_i\}$. Seja

$$A = \limsup_{m \to \infty} A_m = \cap_m^{\infty} \cup_{j \geq m} A_j,$$

e

$$B = \liminf_{m \to \infty} A_m = \cup_m^{\infty} \cap_{j \geq m} A_j.$$

Afirmamos que neste caso, $A \in \mathcal{T}$. De fato, para cada m fixo $A \in \sigma(\mathcal{A}_m, \mathcal{A}_{m+1}, ...)$.

Da mesma forma, $B \in \mathcal{T}$.

Neste exemplo fica bem claro que determinar se um certo conjunto C está ou não na σ-álgebra cauda, é algo que não pode ser determinado por um número finito de informações.

$\diamondsuit$

Definição 5.37. *Os conjuntos em* $\mathcal{T}$ *são chamados de eventos cauda e um qualquer destes conjuntos não pode depender do que acontece em finitos* $\mathcal{A}_n$.

Exemplo 5.18. Vamos exemplificar através de um problema interessante nossa afirmação: considere $S = \{1, 2\}$, $\mathcal{F}$ a σ-álgebra gerada pelos cilindros em $S^{\mathbb{N}}$, e

$$\mathcal{A}_i = \{\, \emptyset,\, S^{\mathbb{N}},\, \{\omega \,|\, X_i(w) = 1\}, \{\omega \,|\, X_i(w) = 0\}\, \}.$$

Associe 1 à cara e 0 à coroa. Vamos supor que P sobre a σ-álgebra gerada pelos cilindros advém de supor que temos um processo independente e que $P(X_1(w) = 1) = 1/2$ e $P(X_1(\omega) = 0) = 1/2$.

Uma pergunta natural é saber se fixado $\omega = (w_1, w_2, ..., w_n, ...)$, vale ou não a afirmação

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} w_i \;=\; \frac{1}{2}.$$

Esperaríamos que para a "maioria" dos $\omega \in \{1, 2\}^{\mathbb{N}}$ fosse válida a afirmação acima. De fato, para um evento ω fixo, a existência do limite acima traduziria o fato que se jogamos uma moeda (honesta) infinitas vezes e w_i descreve a face que sai na i-ésima jogada, então a média de vezes que sai cara em n jogadas converge a $1/2$.

Considere então o conjunto

$$A = \{\omega = (w_1, w_2, ..., w_n, ...) \,|\, \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} w_i \;=\; \frac{1}{2}\}.$$

Afirmamos que A está na σ-álgebra cauda da seqüência

$$\mathcal{A}_i = \{\, \emptyset,\, S^{\mathbb{N}},\, \{\omega = (w_1, w_2, ..) \,|\, w_i = 1\}, \{\omega = (w_1, w_2, ...) \,|\, w_i = 0\}\, \}.$$

A afirmação acima é bem natural, visto que para saber se existe o limite $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} w_i \;=\; \frac{1}{2}$ não basta saber se ω pertence a quais conjuntos de finitas coleções $\mathcal{A}_i$. Ou seja, não podemos determinar se $\omega = (w_1, w_2, ..., w_n, ...) \in A$, olhando apenas um número finito de w_i.

De maneira mais formal,

$$A = \cap_{m=1}^{\infty} \cup_{N=0}^{\infty} \cap_{n>N}^{\infty} \{\omega \mid \ |\frac{1}{n}\sum_{i=0}^{n} w_i - \frac{1}{2}| \leq \frac{1}{m}\}.$$

Na verdade, um forma mais interessante é

$$A = \cap_{m=1}^{\infty} \cup_{N>2\,m^2}^{\infty} \cap_{n>N}^{\infty} \{\omega \mid \ |\frac{1}{n}\sum_{i=m}^{n} w_i - \frac{1}{2}| \leq \frac{1}{m}\}.$$

Desta maneira fica claro que A é um evento cauda.

Será que $P(A) = 1$?

$\diamondsuit$

O próximo resultado dará uma resposta parcial a questão.

Teorema 5.20. *Sejam $A_1,\ A_2, ..., A_n, ... \subset \Omega$ uma seqüência de conjuntos independentes e*

$$A \in \mathcal{T} = \cap_{n=1}^{\infty} \sigma(A_n, A_{n+1}, ...),$$

então $P(A) = 0$ ou $P(A) = 1$.

O resultado acima é conhecido como a Lei zero ou um de Kolmogorov.

Antes de apresentar a demonstração deste teorema precisamos de dois Lemas.

Lema 5.2. *Sob as hipóteses acima, para qualquer n fixo vale que*

$$\sigma(A_1), \sigma(A_2), ..., \sigma(A_{n-1}) \ \ e \ \ \sigma(A_n, A_{n+1}, A_{n+2}, ...),$$

são independentes.

Demonstração: De fato, considere $\mathcal{G}$ o conjunto de todas as interseções finitas de elementos $A_n, A_{n+1}, A_{n+2}, ...$

Sendo assim, as coleções de conjuntos

$$\mathcal{A}_1 = \{A_1, \emptyset, \Omega\},\ \mathcal{A}_2 = \{A_2, \emptyset, \Omega\}, ..., \mathcal{A}_{n-1} = \{A_{n-1}, \emptyset, \Omega\},\ \mathcal{G},$$

são tais que cada um é um sistema-π.

Ainda, o conjunto de n coleções acima é independente. De fato, para $B = A_{i_1} \cap A_{i_2} \cap A_{i_3} \cap ... \cap A_{i_j} \in \mathcal{G}$, onde $A_{i_1}, A_{i_2}, A_{i_3}, ..., A_{i_j}$ foram escolhidos entre $\{A_n, A_{n+1}, A_{n+2}, ...\}$, temos que

$$P(A_1 \cap A_2 \cap A_3 \cap, ..., A_{n-1} \cap B) =$$

$$P(A_1 \cap A_2 \cap A_3 \cap ...A_{n-1} \cap A_{i_1} \cap A_{i_2} \cap A_{i_3} \cap ... \cap A_{i_j}) =$$

$$P(A_1)\, P(A_2)\, P(A_3)...P(A_{n-1})\, P(A_{i_1})\, P(A_{i_2})\, P(A_{i_3}) ... P(A_{i_j}) =$$

$$P(A_1)\, P(A_2)\, P(A_3)...\, P(A_{n-1})\ P(A_{i_1} \cap A_{i_2} \cap A_{i_3} \cap ... \cap A_{i_j}) =$$

$$P(A_1)\, P(A_2)\, P(A_3)...\, P(A_{n-1})\, P(B).$$

Na demonstração acima poderíamos substituir alguns A_i por Ω, e tudo continuaria valendo da mesma forma, ou seja, por exemplo,

$$P(A_1 \cap \Omega \cap A_3 \cap ...A_{n-1} \cap B) =$$

$$P(A_1)\, P(\Omega)\, P(A_3)...\, P(A_{n-1})\, P(B) =$$

$$P(A_1)\, P(A_3)...\, P(A_{n-1})\, P(B).$$

Logo, do Teorema 5.19 acima concluímos que

$$\sigma(\mathcal{A}_1), \sigma(\mathcal{A}_2), ..., \sigma(\mathcal{A}_{n-1}), \sigma(\mathcal{G}),$$

são independentes.

É fácil ver que $\sigma(\mathcal{G}) = \sigma(A_n, A_{n+1}, ...)$ e assim o lema está provado.

$\square$

Lema 5.3. *Seja A um evento cauda, então*

$$\sigma(A) \text{ e } \sigma(A_1, A_2, A_3, ...),$$

são independentes.

Demonstração: Seja $\mathcal{G}$ a coleção das interseções finitas dos A_i, $i \in \mathbb{N}$.

$\mathcal{G}$ é um sistema-π.

Seja $B = A_{i_1} \cap A_{i_2} \cap A_{i_3} \cap ... \cap A_{i_j} \in \mathcal{G}$, onde

$$A_{i_1}, A_{i_2}, A_{i_3}, ..., A_{i_j},$$

foram escolhidos entre $\{A_1, A_2, A_3, ...\}$.

Seja $n+1$ maior que todos estes índices $\{i_1, i_2, ..., i_j\}$.

Ora, $\sigma(A_{n+1}, A_{n+2}, ...)$ é independente de

$$A_1, A_2, ..., A_n,$$

logo temos que $\sigma(A_{n+1}, A_{n+2}, ...)$ é independente de $A_{i_1}, A_{i_2}, A_{i_3}, ..., A_{i_j}$.

Como $A \in \sigma(A_{n+1}, A_{n+2}, ...)$, temos que A é independente de

$$A_{i_1}, A_{i_2}, A_{i_3}, ..., A_{i_j}.$$

Da mesma forma como no lema precedente pode se mostrar que

$$P(A \cap B) = P(A)\, P(B).$$

Como A e $\mathcal{G}$ definem sistemas-π concluímos que $\sigma(A)$ e $\sigma(\mathcal{G})$ são independentes. Finalmente, como

$$\sigma(\mathcal{G}) = \sigma(A_1, A_2, ...),$$

o resultado segue.

$\square$

O teorema 5.20 agora pode ser demonstrado de maneira breve.

Demonstração: Seja A um conjunto na σ-álgebra cauda. Ora, $A \in \sigma(A_1, A_2, ..., A_n)$ e $\sigma(A)$ é independente de $\sigma(A_1, A_2, ..., A_n)$, logo

$$P(A) = P(A \cap A) = P(A)\, P(A) = P(A)^2.$$

Logo, $P(A) = 0$ ou $P(A) = 1$.

$\square$

O Teorema 5.20 acima é de grande utilidade em certas circunstâncias. Se queremos provar que um certo conjunto A tem probabilidade 1 e se por acaso ele está na σ-álgebra cauda, basta mostrar que $P(A) > 0$.

Os resultados acima envolviam probabilidades e conjuntos mensuráveis, vamos agora considerar os resultados análogos para processos estocásticos.

Dado $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{R})$ mensurável é usual denotar $\mathcal{F}_f$ como

$$\mathcal{F}_f = \{f^{-1}(B) \,|\, B \in \mathcal{R}\}.$$

Definição 5.38. *Dizemos que duas funções integráveis $f : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{R})$ e $g : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{R})$ são independentes se para quaisquer $A, B \in \mathcal{R}$,*

$$P(\{\omega \,|\, f(\omega) \in A, g(\omega) \in B\}) = P(\{\omega \,|\, f(\omega) \in A\})\; P(\{\omega \,|\, g(\omega) \in B\}).$$

De outra forma, se as σ-álgebras $\mathcal{F}_f$ e $\mathcal{F}_g$, geradas respectivamente por f e g, são independentes.

Ainda, f é independente da σ-álgebra $\mathcal{G}$ se $\mathcal{G}$ e $\mathcal{F}_f$ são independentes.

Mais geralmente as funções f_i, $i \in \mathbb{N}$, são independentes se a sequência de σ-álgebras $\mathcal{F}_{f_i}$, $i \in \mathbb{N}$, é independente no sentido anterior.

Teorema 5.21. *Sejam $f : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{R})$ e $g : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{R})$ independentes e $h : (\mathbb{R}, \mathcal{R}) \to (\mathbb{R}, \mathcal{R})$ mensurável, então $h \circ f$ e g são independentes.*

Demonstração: Suponha inicialmente que h seja uma função da forma $h = c\, I_C$, para algum $C \in \mathcal{R}$ e $c \in \mathbb{R}$.

Sejam $A, B \in \mathcal{R}$, Ora se $c \in A$, então

$$P(\{\omega \,|\, h \circ f\,(\omega) \in A, g\,(\omega) \in B\}) =$$

$$P(\{\omega \,|\, f\,(\omega) \in C, g\,(\omega) \in B\}) =$$

$$P(\{\omega \,|\, f\,(\omega) \in C\})\; P(\{\omega, |\, g\,(\omega) \in B\}) =$$

$$P(\{\omega \,|\, h \circ f\,(\omega) \in A\})\; P(\{\omega, |\, g(\omega) \in B\}).$$

Se c não está em A, então os dois lados das igualdades acima são nulos.

O argumento facilmente passa para h simples, ou seja h da forma $h = \sum_{i=1}^k c_i\, I_{c_i}$. Finalmente, como toda h mensurável é limite crescente de funções simples h_n, $n \in \mathbb{N}$, podemos considerar o resultado acima para $h_n \circ f$ e g e fazer o limite em n para obter o resultado desejado.

$\square$

Teorema 5.22. *Se as duas funções integráveis* $f : (\Omega, P, \mathcal{F}) \to (\mathbb{R}, \mathcal{R})$ *e* $g : (\Omega, P, \mathcal{F}) \to (\mathbb{R}, \mathcal{R})$ *são independentes então*

$$\int f\,(\omega)\, g\,(\omega)\, dP(\omega) = \int f\,(\omega) dP(\omega) \int g\,(\omega)\, dP(\omega).$$

Demonstração: Quando as funções independentes f e g são funções simples, o resultado é trivial.

É importante agora considerar a seguinte forma canônica de se obter a partir de f uma sequência monótona de funções simples h_n, $n \in \mathbb{N}$, que converge a f

Primeiro, vamos supor que f é não negativa.

Para cada valor n considere a seguinte função mensurável $z_n : \mathbb{R}^+ \to \mathbb{R}$,

a) $z_n(x) = n$, se $x > n$

b) $z_n(x) = \frac{j-1}{2^n}$, se x é tal que $\frac{(j-1)}{2^n} < x \leq \frac{j}{2^n}$, para algum $j \in \{1, 2, 3, 4, ..., 2^n\}$.

Considere agora $h_n = z_n \circ f$. Note que, para n fixo, se *fatiarmos* a imagem de f pela interseção com conjuntos da forma $[\frac{j-1}{n}, \frac{j}{n})$, $j \in \{1, 2, 3, 4, ..., 2^n\}$ então a z_n vai levar cada *fatia* num valor $\frac{j-1}{2^n}$ (que delimita a parte debaixo da *fatia*).

Logo as funções $h_n(\omega)$ são simples e convergem (pontualmente para todo ω, e ainda monotonamente) a $f(\omega)$.

O ponto importante do método acima é que as z_n estão fixas, e o procedimento funciona também para uma g qualquer, ou seja, $v_n = z_n \circ g$ converge a g como acima.

Em resumo, sabemos que

$$\lim_{n\to\infty} \int h_n \, dP = \lim_{n\to\infty} \int (z_n \circ f) \, dP = \int f \, dP,$$

$$\lim_{n\to\infty} \int v_n \, dP = \lim_{n\to\infty} \int (z_n \circ g) \, dP = \int g \, dP,$$

e

$$\lim_{n\to\infty} \int h_n \, v_n \, dP = \lim_{n\to\infty} \int (z_n \circ f) \, (z_n \circ g) \, dP = \int f \, g \, dP.$$

Segue da proposição acima que, fixado n, a função $(z_n \circ f)$ é independente de g. Aplicando mais uma vez, o resultado anterior, temos que $(z_n \circ f)$ é independente de $(z_n \circ g)$.

Ora, $(z_n \circ g)$ e $(z_n \circ f)$ são simples, então, conforme afirmamos no início da demonstração, vale que

$$\int (z_n \circ f) \, (z_n \circ g) \, dP = \int (z_n \circ f) \, dP \, \int (z_n \circ g) \, dP.$$

O resultado segue das três igualdades entre integrais acima.

□

Definição 5.39. *Dado um conjunto de índices* Θ, *e uma família de funções mensuráveis* $X_\theta : (\Omega, P, \mathcal{A}) \to (\mathbb{R}, \mathcal{R})$, *denotamos* $\mathcal{F} = \sigma(X_\theta, \theta \in \Theta) \subset \mathcal{A}$, *a menor* σ*-álgebra contida em* $\mathcal{A}$ *que torna todas as funções* X_θ *mensuráveis.*

Esta coleção é obtida tomando todos os conjuntos possíveis da forma

$$X_{\theta_1}^{-1}(A_1) \cap X_{\theta_2}^{-1}(A_2) \cap ... \cap X_{\theta_n}^{-1}(A_n),$$

$n = 1, 2, ...,$ e A_j boreleano na reta.

Quando Θ é finito ou enumerável usaremos a notação

$$\sigma(X_1, X_2, X_3, ..., X_k),$$

ou

$$\sigma(X_1, X_2, X_3, ..., X_k, ...).$$

A sigma-algebra assim obtida é denominada de sigma-algebra gerada por um conjunto de funções.

Quando S é enumerável e $X_1, X_2, ..., X_k$, são tais que $X_i : (\Omega, \mathcal{F}) \to (S, \mathcal{R})$, então,

$$\sigma(X_1, X_2, X_3, ..., X_k) = \sigma(\mathcal{V}),$$

onde $\mathcal{V}$ são os conjuntos da forma $\{\omega \,|\, X_{i_1} = a_1, X_{i_2} = a_2, ..., X_{i_r} = a_r\}$, e onde $a_j \in S$, $i_j \in \{1, 2, ..., k\}$, $j \in \{1, 2, ..., r\}, r \leq n$.

Sugerimos ao leitor muita cautela ao tentar imaginar as possíveis generalizações da afirmação acima para os outros casos.

Considere um processo estocástico $X_1, X_2, .., X_n, .., n \in \mathbb{N}$, tomando valores em $S = \{1, 2\}$. Denote por $\mathcal{A}$ a sigma algebra $\sigma(X_2, X_3, ..., X_n, ...)$. Assim esta sigma algebra não contem informação da primeira coordenada.

Observe que $\sigma^{-1}(\overline{121}) = \overline{1121} \cup \overline{2121} \in \mathcal{A}$

Segue do teorema 5.15 que as funções que são mensuráveis em relação a sigma-algebra $\mathcal{A}$ são as funções da forma $\varphi \circ \sigma$ onde φ é mensuravel em relação a sigma algebra $\sigma(X_1, X_2, X_3, ..X_n, ...)$.

Se consideramos $\mathcal{A}$ a sigma algebra $\sigma(X_3, X_4, ..., X_n, ...)$, então as funções que são mensuráveis em relação a sigma-algebra $\mathcal{A}$ são as funções da forma

$\varphi \circ \sigma^2$ onde φ é mensuravel em relação a sigma algebra $\sigma(X_1, X_2, X_3, ..., X_n, ...)$ e assim por diante.

Definição 5.40. *Seja X_n, $n = 1, 2, ..., n, ...$, Processo Estocástico com conjunto de estados S. A σ-álgebra cauda do Processo Estocástico X_n, $n \in \mathbb{N}$ é a coleção*

$$\mathcal{T} = \cap_{k=1}^{\infty} \sigma(X_k, X_{k+1}, ...).$$

Fixado o número real a, podemos nos perguntar se existem $\omega \in \Omega = S^{\mathbb{N}}$ tais que

$$\lim_{n \to \infty} \frac{1}{n} \left(X_1(\omega) + X_2(\omega) + ... + X_n(\omega) \right) = a.$$

Denotemos por A_a o conjunto de tais ω.

De maneira mais formal,

$$A_a = \cap_{m=1}^{\infty} \cup_{N=0}^{\infty} \cap_{n>N}^{\infty} \{\omega | \, |\frac{1}{n} \sum_{i=1}^{n} X_i(\omega) - a \,| \leq \frac{1}{m}\}.$$

Uma outra forma equivalente de descrever A_a seria

$$A_a = \cap_{m=1}^{\infty} \cup_{N \geq m}^{\infty} \cap_{n>N}^{\infty} \{\omega | \, |\frac{1}{n} \sum_{i=m}^{n} X_i(\omega) - a \,| \leq \frac{1}{m}\}.$$

Desta forma, se obtém que A_a é um evento cauda do Processo Estocástico X_n, $n \in \mathbb{N}$.

Teorema 5.23. *Se X_n, $n \in \mathbb{N}$, é um Processo Estocástico Independente (Definição 1.11) então, para cada a fixo o conjunto*

$$A_a = \{\omega \,|\, \lim_{n \to \infty} \frac{1}{n} \left(X_1(\omega) + X_2(\omega) + ... + X_n(\omega) \right) = a \,\}$$

tem probabilidade 0 ou 1 para a P associada (ao processo) sobre o conjunto $S^{\mathbb{N}}$.

Esta teorema segue de imediato do seguinte:

Teorema 5.24. *Se X_n, $n \in \mathbb{N}$, é um Processo Estocástico Independente e*

$$A \in \mathcal{T} = \cap_{k=1}^{\infty} \sigma(X_k, X_{k+1}, ...),$$

então A tem probabilidade 0 ou 1 para a P associada (ao processo) sobre $S^{\mathbb{N}}$.

Demonstração: A ideia principal da prova é mostrar que $P(A)^2 = P(A)$. O procedimento é semelhante ao do Teorema 5.18 mas algumas diferenças aparecem na prova.

Primeiro, afirmamos que

$$\cup_{j=1}^{\infty} \sigma(X_1, X_2, ..., X_j),$$

gera a σ-álgebra

$$\sigma(X_1, X_2, ..., X_k, ...).$$

Assumindo isto, vamos mostrar o teorema.

Ora, se $A \in \mathcal{T}$, então, $A \in \sigma(X_{j+1}, X_{j+2}, ...)$, para todo j.

Então para j fixo, como o processo X_n, $n \in \mathbb{N}$, é independente, temos que, A é independente de $\sigma(X_1, X_2, ..., X_j)$. Logo temos que A é independente de

$$\cup_{j=1}^{\infty} \sigma(X_1, X_2, ..., X_j).$$

Como este último conjunto gera $\sigma(X_1, X_2, ..., X_k, ...)$, então, pelo Teorema 5.18, concluímos que A é independente de $\sigma(X_1, X_2, ..., X_k, ...)$.

Ora, A é elemento desta σ-álgebra, logo $P(A) = P(A \cap A) = P(A)\, P(A)$. Logo, o resultado está demonstrado.

Vamos agora mostrar que

$$\sigma(\, \cup_{j=1}^{\infty} \sigma(X_1, X_2, ..., X_j)\,) \,=\, \sigma(X_1, X_2, ..., X_k, ...).$$

Primeiro note que $\cup_{j=1}^{\infty}\sigma(X_1, X_2, ..., X_j)$ é uma álgebra. De fato, se

$$B \in \sigma(X_1, X_2, ..., X_k)$$

e

$$C \in \sigma(X_1, X_2, ..., X_r),$$

então

$$B \cup C \ \in \ \sigma(X_1, X_2, ..., X_{r+k}).$$

Se

$$B \in \sigma(X_1, X_2, ..., X_k),$$

então $X - B$ também pois $\sigma(X_1, X_2, ..., X_k)$ é σ-álgebra. Logo,

$$\cup_{j=1}^{\infty}\sigma(X_1, X_2, ..., X_j),$$

é uma álgebra.

Devemos mostrar finalmente que

$$\sigma(X_1, X_2, ...) \subset \sigma(\, \cup_{j=1}^{\infty}\sigma(X_1, X_2, ..., X_j)\,).$$

Para isto, basta que cada X_r seja mensurável com relação à

$$\sigma(\, \cup_{j=1}^{\infty}\sigma(X_1, X_2, ..., X_j)\,).$$

Mas isto é trivial, pois X_r é mensurável em relação a $\sigma(X_1, X_2, ..., X_r)$

$\square$

Observamos que se X_n, $n \in \mathbb{N}$, for um Processo Estocástico Independente, no máximo para um valor de a, o conjunto A_a (definido no Teorema 5.23) pode ter probabilidade 1.

No resultado acima não usamos o fato que X_n, $n \in \mathbb{N}$, era identicamente distribuído.

Para um Processo Estocástico X_n fixado, e $\omega \in S^{\mathbb{N}}$ fixo, denotamos

$$S_n(\omega) = X_0(\omega) + X_1(\omega) + ... + X_{n-1}(\omega).$$

Note que o teorema acima assegura também que para cada $a \in \mathbb{R}$ fixo o conjunto

$$D_a = \{\omega \mid \lim_{n\to\infty} S_n(\omega) \ = \ \lim_{n\to\infty} (\, X_0(\omega) + X_1(\omega) + ... + X_{n-1}(\omega)) = a\,\}$$

tem probabilidade 0 ou 1 para a P associada sobre $S^{\mathbb{N}}$.

Em princípio, poderia não haver valor $a \in \mathbb{R}$ tal que $P(D_a) = 1$. Vamos tentar dar uma resposta a esta pergunta em um caso particular. Para isto necessitaremos de vários resultados.

A desigualdade de Chebyshev afirma que

$$P(|S_n| \ \geq \ a) \leq \frac{1}{a^2} \, Var(S_n).$$

Desejamos mostrar que sob certas condições, as somas $S_n(\omega)$ obtidas de Processos Estocásticos independentes convergem, para $\omega \in S^{\mathbb{N}}$ P-quase toda parte. Para isto necessitamos a seguinte versão da desigualdade acima:

Teorema 5.25. *Considere $X_0, X_1, ..., X_{n-1}$, $n \in \mathbb{N}$, variáveis aleatórias independentes $X_k : (\Omega, P, \sigma(\mathcal{C})) \to (\mathbb{R}, \mathcal{R})$, tal que a média $E(X_k) = 0$ para todo k e as variâncias $E(X_k^2)$ sejam finitas, $k \in \{0, 1, 2, ..., n-1\}$. Então, para $a \geq 0$,*

$$P(\max_{0\leq k\leq n-1} |S_k| \ \geq \ a) \leq \frac{1}{a^2} \, E(S_n^2) = \frac{1}{a^2} \, Var(S_n).$$

Demonstração: Seja

$$B_k = \{\omega \,|\, |S_k| \ \geq \ a, \text{ mas}, \ |S_j| \ < \ a, \ j \in \{0, 1, 2, ..., k-1\}\,\}.$$

Note que $B_k \cap B_j = \emptyset$, para $j \neq k$.

Ainda,

$$\{\omega \mid \max_{0\leq k\leq n-1} |S_k(\omega)| \geq a\} = \cup_{k=0}^{n-1} B_k \subset \Omega.$$

Logo,

$$E(S_n^2) \geq \sum_{k=0}^{n-1} \int_{B_k} S_n^2 \, dP =$$

$$\sum_{k=0}^{n-1} \int_{B_k} [\, S_k^2 + 2\, S_k\, (S_n - S_k) + (S_n - S_k)^2\,] \, dP \geq$$

$$\sum_{k=0}^{n-1} \int_{B_k} [\, S_k^2 + 2\, S_k\, (S_n - S_k)\,] \, dP.$$

Note que $S_n - S_k = X_k + ... + X_{n-1}$ é $\sigma(X_k, ..., X_{n-1})$ mensurável. S_k, por sua vez é $\sigma(X_0, ..., X_{k-1})$ mensurável. Ainda, como $E(S_n - S_k) = 0$, B_k é conjunto $\sigma(X_0, ..., X_{k-1})$ mensurável, e as σ-álgebras $\sigma(X_k, ..., X_{n-1})$ e $\sigma(X_0, ..., X_{k-1})$ são independentes, temos que

$$\int_{B_k} 2\, S_k\, (S_n - S_k) \, dP = \int 2\, I_{B_k}\, S_k\, (S_n - S_k) \, dP = 0.$$

Acima, usamos o fato que se S_k e I_{B_k} são independentes de $(S_n - S_k)$, então $S_k\, I_{B_k}$ é independente de $(S_n - S_k)$ e ainda o Teorema 5.22.

Finalmente,

$$E(S_n^2) \geq \sum_{k=0}^{n-1} \int_{B_k} [\, S_k^2\,] \, dP \geq$$

$$\sum_{k=0}^{n-1} a^2\, P(B_k) = a^2\, P(\,\{\omega \mid \max_{0\leq k\leq n-1} |S_k(\omega)| \geq a)\}\,).$$

□

O próximo teorema pode ser obtido dividindo

$$\{\omega \mid \max_{0\leq k\leq n-1} |S_k(\omega)| \geq 3\, a\},$$

em conjuntos da forma B_k, $k \in \{0,1,2,...,k-1\}$, de maneira semelhante a anterior e não será demonstrado (prova em [B]).

Teorema 5.26. *Considere* $X_0, X_1, ..., X_{n-1}$, $n \in \mathbb{N}$, *variáveis aleatórias independentes* $X_k : (\Omega, P, \sigma(\mathcal{C})) \to (\mathbb{R}, \mathcal{R})$. *Então, para* $a \geq 0$,

$$P(\max_{0\leq k\leq n-1} |S_k| \geq 3\,a) \leq 3 \max_{0\leq k\leq n-1} P(|S_k| \geq a).$$

O próximo resultado é para Processos Estocásticos Independentes não (necessariamente) identicamente distribuídos.

Teorema 5.27. *Considere* X_n, $n \in \mathbb{N}$, *Processo Estocástico Independente* $X_n : (\Omega, P, \sigma(\mathcal{C})) \to (\mathbb{R}, \mathcal{R})$, *tal que a média* $E(X_n) = 0$ *para todo* $n \in \mathbb{N}$, *e* $\sum_{k=0}^{\infty} E(X_k^2)$ *seja finita. Então,*

$$S_n(\omega) = X_0(\omega) + X_1(\omega) + ... + X_{n-1}(\omega),$$

converge para P*-quase-todo* $\omega \in \Omega = S^{\mathbb{N}}$.

Demonstração: Vamos mostrar que para quase todo ω a seqüência $S_n(\omega)$ é de Cauchy.

Ora, fixado m, então pelo teorema anterior e pelo fato que os X_k são independentes vale que

$$P(\max_{0\leq k\leq v} |(X_{n+1} + X_{n+2} + ... + X_{n+k})| \geq 1/m) =$$

$$P(\max_{0\leq k\leq v} |(S_{n+k} - S_n)| \geq 1/m) \leq$$

$$\frac{1}{(1/m)^2}\, E((X_{n+1} + X_{n+2} + ... + X_{n+v})^2) =$$

$$\frac{1}{(1/m)^2} \sum_{k=1}^{v} E(X_{n+k}^2).$$

Fazendo $v \to \infty$, obtemos que

$$P(\max_{0\leq k} |(S_{n+k} - S_n)| \geq 1/m) \leq \frac{1}{(1/m)^2} \sum_{k=1}^{\infty} E(X_{n+k}^2).$$

Como, $\sum_{k=1}^{\infty} E(X_k^2)$ é finita, $\sum_{k=n}^{\infty} E(X_k^2)$ converge a zero quando $n \to \infty$, logo temos que para m fixo,

$$\lim_{n\to\infty} P(\max_{0\leq k} |(S_{n+k} - S_n)| \geq 1/m) = 0.$$

O conjunto dos ω tais que $S_n(\omega)$ é de Cauchy é dado por

$$\cap_{m\geq 0} \cup_{n\geq 0} \{\omega \,|\, \max_{0\leq k_1 \leq k_2} |(S_{n+k_1}(\omega) - S_{n+k_2}(\omega))| \leq 1/m\,\}.$$

Se ω é tal que $S_k(\omega), k \in \mathbb{N}$, não é de Cauchy, então existe m tal que para qualquer n

$$\omega \in A_{m,n} = \{\omega \,|\, \max_{0\leq k_1 \leq k_2} |(S_{n+k_1}(w) - S_{n+k_2}(w))| \geq 1/m)\,\}.$$

Uma pequena variação do argumento acima mostra que também vale

$$\lim_{n\to\infty} P(\max_{0\leq k_1 \leq k_2} |(S_{n+k_1} - S_{n+k_2})| \geq 1/m) = 0.$$

Sendo assim, mostramos acima que para m fixo vale

$$P(\cap_{n\geq 0} A_{m,n}) = 0.$$

Logo, com probabilidade 1 vale que

$$S_n(\omega) = X_0(\omega) + X_1(\omega) + ... + X_{n-1}(\omega),$$

converge.

$\square$

5.4 Processos Estocásticos Estacionários e Ergódicos

Dizemos que μ é invariante para $T : \Omega \to \Omega$ se para qualquer A mensurável vale que $\mu(A) = \mu(T^{-1}(A))$.

Lembramos que $\mathcal{M}(T)$ é o conjunto das probabilidades invariantes μ para a transformação mensurável $T : (\Omega, \mathcal{A}, \mu) \to (\Omega, \mathcal{A}, \mu)$.

Os resultados a seguir são completamente gerais, mas o leitor pode pensar, se o desejar, que estamos falando sempre do caso em que $X = S^{\mathbb{N}}$, $\mathcal{A}$ é a σ-álgebra gerada pelos cilindros, $\mu = P$ provém de um processo estacionário e T é o shift σ agindo em $S^{\mathbb{N}}$. Neste caso, $\sigma(x_0, x_1, x_2, ...) = (x_1, x_2, x_3, ...)$, para qualquer $x = (x_0, x_1, x_2, ...) \in S^{\mathbb{N}}$.

Outro exemplo importante é $\Omega = S^{\mathbb{Z}}$, $\mathcal{A}$ é a σ-álgebra gerada pelos cilindros, $\mu = P$ provém de um processo estacionário e T é o shift σ agindo em $S^{\mathbb{Z}}$.

Neste caso, denotamos o x geral em $S^{\mathbb{Z}}$, como

$$x = (..., x_{-2}, x_{-1} \,|\, x_0, x_1, x_2, ...),$$

onde | serve para nos dizer onde está o índice zero de $\mathbb{Z}$. Assim,

$$\sigma(..., x_{-2}, x_{-1} \,|\, x_0, x_1, x_2, ...) = (..., x_{-1}, x_0 \,|\, x_1, x_2, x_3, ...).$$

Definição 5.41. *Dizemos que $\mu \in \mathcal{M}(T)$ é ergódica se toda vez que $A \in \mathcal{A}$ é tal que $T^{-1}(A) = A$, então é porque $\mu(A) = 0$ ou $\mu(A) = 1$.*

Note que só faz sentido perguntar se uma probabilidade é ergódica se ela é invariante.

Dizemos que um processo estocástico estacionário $X_n, n \in \mathbb{N}$, onde cada X_n toma valores em S, é ergódico se a medida μ em $\Omega = S^{\mathbb{N}}$, obtida a partir do processo (e que é invariante pelo shift $\sigma : \Omega \to \Omega$) é ergódica.

Definição 5.42. *Diremos que um Processo Estocástico Estacionário é ergódico se a correspondente P sobre $S^{\mathbb{N}}$ for ergódica para o shift T.*

Referimos o leitor para [PY] para uma exposição mais completa sobre Teoria Ergódica.

Definição 5.43. *Dizemos que um conjunto A é invariante para o shift se $T^{-1}(A) = A$.*

Definição 5.44. *Dizemos que um conjunto $A \in \mathcal{A}$ é um conjunto trivial para μ se $\mu(A) = 0$ ou $\mu(A) = 1$.*

Sendo assim, uma probabilidade invariante é ergódica se os únicos conjuntos A invariantes são triviais.

Note que $A = \emptyset$ ou $A = \Omega$ são sempre invariantes, de qualquer forma eles tem probabilidades respectivamente 0 e 1.

Exemplo 5.19. Dada a transformação $T(x) = x + \lambda$ (mod 1), onde λ é uma constante irracional ($X = [0, 1)$, $\mathcal{R}$ é a σ-álgebra) e a probabilidade associada $P = \mu$ tal que $d\mu(x) = dx$, ou seja, a medida de Lebesgue em $[0, 1]$, vamos mostrar que μ é ergódica.

Sabemos já que μ é invariante (ver exemplo 5.13).

Vamos mostrar agora que tal μ é ergódica para $T(x) = x + \lambda$ (mod 1), quando λ é uma constante irracional.

Suponha que $A \in \mathcal{A}$ e $T^{-1}(A) = A$, então $I_A(x) = I_{T^{-1}(A)}(x) = I_A(T(x))$ para todo $x \in [0, 1)$.

Expresse $I_A(x)$ em série de Fourier

$$I_A(x) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n x}.$$

Como $I_A(x) = I_A(T(x))$ temos que

$$I_A(x) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n x} = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n (x+\lambda)} = I_A(T(x)).$$

Portanto

$$\sum_{n=-\infty}^{\infty} a_n e^{2\pi i n x} = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n \lambda} e^{2\pi i n x}.$$

Como os coeficientes de séries de Fourier são únicos, então $a_n e^{2\pi i n \lambda} = a_n$ para todo $n \in \mathbb{Z}$.

Se a_n não for nulo, temos que $1 = e^{2\pi i n \lambda}$.

Como λ é irracional, então $n\lambda$ não pode ser inteiro para qualquer n (a menos de $n = 0$). A conclusão é que $a_n = 0$ para $n \neq 0$. Portanto I_A é constante, Lebesgue quase toda parte, e ainda só assume os valores 0 e 1. Sendo assim I_A tem que ser a função constante igual a 0 ou a função constante igual a 1. Portanto $\mu(A) = \int I_A(x)dx = \int 0dx = 0$ ou $\mu(A) = \int I_A(x)dx = \int 1dx = 1$. Sendo assim concluimos que a probabilidade de Lebesgue é ergódica para a transformação $T(x) = x + \lambda \pmod 1$ quando λ é irracional.

Se $\lambda = 1/2$ é fácil ver que a medida de Lebesgue não é ergódica para a transformação $T(x) = x + \lambda \pmod 1$ (tome o conjunto $[0, 1/4] \cup [1/2, 3/4]$).

$\diamondsuit$

Exemplo 5.20. Seja X_n, $n \in \mathbb{N}$, um Processo Estocástico Independente e Identicamente Distribuído tal que a P associada sobre $\Omega = S^{\mathbb{N}}$ é estacionária. Neste caso, P é ergódica. De fato, se A é tal que $T^{-1}(A) = A$, então $T^{-2}(A) = T^{-1}(\,T^{-1}(A)\,) = T^{-1}(A) = A$ e assim, por indução $T^{-n}(A) = A$. Logo A está na σ-álgebra cauda e assim $P(A) = 0$ ou $P(A) = 1$.

$\diamondsuit$

Teorema 5.28 (Birkhoff). *Seja $(\Omega, \mathcal{A}, \mu)$ um espaço de probabilidade e $T : \Omega \to \Omega$ uma transformação mensurável que preserva μ, isto é, $\mu \in \mathcal{M}(T)$, e suponha ainda que μ é ergódica. Então para qualquer $f : \Omega \to \mathbb{R}$ integrável, ou seja $f \in \mathcal{L}^1(\mu)$, vale que*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(z)) = \int f(x) d\mu(x) \tag{5.1}$$

para $z \in \Omega$, μ-quase toda parte.

O resultado acima é um dos mais importantes teoremas da Matemática mas não apresentaremos a sua demonstração. Este resultado apresenta uma forma estendida da Lei Forte dos Grandes Números para sistemas que sejam ergódicos. Referimos o leitor a [PY] para uma prova. O conhecimento da demonstração deste resultado não é essencial para o que segue.

O conceito de ergodicidade e o Teorema de Birkhoff descrevem em termos matemáticos o análogo da assim chamada hipótese de Boltzmann da Termodinâmica (ver [Au])

Considere acima a função $f : \Omega \to \mathbb{R}$ tal que $f(x_0, x_1, x_2, ..., x_n, ...) = x_0$. Seja $X_n, n \in \mathbb{N}$, onde cada X_n toma valores em $S = \{1, 2, ..., d\}$, um processo estocástico estacionário ergódico e denote por $P = \mu$ probabilidade em $\Omega = S^{\mathbb{N}}$, obtida a partir do processo. Sabemos que $\int X_0 \, dP = \int X_n \, dP$, $\forall n \in \mathbb{N}$. Segue do teorema acima (usando tal f) que para P quase todo $w \in \Omega$ vale que

$$\lim_{n\to\infty} \frac{1}{n}(\, w_0 \, + w_1 \, + w_2 + ... + w_{n-1}\,) \, =$$

$$\lim_{n\to\infty} \frac{1}{n}(\, X_0(w) \, + X_1(w) \, + X_2(w) + ... + X_{n-1}(w)\,) \, = \int X_0 \, dP = \int w_0 \, dP.$$

Ou seja, para P quase toda amostra w a média temporal converge à média espacial.

No caso em que a probabilidade inicial (invariante) for $\pi = (\pi_1, \pi_2, .., \pi_d)$ temos que para P quase todo w

$$\lim_{n\to\infty} \frac{1}{n}(X_0(w) + X_1(w) + X_2(w) + ... + X_{n-1}(w)) = 1\,\pi_1 + 2\,\pi_2 + 3\,\pi_3 + .. + d\,\pi_d.$$

Utilizando agora $f : \Omega \to \mathbb{R}$ tal que $f(x_0, x_1, x_2, ..., x_n, ...) = x_0^2$ obtemos que para P quase todo $w \in \Omega$ vale que

$$\lim_{n\to\infty} \frac{1}{n}((X_0(w))^2 + (X_1(w))^2 + (X_2(w))^2 + ... + (X_{n-1}(w))^2) = \int X_0^2\, dP.$$

Exemplo 5.21. Como aplicação, considere T o shift σ, $P = \mu$ a medida de Bernoulli (independente e identicamente distribuída) sobre $X = \Omega = \{1, 2\}^{\mathbb{N}}$ obtida a partir de p_1 e p_2. Seja a função mensurável $f : \Omega \to \mathbb{R}$, $f(x) = I_{(X_0=2)}(x)$, ou seja $f(x) = 1$ se $x_0 = 2$, e $f(x) = 0$ se $x_0 = 1$, onde $x = (x_0, x_1, x_2, ...)$. Será mostrado em breve que $\mu = P$ é ergódico, assim seguirá do Teorema de Birkhoff que existe um conjunto K tal que $P(K) = 1$ e se $z \in K$, $z = (z_0, z_1, z_2, ...)$, temos que

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} I_{(X_0=2)}(T^j(z)) =$$

$$\lim_{n\to\infty} \frac{1}{n} (I_{(X_0=2)}((z_0, z_1, z_2, ...)) + I_{(X_0=2)}((z_1, z_2, z_3, ...)) + I_{(X_0=2)}((z_2, z_3, z_4, ...)) +$$
$$... + I_{(X_0=2)}((z_{n-1}, z_n, z_{n+1} ...))) =$$
$$\int I_{(X_0=2)}(x) d\,P(x) = P((X_0 = 2)) = p_2.$$

Note que, por exemplo, se $x = (1, 1, 2, 1, 2, 1, 1, 1, 2, 2, ...)$, então

$$\frac{1}{4} [I_{(X_0=2)}((1, 1, 2, 1, 2, 1, 1, 1, 2, 2, ...)) + I_{(X_0=2)}((1, 2, 1, 2, 1, 1, 1, 2, 2, ...)) +$$
$$I_{(X_0=2)}((2, 1, 2, 1, 1, 1, 2, 2, ...)) + I_{(X_0=2)}((1, 2, 1, 1, 1, 2, 2, ...)] = \frac{1}{4},$$

mede o número médio de aparecimento de 2 entre (x_0, x_1, x_2, x_3).

Se $p_1 = 1/2$ e $p_2 = 1/2$, então o processo estocástico acima considerado descreve lançamento de uma moeda (identificando 1 com coroa e 2 com cara). O Teorema de Birkhoff, neste caso, nos assegura algo que nos parece muito natural. Quando jogamos a moeda muitas vezes, na média, aparece cara aproximadamente em metade das vezes.

Voltando ao caso geral em que p_1, p_2 são quaisquer, se desejamos saber o valor médio de aparecimento da seqüência $2, 1$ (nesta ordem), basta tomar

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} I_{(X_0=2,X_1=1)}(T^j(z)) =$$

$$\lim_{n\to\infty} \frac{1}{n} \left(I_{(X_0=2,X_1=1)}((z_0, z_1, z_2, ...)) + I_{(X_0=2,X_1=1)}((z_1, z_2, z_3, ...)) + \right.$$
$$\left. I_{(X_0=2,X_1=1)}((z_2, z_3, z_4, ...)) + ... + I_{(X_0=2,X_1=1)}((z_{n-1}, z_n, z_{n+1}, ..)) \right) =$$
$$\int I_{(X_0=2,X_1=1)}(x) d\,P(x) = P(\,(X_0 = 2, X_1 = 1)\,) = p_2\, p_1,$$

(para P-quase todo z).

Mais geralmente, o Teorema de Birkhof permite calcular aproximadamente (basta tomar apenas **um** $\omega = z = (z_0, z_1, z_2, ...)$ escolhido P-quase toda parte) o valor da medida de qualquer cilindro

$$C = (X_0 = a_0, X_1 = a_1, ..., X_k = a_k),$$

onde $a_i \in \{1, 2\}, i \in \{0, 1, 2, ..., k\}$. Para isto, basta considerar a função mensurável $f = I_C$ e seguir o procedimento acima.

Podemos mais geralmente calcular o valor de uma integral por meio de um limite em n.

$\diamondsuit$

Como dissemos antes, a propriedade ser "ergódico" se aplica apenas a probabilidades invariantes.

Se considerarmos um processo independente, mas não identicamente distribuído, não podemos aplicar o Teorema de Birkhof acima pois P não é invariante.

Note que considerando Processos Estocásticos Independentes e Identicamente Distribuídos, sob certas condições, já sabíamos que a seqüência

$$\frac{1}{n}\,(X_0(\omega)+X_1(\omega)+...+X_{n-1}(\omega)),$$

convergia em P-quase toda parte.

Exemplo 5.22. O Teorema de Birkhof assegura muito mais. Suponha que $S=\{1,2\}$ e a P sobre $S^{\mathbb{N}}$ foi obtida como acima a partir de p_1 e p_2, gerando assim um Processo Independente Identicamente Distribuído. Considere $f: S^{\mathbb{N}} \to \mathbb{R}$ da seguinte forma

$$f(\omega)=\cos(w_0)\,w_1+(w_0+w_1)^2,$$

onde $\omega=(w_0,w_1,w_2,w_3,...,w_n,...)$. Tal f é integrável em relação a P. Considere, para cada $n=0,1,2,...$ uma nova variável aleatória

$$Y_n=\cos(X_n)\,X_{n+1}+(X_n+X_{n+1})^2.$$

O Teorema de Birkhoff assegura que existe para P-quase todo $\omega\in S^{\mathbb{N}}$

$$\lim_{n\to\infty}\frac{1}{n}\,(Y_0(\omega)+Y_1(\omega)+...+Y_{n-1}(\omega))=$$

$$\int f(\omega)\,dP(\omega)=$$

$$[\cos(1)\,1+(1+1)^2\,]\,p_1\,p_1+[\cos(1)\,2+(1+2)^2\,]\,p_1\,p_2+$$

$$[\cos(2)\,1+(2+1)^2\,]\,p_2\,p_1+[\cos(2)\,2+(2+2)^2\,]\,p_2\,p_2.$$

Poderíamos considerar até $f : S^{\mathbb{N}} \to \mathbb{R}$ de natureza mais complexa:

$$f(\omega) = \sum_{j=0}^{\infty} (\frac{1}{7})^j \, w_j,$$

onde $\omega = (w_0, w_1, w_2, ..., w_n, ...)$.

Neste caso, f dependeria de infinitas ordenadas w_i. Da mesma forma, como acima, seja a nova variável aleatória

$$Y_n = \sum_{j=0}^{n} (\frac{1}{7})^j \, X_j(\omega),$$

que depende de n.

O Teorema de Birkhoff assegura, também neste caso, que existe para P-quase todo $\omega \in S^{\mathbb{N}}$

$$\lim_{n\to\infty} \frac{1}{n} \, (Y_0(\omega) + Y_1(\omega) + ... + Y_{n-1}(\omega)).$$

$\diamondsuit$

No próximo Teorema vamos assumir que a medida é invariante mas não necessariamente ergódica.

Teorema 5.29 (Birkhoff). *Seja $(\Omega, \mathcal{A}, \mu)$ espaço de probabilidade $\mu \in \mathcal{M}(T)$, onde T é mensurável , $T : \Omega \to \Omega$. Então para qualquer $f \in \mathcal{L}^1(\mu)$ existe o limite*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(z))$$

para $z \in \Omega$, μ-quase toda parte. Se este limite é denotado por

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) = \tilde{f}(x),$$

então é também verdade que

$$\int \tilde{f}(x) d\mu(x) = \int f(x) d\mu(x).$$

A demonstração deste resultado pode ser encontrada em [PY].

Note que a diferença deste para o Teorema anterior é que no primeiro, quando a probabilidade era ergódica, o $\tilde{f}$ era constante μ-quase toda parte.

O resultado acima é uma versão mais fraca do anterior. Assumimos menos, mas a conclusão não é tão forte.

Este último resultado pode ser aplicado à probabilidade P de qualquer processo estacionário. No entanto, ele não é tão forte quanto o anterior. No presente caso, não basta tomar um evento $\omega = x = (x_1, x_2, x_3, ...)$ ao acaso segundo P para obter um valor aproximado da medida de um cilindro.

Se considerarmos um processo independente, mas não identicamente distribuído, não podemos aplicar o teorema acima pois P não é invariante.

Num problema concreto da vida real muitas vezes o que se tem é uma sequência temporal de dados. É natural supor que estes dados são caminhos amostrais truncados (finitos) provenientes de conjuntos com probabilidade 1. Suponhamos ainda que estes são provenientes de um processo estocástico ergódico. A interseção enumerável de conjuntos de probabilidade 1 tem probabilidade 1. Assim, se pode assumir que estes dados estão entre aqueles em que vale a conclusão do Teorema Ergódico. Desta forma, se a sequência de dados for de um período bem grande de tempo, então eles podem ser usados para calcular o valor aproximado de integrais de funções mensuráveis em relação a probabilidade associada.

Questões mais finas que envolvem o quão grande deve ser o tamanho da amostra e quão boa é esta aproximação da integral envolvem a Teoria dos Grandes Desvios [DZ].

Questões que envolvem a forma como estão dispersos estes dados em torno da integral envolvem o Teorema Central do Limite que foi abordado em um caso particular na seção 2.

Se o Processo Estocástico **não** é ergódico, então o procedimento acima não pode ser garantido. Note que a versão mais fraca do Teorema Ergódico

(Teorema 5.29) não é tão útil. Mesmo que tivéssemos várias sequências de dados distintas não se sabe a priori o valor relativo de cada sequência.

Em breve vamos apresentar uma condição suficiente para que um Processo Estocástico Markoviano seja ergódico (ver Teorema 5.32).

Vamos dar um exemplo agora de um processo estocástico estacionário Markoviano mas que não é ergódico.

Exemplo 5.23. Considere a matriz de transição $\mathcal{P}$ tal que

$$\mathcal{P} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix},$$

e $p = (1/2, 1/2) = (p_1, p_2)$. Temos a partir desta informação um processo de Markov estacionário definido por

$$P(\{X_0(\omega) = a_0, X_1(\omega) = a_1, ..., X_m(\omega) = a_m\}) =$$

$$p_{a_0}\, p_{a_0 a_1}\, p_{a_1 a_2} ... p_{a_{m-2} a_{m-1}}\, p_{a_{m-1} a_m},$$

sobre $S^{\mathbb{N}}$.

Note que o conjunto com um único elemento

$$D_1 = \{(1, 1, 1, 1, 1, 1, 1, ...)\},$$

tem probabilidade 1/2.

A mesma coisa vale para o conjunto

$$D_2 = \{(2, 2, 2, 2, 2, 2, ...)\},$$

que tem probabilidade 1/2.

Considere os conjuntos

$$A = \{x = (x_0, x_1, x_2, x_3, ...) \,|\, \text{tal que existe}\, k \geq 0, \text{ tal que}\, x_i = 1, \forall i \geq k\},$$

e

$$B = \{x = (x_0, x_1, x_2, x_3, ...) \,|\, \text{tal que existe}\, k \geq 0, \text{ tal que}\, x_i = 2, \forall i \geq k\}.$$

Note que cada um destes conjuntos tem probabilidade $1/2$, pois $D_1 \subset A$ e $D_2 \subset B$.

Note que

$$T^{-1}(A) = A,$$

mas $0 < 1/2 = P(A) < 1$, l go, o Processo Estocástico associado a P não é ergódico.

Mais geralmente, suponha que a matriz de transição $\mathcal{P}$, da forma $\#S$ por $\#S$, possa ser escrita como

$$\mathcal{P} = \begin{pmatrix} \mathcal{P}_1 & 0 \\ 0 & \mathcal{P}_2 \end{pmatrix},$$

onde $\mathcal{P}_i$ é da forma $\#S_i$ por $\#S_i$, $i \in \{1, 2\}$, e $S = S_1 \cup S_2$.

Seja um vetor de probabilidade inicial estacionário π sobre S (ou seja, assuma que $\pi\,\mathcal{P} = \pi$), tal que tenha todas as componentes não nulas.

Então, definindo

$$A = \{x = (x_0, x_1, x_2, x_3, ...) \,|\, \text{tal que existe}\, k \geq 0, \text{ tal que}\, x_i \in S_1, \forall i \geq k\},$$

obtemos da mesma forma que $T^{-1}(A) = A$.

Resta mostrar que A tem medida positiva e diferente de 1.

O conjunto $S \times S_1^{\mathbb{N}}$ está contido em A.

Ora, $S \times S_1^{\mathbb{N}} = \cap_{k=1}^{\infty} S \times S_1^k \times S^{\mathbb{N}}$.

Para k fixo, como $P_{c,d}^k = 0$ se $c \in S_1$ e $d \in S_2$, então

$$P(S \times S_1^k \times S^{\mathbb{N}}) = \sum_{i \in S} \sum_{j_1 \in S_1} \sum_{j_2 \in S_1} \pi_i P_{i,j_1}\, P_{j_1,j_2}^k =$$

$$\sum_{j_1\in S_1}\sum_{j_2\in S_1}\pi_{j_1}\,P^k_{j_1,j_2}\;=\sum_{j\in S}\sum_{j_2\in S_1}\pi_j\,P^k_{j,j_2}=\sum_{j_2\in S_1}\pi_{j_2}.$$

Sendo assim, independente de k vale $P(S\times S_1^k\times S^{\mathbb{N}})\;=\;\sum_{j_2\in S_1}\pi_{j_2}$. Como a sequência $S\times S_1^k\times S^{\mathbb{N}}$ é decrescente,

$$P(A)\geq P(S\times S_1^{\mathbb{N}})\;=\sum_{j_2\in S_1}\pi_{j_2}>0.$$

Podemos fazer o mesmo para o conjunto B. Desta forma se mostra que $P(A)<1$

Logo, tal probabilidade P não é ergódica para o shift.

Neste caso **não** podemos usar o Teorema 5.28. Se apenas vale $\pi\,\mathcal{P}=\pi$, o processo é estacionário, e assim, podemos usar o Teorema 5.29.

Uma pequena alteração da demonstração acima permite tratar o caso de matrizes que não são irredutíves com $\#S$ finito.

$\diamondsuit$

Definição 5.45. *Dizemos que a probabilidade $P\in\mathcal{M}(T)$, onde $T:(X,\mathcal{A})\to(X,\mathcal{A})$, é mixing, se para quaisquer conjuntos $A,B\in\mathcal{A}$, vale que*

$$\lim_{n\to\infty}P(T^{-n}(A)\,\cap\,B)=P(A)\,P(B).$$

Um processo Estocástico estacionário X_n, $n\in\mathbb{N}$, é mixing se sua correspondente P o é para o shift.

Uma outra caracterização equivalente seria

$$\lim_{n\to\infty}\int I_A(T^n(\omega))\;I_B(\omega)dP(\omega)=\int I_A(\omega)\,dP(\omega)\;\times\;\int I_B\,(\omega)\,dP(\omega).$$

Isto porque $I_{T^{-n}(A)}(\omega)=I_A(T^n(\omega))$ e

$$\int I_A(T^n(\omega))\;I_B(\omega)dP(\omega)=\int_{T^{-n}(A)\cap(B)}1\;dP(\omega),$$

onde I_A é a função indicador do conjunto A.

Por exemplo, se $X = S^{\mathbb{N}}$, T é o shift e P advém de um processo estocástico X_n, $n \in \mathbb{N}$, temos que se $A = (X_1 = a_1, X_2 = a_2, X_3 = a_3)$ e $B = (X_2 = b_2, X_5 = b_5)$ então

$$\lim_{n\to\infty} P(\,(X_2 = b_2, X_5 = b_5) \cap (X_{n+1} = a_1, X_{n+2} = a_2, X_{n+3} = a_3)\,) =$$

$$\lim_{n\to\infty} P(B \cap T^{-n}(A)\,) = P(B) \times P(A) =$$

$$P(\,(X_2 = b_2, X_5 = b_5)\,) \times P(\,(X_1 = a_1, X_2 = a_2, X_3 = a_3)\,).$$

Observe que, se os eventos B e $T^{-n}(A)$ acima fossem independentes, então

$$P(\,(X_2 = b_2, X_5 = b_5) \cap (X_{n+1} = a_1, X_{n+2} = a_2, X_{n+3} = a_3)\,) =$$

$$P(\,(X_2 = b_2, X_5 = b_5)\,) \times P(\,(X_{n+1} = a_1, X_{n+2} = a_2, X_{n+3} = a_3)\,) =$$

$$P(\,(X_2 = b_2, X_5 = b_5)\,) \times P(\,(X_1 = a_1, X_2 = a_2, X_3 = a_3)\,)$$

Desta forma os eventos $T^{-n}(A)$ e B não estariam correlacionados. Num processo mixing, não necessariamente vale o que foi afirmado acima. O que ocorre é que quando n vai para infinito, a correlação entre os eventos $T^{-n}(A)$ e B vai a zero, para quaisquer A e B.

Em outra palavras, se o Processo estocástico é mixing, então começa a haver perda de memória do que ocorre a medida que n vai a infinito, ou seja, no caso do exemplo acima, o evento $(X_2 = b_2, X_5 = b_5)$ tem cada vez menos influência no evento $(X_{n+1} = a_1, X_{n+2} = a_2, X_{n+3} = a_3)$ a medida que n cresce. Ou seja, o evento $(X_2 = b_2, X_5 = b_5)$ tem cada vez menos influência no aparecimento de uma seqüência tipo (a_1, a_2, a_3) a medida que n cresce.

A probabilidade induzida por certo Processo Markoviano em $\{1, 2, 3\}^{\mathbb{N}}$ (ver exemplo 2.35) é

$$\frac{1}{3}\delta_{z_1} + \frac{1}{3}\delta_{z_2} + \frac{1}{3}\delta_{z_3},$$

onde

$$z_1 = (1,2,3,1,2,3,1,2,3,..),$$

$$z_2 = (2,3,1,2,3,1,2,3,1,..)$$

e

$$z_3 = (3,1,2,3,1,2,3,1,2,..).$$

Esta probabilidade não é mixing embora seja ergódica.

Teorema 5.30. *São equivalentes:*

a) P é mixing, e

b) para quaisquer $f, g \in \mathcal{L}^2(X, P, \mathcal{A})$, vale que

$$\lim_{n\to\infty} \int f(T^n(\omega))\; g(\omega) dP(\omega) = \int f(\omega)\, dP(\omega) \,\times\, \int g\,(\omega)\, dP(\omega).$$

Demonstração: Ora b) implica a) trivialmente, basta tomar $f = I_A$ e $g = I_B$.

Suponha que vale a), então f pode ser aproximado por uma seqüência monótona

$$f = \lim_{i\to\infty} h_i,$$

onde $h_i = \sum_{j=1}^{n_i} a_i^j\, I_{A_i^j}$, é uma função simples. Da mesma forma,

$$g = \lim_{i\to\infty} z_i,$$

onde $z_i = \sum_{j=1}^{m_i} b_i^j\, I_{B_i^j}$, é uma função simples.

Ainda,

$$(f \circ T^n) = \lim_{i\to\infty} (h_i \circ T^n).$$

Ora, a aproximação de $(f \circ T^n)\, g$ por funções simples poderia ser feito pelos produtos

$$(f \circ T^n)\, g = \lim_{i\to\infty} (h_i \circ T^n)\, z_i.$$

O produto de funções simples do lado direito da igualdade acima envolveria soma de produtos tipo

$$a_i^k\, b_i^j\, (I_{A_i^k} \circ T^n)\, I_{B_i^j}.$$

Como vale a) para conjuntos A_i^k e B_i^j, então b) pode ser obtido através de procedimentos clássicos em Teoria da Medida em que se aproxima valores através de limites.

$\square$

É de grande importância em Probabilidade a análise da velocidade que vai à zero a diferença

$$c_n = \int f(T^n(\omega))\; g(\omega) dP(\omega) - (\int f(\omega)\, dP(\omega) \;\times\; \int g\,(\omega)\, dP(\omega)\,).$$

Definição 5.46. *Seja X_n, $n \in \mathbb{N}$, Processo Estocástico Mixing. Dizemos que f e g têm decaimento exponencial de correlação se existe $C > 0$ e $\lambda > 0$ tal que $c_n \leq C\, e^{-\lambda\, n}$.*

Definição 5.47. *Seja X_n, $n \in \mathbb{N}$, Processo Estocástico Mixing. Dizemos que f e g têm decaimento hiperbólico (ou, polinomial) de correlação se existe $C > 0$ e $\lambda > 0$ tal que $c_n \leq C\, n^{\lambda}$.*

No primeiro caso esta velocidade é muito mais rápida. Sob condições bastante gerais pode-se mostrar versões do Teorema Central do Limite para Processos Estocásticos Mixing com decaimento exponencial de correlação (ver [B] fim da seção 35 para um resultado muito geral sobre o assunto).

Teorema 5.31. *Se P é mixing para T então P é ergódica para T.*

Demonstração: Seja um conjunto mensurável A tal que $T^{-1}(A) = A$. Considere A e $B = A$ na definição de mixing, logo como por hipótese $T^{-n}(A) = A$ (segue por indução), temos que

$$P(A) = \lim_{n \to \infty} P(A \cap A)$$

$$\lim_{n\to\infty} P(T^{-n}(A) \cap A) = P(A)\, P(A) = P(A)^2.$$

Os únicos números reais tais que são iguais ao seu quadrado são zero e um. Logo, $P(A) = 0$ ou $P(A) = 1$.

□

Vamos agora analisar a ergodicidade de algumas cadeias de Markov.

Teorema 5.32. *Seja $\mathcal{P}$ matriz estocástica n por n, e p vetor de probabilidade tal que $p\mathcal{P} = p$. Considere o Processo Markoviano Estacionário X_n, $n \in \mathbb{N}$, associado, P a sua probabilidade sobre $S^{\mathbb{N}}$ e $\mathcal{A} = \sigma(\mathcal{C})$ a σ-álgebra gerada pelos cilindros, onde $\#S = n$. Se $\mathcal{P}$ for recorrente, aperiódica e irredutível, então P é ergódica para o shift T.*

Demonstração: Primeiro vamos mostrar que vale

$$\lim_{n\to\infty} P(T^{-n}(A) \cap B) = P(A)\, P(B),$$

quando A e B são cilindros da forma

$$A = (X_1 = j_1, X_2 = j_2, ..., X_s = j_s),$$

$$B = (X_1 = i_1, X_2 = i_2, ..., X_r = i_r).$$

Ora,

$$B \cap T^{-(m+r)}(A) = (X_1 = i_1, X_2 = i_2, ..., X_r = i_r,$$

$$X_{m+r+1} = j_1, X_{m+r+2} = j_2, ..., X_{m+r+s} = j_s) =$$

$$\cup_{x_{r+1},x_{r+2},...,x_{m+r}} (X_1 = i_1, X_2 = i_2, ..., X_r = i_r,$$

$$X_{r+1} = x_{r+1}, X_{r+2} = x_{r+2}, ..., X_{m+r} = x_{m+r},$$

$$X_{m+r+1} = j_1, X_{m+r+2} = j_2, ..., X_{m+r+s} = j_s).$$

Portanto,

$$P(B\cap T^{-(m+r)}(A)) = \sum_{x_{r+1},x_{r+2},...,x_{m+r}} p_{i_1}\, P(i_1, i_2) P(i_2, i_3) ... P(i_{r-1}, i_r)\, P(i_r, x_{r+1}) \times$$

$$P(x_{r+1}, x_{r+2})\, P(x_{r+2}, x_{r+3}) ... \, P(x_{m+r}, j_1) \times$$
$$P(j_1, j_2)\, P(j_2, j_3) ... \, P(j_{s-1}, j_s) =$$
$$P(B)\, P(A)\, \frac{1}{p_{j_1}} \sum_{x_{r+1}, x_{r+2}, ..., x_{m+r}} P(i_r, x_{r+1})\, P(x_{r+1}, x_{r+2}) ... \, P(x_{m+r}, j_1) =$$
$$P(B)\, P(A)\, \frac{(\mathcal{P}^m)_{i_r\, j_1}}{p_{j_1}}.$$

Sendo assim, para i_1 e j_1 fixos, como $\mathcal{P}$ é recorrente, aperiódica e irredutível, então vale que

$$\lim_{m\to\infty} P(T^{-(m+r)}(A) \cap B) =$$
$$\lim_{m\to\infty} P(B)\, P(A)\, \frac{(\mathcal{P}^m)_{i_r\, j_1}}{p_{j_1}} =$$
$$P(B)\, P(A) \lim_{m\to\infty} \frac{(\mathcal{P}^m)_{i_r\, j_1}}{p_{j_1}} =$$
$$P(B)\, P(A) \lim_{m\to\infty} \frac{p_{j_1}}{p_{j_1}} = P(B)\, P(A).$$

Como r está fixo, fazer limite em $n \to \infty$ é a mesma coisa que fazer limite de $n = m + r$ quando $m \to \infty$.

Logo, o resultado que afirmamos acima (tipo mixing) é verdadeiro quando A e B são cilindros da forma acima. Considerando agora somas finitas destes tipos de cilindros obtemos que a afirmação é também verdadeira para qualquer cilindro em $\mathcal{C}$.

Em vez de mostrar que P é mixing, para simplificar a argumentação, vamos mostrar que P é ergódico apenas.

Seja um conjunto mensurável E em $\sigma(\mathcal{C})$ tal que $T^{-1}(E) = E$.

Sabemos pelo Teorema 5.8 que existe uma seqüência de conjuntos A_i, $i \in \mathbb{N}$, em $\mathcal{C}$ tal que $\lim_{i\to\infty} P\,(E \Delta A_i) = 0$.

Primeiro, vamos analisar a questão para um conjunto da forma A_i fixo. Depois vamos elaborar o caso de um conjunto E qualquer via aproximação. Isto será feito a seguir.

Lembre que

$$P(B) = P(B - A) + P(A \cap B)$$

e que

$$T^{-1}(A \Delta B) = T^{-1}(A) \, \Delta \, \sigma^{-1}(B).$$

Ora, $E \Delta A_i = (A_i - E) \cup (E - A_i)$, logo, $\lim_{i\to\infty} P\,(E - A_i) = 0$ e $\lim_{i\to\infty} P\,(A_i - E) = 0$.

Suponha que $P(E) - P(A_i) \geq 0$.

Portanto,

$$| \, P(E) - P(A_i) \, | =$$

$$= P(E \cap A_i) + P(E - A_i) - P(A_i \cap E) - P(A_i - E) =$$

$$P(E - A_i) - P(A_i - E).$$

Estes dois termos vão a zero com $n \to \infty$.

O caso $P(E) - P(A_i) \leq 0$ se trata de forma análoga.

Sendo assim,

$$\lim_{i\to\infty} \, | \, P(E) - P(A_i) \, | = \, 0.$$

Seja m e i fixos, e considere $T^{-m}(A_i)$. Sabemos então que para i fixo

$$\lim_{m\to\infty} P(T^{-m}(A_i) \cap A_i) \, = \, P(A_i)^2,$$

pois $A_i \in \mathcal{C}$.

Note que $T^{-m}(A \, \Delta \, B) = T^{-m}(A) \, \Delta \, T^{-m}(B)$.

Como

$$P(E \, \Delta \, T^{-m}(A_i)) \, = P(T^{-m}(E) \, \Delta \, T^{-m}(A_i)) = P(E \, \Delta \, A_i),$$

e

$$E \, \Delta \, (A_i \, \cap \, T^{-m}(A_i)) \, \subset (E \, \Delta \, A_i) \, \cup \, (E \, \Delta \, T^{-m}(A_i)),$$

temos que

$$P(\, E\,\Delta\,(A_i \cap T^{-m}(A_i)\,)\, \leq\, 2\, P(\, E\,\Delta\, A_i\,).$$

Desejamos mostrar que $P(E) = P(E)^2$.

Lembre que

$$|P(C) - P(D)|\, \leq P(C-D) + P(D-C).$$

Da mesma maneira como mostramos antes, obtém-se que

$$\left|P(E)\, -\, P(A_i\, \cap\, T^{-m}(A_i))\right| \leq$$

$$P(E - (A_i \cap T^{-m}(A_i)) + P((A_i \cap T^{-m}(A_i) - E) \leq\, 4\, P(\, E\,\Delta\, A_i\,).$$

Finalmente, para m e i fixos, vale que

$$\left|P(E) - P(E)^2\right| \leq$$

$$\left|P(E) - P(A_i\, \cap\, T^{-m}(A_i))\right| + \left|P(A_i\, \cap\, T^{-m}(A_i)) - P(E)^2\right| \leq$$

$$4\, P(\, E\,\Delta\, A_i\,)\, +\, \left|P(A_i\, \cap\, T^{-m}(A_i)) - P(E)^2\right| \leq$$

$$4\, P(\, E\,\Delta\, A_i\,)\, +\, \left|P(A_i\, \cap\, T^{-m}(A_i)) - P(A_i)^2\right| + \left|P(A_i)^2 - P(E)^2\right|.$$

Pelo que assumimos acima, as três parcelas vão a zero quando $i \to \infty$. Desta forma, $P(E) = P(E)^2$, e assim só resta a alternativa que $P(E) = 0$ ou $P(E) = 1$.

Logo, P é ergódica.

$\square$

Note que uma condição necessária (mas não suficiente) para P ser ergódica é que P seja estacionária. Assim, esta última propriedade, no caso de um processo estocástico Markoviano, requer que consideremos a condição inicial π tal que $\pi\mathcal{P} = \pi$.

Segue do resultado acima que se $\mathcal{P}$ for recorrente aperiódica e irredutível e considerarmos o Processo Markoviano **Estacionário** X_n associado, com probabilidade inicial π, então dadas $f : S \to \mathbb{R}$ e $g : S \to \mathbb{R}$ vale

$$\lim_{n\to\infty} \int f(X_n(\omega))\ g(X_n(\omega))dP(\omega) = \int f(s)\,d\pi(s)\ \times \int g\,(s)\,d\pi(s).$$

Ainda, dadas $f : S \times S \to \mathbb{R}$ e $g : S \to \mathbb{R}$ vale

$$\lim_{n\to\infty} \int f(X_n(\omega))\ g(X_n(\omega))dP(\omega) = \int f(w)\,dP(w)\ \times \int g\,(s)\,d\pi(s) =$$

$$\int f(w_0, w_1)\,dP(w)\ \times \int g\,(s)\,d\pi(s).$$

Note que segue do que foi mostrado acima que o processo de jogar uma moeda de forma independente com probabilidade $1 > p_1 > 0$ e $p_2 = 1 - p_1$ é ergódico. Isto é consequencia do fato que ele pode ser considerado markoviano recorrente aperiódica e irredutível com a matriz de transição

$$\mathcal{P} = \begin{pmatrix} p_1 & p_2 \\ p_1 & p_2 \end{pmatrix}.$$

Considere $S = \{1, 2\}$, uma matriz estocástica $\mathcal{P}$ do tipo 2 por 2, com todas as entradas positivas, e um vetor de probabilidade inicial $\pi \in \mathbb{R}^2$ invariante para $\mathcal{P}$. Neste caso, pelo resultado acima, podemos usar o Teorema 5.28. Se estivermos interessados em calcular a frequência de aparecimento do bloco $1, 2, 2$ numa sequencia w escolhida ao acaso, de acordo com a P associada, devemos considerar a função $I_{\overline{1,1,2}}$.

Segue do fato do sistema em consideração ser ergódico que existe um conjunto K, tal que $P(K) = 1$, e para todo $w = (w_0, w_1, w_2, w_3, ...) \in K$, vale

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} I_{\overline{1,1,2}}(w_j, w_{j+1}, w_{j+2}, w_{j+3}, ...) = \int I_{\overline{1,1,2}}\,dP = P(\overline{1,1,2}) = \pi_1 \mathcal{P}_{1,2} \mathcal{P}_{2,2}.$$

5.5 Esperança e Probabilidade Condicional

O leitor poderá encontrar em [B] uma versão mais completa do material apresentado aqui.

Seja $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{R})$ integrável e $\mathcal{G}$ σ-álgebra contida em $\mathcal{A}$. Desta forma podemos calcular $\int_G X dP = \int I_G X dP$, para G em $\mathcal{G}$. Note no entanto que X pode não ser $\mathcal{G}$-mensurável.

Definição 5.48. *Seja $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{R})$ que é P-integrável e $\mathcal{G}$ uma σ-álgebra contida em $\mathcal{A}$. Diremos que $Y : (\Omega, \mathcal{G}, P) \to (\mathbb{R}, \mathcal{R})$ (assim $\mathcal{G}$-mensurável) é a Esperança Condicional (ou, a função Valor Esperado de X condicionado a $\mathcal{G}$) de X dada a σ-álgebra $\mathcal{G}$, se para qualquer conjunto $G \in \mathcal{G}$, vale*

$$\int_G X(\omega) d\, P(\omega) = \int_G Y(\omega) d\, P(\omega).$$

Note que Y é uma função e não um número.

De outra forma, Y é uma versão de X que é $\mathcal{G}$-mensurável, e que em termos de cálculo de integral sobre qualquer conjunto G de $\mathcal{G}$, não tem diferença do que acontece para X.

Primeiro, note que tal Y sempre existe. De fato, considere primeiro o caso em que X é não-negativa. Seja μ, medida sobre $\mathcal{G}$ tal que

$$\mu(G) = \int_G X\, dP.$$

Então, μ é absolutamente contínua em relação a P, quando consideramos a σ-álgebra $\mathcal{G}$. Ou seja, estamos aqui restringindo a ação de P a $\mathcal{G}$. Ora, pelo Teorema de Radon-Nykodin sobre a σ-álgebra $\mathcal{G}$, temos que existe uma função Y, $\mathcal{G}$-mensurável, tal que $\mu(G) = \int_G Y\, dP$, para todo $G \in \mathcal{G}$.

Logo,

$$\int_G Y\, d\, P = \mu(G) = \int_G X\, dP.$$

A função Y é única, a menos do que acontece em um conjunto de medida P nula.

Se X não é positiva, podemos escrever X como $X = X^+ - X^-$, onde X^+ e X^- são não negativas, e a seguir aplicar o procedimento anterior a cada uma das duas. O resultado segue trivialmente.

Usaremos a notação $Y = E[X \,|\, \mathcal{G}]$.

Ainda, afirmamos que para qualquer função g que seja $\mathcal{G}$ mensurável vale

$$\int g\, Y\, dP = \int g\, X\, dP.$$

Isto segue do procedimento usual de aproximar g por soma de funções simples usando indicadores de conjuntos $\mathcal{G}$ mensuráveis. Note que $\int_G Y\, d\,P = \int_G X\, dP$ significa $\int I_G\, Y dP = \int I_G X\, dP$.

É equivalente dizer que $Y = E[X \,|\, \mathcal{G}]$ e ser válido: para qualquer função g que seja $\mathcal{G}$ mensurável temos

$$\int g\, Y\, dP = \int g\, X\, dP.$$

A partir das considerações que fizemos antes da Definição 5.40 se consideramos a sigma algebra $\mathcal{A} = \sigma(X_3, X_4, ..., X_n, ...)$, então $Y = E[X \,|\, \mathcal{A}]$ é equivalente a: para qualquer função g que seja mensurável em relação a sigma algebra $\sigma(X_1, X_2, X_3, X_4, ..X_n, ...)$ vale

$$\int (g \circ \sigma^2)\, Y\, dP = \int (g \circ \sigma^2)\, X\, dP.$$

Exemplo 5.24. Considere sobre o intervalo $[0, 1]$, a σ-álgebra de Borel $\mathcal{R}$ e P a probabilidade de Lebesgue. Seja $X = I_B$ onde $B = [0.4, 0.8]$ e $\mathcal{G}$ a σ-álgebra gerada pelos conjuntos A_1, A_2, A_3 onde $A_1 = [0, 0.2), A_2 = [0.2, 0.5), A_3 = [0.5, 1]$. A σ-álgebra $\mathcal{G}$ tem um número finito de elementos.

Ora, para $Y = E[X \mid \mathcal{G}]$ ser $\mathcal{G}$ mensurável, ela tem que ser constante em cada um dos conjuntos A_i, $i \in \{1,2,3\}$. A demonstração deste fato é similar a apresentada na observação feita antes do Teorema 5.3.

Vamos denotar por a_i, $i \in \{1,2,3\}$ o valor $Y(x)$, para $x \in A_i$.

Da relação

$$\int_{A_i} X dP = \int_{A_i} Y dP,$$

obtemos que

$$P(A_i \cap B) = \int I_{A_i}\, I_B\, dP = \int I_{A_i}\, X\, dP = P(A_i)\, a_i,$$

$i \in \{1,2,3\}$.

Resulta então que

$$a_i = \frac{P(A_i \cap B)}{P(A_i)} = P(B \mid A_i),$$

$i \in \{1,2,3\}$.

Ou seja, $a_1 = 0$, $a_2 = \frac{0.1}{0.3}$, $a_3 = \frac{0.3}{0.5}$.

Finalmente,

$$Y = a_1\, I_{A_1} + a_2\, I_{A_2} + a_3\, I_{A_3}.$$

Considerando o ponto de vista que uma σ-álgebra é uma informação de que dispomos, podemos pensar, usando a notação da definição acima de valor esperado, que existe certa função X que é $\mathcal{A}$-mensurável (muito rica em termos de informação, mas desconhecida para nós). Suponhamos que nossa informação se restringe a $\mathcal{G}$. Então $Y = E[X \mid \mathcal{G}]$ é o que podemos saber de tal função X através do conhecimento de $\mathcal{G}$.

$\diamondsuit$

Um outro exemplo que é uma variação do acima é o seguinte: considere sobre o intervalo $[0,1]$, a σ-álgebra de Borel $\mathcal{R}$ e P a probabilidade de Lebesgue.

Seja $\mathcal{G}$ a σ-álgebra gerada pelos conjuntos A_1, A_2, A_3 onde $A_1 = [0, 0.2), A_2 = [0.2, 0.5), A_3 = [0.5, 1]$. Seja $X : [0,1] \to \mathbb{R}$ uma função contínua. A função valor esperado $Y = E[X \,|\, \mathcal{G}]$ precisa ser constante sobre cada conjunto A_1, A_2, A_3. É fácil ver que o valor de Y sobre A_j, $j = 1, 2, 3$, é $\frac{\int_{A_j} X\,dx}{\text{comprimento de } A_j}$, isto é, o valor médio de X em A_j.

No caso de uma certa função $f : [0,1] \to \mathbb{R}$ exibimos um exemplo (ver figura 5.3) de função esperança condicional segundo a partição (sigma-algebra) dos intervalos $[0, 1/4], (1/4, 1/2], (1/2, 3/4], (3/4, 1]$. Neste caso a função esperança condicional é constante em intervalos e definida pelos valores $a_{i,j}$.

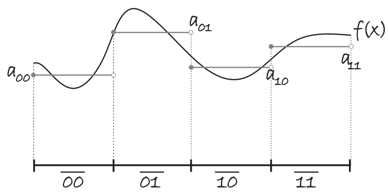


Figura 5.3: Uma função contínua $f : [0.1] \to \mathbb{R}$ e a esperança condicional em relação a certa partição.

Exemplo 5.25. Seja $\Omega = \{1,2\}^{\mathbb{N}}$ e P obtida de um Processo Estocástico X_n, $n \in \mathbb{N}$, com valores em $\{1, 2\}$. Seja $\mathcal{A}$, a σ-álgebra gerada pelos cilindros. Tipicamente, podemos pensar que estamos jogando uma moeda e que cara significa 1 e coroa significa 2. Não necessariamente estamos considerando um Processo Estocástico Independente. Suponhamos $X = I_B$, onde $B \in \mathcal{A}$ poderia depender de infinitas coordenadas, ou seja B não seria um cilindro. Não sabemos quem é B. Quando consideramos I_B, é porque por algum mecanismo, vai ocorrer B e nada mais.

Quando jogamos a moeda duas vezes, nosso conhecimento do jogo é a σ-álgebra $\mathcal{G}$ gerada pelos cilindros

$$A_1 = \{\omega \,|\, X_1 = 1, X_2 = 1\},\ A_2 = \{\omega \,|\, X_1 = 1, X_2 = 2\},$$

e

$$A_3 = \{\omega \,|\, X_1 = 2, X_2 = 1\},\ A_4 = \{\omega \,|\, X_1 = 2, X_2 = 2\}.$$

Uma função F é mensurável em relação a sigma álgebra gerada por estes conjuntos, se e só se, é constante nestes conjuntos. A demonstração deste fato é similar a apresentada na observação feita antes do Teorema 5.3.

Da mesma forma como no exemplo anterior, temos que se

$$c_i = \frac{P(B \cap A_i)}{P(A_i)},$$

$i \in \{1, 2, 3, 4\}$, então

$$Y = E[X \,|\, \mathcal{G}] = \sum_{i=1}^{4} c_i I_{A_i}.$$

A função Y acima traduz o conhecimento de X após duas jogadas da moeda. Note que a medida que jogamos a moeda três ou mais vezes, vamos tendo maior informação através da correspondente $\mathcal{G}$, e assim obtendo uma Y mais e mais rica em termos de informação.

Muitas vezes, a informação da σ-álgebra $\mathcal{G}$ provém de uma seqüência de funções mensuráveis $X_1, X_2, X_3, ..., X_n$, onde $X_i : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{R})$, ou seja, $\mathcal{G} = \sigma(X_1, X_2, ..., X_n)$.

Por exemplo se $S = \{1, 2\}$, $\Omega = \{1, 2\}^{\mathbb{N}}$, os cilindros de ordem 3 são

$$\overline{1,1,1},\ \overline{1,1,2},\ \overline{1,2,1},\ \overline{1,2,2},\ \overline{2,1,1},\ \overline{2,1,2},\ \overline{2,2,1},\ \overline{2,2,2}.$$

Por exemplo, temos que $\{X_1 = a_1, X_2 = a_2, X_3 = a_3\} = \overline{a_1, a_2, a_3}$.

A sigma algebra $\mathcal{F}_3$ é a gerada por tais cilindros. Ela descreve a "informação" que obtemos a partir de X_1, X_2, X_3. Ou seja, se 1 e 2 estão associados a cara e coroa, a sigma-algebra $\mathcal{F}_3$ é a informação do jogo definido por lançar a moeda tres vezes. Neste caso, $\mathcal{F}_3 = \sigma(X_1, X_2, X_3)$

Generalizando o teorema 5.15, pode se mostrar que se Y é $\mathcal{G}$-mensurável, então existe $h(x_1, x_2, ..., x_n)$ mensurável, $h : (\mathbb{R}^n, \mathcal{R}^n) \to (\mathbb{R}, \mathcal{R})$, tal que,

$$Y(\omega) = h(X_1(\omega), X_2(\omega), ..., X_n(\omega)).$$

Ou seja, $Y = E[X \,|\, \mathcal{G}]$, é tal que $Y : (\Omega, \mathcal{D}) \to (\mathbb{R}, \mathcal{R})$, é uma função dos valores $X_1(\omega), X_2(\omega), ..., X_n(\omega)$.

Aproveitando o exemplo acima, em que $S = \{1, 2\}$, e $\mathcal{G} = \sigma(X_1, X_2)$, a $h(x_1, x_2)$, neste caso, pode ser qualquer função tal que

$$h(1,1) = c_1,\ h(1,2) = c_2,\ h(2,1) = c_3,\ h(2,2) = c_4.$$

Alguns textos em probabilidade chamam a função

$$h(x_1, x_2, ..., x_n),$$

acima obtida de valor esperado, ou seja de

$$E(X \,|\, \sigma(X_1, X_2, ..., X_n)) = E(X \,|\, X_1, X_2, ..., X_n),$$

e não a função

$$Y = h(X_1, X_2, ..., X_n).$$

Não poderá haver confusão nisso, pois mantém-se sempre a notação de se utilizar letras minúsculas para denotar pontos $x_i \in \mathbb{R}$ e letras maiúsculas para as funções X_i.

Se as variáveis aleatórias X_i acima, $i \in \{1, 2, 3, ..., n\}$, tomassem valores em S, então h poderia ser tomada como uma função de

$$h : S^n \to \mathbb{R}.$$

$\diamondsuit$

Note que se X for $\mathcal{G}$-mensurável então $X = E(X \mid \mathcal{G})$.

Ainda, dada a σ-álgebra $\mathcal{G}$ tal que $\mathcal{G} = \{\emptyset, \Omega\}$, então, as únicas funções $\mathcal{G}$-mensuráveis são as funções constantes.

Então, neste caso, se $Y = E(X \mid \mathcal{G})$ e $Y = c$, onde c é constante, temos que $c = \int Y\,dP = \int X dP$. Desta forma $E[X] = E[X \mid \{\emptyset, \Omega\}\,]$.

Teorema 5.33. *Seja $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{R})$, mensurável e $\mathcal{G}_1$, $\mathcal{G}_2$ σ-álgebras contidas em $\mathcal{A}$ tais que $\mathcal{G}_1$ está contida em $\mathcal{G}_2$.*

Então,

$$E[\,(\,E[\,X \mid \mathcal{G}_2\,]\,) \mid \mathcal{G}_1\,] = E[\,E[\,X \mid \mathcal{G}_2\,] \mid \mathcal{G}_1\,] = E[\,X \mid \mathcal{G}_1].$$

Demonstração: Denote por $Y = E[\,X \mid \mathcal{G}_2\,]$, $Z = E[\,X \mid \mathcal{G}_1\,]$.

Desejamos mostrar que $E[\,Y \mid \mathcal{G}_1\,] = Z$.

Basta mostrar que para todo $G \in \mathcal{G}_1$, vale que

$$\int_G Y\,dP = \int_G Z\,dP.$$

Ora, todo G em $\mathcal{G}_1$ está em $\mathcal{G}_2$, logo como $Y = E[\,X \mid \mathcal{G}_2\,]$, então,

$$\int_G Y\,dP = \int_G X\,dP.$$

Por outro lado, como $Z = E[\,X \mid \mathcal{G}_1\,]$, então,

$$\int_G Z\,dP = \int_G X\,dP.$$

Logo, o resultado está demonstrado.

□

Teorema 5.34. *Seja $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{R})$, mensurável e $\mathcal{G}_1$, $\mathcal{G}_2$ σ-álgebras contidas em $\mathcal{A}$ tais que $\mathcal{G}_1$ está contida em $\mathcal{G}_2$.*

Então,

$$E[\,E[\,X \mid \mathcal{G}_1\,] \mid \mathcal{G}_2\,] = E[\,X \mid \mathcal{G}_1].$$

Demonstração: Denote por $Y = E[\,X\,|\,\mathcal{G}_1\,]$, então Y é $\mathcal{G}_1$-mensurável, e assim, $\mathcal{G}_2$-mensurável. Logo $Y = E[\,Y\,|\,\mathcal{G}_2\,]$.

$\square$

Os dois resultados acima indicam o seguinte, se $\mathcal{G}_1 \subset \mathcal{G}_2$, então $\mathcal{G}_1$ tem menos informação que $\mathcal{G}_2$. Sendo assim, independente da ordem em que se condiciona em $\mathcal{G}_1$ e $\mathcal{G}_2$, o que resulta é uma função com menos informação.

Note que se $Y = E[\,X\,|\,\mathcal{G}\,]$, então, como $\Omega \in \mathcal{G}$, temos que

$$\int X\,dP = \int_\Omega X\,dP = \int_\Omega Y\,dP = \int Y\,dP.$$

Logo, para qualquer $\mathcal{G}$,

$$E[X] = E[\,E[\,X\,|\,\mathcal{G}\,]\,].$$

Teorema 5.35. *Seja o Processo Estocástico X_n, $n \in \mathbb{N}$, independente e identicamente distribuído, $X_n : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{R})$. Assumimos que para qualquer n, vale que X_n toma valores em $\mathbb{N}$. Considere agora uma outra variável aleatória $N : (\Omega, \mathcal{F}) \to \mathbb{N}$ que é independente das X_n, $n \in \mathbb{N}$. Seja $\mathcal{G}$ a σ-álgebra $\sigma(N)$ gerada por N sobre o conjunto Ω. Considere a variável aleatória U sobre $(\Omega, \mathcal{F}, P)$ dada por*

$$U(\omega) = X_1(\omega) + X_2(\omega) + ... + X_{N(w)}(\omega).$$

Então,

$$E[\,U\,|\,\mathcal{G}\,] = Z,$$

onde $Z(\omega) = \int (X_1 + X_2 + ... + X_{n_0})\,dP$, para todo ω tal que $N(\omega) = n_0$.

Demonstração: Ora, a função Z definida sobre $(\Omega, \mathcal{F}, P)$ é claramente $\mathcal{G}$ mensurável.

Ainda, dado $G \subset \mathbb{N}$, como $X_1, X_2, ..., X_n$ é independente de N,

$$\int_{N(\omega)\in G} Z(\omega)\, dP(\omega) = \sum_{n\in G} \; P(N=n) \; \int (X_1 + X_2 + ... + X_n)\, dP \; =$$

$$\sum_{n\in G} \int (X_1 + X_2 + ... + X_n)\,(w)\, dP(w) \; \int I_{N(w)=n}\, dP(w) =$$

$$\sum_{n\in G} \int (X_1 + X_2 + ... + X_n) \; I_{N=n} \; dP =$$

$$\sum_{n\in G} \int U \; I_{N=n} \; dP = \int_{N(\omega)\in G} U \;\; dP.$$

$\square$

A partir do resultado acima se pode mostrar facilmente que para $s > 0$ fixo vale

$$E[\, s^U \,|\, \mathcal{G}\,] = Z,$$

onde $Z(\omega) = \int \, s^{(X_1+X_2+...+X_{n_0})}\, dP$, para todo ω tal que $N(\omega) = n_0$.

Teorema 5.36. *Considere o Processo Estocástico X_n, $n \in \mathbb{N}$, independente e identicamente distribuído, $X_n : (\Omega, \mathcal{F}, P) \to (\mathbb{N}, \mathcal{P}(\mathbb{N}))$. Assumimos que para qualquer n, vale que X_n toma valores em $\mathbb{N}$ (definição 3.9). Seja $G_X(s)$ a função geradora de probabilidade de X_1. Considere uma outra variável aleatória $N : (\Omega, \mathcal{F}) \to \mathbb{N}$ que é independente das X_n, $n \in \mathbb{N}$ e tem função geradora $G_N(s)$. Então, a variável aleatória*

$$U\,(\omega) \; = \; X_1\,(\omega) + X_2\,(\omega) + X_3\,(\omega) + ... + X_{N(\omega)}\,(\omega),$$

tem função geradora G_U que satisfaz

$$G_U(x) \; = G_N\,(\, G_X(s)\,).$$

Demonstração: Ora, usando a observação que fizemos antes deste teorema

$$G_U(s) = E[s^U] = \sum_n^\infty E[s^U \mid N = n]\, P(N = n) =$$

$$\sum_n^\infty E[s^{X_1+X_2+...+X_N} \mid N = n]\, P(N = n) =$$

$$\sum_n^\infty E[s^{X_1+X_2+...+X_n}]\, P(N = n) =$$

$$\sum_n^\infty E[s^{X_1}]\, E[s^{X_2}] \ldots E[s^{X_n}]\ P(N = n) =$$

$$\sum_n^\infty (E[s^{X_1}])^n\ P(N = n)\ =\ G_N(G_X(s)).$$

□

Exemplo 5.26. Como aplicação do Teorema acima podemos mencionar o seguinte problema: suponha que uma galinha coloque ovos em um número N, tal que N tenha uma distribuição exponencial de parâmetro λ. A seguir, supomos que cada ovo gera um pintinho com uma certa probabilidade. Neste caso, definimos $X = 0$ ou $X = 1$ de acordo com a possibilidade respectivamente de nascer ou não o pintinho. Então, supondo independência entre todas as variáveis envolvidas, para calcular o número de pintinhos que nascem, devemos calcular a probabilidade de

$$U\ =\ X_1 + X_2 + ... + X_N.$$

Neste caso, esta probabilidade pode ser facilmente descrita a partir da função geradora $G_U(s)$, que pode ser obtida a partir da afirmação do teorema acima.

♢

Teorema 5.37. *Seja* $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{R})$*, mensurável e* $\mathcal{G}$ σ*-álgebra contida em* $\mathcal{A}$*. Suponha que* X *seja* $\mathcal{G}$ *mensurável e considere* $Z : \Omega \to \mathbb{R}$ *tal que é* P*-integrável e ainda que valha que* $X\,Z$ *é também* P*-integrável. Então,*

$$E(X\,Z\,|\,\mathcal{G}) = X\,E(Z\,|\,\mathcal{G}).$$

Demonstração: Primeiro vamos mostrar o resultado quando X é uma função indicador, ou seja, $X = I_K$, $K \in \mathcal{G}$. Seja $Y = E(Z\,|\,\mathcal{G})$.

Seja G conjunto $\mathcal{G}$-mensurável. Então,

$$\int_G X\,Y\,dP = \int I_G\,I_K Y\,dP = \int_{G\cap K} Y\,dP =$$

$$\int_{G\cap K} Z\,dP = \int_G I_K Z\,dP = \int_G X\,Z\,dP.$$

Desta forma

$$E(X\,Z\,|\,\mathcal{G}) = X\,Y \;=\; X\,E(Z\,|\,\mathcal{G}).$$

Sendo assim, a afirmação esta demonstrada neste caso.

O resultado se estende linearmente para X do tipo

$$X = \sum_{i=1}^{k} a_i\,I_{K_i},$$

$K_i \in \mathcal{G}$ e $a_i \in \mathbb{R}$.

Como toda X que é $\mathcal{G}$ mensurável é limite crescente de uma seqüência de funções simples X_k, $k \in \mathbb{N}$, como do tipo descrita acima, o resultado segue do teorema da convergência monótona.

Note que

$$\lim_{k\to\infty} X_k\,Z \;=\; X\,Z.$$

□

Lembre que dizemos que a variável aleatória $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{R})$ é independente da sigma álgebra $\mathcal{G}$ se a σ-álgebra $\mathcal{F} = \{X^{-1}(B)\,|\,B$ conjunto de Borel $\mathcal{R}\}$ é independente da σ-álgebra $\mathcal{G}$.

Teorema 5.38. *Seja $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{R})$ mensurável e $\mathcal{G}$ σ-álgebra contida em $\mathcal{A}$. Suponha que X seja independente de $\mathcal{G}$ então*

$$E[\,X\,|\,\mathcal{G}\,] = \int X\,dP = E[X].$$

Demonstração: Por hipótese, para todo $G \in \mathcal{G}$, temos que I_G é $\mathcal{G}$ mensurável, logo

$$\int_G X\,dP = \int I_G\,X\,dP = \int I_G\,dP\,\int X\,dP =$$

$$\int I_G\,(\int X\,dP)\,dP = \int_G (\int X\,dP)\,dP.$$

Logo, o resultado está demonstrado.

□

Teorema 5.39. *Seja $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{R})$, integrável e $\mathcal{G}$ σ-álgebra contida em $\mathcal{A}$.*

Suponha que $\mathcal{V}$ seja um sistema-π e que $\mathcal{G} = \sigma(\mathcal{V})$. Assuma ainda que Ω seja uma união finita ou enumerável de elementos de $\mathcal{V}$. Então Y será igual a $E(X\,|\,\mathcal{G})$, se e só se, vale que Y é $\mathcal{G}$-mensurável e para qualquer elemento $V \in \mathcal{V}$, vale que

$$\int_V Y\,dP = \int_V X\,dP.$$

Demonstração: Suponha primeiro que X e Y são positivos.

Considere duas probabilidades sobre $\mathcal{G}$, que são μ_1 e μ_2, onde, respectivamente, $\mu_1(G) = \int_G Y dP$ e $\mu_2(G) = \int_G X\,dP$, para todo $G \in \mathcal{G}$. Quando estas são restritas ao sistema-π $\mathcal{V}$, temos que μ_1 e μ_2 coincidem. Concluímos assim que μ_1 e μ_2 coincidem sobre $\mathcal{G}$.

O caso geral segue de considerar a parte positiva e negativa de X, e, assim por diante, como em ocasiões anteriores.

□

O resultado acima afirma que se $\mathcal{G} = \sigma(\mathcal{V})$, então basta testar nos conjuntos geradores $V \in \mathcal{V}$, para saber se Y é o valor esperado de X dado a σ-álgebra $\mathcal{G}$.

Exemplo 5.27. Vamos apresentar agora um exemplo em que a σ-álgebra $\mathcal{G}$ é mais interessante. Considere um Processo Estocástico X_n, onde $n \in \{1, 2, 3, ...\}$, tomando valores em $S = \{1, 2\}$ com a σ-álgebra $\mathcal{A}$ gerada pelos cilindros sobre $S^{\mathbb{N}}$. Suponha que a correspondente P seja independente e identicamente distribuída de acordo com $P(X_1 = 1) = p_1$ e $P(X_1 = 2) = p_2$. Seja $\mathcal{G} = \sigma(X_2, X_3, X_4, ..., X_n, ...)$. Os conjuntos $\mathcal{G}$-mensuráveis serão da forma

$$(\{1\} \times B) \cup (\{2\} \times B),$$

onde B está na σ-álgebra $\mathcal{A}$.

Logo, uma função Y que seja $\mathcal{G}$ mensurável terá que ter o mesmo valor em cada ponto da forma $(1, x_2, x_3, x_4, ...)$ e $(2, x_2, x_3, x_4, ...)$. Ou seja, Y não sabe distinguir a primeira coordenada x_1. Ela não contém esta informação.

Os conjuntos da forma

$$(\{1\} \times C) \cup (\{2\} \times C),$$

onde C é cilindro da forma

$$C = \{\omega = (x_1, x_2, x_3, ...) \,|\, x_1 = s_1, x_2 = s_2, ..., x_k = s_k\,\} = \overline{s_1, s_2, ..., s_k},$$

geram a σ-álgebra $\mathcal{A}$. As uniões finitas de tais conjuntos definem um sistema-π e estão sob as hipóteses do último teorema.

Vamos calcular quem é $Y = E[X_1\, X_2 \,|\, \mathcal{G}]$. Seja $G = (\{1\} \times C) \cup (\{2\} \times C)$, onde C é cilindro da forma acima, então

$$\int_G X_1\, X_2\, dP = \int_{\{1\} \times C} 1\, X_2\, dP + \int_{\{2\} \times C} 2\, X_2\, dP =$$

$$s_1\, P(\{1\} \times C) + 2 s_1 P(\{2\} \times C).$$

Sendo assim, o valor acima é igual a $\alpha_1 = P(\{1\} \times C) + 2P(\{2\} \times C)$ se $s_1 = 1$, e $\alpha_2 = 2\,P(\{1\} \times C) + 4P(\{2\} \times C)$ se $s_1 = 2$.

Afirmamos que Y tem os seguintes valores:

a) $Y(\omega)$ é constante igual a $c_1 = \frac{\alpha_1}{P(\overline{1, s_2, s_3, ..., s_k})}$ em $\overline{11}$ e $\overline{21}$.

b) $Y(\omega)$ é constante igual a $c_2 = \frac{\alpha_2}{P(\overline{2, s_2, s_3, ..., s_k})}$ em $\overline{12}$ e $\overline{22}$.

Isto segue trivialmente do fato que, neste caso, se $s_1 = 1$, então

$$\int_G Y\,dP = \int_{\overline{11} \times \overline{s_2, s_3, ..., s_k}} c_1\,dP + \int_{\overline{21} \times \overline{s_2, s_3, ..., s_k}} c_1\,dP = c_1\,P(\overline{1, s_2, s_3, ..., s_k}) = \alpha_1.$$

Ainda, se $s_2 = 2$, então,

$$\int_G Y\,dP = \int_{\overline{12} \times \overline{s_2, s_3, ..., s_k}} c_2\,dP + \int_{\overline{22} \times \overline{s_2, s_3, ..., s_k}} c_2\,dP = c_2\,P(\overline{2, s_2, s_3, ..., s_k}) = \alpha_2.$$

Fica assim determinada, neste caso, quem é $E[X_1\,X_2\,|\,\mathcal{G}]$ pelo teorema acima.

$\diamondsuit$

Exemplo 5.28. Considere $S = \{1, 2, ..., d\}$, uma matriz estocástica $\mathcal{P}$ do tipo d por d e um vetor de probabilidade inicial $\pi \in \mathbb{R}^d$. Ficam assim definidas as probabilidades dos cilindros $\overline{a_1, a_2, ..., a_k}$ de tamanho k. Denotamos por X_n, $n \in \mathbb{N}$, $X_n : \{1, 2, ..., d\}^{\mathbb{N}} \to \mathbb{R}$ o processo estocástico Markoviano associado.

Denote por $\mathcal{F}_n$, $n \in \mathbb{N}$, a sigma algebra gerada pelas variáveis $X_1, X_2, ..., X_n$. Observe que uma função é $\mathcal{F}_n$-mensuravel, se e só se, é constante nos cilindros de tamanho n.

Seja F uma função definida sobre o conjunto $S = \{1, 2, 3, ..., d\}$. Para n fixo considere a função mensurável $F(X_{n+1}) = G : \{1, 2, ..., d\}^{\mathbb{N}} \to \mathbb{R}$. Desta forma, G tem o valor $F(j)$ para todo $w \in \overline{a_1, a_2, ..., a_n, j}$.

A função $E(F(X_{n+1})|\mathcal{F}_n) = g$ será constante em cilindros de tamanho n.

O valor de G no cilindro $\overline{a_1, a_2, ..., a_k}$ é

$$\sum_{j=1}^{j} F(j)\, P(\overline{a_1, a_2, ..., a_k, j}) = \sum_{j=1}^{j} F(j)\, \pi_{a_1}\, \mathcal{P}_{a_1,a_2}\, \mathcal{P}_{a_2,a_3} ... \mathcal{P}_{a_k,j}.$$

$\diamondsuit$

Agora vamos introduzir o conceito de probabilidade condicional no seu sentido mais amplo.

Seja $(\Omega, \mathcal{A}, P)$, espaço de medida e $\mathcal{G}$ σ-álgebra que está contida em $\mathcal{A}$. Gostaríamos de dar sentido ao conceito de probabilidade condicional $Y = Y_A$ de um conjunto $A \in \mathcal{A}$ dada a σ-álgebra $\mathcal{G}$. Ou seja, qual seria a probabilidade de A (desconhecido) se tudo que sabemos é $\mathcal{G}$?

Definição 5.49. *Considere um conjunto A, então $Y = E[I_A \,|\, \mathcal{G}] : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{R})$ será a probabilidade condicional de A dado $\mathcal{G}$. De outra forma, para qualquer $G \in \mathcal{G}$, vale que*

$$\int_G Y\, dP = \int_G I_A\, dP = P(A \cap G).$$

Se $\mathcal{G}$ fosse gerada por uma partição, por exemplo, suponha que $\Omega = \cup_{j=1}^{\infty} C_j$ e $\mathcal{G} = \sigma(C_1, C_2, ..., C_j, ...)$.

Ora, se Y é $\mathcal{G}$ mensurável, ela tem que ser constante nos conjuntos C_j. Denote por c_j o correspondente valor. Sendo assim, para todo j

$$c_j\, P(C_j) = \int_{C_j} Y\, dP = P(A \cap C_j).$$

Logo, para todo j

$$c_j = \frac{P(A \cap C_j)}{P(C_j)} = P(A \,|\, C_j).$$

Desta forma, o conceito introduzido generaliza o que já nos era conhecido para conjuntos.

Agora, fixada a σ-álgebra $\mathcal{G}$ podemos pensar que $A \in \mathcal{G}$ é um conjunto variável.

Fica assim definida uma lei $P_{\mathcal{G}}$ sobre $\mathcal{G}$ tal que

$$P_{\mathcal{G}}(A) = \int E[I_A \,|\, \mathcal{G}] \; dP.$$

Note que, se A_k, $k \in \mathbb{N}$, é uma família disjunta de elementos em $\mathcal{G}$, então

$$\sum_{k=1}^{\infty} P_{\mathcal{G}}(A_k) = \sum_{k=1}^{\infty} \int E[I_{A_k} \,|\, \mathcal{G}\ ] \; dP = \int E[\, \sum_{k=1}^{\infty} I_{A_k} \,|\, \mathcal{G}\ ] \; dP.$$

$$\int E[\ I_{\cup_{k=1}^{\infty} A_k} \,|\, \mathcal{G}\ ] \; dP = P_{\mathcal{G}}\,(\,\cup_{k=1}^{\infty} A_k\,).$$

As outras propriedades de probabilidade seguem de maneira fácil.

Logo, $P_{\mathcal{G}}$ é uma probabilidade.

5.6 Martingale e tempo de parada

Seja P uma probabilidade na σ-algebra $\mathcal{F}$. Se $\mathcal{A}$ é uma σ-algebra tal que $\mathcal{A} \subset \mathcal{F}$, então para toda f que seja $\mathcal{F}$-mensurável definimos $\mathbb{E}[f|\mathcal{A}] = g$ que é $\mathcal{A}$-mensurável e ainda tal que

$$\forall B \in \mathcal{A},\ \text{temos}\ \int_B f dP = \int_B g dP.$$

Definição 5.50. *Seja $(X_n, n \in \mathbb{N})$ um processo estocástico com um espaço de estados $S \subset \mathcal{R}$ finito ou enumerável, ou seja, $X_n : \Omega \to \mathbb{R}$ é mensurável e consideramos uma probabilidade P sobre a sigma algebra $\mathcal{F}$ gerada pelos cilindros. Então o processo estocástico é uma Martingale se:*

(i) $\mathbb{E}[|X_n|] < \infty$,

(ii) $\mathbb{E}[X_{n+1}|X_0, X_1, \cdots, X_n] = X_n$.

O valor esperado acima é tomado em relação a probabilidade P.

*Dizemos que um processo estocástico é **supermartingale** se ao invés da condição (ii) tivermos $\mathbb{E}[X_{n+1}|X_0, X_1, \cdots, X_n] \leq X_n$, e, finalmente, dizemos que é **submartingale** se tivermos em (ii) $\mathbb{E}[X_{n+1}|X_0, X_1, \cdots, X_n] \geq X_n$*

Um processo Martingale pode ser pensado como um jogo onde é esperado que um jogador mantenha sua fortuna acumulada; no caso supermartingale, é esperado que a fortuna acumulada diminua e no caso de submartingale que aumente.

Como veremos, a técnica do uso de Martingales permite obter o cálculo exato de valores esperados e de probabilidades em distintas situações. O texto [NP] (ver seção 11) descreve o uso de martingales para cadeias de Markov a tempo discreto.

Exemplo 5.29. Sejam Y_0 e $(Y_n, n \in \mathbb{N})$ variáveis aleatórias independentes com $\mathbb{E}[|Y_n|] < \infty$ e $\mathbb{E}[Y_n] = 0$, $\forall n$. Se $X_0 = 0$ e $X_n = Y_1 + Y_2 + \cdots + Y_n$ então X_n é Martingale para $\mathcal{F}_n = \sigma(Y_1, \cdots, Y_n)$

Demonstração.

$$\begin{aligned}\mathbb{E}[X_{n+1}|Y_0, \cdots, Y_n] &= \mathbb{E}[X_n + Y_{n+1}|Y_0, \cdots, Y_n] \\ &= \mathbb{E}[X_n|Y_0 \cdots, Y_n] + \mathbb{E}[Y_{n+1}|Y_0, \cdots, Y_n]\end{aligned}$$

Observe que X_n é $\sigma(Y_0, \cdots, Y_n)$-mensurável logo $\mathbb{E}[X_n|Y_0, \cdots, Y_n] = X_n$ e como os Y_i são independentes temos $\mathbb{E}[Y_{n+1}|Y_0 \cdots, Y_n] = \mathbb{E}[Y_{n+1}] = 0$, logo $\mathbb{E}[X_{n+1}|Y_0, \cdots, Y_n] = X_n$. $\square$

Exemplo 5.30. Sejam $S = \{1, \ldots, d\}$ e $Y_0, Y_1, Y_2, \ldots$ um processo de Markov em S, com matriz de transição $P = (P_{ij})$. Se $f : S \to \mathbb{R}$ (ou ainda $f = (f_1, \ldots, f_d)$) é uma função com $Pf = f$, isto é, $f(i) = \sum_j P_{ij} f(j)$, então: $\{X_n = f(Y_n)\}$ é Martingale com respeito à $\{Y_n\}$.

De fato, como f é limitada temos que $E[|X_n|] < \infty$. Além disso,

$$E[X_{n+1}|Y_0, \ldots, Y_n] = E[f(Y_{n+1})|Y_0, \ldots, Y_n] = E[f(Y_{n+1})|Y_n] =$$

$$\sum_j P_{Y_n j} f(j) = f(Y_n) = X_n.$$

Ainda, podemos considerar o mesmo processo de Markov do exemplo anterior, mas com uma função $f : S \to \{1, \ldots, d\}$ satisfazendo $P(f) = \lambda f$. Nesse caso tomando $X_n = \lambda^{-n} f(Y_n)$ também obtemos que $X_0, X_1 \ldots$ é martingale com respeito à $\{Y_n\}$.

Exemplo 5.31. Seja Y_0 e $(Y_n, n \in \mathbb{N})$ variáveis aleatórias, independentes e indenticamente distribuídas, tal que $\mathbb{E}[Y_k] = 0$ e $\mathbb{E}[Y_k^2] = \sigma^2$, $\forall k \in \mathbb{N}$. Seja $X_0 = 0$ e $X_n = (\sum Y_k)^2 - n\sigma^2$. Afirmamos então que X_n é Martingale com relação à $\mathcal{F}_n$, $n \in \mathbb{N}$, gerada por $\{Y_0, ..., Y_n\}$,.

$$\begin{aligned}\mathbb{E}[X_{n+1}|Y_0, \cdots, Y_n] &= \mathbb{E}[(Y_{n+1} + \sum_{k=1}^{n} Y_k)^2 - (n+1)\sigma^2 | Y_0, \cdots, Y_n] \\ &= \mathbb{E}[Y_{n+1}^2 + 2Y_{n+1}(\sum_{k=1}^{n} Y_k) + (\sum_{k=1}^{n} Y_k)^2 - (n+1)\sigma^2 | Y_0, \cdots, Y_n].\end{aligned}$$

Observe que $\mathbb{E}[Y_{n+1}^2 | Y_0, \cdots, Y_n] = \sigma^2$, além disso

$$\mathbb{E}[2Y_{n+1}(\sum_{k=1}^{n} Y_k)|Y_0, \cdots, Y_n] = 2(\sum_{k=1}^{n} Y_k)\mathbb{E}[Y_{n+1}|Y_0, \cdots, Y_n] = 0,$$

temos também que $\mathbb{E}[(\sum_{k=1}^{n} Y_k)^2|Y_0,\cdots,Y_n] = (\sum_{k=1}^{n} Y_k)^2$ logo

$$\mathbb{E}[X_{n+1}|Y_0,\cdots,Y_n] = \sigma^2 + (\sum_{k=1}^{n} Y_k)^2 - (n+1)\sigma^2 = (\sum_{k=1}^{n} Y_k)^2 - n\sigma^2 = X_n$$

Podemos também tomar uma definição mais geral para os processos de Martingale:

Definição 5.51. *Seja $X_n : \Omega \to \mathbb{R}$, $n \in \mathbb{N}$, mensurável como na definição anterior e seja $\mathcal{F}_1 \subset .... \subset \mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \cdots \subset \mathcal{F}$, uma sequência crescente de σ-álgebras, então X_n é martingale se $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$.*

Seja $X_n : \Omega \to \mathbb{R}$, $n \in \mathbb{N}$, e $Y_n : \Omega \to \mathbb{R}$, $n \in \mathbb{N}$, dois processos. Denote $\mathcal{F}_n$ a sigma-algebra gerada por $Y_0, Y_1, .., Y_n$. Se X_n, $n \in \mathbb{N}$, é martingale em relação a $\mathcal{F}_n$, $n \in \mathbb{N}$, é usual dizer que X_n, $n \in \mathbb{N}$, é martingale em relação a $Y_n : \Omega \to \mathbb{R}$, $n \in \mathbb{N}$.

Exemplo 5.32. Seja $Y_0 = 0$ e $Y_n, n = 1, 2, \ldots$, uma sequência de variáveis aleatórias independentes e identicamente distribuídas tomando valores em $\{-1, 1\}$ com $P(Y_n = 1) = p$ e $P(Y_n = -1) = q$, com $p + q = 1$. Defina $X_n = S_n - n\mu$, $n \in \mathbb{N}$, onde $\mu = E[Y_n] = p - q$ e $S_n = \sum_{j=1}^{n} Y_j$.

Considere a sequencia de sigma-algebras $\mathcal{F}_n$, $n \in \mathbb{N}$, tal que cada $\mathcal{F}_n$ é gerada por $Y_0, \ldots, Y_n$.

Vamos mostrar que X_n , $n \in \mathbb{N}$, é martingale em relação a $\mathcal{F}_n$, $n \in \mathbb{N}$.

De fato,

$$E(|X_n|) = E(|(Y_1 - \mu) + \ldots + (Y_n - \mu)|) \leq \sum_{i=0}^{n} E|(Y_i - \mu)) = 0 < \infty.$$

Além disso,

$$E(X_{n+1}|Y_0, \ldots, Y_n) = E(X_n + Y_{n+1} - \mu|Y_0, \ldots, Y_n) =$$

$$X_n \; + \; E(Y_{n+1}) - \mu = X_n.$$

Logo, X_n é martingale com respeito a $\mathcal{F}_n$

Além disso, $X'_n = \left(\frac{q}{p}\right)^{S_n}$, $n \in \mathbb{N}$, também é martingale com respeito a $\{Y_n\}$, pois

$$E\left(X'_{n+1}|Y_0,\ldots,Y_n\right) = E\left(\left(\frac{q}{p}\right)^{S_{n+1}}|Y_0,\ldots,Y_n\right) =$$

$$\left(\frac{q}{p}\right)^{S_n} E\left(\left(\frac{q}{p}\right)^{Y_{n+1}}|Y_0,\ldots,Y_n\right) = \left(\frac{q}{p}\right)^{S_n}\left(p\frac{q}{p}+q\frac{p}{q}\right)$$

Como $p+q=1$, concluímos que

$$E\left(X'_{n+1}|Y_0,\ldots,Y_n\right) = \left(\frac{q}{p}\right)^{S_n} = X'_n. \tag{5.2}$$

Portanto, X'_n também é martingale com respeito a Y_n.

$\diamondsuit$

Exemplo 5.33. Considere fixada uma função integrável $X:\Omega\to\mathbb{R}$ limitada sobre o espaço de medida $(P,\Omega,\mathcal{F})$. Seja $\mathcal{F}_n$, $n\in\mathbb{N}$, uma sequencia crescente de sigma-algebras $\mathcal{F}_n\subset\mathcal{F}$. Defina $X_n = E[\,X\,|\,\mathcal{F}_n\,]$. É fácil ver que X_n, $n\in\mathbb{N}$, é uma martingale.

De fato,

1) $\forall n$, $E(|X_n|)<\infty$ (Lemma 5.11 in [Du])

2) X_n is $\mathcal{F}_n$ integrável por definição

3) $E[X_{n+1}\,|\,\mathcal{F}_n] = E[E[X\,|\,\mathcal{F}_{n+1}\,]\,|\,\mathcal{F}_n] = E[X\,|\,\mathcal{F}_n] \;=\; X_n$

Observe que se a sigma algebra gerada pela união $\cup_n \mathcal{F}_n$ for igual a $\mathcal{F}$, então $X_n = E[\,X\,|\,\mathcal{F}_n\,]$, para n grande, vai estar muito "parecida"com X. O próximo teorema é uma espécie de versão inversa do que exemplificamos acima.

Lema 5.4. *Suponha que X_n, $n \geq 0$, seja Martingale com relação a Y_n, $n \geq 0$. Então dados $k, n > 0$ temos que*

$$\mathbb{E}[X_{n+k}|Y_0, Y_1, \cdots, Y_n] = X_n.$$

Demonstração Seja n fixo. Para $k = 1$ o resultado segue da definição. Vamos supor que a propriedade vale para qualquer n e também para k, e então mostrar que vale para $k+1$.

Ora,

$$\mathbb{E}[X_{n+k+1}\,|\,Y_0, Y_1, \cdots, Y_n] =$$
$$\mathbb{E}[\ \mathbb{E}[X_{n+k+1}\,|Y_0, Y_1, \cdots, Y_{n+k}]\,|\,Y_0, Y_1, ..., Y_n]\ ] =$$
$$\mathbb{E}[X_{n+k}\,|\,Y_0, Y_1, \cdots, Y_n\,] = X_n.$$

$\square$

Uma sequencia X_n, $n \in \mathbb{N}$, é uniformemente integrável se

$$\lim_{a \to \infty} \sup_{n \geq 0} \int |\,X_n\,|\ I_{\{\,|X_n|\,>\,a\}}\,dP = 0.$$

O próximo resultado é conhecido como o Teorema de Convergencia de Martingales.

Teorema 5.40. *Seja X_n, $n \in \mathbb{N}$, uma Martingale (ou, submartingale) satisfazendo a propriedade*

$$\sup_{n \geq 0} E[\,|X_n\,|\,] < \infty.$$

Então existe $X : \Omega \to \mathbb{R}$ tal que X_n converge a X em probabilidade, ou seja, $P(|\,X_n - X\,| > \epsilon\,) \to 0$, quando $n \to \infty$.

Ainda, se X_n é uma Martingale uniformemente integrável, então X_n converge a X em $\mathcal{L}^1$, isto é,

$$\lim_{n\to\infty} \int |X - X_n|\, dP = 0.$$

Além disso, $\int X_n dP = \int X dP$, para todo número natural $n \in \mathbb{N}$.

Exemplo 5.34. Considere o conjunto $S = \{1,2\}^{\mathbb{N}}$. Seja $\mathcal{F}_n$ a σ-álgebra gerada pelos cilindros de tamanho n. Temos que $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ e $\mathcal{F}_n$ converge à σ-álgebra de Borel. Seja $Y : S \to \mathbb{R}$ uma função Borel mensurável e defina a sequência de variáveis aleatórias $\{X_n\}$ por $X_n = E(Y|\mathcal{F}_n)$. Como

$$E(X_{n+1}|\mathcal{F}_n) = E(E(Y|\mathcal{F}_{n+1})|\mathcal{F}_n) = E(Y|\mathcal{F}_n) = X_n, \tag{5.3}$$

temos que $\{X_n\}$ é martingale com respeito a $\mathcal{F}_n$. Pelo teorema anterior, X_n converge a uma variável aleatória X_∞. Nesse caso, $X_\infty = Y$.

$\diamondsuit$

Exemplo 5.35. Dada uma matriz linha estocástica P de uma cadeia de Markov com estados finitos, denote por u o vetor com todas as entradas iguais a 1. Então $P(u) = u$. O mesmo vale se u tiver todas as entradas iguais a mesma constante c. Uma pergunta natural é se poderiam existir outros vetores u diferentes destes tais que $P(u) = u$.

Suponha que P seja recorrente e irredutível $\{Y_n\}$. Vamos usar o teorema de convergência de martingales para mostrar que todo vetor coluna limitado $\gamma = \{\gamma(i)\}$ solução da equação $\gamma = P\gamma$ é constante. Como $X_n = \gamma(Y_n)$ é martingale, temos que $\lim_{n\to\infty} X_n = \lim_{n\to\infty} \gamma(Y_n)$ existe com probabilidade 1. Como a cadeia é recorrente, temos que ambos os eventos $\{X_n = \gamma(i)\}$ e $\{X_n = \gamma(j)\}$ ocorrem para infinitos n. Mas como o limite $\lim_{n\to\infty} X_n$ existe, concluímos que $\gamma(i) = \gamma(j)$ para todo i, j e portanto, γ é constante.

♢

Definição 5.52. *Lembre que uma variável aleatória T é chamada de tempo de parada com respeito a $\{X_n\}$ se T toma valores em $\{0, 1, \ldots, \infty\}$ e se, para cada $n = 0, 1, \ldots$, temos que o evento $T = n$ é determinado por $\{X_0, \ldots, X_n\}$. Mais precisamente, a função indicadora do evento $\{T = n\}$ pode ser escrita como uma função de $X_0, X_1, \ldots, X_n$.*

Definição 5.53. *Dado um processo estocástico $X_n : \Omega \to \mathbb{R}$, (uma probabilidade sobre $S^{\mathbb{N}}$) onde para cada $w = (w_0, w_1, w_2, ..., w_n, ...) \in S^{\mathbb{N}}$, temos que $X_n(w) = w_n$, e um tempo de parada T. O processo estocástico X_T é naturalmente definido através das funções $(X_T)_n$, $n \in \mathbb{N}$:*

$(X_T)_n\,(w) = X_n(w)$, para $n < T(w)$, e $(X_T)_n(w) = X_{T(w)}$ para $n \geq T(w)$.

Dito de outra forma: $(X_T)_n = X_{T \wedge n}$, onde $T \wedge n$ é o mínimo de T e n.

Fica assim definida uma função $G : \Omega \to \Omega$ tal que

$$G(w) = (w_0, w_1, .., w_{T(w)-1}, w_{T(w)}, w_{T(w)}, ..., w_{T(w)}, ...).$$

A um processo estocástico está associada uma probabilidade P_T sobre $S^{\mathbb{N}}$.

Assim, precisamos definir P_T sobre cilindros.

Suponha que P seja a probabilidade sobre $S^{\mathbb{N}}$ associada ao processo estocástico $X_n : \Omega \to \mathbb{R}$.

Assim, estabelecemos que

$$P_T\,((X_T)_0 = a_0, (X_T)_1 = a_1, .., (X_T)_k = a_k) =$$

$$P(\,G^{-1}\,(\{X_0 = a_0, X_1 = a_1, .., X_k = a_k\,\}\,)\,).$$

Fica assim definida a probabilidade P_T sobre $S^{\mathbb{N}}$.

Exemplo 5.36. Seja $S = \mathbb{Z}$, X_n, $n \in \mathbb{N}$, o random walk $(1/2, 1/2)$, onde $X_0 = 0$, e $T(w)$ a primeira vez que w atinge o valor 5 ou -5. Com probabilidade um vale que $T(w)$ é finito.

Assim, $T(0, 1, 2, 1, 2, 3, 4, 5, 4, 3, 4, ..) = 8$.

$$X_T(0, 1, 2, 1, 2, 3, 4, 5, 4, 3, 4, ..) = (0, 1, 2, 1, 2, 3, 4, 5, 5, 5, 5, ..).$$

$\diamondsuit$

Lema 5.5. *Se S, T são tempos de parada então $S + T$ e $S \wedge T = \min\{S, T\}$ são tempos de parada também.*

Demonstração: Basta observar que $I_{\{S+T=n\}} = \sum_{k=0}^{n} I_{\{S=k\}} I_{\{T=n-k\}}$ e $I_{\{S \wedge T > n\}} = I_{\{S>n\}} I_{\{T>n\}}$.

$\square$

Lema 5.6. *Suponha que X_n é martingale relativo a Y_n, $n \in \mathbb{N}$ e T tempo de parada com respeito à Y_n, então $\forall n \geq k, E[X_n I_{\{T=k\}}] = E[X_k I_{\{T=k\}}]$.*

Demonstração:

Dizer X_n é martingale relativo a Y_n, $n \in \mathbb{N}$, significa que estamos considerando a sequencia crescente de sigma-algebras $\mathcal{F}_n = \sigma(Y_1, Y_2, .., Y_n)$.

Sendo assim,

$$E[X_n I_{\{T=k\}}] = E[E[X_n I_{\{T=k\}} | Y_0, \ldots, Y_k]] =$$
$$E[I_{\{T=k\}} E[X_n | Y_0, \ldots, Y_k]] = E[X_k I_{\{T=k\}}].$$

$\square$

Lema 5.7. *Se X_n é martingale e T tempo de parada, então para todo $n \geq 1$ vale $E[X_n] = E[X_0] = E[X_{T \wedge n}]$.*

Demonstração:

$$E[X_{T\wedge n}] = E[X_n I_{\{T\geq n\}}] + \sum_{k=0}^{n-1} E[X_T I_{\{T=k\}}] =$$

$$E[X_n I_{\{T\geq n\}}] + \sum_{k=0}^{n-1} E[X_k I_{\{T=k\}}] = E[X_n I_{\{T\geq n\}}] + \sum_{k=0}^{n-1} E[X_n I_{\{T=k\}}] = E[X_n].$$

□

Vale ainda o seguinte teorema, que não iremos demonstrar.

Teorema 5.41. *Suponha que X_n, $n \geq 0$, seja martingale e T seja um tempo de parada.*

Se $P(T < \infty) = 1$ e $E\left(\sup_{n\geq 0} |X_{T\wedge n}|\right) < \infty$, então $E(X_T) = E(X_0)$

Este teorema - Teorema do Martingale stopping time - possui muitas aplicações interessantes (ver seção 6.4 em [KT]) como veremos.

Exemplo 5.37. Seja $Y_0 = 0$ e $Y_n; n = 1, 2, \ldots$ uma sequência de variáveis aleatórias independentes e identicamente distribuídas tomando valores em $\{-1, 1\}$ com $P(Y_n = 1) = p$ e $P(Y_n = -1) = q$, com $p + q = 1$. Sabemos que $X_n = S_n - n\mu$, $n \in \mathbb{N}$ é uma Martingale.

Seja T tempo de parada que esteja nas condições do Teorema acima. Por exemplo, se $E(T) < \infty$ então se pode usar o teorema (ver Corollary 3.3 na seção 6 de [KT1]).

Então

$$0 = E(X_0) = E[X_T] = E[S_T] - \mu E[T].$$

Desta forma, sabendo calcular $E(T)$ podemos obter o valor $E[S_T]$.

Suponha que $q = p = 1/2$. Neste caso, $\mu = 0$ e $E(S_T) = 0$.

Considere agora o tempo de parada

$$T(w) = \min_n \{S_n(w) = -c \quad \text{ou} \quad S_n(w) = d\}$$

onde $c, d > 0$.

Estamos supondo que $q = p = 1/2$.

Seja $p(c)$ a probabilidade que S_n atinja $-c$ antes de d.

Do teorema acima temos que

$$0 = E[S_T] = -c\, p(c) \, + \, d\, (1 - p(c)).$$

Assim, obtemos que

$$p(c) = \frac{d}{c+d}.$$

Sabemos também que o processo $V_n = S_n^2 - n\, var(S_n) = S_n^2 - n$, $n \in \mathbb{N}$, é uma Martingale.

Ora, pelo teorema acima

$$0 = E(V_0) = E[S_T^2] - E[T] = c^2\, p(c) \, + \, d^2\, (1 - p(c)) - E[T].$$

Como $p(c) = \frac{d}{c+d}$ obtemos que $E[T] = c\, d$.

$\diamondsuit$

Exemplo 5.38. Seja $Y_0 = 0$ e $Y_n; n = 1, 2, \ldots$ uma sequência de variáveis aleatórias independentes e identicamente distribuídas tomando valores em $\{-1, 1\}$ com $P\,(Y_n = 1) = p$ e $P\,(Y_n = -1) = q$, com $p + q = 1$. Assim, temos que $E\,(Y_n) = p - q = \mu$. Definimos $X_n = S_n - n\mu$, $n \in \mathbb{N}$.

Lembre que definimos $X_n' = \left(\frac{q}{p}\right)^{S_n}$, $n \in \mathbb{N}$, e mostramos que é uma Martingale com relação a sequencia de sigma-algebras $\mathcal{F}_n$, $n \in \mathbb{N}$, tal que cada uma é gerada por $\{Y_0, ..., Y_n\}$.

Defina agora o tempo de parada T como sendo $T = min\{n : S_n = -a \quad ou \quad S_n = b\}$.

Sabemos que $E(S_T) = \mu E(T)$ pelo que foi explicado acima.

Seja $\nu_a = P(S_T = -a)$. Então, vale que

$$E\left(X'_T\right) = \nu_a \left(\frac{q}{p}\right)^{-a} + (1-\nu_a)\left(\frac{q}{p}\right)^{b}. \tag{5.4}$$

Mas do teorema acima, $E\left(X'_T\right) = E\left(X'_0\right) = 1$. Assim, concluímos que

$$\nu_a = \frac{1 - \left(\frac{q}{p}\right)^b}{\left(\frac{q}{p}\right)^{-a} - \left(\frac{q}{p}\right)^b} \tag{5.5}$$

Desta forma graças ao resultado acima (para tempo de parada) e o fato que X'_n $n \in \mathbb{N}$, é Martingale, conseguimos calcular algo não trivial, o valor $P(S_n = -a)$

♢

Lema 5.8. *Se X_n é martingale relativo a Y_n e ϕ uma função convexa tal que $E[\phi(X_n)^+] < \infty$ para todo n, então $\phi(X_n)$ é submartingale com respeito à Y_n.*

Pela desigualdade de Jensen $E[\phi(X_{n+1})|Y_0, \ldots, Y_n] \geq \phi(E[X_{n+1}|Y_0, \ldots, Y_n]) = \phi(X_n)$, o que prova o enunciado.

□

5.7 O movimento Browniano

Agora vamos apresentar uma introdução ao movimento Browniano, que é um exemplo de processo de Markov com tempo e espaço contínuos. O espaço de estados $S = \mathbb{R}^n$. Então, o processo estocástico $X(t)$, $t \in [0, \infty) \subset \mathbb{R}$, vai descrever a posição de uma partícula em $\mathbb{R}^n$ através do movimento Browniano.

Seja Ω o espaço de Skhorohod:

$$\Omega = \{w : [0, \infty) \to \mathbb{R}^n \text{ contínuas a direita e com limite a esquerda }\}$$

Como sempre, assumimos que $X(t)$, $t \geq 0$, é tal que os caminhos amostrais $w \in \Omega$ satisfazem $w(t) = X_t(w)$.

Um processo estocástico a tempo contínuo tomando valores em $\mathbb{R}^n$ a grosso modo seria uma probabilidade sobre o espaço de Skhorodod. Referimos o leitor a [EK] para uma descrição bastante completa destes espaços.

O Processo de Wiener (Movimento Browniano) vai definir uma probabilidade P sobre Ω tal que com probabilidade 1 todos os caminhos são contínuos.

Alertamos o leitor que para colocar em base matematicamente sólida os resultados desta seção seria necessário uma versão muito mais apurada e sofisticada do que pretendemos fazer aqui. Nosso objetivo é tão somente descrever as propriedades mais importantes desta teoria de forma intuitiva e matematicamente inteligível. Referimos o leitor a [KS] e [KT2] para exposições matematicamente rigorosas do assunto.

Definição 5.54. *Um movimento Browniano é um processo estocástico X_t; $t \geq 0$, tomando valores em $\mathbb{R}^n$, com as seguintes propriedades:*

(a) *Cada incremento $X(t+s) - X(s)$, $t, s \geq 0$, tem distribuição normal com média $\mu = 0$ e variância $\sigma\, t$, onde σ é um parâmetro fixo, e ainda os caminhos amostrais $w_t = X_t(w)$ são contínuos em $t \geq 0$.*

(b) $X(0) = x_0 \in \mathbb{R}^n$.

(c) *Para cada par de intervalos de tempo disjuntos $[t_1, t_2], [t_3, t_4]$ com $t_1 < t_2 \leq t_3 < t_4$, os incrementos $X(t_4) - X(t_3)$ e $X(t_2) - X(t_1)$ são variáveis aleatórias independentes dadas em (a), e similarmente para n intervalos disjuntos, onde n é um inteiro positivo.*

Segue de (a) que no caso unidimensional ($n = 1$) a Probabilidade P associada ao processo é tal que para qualquer $t \geq 0$

$$P(X_t \in (a,b)) = \int_a^b \frac{e^{-\frac{(x-x_0)^2}{2\,\sigma^2\,t}}}{\sqrt{2\,\pi\,\sigma^2\,t}}.dx.$$

É usual tambem chamar de movimento Browniano um processo X_t, $t \geq 0$, tal que satisfaz apenas (a) e (c). Neste caso não se diz nada sobre a distribução da posição inicial X_0.

No caso em que vale (b), dizemos que X_t, $t \geq 0$, é o movimento Browniano condicionado a começar em $x_0 \in \mathbb{R}^n$.

Referimos o leitor para [KT1] e [Nel] para resultados mais detalhados sobe o Movimento Browniano.

No caso bidimensional $X_t \in \mathbb{R}^2$ este processo pode ser descrito através das observações empíricas que foram feitas inicialmente pelo biólogo R. Brown (que dá nome ao processo) que viveu de 1773 a 1858 sobre a posição de uma partícula de pólen sobre a superfície de um lago. Assuma que $\mathbb{R}^2$ vai descrever a superfície de um lago e que uma partícula de pólen será colocada na posição x_0 no tempo 0. Esta partícula de pólen estará sujeita a colisões com muitas moléculas (de dimensão muito menor que a de pólen e que se encontram na superfície do lago) que se movem e colidem ao acaso com a partícula de pólen ao longo do tempo $t \geq 0$. No tempo $t \geq 0$ a posição da partícula de pólen será descrita por w_t que define assim um caminho amostral w_t, $t \geq 0$. Na figura 5.4 mostramos vários possíveis caminhos que são observados. Ou seja, se coloca uma vez a partícula de pólen em x_0 e se observa um certo caminho amostral. Depois se repete o experimento e, sem que as condições externas sejam modificadas, se coloca de novo a partícula de pólen em x_0; então se observa um outro caminho amostral diferente do primeiro. Isto caracteriza o fato que o problema em consideração é não determinístico. Será descrito por um processo estocástico X_t, $t \geq 0$. O que faz sentido é perguntar: dado uma região $A \subset \mathbb{R}^2$, qual a probabilidade de encontrar a partícula de polen na região A no tempo t_0, dado que estava no tempo 0 na posição x_0. Ou seja, desejamos saber o valor $P(X_{t_0} \in A)$. Isto dá uma ideia em termos práticos do que é o movimento Browniano.

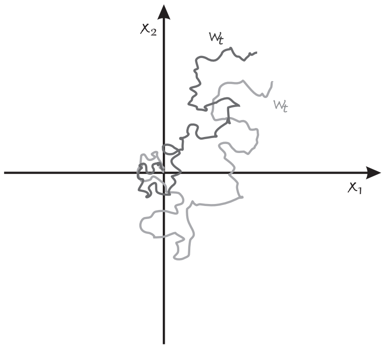


Figura 5.4: caminhos amostrais w_t, $t \geq 0$, do movimento Browniano no plano

As origens físicas do movimento Browniano sugerem que os caminhos amostrais $w_t, t \geq 0$, sejam caminhos contínuos mas não diferenciáveis.

Vamos aqui nos concentrar no caso unidimensional em que $X_t \in \mathbb{R}$.

No caso do movimento Browniano unidimensional com probabilidade 1 os caminhos amostrais serão funções contínuas $w : [0, \infty) \to \mathbb{R}$ tal que $w(0) = x_0$.

Vamos supor por um momento que $x_0 = 0$. Assim, os $w \in \Omega$ são tais que $w(0) = 0$.

Desejamos obter uma probabilidade sobre o conjunto de tais caminhos Ω. Vamos ser mais precisos no que desejamos obter.

Um conjunto cilindro, por exemplo, seria uma especificação do tipo

$$\{X_0 = 0, X_{t_1} \in (a_1, b_1), X_{t_2} \in (a_2, b_2), X_{t_3} \in (a_3, b_3)\}$$

onde $0 < t_1 < t_2 < t_3$, $a_1 < b_1, a_2 < b_2, a_3 < b_3$.

Ou seja, todas as funções contínuas w tais que $w(t_1) \in (a_1, b_1), w(t_2) \in (a_2, b_2), w(t_3) \in (a_3, b_3)$ (ver figura 5.5).

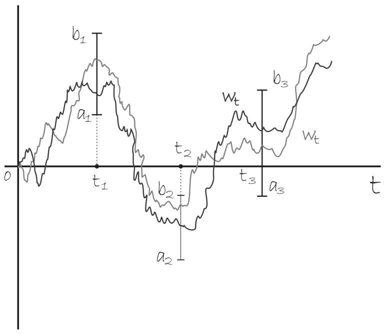


Figura 5.5: caminhos amostrais w_t, $t \geq 0$, dentro de um cilindro

Por exemplo,

$$\{X_0 = 0, X_{2.1} \in (5,7), X_{4.7} \in (-3,1.2), X_{10.9} \in (-3,8.4)\}$$

é um cilindro.

Mais geralmente um cilindro seria um conjunto da forma

$$\{X_0 = 0, X_{t_1} \in (a_1,b_1), X_{t_2} \in (a_2,b_2), X_{t_3} \in (a_3,b_3), ..., X_{t_n} \in (a_n,b_n)\}$$

onde $n \in \mathbb{N}$, $0 < t_1 < t_2 < t_3 < ... < t_n$, $a_1 < b_1, a_2 < b_2, a_3 < b_3, ..., a_n < b_n$.

Primeiramente desejamos obter a expressão analítica para o valor

$$P(\{X_0 = 0, X_{t_1} \in (a_1,b_1), X_{t_2} \in (a_2,b_2), X_{t_3} \in (a_3,b_3), ..., X_{t_n} \in (a_n,b_n)\})$$

no caso do movimento Browniano. A partir desta caracterização de P através das distribuições finito-dimensionais se poderá estender P a conjuntos mais complexos na sigma-algebra gerada pelos cilindros. Isto vai definir P sobre um certa sigma-algebra sobre Ω. Uma boa descrição da teoria necessária para a correta formalização destas afirmações aparece em [EK]. Existem várias questões técnicas nesta formalização e isto está acima do escopo do presente texto (ver [KS]).

Vamos voltar agora ao caso geral de um x_0 inicial qualquer.

Esta probabilidade P vai fazer com que sejam verdadeiras a seguintes afirmações:

$$E(X_t) = x_0,\ Var(X_t) = \sigma^2\, t \ \text{ e } \ E(X_t\, X_s) = \min\{t, s\}\,.$$

Seja x_0 a posição da partícula no tempo 0, ou seja, $X(0) = x_0$. Seja $p_t(x_0, x)$ a densidade de probabilidade condicional de $X(t)$ dado que $X(0) = x_0$. Ou seja, $p_t(x_0, x)$ é a densidade de $P(X_t \,|\, X(0) = x_0)$.

Ou seja, $p(t, x_0, x)$, onde $t \geq 0$, $x_0, x \in \mathbb{R}$, é a densidade da probabilidade de estando em x_0 no tempo 0, estar em x no tempo t.

Com isto queremos dizer que se a partícula no tempo 0 está em x_0, então a probabilidade de encontrá-la no intervalo (a, b) no tempo t será

$$\int_a^b p_t(x_0, x)\, dx.$$

Como $p_t(x_0, x)$ é uma função de densidade em x, temos que $p_t(x_0, x) \geq 0$ e $\int_{-\infty}^{+\infty} p_t(x_0, x) dx = 1$. Além disso, vamos estipular que para t próximo de t_0, $X(t + t_0)$ está perto de x_0, ou seja,

$$\lim_{t\to 0} p_t(x_0, x) = 0 \qquad se \qquad x \neq x_0. \tag{5.6}$$

Pelas hipóteses mencionadas no começo desta seção sabemos que o movimento Browniano X_t, $t \geq 0$, satisfaz

$$p_t(x_0, x) = p(x, t|x_0) = \frac{e^{-\frac{(x-x_0)^2}{2\,\sigma^2\, t}}}{\sqrt{2\,\pi\,\sigma^2\, t}}. \tag{5.7}$$

No caso em que $x_0 = 0$ usamos a notação $p(t, x)$ para tal densidade. Assim,

$$p(t, x) = \frac{e^{-\frac{x^2}{2\,\sigma^2\, t}}}{\sqrt{2\,\pi\,\sigma^2\, t}}. \tag{5.8}$$

Note que $p(t,x)$ satisfaz a equação de difusão

$$\frac{\partial}{\partial t}\,p(t,x)=\frac{1}{2}\,\sigma^2\frac{\partial^2 p(t,x)}{\partial^2\,x}, \tag{5.9}$$

tomando como condição inicial $p(0,.)$ a Delta Dirac no ponto 0 e onde σ é uma constante positiva.

No tempo t a distribuição definida pela densidade $p_t(x_0,x)$ tem média x_0 e variância $\sigma^2\,t$.

Vamos assumir a partir de agora que $\sigma=1$.

A expressão $\int_a^b p(x,t|x_0)dx$ denota a probabilidade de a partícula estar entre a e b no instante t dado que estava em x_0 no tempo 0.

Supondo que $X(0)=0$, pode se mostrar que a probabilidade P do movimento Browniano satisfaz, por exemplo,

$$P(X_0=0,X_{t_1}\in(a_1,b_1)\,,X_{t_2}\in(a_2,b_2)\,,X_{t_3}\in(a_3,b_3))=$$

$$\int_{a_3}^{b_3}\int_{a_2}^{b_2}\int_{a_1}^{b_1} p_{t_1}(0,x_1)\;p_{t_2-t_1}(x_1,x_2)\;p_{t_3-t_2}(x_2,x_3)\,dx_1\,dx_2\,dx_3.$$

$$\int_{a_3}^{b_3}\int_{a_2}^{b_2}\int_{a_1}^{b_1}\frac{e^{-\frac{x_1^2}{2t_1}}}{\sqrt{2\,\pi t_1}}\;\frac{e^{-\frac{(x_2-x_1)^2}{2\,(t_2-t_1)}}}{\sqrt{2\,\pi(t_2-t_1)}}\;\frac{e^{-\frac{(x_3-x_2)^2}{2\,(t_3-t_2)}}}{\sqrt{2\,\pi(t_3-t_2)}}\,dx_1\,dx_2\,dx_3.$$

Fica assim definida para o movimento Browniano com $x_0=0$ e $\sigma=1$ o valor

$$P(X_0=0,X_{t_1}\in(a_1,b_1),X_{t_2}\in(a_2,b_2),X_{t_3}\in(a_3,b_3)).$$

Para um cilindro geral

$$P(X_0=0,X_{t_1}\in(a_1,b_1),X_{t_2}\in(a_2,b_2),X_{t_3}\in(a_3,b_3),...,X_{t_n}\in(a_n,b_n)))=$$

$$\int_{a_n}^{b_n}\cdots\int_{a_3}^{b_3}\int_{a_2}^{b_2}\int_{a_1}^{b_1}\frac{e^{-\frac{x_1^2}{2t_1}}}{\sqrt{2\,\pi t_1}}\;\frac{e^{-\frac{(x_2-x_1)^2}{2\,(t_2-t_1)}}}{\sqrt{2\,\pi(t_2-t_1)}}\;\frac{e^{-\frac{(x_3-x_2)^2}{2\,(t_3-t_2)}}}{\sqrt{2\,\pi(t_3-t_2)}}\cdots\frac{e^{-\frac{(x_n-x_{n-1})^2}{2\,(t_n-t_{n-1})}}}{\sqrt{2\,\pi(t_n-t_{n-1})}}\,dx_1\,dx_2\,dx_3\ldots dx_n.$$

O processo X_t com $X_0 = 0$ é claramente não estacionário.

No caso em que a condição inicial é ela mesmo não determinística, e descrita pela densidade $f_0 : \mathbb{R} \to [0, \infty)$, onde $\int_{-\infty}^{\infty} f_0(x)\, dx\ = 1$, o Processo Estocástico movimento Browniano X_t, $t \geq 0$, associado é tal que

$$P(X_0 \in (a_0, b_0), X_{t_1} \in (a_1, b_1)\,, X_{t_2} \in (a_2, b_2)\,, X_{t_3} \in (a_3, b_3)) =$$

$$\int_{a_3}^{b_3} \int_{a_2}^{b_2} \int_{a_1}^{b_1} \int_{a_0}^{b_0} p_{t_1}(x_0, x_1) p_{t_2 - t_1}(x_1, x_2) p_{t_3 - t_2}(x_2, x_3)\, f(x_0)\, dx_0\, dx_1\, dx_2\, dx_3.$$

$$\int_{a_3}^{b_3} \int_{a_2}^{b_2} \int_{a_1}^{b_1} \int_{a_0}^{b_0} \frac{e^{-\frac{(x_1 - x_0)^2}{2\, t_1}}}{\sqrt{2\, \pi t_1}} \frac{e^{-\frac{(x_2 - x_1)^2}{2\, (t_2 - t_1)}}}{\sqrt{2\, \pi (t_2 - t_1)}} \frac{e^{-\frac{(x_3 - x_2)^2}{2\, (t_3 - t_2)}}}{\sqrt{2\, \pi (t_3 - t_2)}}\, f(x_0)\, dx_0\, dx_1\, dx_2\, dx_3.$$

Como nos processos Markovianos a tempo discreto, podemos nos perguntar se existe uma função densidade inicial estacionária $f(x)$ com respeito ao processo X_t descrito acima, ou seja, se existe solução f para o problema

$$\int_a^b f(x)\, dx = \int_a^b \int_{-\infty}^{\infty} f(x_0) p_t(x_0, x_1)\, dx_0 dx_1\,. \tag{5.10}$$

Neste caso valeria para qualquer $t \geq 0$

$$P(\, X_0 \in (a, b))\,) = P(\, X_t \in (a, b)\,)$$

A resposta para tal pergunta é não. Para entender esta questão de forma mais profunda vamos explicar a relação do movimento Browniano com equações diferenciais parciais.

Fixada a condição inicial f_0 se deseja obter a densidade $f_t : \mathbb{R} \to \mathbb{R}$ tal que para o processo acima vale para qualquer intervalo (a, b) e $t \geq 0$

$$\int_a^b f_t(x)\, dx = \int_a^b \int_{-\infty}^{\infty} f_0(x_0) p_t(x_0, x_1)\, dx_0 dx_1\,. \tag{5.11}$$

Ou seja, desejamos obter a densidade f_t da variável aleatória X_t. Vamos denotar $f_t(x) = f(t, x)$.

É possível mostrar que $f(t,x)$, onde $f:[0,\infty)\times\mathbb{R}\to\mathbb{R}$, satisfaz a equação da difusão

$$\frac{\partial f(t,x)}{\partial t} = \frac{\sigma}{2}\frac{\partial^2 f(t,x)}{\partial^2 x}, \text{ com a condição inicial } f(0,.) = f_0(.) \tag{5.12}$$

onde σ é a constante positiva associada.

Existe uma analogia da evolução dinâmica da densidade f_t, $t\geq 0$, a partir da condição inicial f_0, com a difusão do calor numa barra $\mathbb{R}$.

Considere uma barra unidimensional de comprimento infinito. Seja $u(t,x)$ a temperatura do ponto x da barra no instante t. Sabemos que u deve satisfazer a equação do calor $u_t = u_{xx}$, que é a equação de difusão para $D=1$.

Mais precisamente, se a temperatura no tempo $t=0$ for dada em cada ponto x da barra $(-\infty,\infty)$ por $u_0(x)$, onde $u_0:\mathbb{R}\to[0,\infty)$, então, a equação do calor $u_t = u_{xx}$ (que tem solução denotada por $u_t(.)$) é tal que $u(t,x)=u_t(x)$ é a temperatura no tempo t no ponto x.

Dada a condição de contorno a delta Dirac em 0 (temperatura super concentrada em 0) obtemos que, fixado t, a solução $u(t,.)$ tem distribuição normal com média 0 e variância t.

A condição inicial f_0 (respectivamente a densidade f_t) e os valores da temperatura no tempo $t=0$ dados por $u_0(x)$ (respectivamente os valores de temperatura $u_t(x)$ no tempo t) desempenham um papel semelhante.

Definição 5.55. *Considere uma sequencia crescente de sigma-algebras $\mathcal{F}_t$, $t\geq 0$, sobre Ω, isto é*

$$\mathcal{F}_t \subset \mathcal{F}_s,$$

para $0\leq t\leq s$.

Dado um processo estocástico X_t, $t\geq 0$, tomando valores em $\mathbb{R}$, dizemos que é uma martingale com respeito a sequencia $\mathcal{F}_t$, $t\geq 0$ se

(a) X_t é $\mathcal{F}_t$ mensurável para qualquer $t\geq 0$,

(b) $E[\,X_{t+s}\,|\,\mathcal{F}_t\,] = X_t$, para qualquer $s,t\geq 0$

(c) $E[X_t^+] < \infty$ *(onde para qualquer* $t \geq 0$*, temos que* $X_t^+(w) = \sup\{X_t(w), 0\}$*).*

Dados dois processos $Y_t : \Omega \to \mathbb{R}$, $t \geq 0$, $X_t : \Omega \to \mathbb{R}$, $t \geq 0$, tomando valores em $\mathbb{R}$, considere $\mathcal{G}_t$ a sigma algebra gerada pelas funções $X_s : \Omega \to \mathbb{R}$, $s \leq t$.

Se $Y_t : \Omega \to \mathbb{R}$, $t \geq 0$ for uma martingale em relação a $\mathcal{G}_t$ descrita acima, $t \geq 0$, diremos que $Y_t : \Omega \to \mathbb{R}$, $t \geq 0$, é uma martingale em relação a $X_t : \Omega \to \mathbb{R}$, $t \geq 0$.

Agora vamos usar martingales para descobrir resultados interessantes sobre o movimento Browniano.

Lema 5.9. *Os processos* $U(t) = X^2(t) - t$ *e* $V(t) = e^{\lambda X(t) - \frac{1}{2}\lambda^2 t}$*, onde* $\lambda \in \mathbb{R}$ *é uma constante arbitrária, são martingales com respeito ao movimento Browniano* X_t*.*

De fato, para $0 \leq t_1 \leq \ldots \leq t_n = t$ e $s > 0$, temos que

$$E[U(t+s)|X(t_1), \ldots, X(t_n)] = E[X^2(t+s)|X(t)] - (t+s) =$$

$$E[(X(t+s) - X(t))^2\, |X(t)] + E[2X(t+s)X(t) - X^2(t)|X(t)] - (t+s) =$$

$$E[(X(t+s) - X(t))^2\, |X(t)] + 2E[X(t)(X(t+s) - X(t))|X(t)] + E[X^2(t)|X(t)]$$

$$-(t+s) = \quad s + 2 \times 0 + X^2(t) - (t+s) = U(t).$$

Segue que U é martingale com respeito a X_t. Similarmente,

$$E[V(t+s)|X(t_1), \ldots, X(t_n)] = V(t) E\left[e^{\lambda[X(t+s) - X(t)] - \frac{\lambda^2 s}{2}}\right] = V(t).$$

Definição 5.56. *Considere uma sequencia crescente de sigma-algebras* $\mathcal{F}_t$*,* $t \geq 0$*, sobre* Ω*, isto é*

$$\mathcal{F}_t \subset \mathcal{F}_s,$$

para $0 \leq t \leq s$.

Dado um processo estocástico X_t, $t \geq 0$, *tomando valores em* $\mathbb{R}$, *uma variável aletória* $T : \Omega \to [0, \infty]$ *é chamada de um tempo de parada para* X_t, $t \geq 0$, *se o evento* $\{T \leq t\}$ *está em* $\mathcal{F}_t$, *para todo* $t \geq 0$.

Definição 5.57. *Dado um processo estocástico* $X_t : \Omega \to \mathbb{R}$, *(uma probabilidade* P *sobre* Ω*) seja* T *um tempo de parada para* X_t, $t \geq 0$. *O processo estocástico* X_T *é naturalmente definido através das funções* $(X_T)_t$, $t \in [0, \infty)$*:*

$(X_T)_t\,(w) = X_t(w)$, *para* $t < T(w)$, *e* $(X_T)_t(w) = X_{T(w)}$ *para* $t \geq T(w)$.

Desta forma se obtém de forma semelhante ao caso em que o tempo é discreto uma medida P_T *sobre* Ω.

Muitas propriedades de tempo de parada quando o Processo Estocástico é a tempo discreto valem da forma análoga para o caso em que $t \in [0, \infty)$ (ver [KT1]).

Lema 5.10. *Se* S, T *são tempos de parada então* $S + T$ *e* $S \wedge T = \min\{S, T\}$ *são tempos de parada também.*

Lema 5.11. *Se* X_t, $t \geq 0$, *é martingale e* T *tempo de parada, então para todo* $t \geq 0$ *vale* $E[X_t] = E[X_0] = E[X_{T \wedge t}]$.

Exemplo 5.39. Seja X_t, $t \geq 0$, o movimento Browniano unidimensional.

Sejam $a < 0 < b$ constantes dadas e $T = T_{ab}$ o primeiro instante no qual X_t atinge a ou b:

$$T_{ab} = \inf\{t \geq 0 : X(t) = a \text{ ou } X(t) = b\}. \tag{5.13}$$

Este T é um tempo de parada.

Seja $\mu = P\,(X(T_{ab}) = b\,)$. Como $\{X_t\}$ é um martingale, e $T = T_{ab}$ é um tempo de parada, temos que $E(X_T) = E(X(0)) = 0$. Logo,

$$\mu b + (1-\mu)a = 0. \tag{5.14}$$

E portanto

$$\mu = \frac{|a|}{|a|+b}. \tag{5.15}$$

Agora, como $\{U(t) = X^2(t) - t\}$ é martingale e $E[U(T)] = E[U(0)] = X^2(0) - 0 = 0$, temos que

$$0 = E[U(T)] = E[X^2(T) - T]. \tag{5.16}$$

E portanto

$$E[T] = E[X^2(T)] = a^2(1-\mu) + b^2\mu = |a|b. \tag{5.17}$$

Mais detalhes podem ser encontrados em [KT].

$\diamondsuit$

É usual denotar por P_x a probabilidade sobre Ω sujeita a condição inicial $X_0 = x$. Desta forma E_x vai denotar o valor esperado em relação a tal probabilidade P_x.

Seja

$$p_t(x,y) = p(t,x,y) = \frac{e^{-\frac{(y-x)^2}{2\,\sigma^2\,t}}}{\sqrt{2\,\pi\,\sigma^2\,t}}. \tag{5.18}$$

Para $t \geq 0$ vamos definir U_t que vai levar funções em funções, assim $U_t(f) = g$.

Para $t \geq 0$ e $f : \mathbb{R} \to \mathbb{R}$ denote

$$U_t(f)(x) \;=\; g(x) = \; E_x[f\,(X_t)\,] = \int p(t,x,y) f(y) dy.$$

Note que U_t está bem definido para funções f integráveis em relação a densidade $p_t(x,.)$.

Uma propriedade extremamente importante do Movimento Browniano é a propriedade de Markov que é descrita por

$$U_{t+s} = U_t \circ U_s.$$

Ou seja, U_t, $t \geq 0$, define um semigrupo.

Outra maneira de descrever a propriedade acima é

$$U_{t+s}(f)(x) = E_x[f(X_{t+s})] = E_x\,[\,E[f(X_{t+s})\,|\,X_t\,]\,].$$

Ou seja, para o cálculo de $E_x[f(X_{t+s})]$, ao condicionarmos num certo tempo $t > 0$, o processo "esquece todo tempo anterior t", guardando somente a informação X_t (ou seja a posição no tempo t).

Disto se pode deduzir a equação de Chapman-Kolmogorov:

$$p(t+s,x,y) = \int p(t,x,z)\,p(s,z,y)\,dz.$$

Ou seja, condicionamos a passagem de x a y nos valores z em um tempo intermediário.

Podemos nos perguntar se vale alguma propriedade análoga ao que acontece no caso de cadeias de Markov a tempo contínuo com conjunto de estados finitos $S = \{1,2,..,d\}$. Neste caso, o semigrupo era dado por $e^{t\,\mathcal{L}}$, onde $\mathcal{L}$ era uma matriz d por d tipo linha soma zero. É natural pensar que U_t, $t \geq 0$, desempenha um papel análogo a $e^{t\,\mathcal{L}}$, $t \geq 0$. Assim, $\mathcal{L}$ é denominado o gerador infinitesimal do processo Markoviano $X_t, t \geq 0$, onde X_t toma valores em $S = \{1,2,..,d\}$.

Lembre que

$$\lim_{t\to 0} \frac{e^{t\,\mathcal{L}} - I}{t} = \mathcal{L}.$$

Vamos mostrar (mais detalhes no capítulo 15 seção 11 em [KT2]) que se $f : \mathbb{R} \to \mathbb{R}$ é diferenciável de classe C^∞ e integrável para a densidade $p_t(x,.)$,

$t \geq 0$, então

$$\lim_{t \to 0} \frac{(U_t \, - \, I)\,(f)}{t}\,(x) \;=\; \frac{\sigma^2}{2}\, f''(x). \tag{5.19}$$

Como para x, t fixo

$$U_t(f)(x) = \int \frac{e^{-\frac{(y-x)^2}{2\,\sigma^2\,t}}}{\sqrt{2\,\pi\,\sigma^2\,t}}\, f(y)\, dy, \tag{5.20}$$

então para poder usar de forma natural a serie de Taylor de ordem dois façamos a mudança de variável $v \to y(v) = x + v\sqrt{t}$. Neste caso, $dy = \sqrt{t}dv$.

Assim, temos que

$$U_t(f)(x) = \int \frac{e^{-\frac{v^2}{2\,\sigma^2}}}{\sqrt{2\,\pi\,\sigma^2}}\, f(x + v\sqrt{t})\, dv.$$

A fórmula de Taylor de ordem dois de $v \to f(x + v\,\sqrt{t})$ em termos de $v\,\sqrt{t}$ em torno de x é dada por

$$f(x + v\,\sqrt{t}) - f(x) \;=\; f'(x)\sqrt{t}\,v + \frac{1}{2} f''(x) t\, v^2 + R(v\,\sqrt{t}), \tag{5.21}$$

onde

$$\lim_{t \to 0} \frac{R(z\,\sqrt{t})}{t^{3/2}} = 0. \tag{5.22}$$

Agora, usando o fato que

$$\int v\, \frac{e^{-\frac{v^2}{2\,\sigma^2}}}{\sqrt{2\,\pi\,\sigma^2}}\, dv = 0,$$

$$\int \frac{e^{-\frac{v^2}{2\,\sigma^2}}}{\sqrt{2\,\pi\,\sigma^2}}\, dv = 1,$$

$$\int v^2\, \frac{e^{-\frac{v^2}{2\,\sigma^2}}}{\sqrt{2\,\pi\,\sigma^2}}\, dv = \sigma^2,$$

e ainda integrando dos dois lados de (5.21) em relação a v obtemos

$$\frac{1}{t}[\,U_t(f)(x)-f(x)\,] \;=\; \frac{1}{t}\int [f(x+v\sqrt{t})-f(x)]\,\frac{e^{-\frac{v^2}{2\,\sigma^2}}}{\sqrt{2\,\pi\,\sigma^2}}dv =$$

$$\frac{1}{2}\,\sigma^2\,f''(x)+\frac{\int R(z\sqrt{t})\,dv}{t}.$$

Usando (5.22) obtemos que o último termo da expressão acima vai a zero quanto $t\to 0$. Desta forma mostramos que a expressão (5.19) é verdadeira para funções f diferenciáveis.

Assim, se A é o operador linear tal que $A(f)=\frac{\sigma^2}{2}\,f''$, é natural afirmar que A é o gerador infinitesimal do movimento Browniano. Note que A (e assim $e^{t\,A}$) não está definido para funções integráveis, mas apenas para funções duas vezes diferenciáveis. Em geral os semigrupos importantes não estão definidos para todas as funções, mas apenas para uma classe "grande"de funções (ver [KT2]). Observamos que tal $e^{t\,A}$ pode ser extendido para agir em uma classe de funções mais ampla (não só diferenciáveis) via o kernel integral

$$K_t(x,y)=\frac{e^{-\frac{(y-x)^2}{2\,\sigma^2\,t}}}{\sqrt{2\,\pi\,\sigma^2\,t}}$$

conforme expressão (5.20), ou seja,

$$U_t(f)(x)=\int K_t(x,y)\;f(y)\,dy. \tag{5.23}$$

Lembre que $p(t,x,y)=p(t,y,x)$ para todo $t\geq 0$ e $x,y\in\mathbb{R}$. Observamos que encontrar f tal que para qualquer intervalo (a,b) e $t\geq 0$

$$\int_a^b f(x)\,dx=\int_a^b\int_{-\infty}^{\infty}\;f(y)p_t(y,x)\,dydx\;=\int_{-\infty}^{\infty}\int_a^b\;\;f(y)p_t(x,y)\,dxdy \tag{5.24}$$

equivale a resolver $U_t(f) = f$ para todo $t \geq 0$.

Note que

$$\frac{\partial}{\partial t} p(t,x) = \frac{\sigma^2}{2} \frac{\partial^2 p(t,x)}{\partial^2 x} = A(p(t,x)). \tag{5.25}$$

O operador A agindo em funções p define o gerador infinitesimal associado.

No caso de cadeias de Markov em tempo contínuo e estado discreto (obtidos a partir de uma matriz linha soma zero $\mathcal{L}$) sabemos que a evolução de $\pi(t)$ a partir da condição inicial π_0 era dada por

$$\frac{d}{dt} \pi(t) = \pi(t)\, \mathcal{L}, \quad \text{onde } \pi(0) = \pi_0.$$

Esta propriedade é análoga a (5.25). Ou seja, o A faz o papel do antigo $\mathcal{L}$. Aqui o espaço de estados está em um contínuo (não enumerável).

É importante tentar obter os estados estacionários $\pi \in \mathbb{R}^d$ no caso de cadeias de Markov com gerador infinitesimal $\mathcal{L}$.

Lembre que para encontrar π tal que $\pi\, e^{t\,\mathcal{L}} = \pi$ para todo $t \geq 0$, é equivalente a resolver $\pi L = 0$.

De forma análoga, em função da expressão (5.19), se desejamos encontrar uma densidade f tal que $f = U_t(f)$, para todo $t \geq 0$, somos levados a resolver a equação

$$0 = A(f) = \frac{\sigma^2}{2} f''(x).$$

As funções diferenciáveis f que satisfazem $f''(x) = 0$, são necessariamente lineares. Desta forma não vai poder se obter com elas a propriedade $\int f(x)dx = 1$. Sendo assim, não existem funções f de classe C^2 tais que determinam densidade iniciais invariantes para o movimento Browniano. Isto nos dá uma resposta parcial a pergunta sobre a existencia de densidades invariantes: não existem! Esta questão é análoga àquela que enfrentamos na análise do random walk $(1/2, 1/2)$ (que também não possui vetor inicial de probabilidade invariante).

5.8 Processos de Difusão

Agora vamos apresentar uma breve introdução as difusões. Um exposição bastante completa sobre o assunto pode ser encontrada em [RW], [KT2], [Nel] ou [KS].

De forma similar ao movimento Browniano, o espaço de estados é $S = \mathbb{R}^n$. O processo estocástico de difusão $X(t)$, $t \in [0, \infty) \subset \mathbb{R}$, vai descrever a posição de uma partícula em $\mathbb{R}^n$.

Seja Ω o espaço de Skhorohod.

O processo de difusão X_t, $t \geq 0$, vai determinar uma probabilidade P em Ω tal que

$$P\{w : [0, \infty) \to \mathbb{R}^n \text{ contínuas } \} = 1.$$

Como sempre, assumimos que $X(t)$, $t \geq 0$, é tal que os caminhos amostrais $w \in \Omega$ satisfazem $w(t) = X_t(w)$.

Considere uma função $u : \mathbb{R}^n \to \mathbb{R}^n$ (com alguma regularidade) e outra função positiva $\sigma^2 : \mathbb{R}^n \to \mathbb{R}$ (com alguma regularidade). A difusão ficará determinada a partir de tais funções. Um caso particular importante é quando estas funções são constantes.

Considere, para $t \geq 0$, o processo estocástico X_t, $t \geq 0$.

Uma *difusão* é um processo que vai cumprir as seguintes propriedades: a probabilidade P associada ao processo é tal que

$$\lim_{h \to 0} \frac{1}{h}\, \mathbb{E}(X_{t+h} - X_t | X_t = x) = u(x), \text{ e} \tag{5.26}$$

$$\lim_{h \to 0} \frac{1}{h}\, \mathbb{E}(|\,(X_{t+h} - X_t) - u(x)\,|^2 | X_t = x) = \sigma^2(x). \tag{5.27}$$

Se $u = 0$ e $\sigma^2 = \sigma$ (constante) o processo X_t é o processo Browniano.

O valor $u(x)$ é denominado de drift. Ele indica a direção (aleatória) média que os caminhos amostrais vão seguir quando na posição x.

Por sua vez, o valor $\sigma^2(x)$ descreve a maior ou menor dispersão desta direção em torno do ponto x. Assim, $\sigma^2(x)$ é denominado de dispersão infinitesimal em torno de x. Se $\sigma^2(x)$ é pequeno no ponto x os caminhos amostrais tem pouca volatilidade (na média) quando posicionados perto de x.

O movimento Browniano corresponde ao caso em que $u = 0$ e não possui uma densidade inicial invariante. O drift u vai possibilitar em alguns casos a obtenção de uma densidade inicial invariante (ver (5.37)).

Supondo que as funções u e σ tem boa regularidade (por exemplo de classe C^∞) então existe um processo estocástico X_t, $t \geq 0$, tal que satisfaz (5.26) e (5.27) e que com probabilidade 1 os caminhos amostrais $w \in \Omega$ são contínuos (ver [SW]).

Vamos nos restringir a partir de agora ao caso unidimensional (em que $n = 1$).

Com probabilidade um os caminhos $w : [0, +\infty) \to \mathbb{R}$ são contínuos mas não diferenciáveis.

No caso em que $X_0 = x_0 \in \mathbb{R}$ dizemos que a difusão está condicionada a começar em x_0. Neste caso, a probabilidade associada será denotada por P_{x_0} e o correspondente valor esperado por E_{x_0}.

Suponha que X_t, $t \geq 0$, descreva uma difusão com parâmetros $u(x)$ e $\sigma(x)$, onde $x \in \mathbb{R}$.

De forma similar ao caso do movimento Browniano considere para $t \geq 0$ e $f : \mathbb{R} \to \mathbb{R}$

$$f_t = U_t(f)(x) \;=\; g(x) = \; E_x[f\,(X_t)\,] = \int p(t, x, y) f(y) dy. \tag{5.28}$$

Então U_t, $t \geq 0$ define um semigrupo agindo em funções f. Assim, $U_t(f) = g_t$, $t \geq 0$, onde f foi tomada como condição inicial.

Denote para f fixo

$$f^t(x) = v(t, x) = E_x[f(X_t)] = U^t(f)(x).$$

Pode-se mostrar que se f for duas vezes diferenciável vale que

$$\lim_{t\to 0}\frac{U^t(f)(x)-f(x)}{t}=\frac{1}{2}\sigma^2(x)\frac{\partial^2 f}{\partial x^2}(x)+u(x)\frac{\partial f}{\partial x}(x).$$

A expressão a direita determina o gerador infinitesimal $\mathcal{L}$ agindo em funções f.

É possivel também mostrar (ver (5.4) na seção 15 em [KT2]) que v satisfaz a edp

$$\frac{\partial v}{\partial t}(t,x)=\frac{1}{2}\sigma^2(x)\frac{\partial^2 v}{\partial x^2}(t,x)+u(x)\frac{\partial v}{\partial x}(t,x), \tag{5.29}$$

com a condição inicial $v(0,x)=f(x)$.

Esta equação é denominada de **equação backward** de Kolmogorov. Ela descreve a **evolução temporal da função** f **que está definida no espaço de estados** $\mathbb{R}$.

O operador tal que

$$v(x)\to\frac{1}{2}\sigma^2(x)\frac{\partial^2 v}{\partial x^2}(x)+u(x)\frac{\partial v}{\partial x}(x)$$

é denominado de **operador backward** associado a difusão determinada por u e σ. Ele é o gerador infinitesimal do processo estocástico com difusão descrita por σ^2 e drift descrito por u.

Algumas propriedades das difusões são semelhantes às do movimento Browniano. Desejamos encontrar $p(t,x,y)$, onde $t\geq 0$, $x,y\in\mathbb{R}$, (a densidade da probabilidade de estando em x no tempo 0, estar em y no tempo t) que satisfaz

$$P_x(X_t\in(a,b))=\int_a^b p(t,x,y)dy.$$

Além disso, desejamos que

$$P_{X_0=x_0}(X_{t_1}\in(a_1,b_1),X_{t_2}\in(a_2,b_2),X_{t_3}\in(a_3,b_3))$$

$$= \int_{a_3}^{b_3} \int_{a_2}^{b_2} \int_{a_1}^{b_1} p(t_1, x_0, x_1) p(t_2 - t_1, x_1, x_2) p(t_3 - t_2, x_2, x_3)\, dx_1\, dx_2\, dx_3.$$

Uma das utilidades da equação (5.29) é que se pode tentar obter a probabilidade de transição $p(t, x, y)$ resolvendo uma equação diferencial parcial.

De fato, fixe um y e tome como condição inicial $f(z)$ tal que $f(z) = 1$ se $z < y$ e $f(z) = 0$ se $z \geq y$. A função $p(t, x, y)$ quando existe (a existencia requer certas hipóteses técnicas sobre σ e u) satisfaz uma equação diferencial parcial de segunda ordem. Mais precisamente, $p(t, x, y)$ satisfaz a equação

$$\frac{\partial p}{\partial t}(t, x, y) = \frac{1}{2} \sigma^2(x) \frac{\partial^2 p}{\partial x^2}(t, x, y) + u(x) \frac{\partial p}{\partial x}(t, x, y). \tag{5.30}$$

Observe que y está fixo e derivamos em x.

Neste caso, para todo $t \geq 0$ e x fixo, tem-se $\int_{-\infty}^{+\infty} p(t, x, y) dy = 1$.

Assim, a função $p(t, x, y)$ que soluciona tal equação diferencial parcial satisfaz

$$\int_a^b p(t, x, y) dy = P(X_t \in (a, b) \,|\, X_0 = x).$$

Quando $u = 0$ e $\sigma(x)^2 = \sigma^2$ é constante obtemos a equação associada ao movimento Browniano e a solução é

$$\frac{e^{-\frac{(x-y)^2}{2\,\sigma^2\,t}}}{\sqrt{2\,\pi\,\sigma^2\,t}}.$$

Exemplo 5.40. Suponha que $u(x) = -\alpha x$, onde α é constante, e $\sigma^2(x) = \sigma^2$ é constante. Seja $\phi(t, x, y) = \frac{e^{-\frac{(x-y)^2}{2\,t}}}{\sqrt{2\,\pi\,t}}$. A difusão X_t, $t \geq 0$ associada a tais funções é denominado de Ornstein-Uhlenbeck (ver [Nel]).

Pode-se mostrar que

$$p(t, x, y) = \phi\Big(\frac{\sigma^2}{2\,\alpha}\,(1 - e^{-2\,\alpha\,t})\,, x\,e^{-\alpha t}, y\,\Big)$$

satisfaz a equação backward de Kolmogorov

$$\frac{\partial p}{\partial t}(t,x,y) = \frac{1}{2}\sigma^2 \frac{\partial^2 p}{\partial x^2}(t,x,y) - \alpha x \frac{\partial p}{\partial x}(t,x,y), \tag{5.31}$$

e portanto descreve a probabilidade de transição a difusão X_t, $t \geq 0$, associada.

Mais detalhes podem ser encontrados na seção 15 de [KT2].

$\diamondsuit$

Vamos supor agora que a difusão X_t, $t \geq 0$, seja tal que

$$P(X_0 \in (a,\, b)) = \int_a^b \rho_0(x)\, dx$$

Ou seja, $\rho_0 : \mathbb{R} \to \mathbb{R}$ descreve a densidade da variável X_0 (a posição inicial da difusão no tempo 0 é aleatória).

Desejamos saber neste caso que é ρ_t tal que

$$\int_a^b \rho_t(x)\, dx \;=\; P(X_t \in (a,\, b)) = \int \int_a^b \rho_0(x)\, p(t,x,y)\, dy\, dx.$$

Em outras palavras, a densidade da variável X_t, $t \geq 0$, seria dada por

$$\rho_t(y) = \int\, p(t,x,y)\, \rho_0(x)\, dx. \tag{5.32}$$

Note que para y fixo o limite quando $t \to 0$ da probabilidade associada a densidade $p(t,x,y)$ é a delta Dirac $\delta_y(dx)$.

A **equação forward** de Kolmogorov para $\rho(t,y)$ é dada por

$$\frac{\partial \rho}{\partial t}(t,y) = \frac{1}{2}\frac{\partial^2}{\partial y^2}[\sigma^2(y)\, \rho(t,y)] - \frac{\partial}{\partial y}[u(y)\, \rho(t,y)] \tag{5.33}$$

onde $\rho(0,y) = \rho(y)$.

Esta evolução temporal descreve a **evolução temporal de probabilidades** no espaço de estados $\mathbb{R}$.

Usando (5.33) existe uma outra edp associada que permite calcular (de outra forma diferente de (5.30)) a probabilidade de transição $p(t,x,y)$ da difusão com parâmetros u e σ.

Na evolução temporal descrita por (5.32) tome como condição inicial ρ_0 a delta Dirac no ponto $x \in \mathbb{R}$. Então a solução $\rho(t,y) = p(t,x,y)$.

A **equação forward** de Kolmogorov para $p(t,x,y)$ é dada por

$$\frac{\partial p}{\partial t}(t,x,y) = \frac{1}{2}\frac{\partial^2}{\partial y^2}[\sigma^2(y)p(t,x,y)] - \frac{\partial}{\partial y}[u(y)p(t,x,y)]. \tag{5.34}$$

A função $p(t,x,y)$ satisfaz tal equação sujeita a condição inicial ser a delta de Dirac em x.

Observe a diferença entre (5.28) (onde se considera $p(t,x,y)\, f(y)dy$) e (5.32) (onde se considera $p(t,x,y)\, \rho_0(x)\, dx$). Isto justifica a diferença de terminologia **backward** and **forward**.

A partir de (5.33) é natural dizer que o **operador forward** A, associado a u e σ^2, é dado por

$$\rho \to A(\rho) = \frac{1}{2}\frac{\partial^2}{\partial y^2}[\sigma^2(y)\rho(y)] - \frac{\partial}{\partial y}[u(y)\,\rho(y)]. \tag{5.35}$$

A escolha da ρ_0 inicial torna o processo de difusão estacionário se, e somente se,

$$\rho_0(y) = \int \rho_0(x)\, p(t,x,y)\; dx.$$

para todo $t \geq 0$ e para todo y.

Se o drift $u(x)$ for zero e σ^2 constante obtemos o processo Browniano que como vimos não possui estado estacionário. Assim, a equação forward Kolmogorov não tem estado estacionário. No entanto para muitos possiveis $u \neq 0$ a densidade invariante existe.

A função densidade inicial ρ_0 define um estado estacionário se

$$0 = \frac{1}{2}\frac{\partial^2}{\partial x^2}[\sigma^2(x)\rho_0(x)] - \frac{\partial}{\partial x}[u(x)\,\rho_0(x)] \;= A(\rho_0)(x). \tag{5.36}$$

Assim, para encontrar ρ_0 é necessário resolver uma equação diferencial ordinária de segunda ordem.

Suponha que está bem definido $a(y) = e^{-\int_{-\infty}^{y} \frac{2\,u(z)}{\sigma^2(z)}\,d\,z}$ e $B(x) = \int_{-\infty}^{x} a(y)\,dy$. Se ρ for da forma

$$\rho(x) = k_1\,\frac{B(x)}{a(x)\,\sigma^2(x)} + \,k_2\,\frac{1}{a(x)\,\sigma^2(x)}, \tag{5.37}$$

então se pode mostrar que ρ satisfaz $A(\rho) = 0$ (equação (5.36)).

Em várias situações, dependendo de $u(x)$ e $\sigma^2(x)$, é possível encontrar constantes k_1, k_2 tais que tal ρ satisfaz $\rho(x) \geq 0$, para todo x, e ainda que $\int \rho(x)dx \;=\; 1$. Desta forma se pode obter explicitamente a densidade ρ_0 (condição inicial) que deixa a difusão associada X_t, $t \geq 0$, estacionária.

Note que a expressão para $\rho(x)$ em (5.37) pode ser infinito se $a(x)$ ou $\sigma(x)$ for 0 e pode ser 0 se $a(x)$ ou $\sigma(x)$ for ∞.

Observe que se $u(z)$ for a zero rapidamente quando $z \to \infty$ e $z \to -\infty$ então a integral $\int_{-\infty}^{y} \frac{2\,u(z)}{\sigma^2(z)}\,d\,z$ vai convergir e $a(y)$ sera finito. Depois precisa ainda que $B(x)$ seja finito.

Mais detalhes sobre este tópico podem ser encontrados em [KT2] seção 15.

Dependendo do drift u a existência de um estado estacionário se torna possível. Por exemplo, se u for tal que $u(x) < 0$ se $u > 0$, e ainda $u(x) > 0$ se $u < 0$, então ele age no sentido de que, heuristicamente falando, a massa da probabilidade fique confinada e não escape para infinito.

Exemplo 5.41. Suponha que $u(x) = -\alpha x$, onde $\alpha > 0$ é constante, e $\sigma^2(x) = \sigma^2$ é constante. Seja $\phi(t, x, y) \;=\; \frac{e^{-\frac{(x-y)^2}{2\,t}}}{\sqrt{2\,\pi\,t}}$. Assim X_t, $t \geq 0$, é o processo de Ornstein-Uhlenbeck. Um cálculo simples mostra que a função a descrita acima é tal que $a(x) = e^{\frac{\alpha}{\sigma^2}x^2}$. A partir disso se obtém que $\rho_0(x) = \frac{e^{\frac{\alpha}{\sigma^2}x^2}}{\int e^{\frac{\alpha}{\sigma^2}y^2}dy}$ é a densidade inicial invariante que torna o processo X_t, $t \geq 0$, estacionário.

$\diamondsuit$

Suponha que o processo estocástico X_t, $t \geq 0$, seja uma difusão com $\sigma^2(x)$ e o drift $u(x)$.

Seja $U_t : f \mapsto g$, onde para fixado $t > 0$, $U_t(f(x)) = \mathbb{E}(f(X(t))|X(0) = x) = g(x)$.

Então, da mesma forma que no caso do movimento Browniano, temos que U_t é semi-grupo: $U_{t+s} = U_t \circ U_s$. Ou seja, o processo é Markoviano.

Considere uma difusão unidimensional X_t, $t \geq 0$, associada a $u(x)$ e $\sigma^2(x)$.

Suponha que $X_0 = x$.

Fixados $l < r$, tal que $l < x < r$, definimos o tempo de parada T tal que

$$T(w) = \min\{t \text{ tal que } w_t = l \text{ ou } w_t = r\}.$$

Ainda,

$$T_l(w) = \min\{t \text{ tal que } w_t = l\}$$

e

$$T_r(w) = \min\{t \text{ tal que } w_t = r\}$$

são também tempos de parada.

Assuma que T é finito com probabilidade 1.

Denote por $v(x) = P_x(T(r) < T(l))$, $x \in [l, r]$. Este valor nos dá a probabilidade de que o caminho amostral w_t, $t \geq 0$, ao sair do intervalo $[l, r]$, o faça pelo ponto r.

Pode se mostrar (ver seção 15 [KT2]) que $v(x)$ satisfaz a equação diferencial ordinária

$$0 = u(x)\frac{d\,v(x)}{d\,x} + \frac{1}{2}\sigma^2(x)\,\frac{d^2\,v(x)}{d\,x^2}$$

com a condição de fronteira $v(l) = 0$ e $v(r) = 1$.

Desta forma através de problemas do tipo Sturm-Liouville podemos obter informações importantes sobre a difusão X_t, $t \geq 0$.

Exemplo 5.42. Medidas invariantes para difusões como estados de equilíbrio da Mecânica Estatística - Considere fixada uma função $U : \mathbb{R} \to \mathbb{R}$ que desempenha o papel de um potencial. Dada uma densidade $\rho : \mathbb{R} \to \mathbb{R}^+$ (isto é $\int_{-\infty}^{\infty} \rho(x)\, dx = 1$) o valor esperado de U dado ρ é o valor

$$\int U(x)\, \rho(x)\, dx.$$

A entropia da densidade ρ é o valor

$$-\int \rho(x)\, \log \rho(x)\, dx.$$

Denotamos por T a temperatura de um certo sistema que é regido pelas leis da Mecânica Estatística e que está sob a influencia do potencial U. Desejamos descrever uma partícula que esta na reta real $\mathbb{R}$ no equilíbrio termodinâmico.

A energia livre de uma possível densidade ρ que esta sujeita a tal sistema é dada por

$$\int U(x)\, \rho(x)\, dx + \int \rho(x)\, \log \rho(x)\, dx.$$

Um dos princípos básicos da Mecânica Estatística é que o estado de equilíbrio observado no sistema físico sob a influência do potencial U deveria ser a densidade ρ_0 tal que minimiza

$$\int U(x)\, \rho(x)\, dx + \int \rho(x)\, \log \rho(x)\, dx.$$

entre todas as densidades ρ possíveis.

Ou seja, a descrição de uma partícula sob a ação de tal potencial U (no equilíbrio) tem uma característica aleatória. Podemos apenas falar na probabilidade de encontrá-la num certo intervalo $(a,b) \subset \mathbb{R}$. Esta probabilidade será dada por $\int_a^b \rho_0(x)\,dx$.

Sendo assim o equilíbrio ρ_0 satisfaz um princípio variacional (minimiza energia livre entre as outras possíveis densidades ρ).

Sabe-se que a solução ρ_0 (ver [Gar], [Ris], [You] ou [JKO]) é da forma

$$\rho_0(x) = \frac{1}{Z}\, e^{-\frac{1}{T}\,U(x)}$$

onde

$$Z = \int e^{-\frac{1}{T}\,U(x)}.$$

Naturalmente condições sobre U são necessárias (por exemplo Z ser finito) mas não vamos elaborar sobre isto.

Note que os valores de x tais que $V(x) > 0$ é grande determinam regiões de menor probabilidade de encontrar a partícula.

Pode se mostrar (ver [Gar] ou [Ris]) que tal ρ_0 é uma solução estacionária para a equação diferencial

$$\frac{\partial\,\rho}{d\,t} \;=\; -\frac{\partial}{\partial\,x}\,(\,\frac{d\,U(x)}{d\,x}\,\rho(x)\,)\,+\,\frac{\partial^2\;\rho(x)}{\partial\,x^2} =$$
$$-\frac{d^2\,U(x)}{d\,x^2}\,\rho(x) - \;\frac{d\,U(x)}{d\,x}\frac{\partial\,\rho(x)}{\partial\,x}\;+\;\frac{\partial^2\;\rho(x)}{\partial\,x^2}$$

para $\rho(x,t)$.

Ou seja,

$$0 \;=\; -\frac{d}{d\,x}\,(\,\frac{d\,U(x)}{d\,x}\,\rho_0(x)\,)\,+\,\frac{d^2\;\rho_0(x)}{d\,x^2}.$$

A equação acima é da forma da expressão $A(\rho_0) = 0$ em (5.35) onde $\rho^2 = 1$, $T = 2$ e $u(x) = -\frac{d\,U(x)}{d\,x}$.

Obtemos assim uma relação interessante entre estados de equilíbrio para processos de difusão e problemas importantes em Mecânica Estatística.

$\diamondsuit$

Suponha que X_t, $t \geq 0$, seja um processo de difusão associado aos parâmetros $\mu(x)$ e $\sigma^2(x)$, onde $x \in \mathbb{R}$.

Considere uma função contínua não negativa $A : [a, b] \to \mathbb{R}$ e também uma outra função contínua $f : [a, b] \to \mathbb{R}$.

A seguir definimos para $x \in \mathbb{R}$ e $t \geq 0$.

$$w(x,t) = E_x[e^{-\int_0^t A(X(s))\,ds} f(X(t))].$$

Teorema 5.42. *Feynman-Kac - Sob as hipóteses acima w satisfaz a edp*

$$\frac{\partial w}{\partial t}(x,t) = -A(x)\,w(x,t) + \mu(x)\frac{\partial w}{\partial x}\,(x,t) + \frac{1}{2}\sigma^2(x)\,\frac{\partial^2 w}{\partial x^2}\,(x,t),$$

com a condição inicial $w(x, 0) = f(x)$.

Para uma prova ver [KT2] ou [PDM].

A função $w(x,t)$ é obtida via integral de caminhos (da difusão X_t, $t \geq 0$) e "lembra"vagamente a integral de Feynman (que requer o uso do numero complexo i). Por outro lado a equação acima "lembra"vagamente a equação de Schrodinger (que requer o uso do número complexo i).

Bibliografia

[Au] K. Aurani, Boltzmann e a matematização da Física no Seculo XIX, Educação Matemática - on line.

[B] P. Billingsley, Probability and Measure, Wiley, 1995.

[Ba] R. Bartle, The Elements of Integration, Wiley, 1966.

[Ba2] R. Bartle, The Elements of Real Analysis, Wiley, 1964.

[CD] B. Clarke and R. Disney, Probabilidade e Processos Estocásticos, Livros Técnicos e Científicos, 1979.

[CM] D.R. Cox and H. D. Miller, The Theory of Stochastics Processes (1965)

[DZ] A. Dembo and O. Zeitouni, Large Deviations Techniques and Applications, Springer Verlag, 1998.

[DL] C. Doering e A. O. Lopes, Equações Diferenciais Ordinárias, Mat. Univ. IMPA, 2012.

[Du] R. Durret, Probability: Theory and Examples, Duxbury Press, 1995.

[EK] S. Ethier and T. Kurtz, Markov Processes, John Wiley, 1986.

[Fe] P. Fernandez, Teoria da Medida, Projeto Euclides, IMPA, 1976.

[F] P. Fernandez, Introdução aos Processos Estocásticos, Coloq. Bras. de Matemática, 1975.

[Gar] C.W. Gardiner, Handbook of stochastic methods, 2nd Ed. (Springer, Berlin, 1985).

[GP] V. Guilhemin and A. Pollack, Differential Topology, AMS Chelsea Publishing, 2010.

[GS] G. Grimmett and D. Stirzaker, Probability and Random Processes, Oxford Press, 1994.

[Gu] P. Guttorp, Stochastic modeling of scientific data, Chapmann and Hall, 1995.

[HPS] P. Hoel, S. Port and G. Stone, Introduction to Stochastic Processes, Mifflin Ed, 1972.

[I] D. Isaacson, Markov Chains, Theory and Applications Wiley, 1976.

[JKO] R. Jordan, D. Kinderlehrer and F. Otto, Free energy and the Fokker-Planck equation. Phys. D 107 (1997), no. 2-4, 265–271.

[KS] I. Karatzas and S. Shreve, Brownian Motion and Stochastic Calculus, Springer Verlag

[KT] S. Karlin and H. Taylor, A First Course in Stochastic Processes, Academic Press, 1975.

[KT2] S. Karlin and H. Taylor, A Second Course in Stochastic Processes, Academic Press, 1975.

[KSK] J. G. Kemeny, J. L. Snell and A. W. Knapp, Denumerable Markov Chains, second edition, Springer Verlag

[Kun] J. Kuntz Markov chains revisited, arXiv (2020)

[Lal] S. Lalley, Measure-Theoretic Probability I - Lecture Notes (2017)

[Lal1] S. Lalley, Continuous time Markov Chains - Lecture Notes

[Li1] Elon Lima, Algebra Linear, 1980.

[Li2] Elon Lima, Curso de Análise, Vol II, Projeto Euclides, 1981.

[Li3] Elon Lima, Espaços Métricos, Projeto Euclides, 1981.

[Lop] A. O. Lopes, An Introduction to Coupling, "Modeling, Dynamics, Optimization and Bioeconomics II"Editors: Alberto Pinto and David Zilberman, Springer Proceedings in Mathematics and Statistics, pp 307-335, (2017) Springer Verlag

[PDM] Pierre Del Moral, Feynman-Kac Formulae, Springer Verlag (2004)

[Nel] E. Nelson, Dynamical theories of Brownian motion. Princeton, N.J.: Princeton University Press (1967)

[N] J. Norris, Markov Chains, Cambridge Press, (1997).

[NP] N. Privault, Understanding Markov chains, Springer Verlag

[PY] M. Pollicott and M. Yuri, Dynamical systems and Ergodic Theory, Cambrige Press, 1998.

[Ro] J. Rosenthal, A first look at rigorous Probability Theory, World Scientific Publishing, 2006.

[RW] L. Rogers and D. Williams, Diffusions, Markov Processes and Martingales: Volume 1 an 2, Cambridge Press

[Ris] H. Risken, The Fokker-Planck Equation: Methods of Solution and Applications, 2nd ed. (Springer, Berlin, 1989).

[Ru] W. Rudin, Real and Complex Analysis, McGraw Hill, 1970.

[S] D. W. Stroock, An introduction to Markov Processes, Springer Verlag, 2005

[You] L. C. Young, Lectures on the Calculus of Variations and Optimal Control Theory

Índice Remissivo