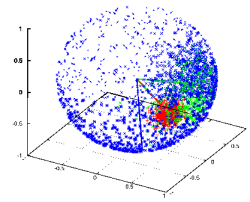
04

Mat02274
Estatística Computacional

Prof. Lorí Viali, Dr.
viali@mat.ufrgs.br
http://www.ufrgs.br/~viali/

Métodos de Geração de Variáveis Aleatórias



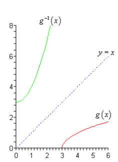
Existem algumas técnicas para a geração de variáveis aleatórias. O tipo de algoritmo a ser utilizado depende da distribuição que se quer gerar.

Contudo, quase todas as técnicas podem ser classificadas conforme suas bases teóricas.

As principais abordagens são:

(i) Inversão;
(ii) Composição;
(iii) Convolução;
(iv) Aceitação e Rejeição;
(v) Propriedades Especiais.

Método da Transformada Inversa



Variáveis Contínuas

Suponha que se queira gerar uma variável contínua X com FDA F(x). Vamos admitir que a inversa existe e que será representada por $F^{-1}(x)$. Então um algoritmo para gerar valores da VAC X com fdp f(x) é:

1. Gerar u ≈ U(0, 1);

2. Fazer x = $F^{-1}(u)$.

Note-se que $F^{-1}(u)$ vai sempre estar definida, pois $0 \leq u \leq 1$ e a imagem da F é o intervalo [0, 1].

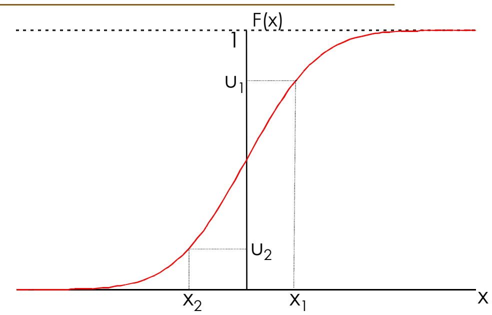


Ilustração do Método da Inversão

Para mostrar que o valor de X obtido com o algoritmo anterior, denominado de *método da transformada inversa*, tem a desejada distribuição F, é preciso mostrar que para qualquer número real x, $P(X \leq x) = F(x)$.

Mas, como F é inversível, tem-se:

$$P(X \leq x) = P[F^{-1}(U) \leq x] = P[F(F^{-1}(U)) \leq F(x)]$$
$$= P[U \leq F(x)] = P[0 \leq U \leq F(x)] =$$
$$= F(x) - 0 = F(x).$$

Onde a última igualdade segue do fato de que U ≈ U(0, 1) e $0 \leq F(x) \leq 1$.

Exemplo

Seja X uma VAC com uma distribuição W(α, β). Assim a fdp de X é:

$$f(x) = \begin{cases} \alpha\beta^{-\alpha}x^{\alpha-1}e^{-(x/\beta)^{\alpha}} & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases}$$

A FDA de X é:

$$F(x) = \int_0^x f(u)du = \begin{cases} 1-e^{-(x/\beta)^{\alpha}} & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases}$$

Exemplo

Resolvendo U = F(x) para X, tem-se:

$$U = 1 - e^{-(x/\beta)^{\alpha}}$$
$$e^{-(x/\beta)^{\alpha}} = 1 - U$$
$$-(x/\beta)^{\alpha} = \ln(1-U)$$
$$X = \beta[-\ln(1-U)]^{1/\alpha}$$

Uma vez que 1 – U ~U(0, 1) pode-se escrever: $X = \beta[-\ln(U)]^{1/\alpha}$

Exercício

Gere 10000 valores de uma W(2, 3).

Represente graficamente a distribuição e o modelo. Compare os parâmetros do modelo e estime os seus valores com os dados obtidos, determinando as seguintes medidas: média, desvio, mediana, assimetria e curtose.

Variáveis Discretas

Seja X uma VAD onde $p(x_i)$ é a função de probabilidade, isto é, $p(x_i) = P(X = x_i)$.

Vamos supor que X possa assumir os valores $x_1, x_2, ..., x_n, ...$, onde $x_1 < x_2 < ... < x_n < ....$ Então o algoritmo é:

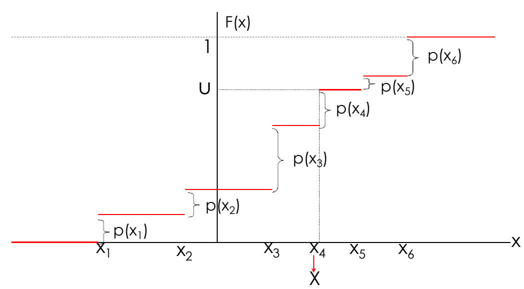
Prof. Lori Viali, Dr. – UFRGS – Instituto de Matemática - Departamento de Estatística

1. Gerar $U \approx U(0, 1)$;

2. Determinar o menor inteiro positivo I tal que $U \leq F(x_I)$ e retornar $X = x_I$.

Para verificar que o método da transformada inversa discreta é válido, devemos mostrar que $P(X = x_i) = p_i$ para todo i.

Prof. Lori Viali, Dr. – UFRGS – Instituto de Matemática - Departamento de Estatística



Método da Inversão para variáveis discretas

Prof. Lori Viali, Dr. – UFRGS – Instituto de Matemática - Departamento de Estatística

Para i = 1, teremos $X = x_1$ se e só se $U \leq F(x_1) = p(x_1)$, uma vez que os valores x_i estão em ordem crescente. Como $U \approx U(0, 1)$, $P(X = x_1) = p(x_1)$ como o requerido. Para $i \geq 2$, o algoritmo coloca $X = x_i$ se e só se $F(x_{i-1}) < U \leq F(x_i)$, já que o i determinado pelo algoritmo é o menor inteiro positivo tal que $U \leq F(x_i)$.

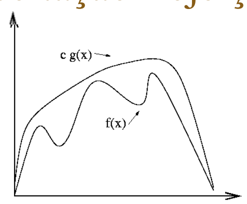
Prof. Lori Viali, Dr. – UFRGS – Instituto de Matemática - Departamento de Estatística

Exercício

Seja X uma VAD assumindo os valores: 1, 2, ..., 10 com probabilidades 1/10 para x = 1, 2, ..., 10. Gerar 5000 valores dessa distribuição. Representar graficamente e determinar: média, desvio, mediana, assimetria e curtose.

Prof. Lori Viali, Dr. – UFRGS – Instituto de Matemática - Departamento de Estatística

Método da Aceitação Rejeição



Suponha que desejamos uma amostra de uma VAC com fdp f(x) e que isso não possa ser feito pelo método da Inversão.

Suponha que sejam válidas as seguintes hipóteses:

---

1. Existe uma função r(x) que domina f(x), isto é, $r(x) \geq f(x)$ para todo x.

2. É possível gerar pontos uniformente espalhadas sob o gráfico da r(x), acima do eixo x. Representa-se as coordenadas de um desses pontos por (X, Y).

---

3. Se o gráfico da r(x) é esboçado no mesmo diagrama, os pontos (X, Y) estarão acima ou abaixo dele de acordo com $Y > f(X)$ ou $Y \leq f(X)$.

4. Se a função r(x) não for uma fdp então fazer $g(x) = r(x)/c$, onde c é a área da r(x).

O Algoritmo

1. Gerar Y tendo uma densidade g(x).

2. Gerar U(0, 1) (independente de Y em um).

3. Se $U \leq f(Y)/r(Y)$, retorna X = Y e pare; senão volte para o passo um e tente novamente (repita até que uma aceitação aconteça no passo 3).

Exemplo

Seja B(4, 3), isto é, com fdp dada por $f(x) = 60x^3(1 - x)^2$ se $0 \leq x \leq 1$ e 0 cc.

O topo da densidade é f(0,6) = 2,0736. Vamos fazer r(x) = 2,0736 se $0 \leq x \leq 1$. Assim c = 2,0736 e g(x) = r(x)/c e, portanto, g(x) é uma U(0,1).

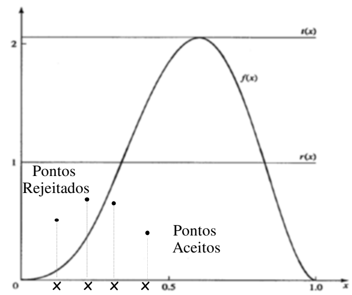


Ilustração do Método da Aceitação-Rejeição

O Algoritmo

1. Gerar Y ~ U(0, 1).

2. Gerar U ~ U(0, 1) (independente de Y).

3. Se $U \leq 60Y^3(1 - Y)^2/2{,}0736$, retorna X = Y e para, senão volta ao passo 1.

Prof. Lori Viali, Dr. – UFRGS – Instituto de Matemática - Departamento de Estatística

Exercício

Gerar 10000 valores da variável aleatória abaixo, utilizando o método da Aceitação/Rejeição.

$$f(x) = \begin{cases} 4x\,e^{-2x} & \text{se } x \geq 0 \\ 0 & \text{c.c.} \end{cases}$$

Prof. Lori Viali, Dr. – UFRGS – Instituto de Matemática - Departamento de Estatística

Solução

Para obter uma função r(x) simples para simular valores de f(x) vamos desconsiderar valores acima de x = 5, pois se x < 5, F(x) = 0,9995.

$$f(x) = \begin{cases} 4x\,e^{-2x} & \text{se } x \geq 0 \\ 0 & \text{c.c.} \end{cases}$$

Prof. Lori Viali, Dr. – UFRGS – Instituto de Matemática - Departamento de Estatística

Então uma função r(x) poderá ser:

$$r(x) = \begin{cases} 1 - x/5 & \text{se } 0 \leq x \leq 5 \\ 0 & \text{c.c.} \end{cases}$$

A função g(x) será obtida integrando a r(x) no intervalo considerado.

Prof. Lori Viali, Dr. – UFRGS – Instituto de Matemática - Departamento de Estatística

Então uma função r(x) poderá ser:

$$r(x) = \begin{cases} 1 - x/5 & \text{se } 0 \leq x \leq 5 \\ 0 & \text{c.c.} \end{cases}$$

A função g(x) será obtida integrando a r(x) no intervalo considerado.

Prof. Lori Viali, Dr. – UFRGS – Instituto de Matemática - Departamento de Estatística

Assim a função g(x) será:

$$g(x) = \begin{cases} \frac{2}{5}(1 - x/5) & \text{se } 0 \le x \le 5 \\ 0 & \text{c.c.} \end{cases}$$

A função g(x) é obtida integrando a r(x) no intervalo considerado e dividindo r(x) pela área obtida.

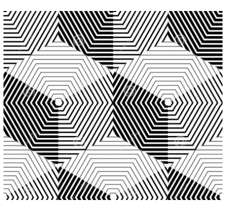
Então a G(x) será:

$$G(x) = \begin{cases} 0 & \text{se } x < 0 \\ (10x - x^2)/25 & \text{se } 0 \le x \le 5 \\ 1 & \text{se } x > 5 \end{cases}$$

Igualando a expressão de G(x) a U e isolando X, obtém-se: $X = 5(1 - \sqrt{U})$

que é o gerador da variável com fdp g(x).

Método da Composição



O método da composição pode ser aplicado quando a FDA da qual precisamos gerar valores pode ser expressa como uma combinação convexa de outras FDAs F_1, F_2, ... Com isso, espera-se poder determinar valores das F_is de uma forma mais simples do que da F original.

Especificamente, assume-se que para todo x a F(x) pode ser escrita como:

$$F(x) = \sum_{i=1}^{\infty} p_i F_i(x)$$

onde os pesos, p_i, satisfazem $p_i > 0$ e $\Sigma p_i = 1$ e cada F_i é uma FDA.

De forma equivalente se X tem densidade f então:

$$f(x) = \sum_{i=1}^{\infty} p_i f_i(x)$$

onde as f_i são outras densidades.

O caso discreto é análogo.

O Algoritmo

O algoritmo para o método da composição é, então:

(i) Gerar um número inteiro aleatório I tal que: $P(I = i) = p_i$ para i = 1, 2, ...

(ii) Retornar X com FDA F_I.

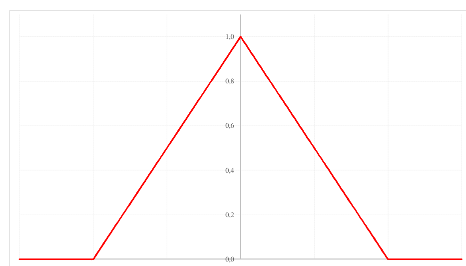
O problema é encontrar F_i que possibilite uma geração rápida e fácil.

Algumas vezes a geometria da distribuição pode dar uma ideia dessa decomposição.

Exemplo 1

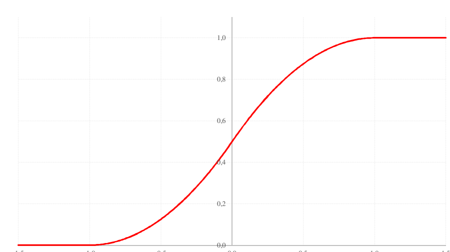
Considere a distribuição triangular simétrica em [-1, 1].

$$f(x) = \begin{cases} x+1 & \text{se } -1 \le x \le 0 \\ -x+1 & \text{se } 0 < x \le 1 \\ 0 & \text{c.c.} \end{cases}$$



A distribuição acumulada é:

$$F(x) = \begin{cases} 0 & \text{se } x < -1 \\ \dfrac{x^2}{2} + x + \dfrac{1}{2} & \text{se } -1 \le x \le 0 \\ -\dfrac{x^2}{2} + x + \dfrac{1}{2} & \text{se } 0 < x \le 1 \\ 1 & \text{se } x > 1 \end{cases}$$



A transformação inversa será feita por:

$$U = F(x) = \begin{cases} \dfrac{x^2}{2} + x + \dfrac{1}{2} & \text{se } U \le 1/2 \\ -\dfrac{x^2}{2} + x + \dfrac{1}{2} & \text{se } U \ge 1/2 \end{cases}$$

Então:

$$X = \begin{cases} \sqrt{2U} - 1 & \text{se } U \le 1/2 \\ 1 - \sqrt{2(1-U)} & \text{se } U > 1/2 \end{cases}$$

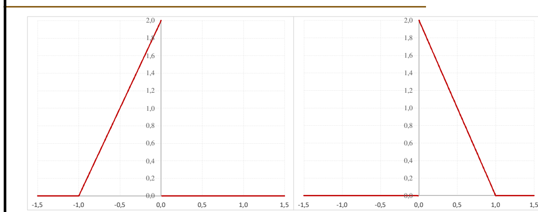
A Composição

Definir a função indicadora para o conjunto A como:

$$I_A(x) = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{se } x \notin A \end{cases}$$

Assim:

$$f(x) = (x+1)\,I_{[-1,0]}(x) + (-x+1)\,I_{[0,1]}(x) =$$
$$= 0{,}5[2(x+1)\,I_{[-1,0]}(x)] + 0{,}5[2(-x+1)\,I_{[0,1]}(x)] =$$
$$= p_1 f_1(x) + p_2 f_2(x)$$



$F_1(x) = x^2 + 2x + 1$

$F_1^{-1}(U) = \sqrt{U} - 1$

$F_2(x) = -x^2 + 2x$

$F_2^{-1}(U) = 1 - \sqrt{1-U}$

Obs. Agora não é necessário somar ½ na segunda função, pois a $f_2(x)$ é agora uma fdp e assim $F_2(x)$ é uma FDA sem a necessidade de somarmos mais 0,5.

O Algoritmo

O algoritmo da composição será:

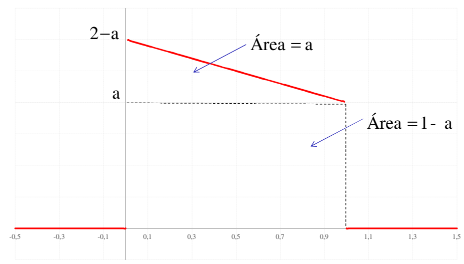
1. Gerar U_1 e U_2 ~ U(0, 1) de forma independente;
2. Se U_1 < ½, retorna $X = \sqrt{U_2} - 1$

   senão, retorna $X = 1 - \sqrt{1 - U_2}$

Exercício

Considere a distribuição trapezoidal em [0; 1] com parâmetro **a** (0 < a < 1). Expresse a distribuição como uma soma de duas densidades mais simples.

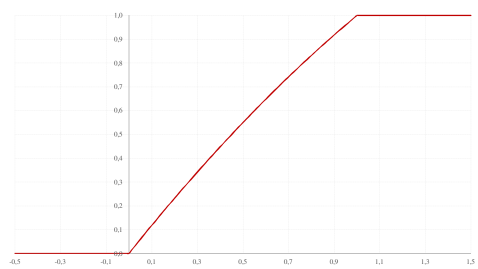
$$f(x) = \begin{cases} 2 - a - 2(1-a)x & \text{se } 0 \le x \le 1 \\ 0 & \text{c.c.} \end{cases}$$



A distribuição acumulada é:

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ (2-a)x - (1-a)x^2 & \text{se } 0 \le x \le 1 \\ 1 & \text{se } x > 1 \end{cases}$$

A representação é:



A transformação inversa é dada por:

$U = F(X) = (2 - a)X - (1 - a)X^2$.

Então:

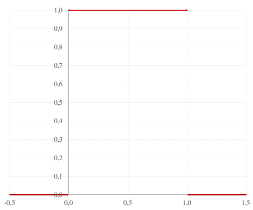
$$X = \frac{2-a}{2(1-a)} - \sqrt{\frac{(a-2)^2}{4(1-a)^2} - \frac{U}{1-a}}$$

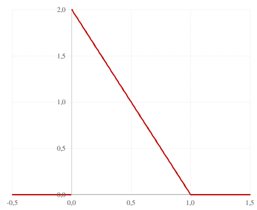
A composição

Definir a função indicadora para o conjunto A como:

$$f(x) = a[I_{[-1,0]}(x)] + (1-a)[2(1-x)I_{[0,1]}(x)] =$$
$$= p_1 f_1(x) + p_2 f_2(x)$$

onde $p_1 = a$ e $f_1(x) = I_{[-1,0]}(x)$, $p_2 = 1 - a$ e $f_2(x) = 2(1 - x)I_{[0,\ 1]}(x)$





$F_1(x) = x$

$F_1^{-1}(U) = U$

$F_1(x) = -x^2 + 2x$

$F_1^{-1}(U) = 1 - \sqrt{1-U}$

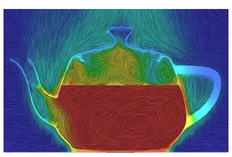
O Algoritmo

O algoritmo da composição:

1. Gerar U_1 e $U_2 \sim U(0, 1)$ de forma independente.
2. Se $U_1 < a$, retorna $X = U_2$ senão, retorna $X = 1 - \sqrt{1 - U_2}$.

Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

Método da Convolução



Suponha que a VA desejada tem a mesma distribuição que $Y_1 + Y_2 + ... + Y_n$, onde as Y_i são IID.

$X \sim Y_1 + Y_2 + ... + Y_n$, é denominado uma convolução das Y_i.

Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

Resumindo tem-se:

Composição: a função de distribuição (a fp ou fdp) é expressa como uma soma (ponderada) de outras funções de distribuição (a fd ou fdp).

Convolução: expressa a própria variável como a soma de outras variáveis.

Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

O Algoritmo

01. Gerar $Y_1 + Y_2 + ... + Y_n$, independente da sua distribuição.
02. Retornar $X \sim Y_1 + Y_2 + ... + Y_n$.

Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

Exemplo 1

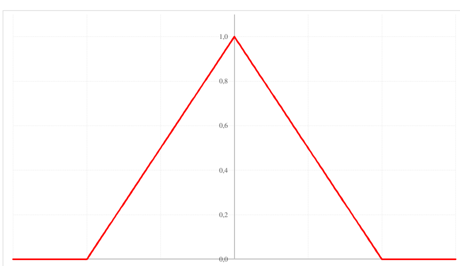
Se X é uma Erlang de parâmetros r inteiro e λ, isto é, $X \sim E(r, \lambda)$.

Expresse $X \sim Y_1 + Y_2 + \dots + Y_r$ onde as Y_i são variáveis IDD exponenciais com média λ. Faça $r = 5$ e $\lambda = 2$.

Exemplo 2

Considere a distribuição triangular simétrica em [-1, 1]. A densidade é:

$$f(x) = \begin{cases} x + 1 & \text{se } -1 \le x \le 0 \\ -x + 1 & \text{se } 0 \le x \le 1 \\ 0 & \text{c.c.} \end{cases}$$



Por uma probabilidade condicional: se U_1 e U_2 são IID U(0, 1), então $U_1 + U_2 \sim$ é triangular simétrica em [0, 2], basta então deslocar por 1:

$X = U_1 + U_2 - 1 = U_1 - 0{,}5 + U_2 - 0{,}5 = Y_1 + Y_2$.

Referências

CARLO, David. Random Number Generation: Types and Techniques, 2012.

FISHMAN, George S. ***Monte Carlo: Concepts, Algortihms, and Applications.*** New York (NY): Springer, 1996.

KNUTH, Donald E. ***The Art of Computer Programing. Volume 2 - Seminumerical Algorithms.*** Reading (Massachusetts): Addison Wesley, 1981.

LEWIS, P. A. W., ORAV, E. J. ***Simulation Methodology for Statisticians, Operations Analysts and Engineers. Volume I.*** Belmont (California): Wadsworth, Inc., 1989.

MADRAS, Neal. ***Lectures on Monte Carlo Methods.*** Providence (RI): American Mathematical Society, 2002.