Decision Theory and Large Deviations for Dynamical Hypotheses Test: Neyman-Pearson, Min-Max and Bayesian Tests

Hermes H. Ferreira, Artur O. Lopes[‡] and Sílvia R.C. Lopes

Mathematics and Statistics Institute
Federal University of Rio Grande do Sul
Porto Alegre, RS, Brazil

January 20, 2021

Abstract

We analyze hypotheses tests via classical results on large deviations for the case of two different Hölder Gibbs probabilities. The main difference for the the classical hypotheses tests in Decision Theory is that here the two considered measures are singular with respect to each other. We analyze the classical Neyman-Pearson test showing its optimality. This test becomes exponentially better when compared to other alternative tests, with the sample size going to infinity. We also consider both, the Min-Max and a certain type of Bayesian hypotheses tests. We shall consider these tests in the log likelihood framework by using several tools of Thermodynamic Formalism. Versions of the Stein's Lemma and the Chernoff's information are also presented.

Keywords: Decision Theory, Large Deviations Properties, Rejection Region, Neyman-Pearson Hypotheses Test, Min-Max Hypotheses Test, Bayesian Hypotheses Test, Thermodynamic Formalism, Gibbs Probabilities.

2010 Mathematics Subject Classification: 62C20, 62C10, 37D35.

1 Introduction

The problem we are interested in here can be simply expressed as the following: there are two measures μ_0 and μ_1 that we know in advance what they are. A data set is obtained by sampling but we do not know, in advance, if it was originated from μ_0 or μ_1. Suppose it comes from μ_1. From this data set, we need to decide on which one of the two generated this sampling data. A hypotheses test is a method that helps us to make such a choice. Taking large samples from the random process we will be able to make the right decision, that is, to choose the alternative μ_1. Classical results on Large Deviations properties can estimate the risk of a wrong decision. In the Bayesian point of view, we should attach to μ_0 a probability π_0 and to μ_1 a probability π_1, where $\pi_0 + \pi_1 = 1$.

We shall extend the reasoning described on page 91 of section VI in [9], where the author considers LDP properties. However, we point out that in [9] there is no dynamics involved in the process.

[‡]Corresponding author. E-mail: arturoscar.lopes@gmail.com

We are interested in probabilities on the symbolic space $\Omega = \{1, 2, ..., d\}^{\mathbb{N}}$. The shift transformation σ is given by $\sigma(b_0, b_1, b_2, b_3, ..., b_n, ...) = (b_1, b_2, b_3, ..., b_n, ...)$.

A nice reference for Thermodynamic Formalism is [29] (see also [13] and [27]). For results on Large Deviations for Thermodynamic Formalism we refer the reader to [22], [25], [26] and [27]. Important references for basic results in Hypotheses tests are Sections 3.4 and 3.5 in [15], [1], [30], [9], [6], [5], [21], [11], [12], [14], where some of these references use Large Deviations techniques. For additional results on the Bayesian point of view in Thermodynamic Formalism, we refer the reader to [16], [23] and [28].

Invariant probabilities for the shift transformation correspond to stationary processes X_n, for $n \in \mathbb{N}$, with values on $\{1, 2, ..., d\}$.

Given a Hölder potential $A : \Omega \to \mathbb{R}$ *the pressure of* A is defined as

$$P(A) = \sup_{\mu \quad \textit{invariant for the shift}} \left\{ \int A d\mu + h(\mu) \right\},$$

where $h(\mu)$ is *the Shannon-Kolmogorov entropy* for the invariant probability measure μ. The unique probability which realizes such supremum is called the Hölder equilibrium probability for the potential A. It's known that $P(A)$ is an analytic function on the potential A (see [29]). This property is quite useful to obtain good large deviation properties (see, for instance, [26] and [27]).

Consider a Hölder continuous function $\log J : \Omega \to \mathbb{R}$, where $J > 0$, such that, $\sum_{a=1}^{d} J(a, b_0, b_1, b_2, b_3, \cdots) = 1$, for all $x = (b_0, b_1, b_2, b_3, \cdots) \in \Omega$.

In this case $P(\log J) = 0$ and the Hölder equilibrium probability will be called a *Hölder Gibbs equilibrium probability for* $\log J$. For instance, when $\log J = -\log d$, the corresponding equilibrium probability will be the maximum entropy probability, which is the independent probability with weights $1/d$.

Equilibrium probabilities play a central role in several problems in Statistical Physics and Information Theory. Hypotheses tests are relevant in all these domains (see, for instance, [10], [20], [31], [8], [33], [32] and [4]).

To each Hölder Gibbs probability μ, one can associate a unique Hölder continuous function $\log J : \Omega = \{1, 2..., d\}^{\mathbb{N}} \to \mathbb{R}$ (see [29]). We call J the *Jacobian* of μ. All Jacobian functions considered here are of Hölder class while all measures are ergodic when $\log(J)$ is in the Hölder class. The Shannon-Kolmogorov entropy of such μ is given by the formula $h(\mu) = -\int \log J \, d\mu$. Two **different** Hölder Gibbs probabilities are **singular** with respect to each other (see [29]). We point out that, in most of the cases in the classical setting of Hypotheses tests, the researchers consider families of probabilities which are **absolutely continuous** with respect to each other. The set of Hölder Gibbs probabilities is dense in the set of invariant probabilities (see, for instance, [26]).

Here we will just consider probabilities μ on Ω of Hölder Gibbs type. The associated stochastic process $\{X_n\}_{n \in \mathbb{N}}$, taking values on $\{1, 2, \cdots, d\}$, is described by

$$\mathbb{P}(X_0 = a_0, X_2 = a_2, \cdots, X_n = a_n) = \mu(\overline{a_0, a_2, \cdots, a_n}),$$

where $\overline{a_0, a_2, \cdots, a_n} \subset \Omega$ is a general cylinder set.

For instance, if μ is a Markov measure associated to a line stochastic matrix $\mathcal{P} = (p_{i\,j})_{i,j=1}^{d}$, then the function J on the cylinder $\overline{i\,j}$ has the constant value $\frac{\pi_i \, p_{ij}}{\pi_j}$, where $\pi = (\pi_1, \pi_2, .., \pi_d)$ is the initial stationary vector for $\mathcal{P}$. The references [3], [18] and [19] consider statistical tests for Markov Chains. In the two by two case, we get that

$\frac{\pi_i\, p_{ij}}{\pi_j} = p_{j\,i}$, for $i, j = 1, 2$. The Jacobian is the natural extension of the concept of stochastic matrix (see page 27, in [29] or example 1, in [27]).

The paper is organized as follows: in Section 2 we present the basic idea of two simple hypotheses test in the thermodynamical formalism sense, where the definition of the *type I* and *type II errors* are stated. In Section 3, Large Deviation properties and some basic results are presented. The Neyman-Pearson hypotheses test and its main result are considered in Section 4. The Min-Max hypotheses test is presented in Section 5, while the Bayesian hypotheses test is in Section 6. Finally, Section 7 presents an example based on the Min-Max hypotheses test.

2 Preliminaries on Hypotheses Tests

In this section, we set the preliminaries and basics concepts to consider a simple hypotheses test in thermodynamic formalism sense.

Let $\{X_n\}_{n\in\mathbb{N}}$ be a stochastic process, defined in a probability on $\Omega = \{1, \cdots, d\}^{\mathbb{N}}$. We can test two simple hypotheses in the following way:

H_0**:** $\{X_n\}_{n\in\mathbb{N}}$ is described by μ_0 with Jacobian J_0

H_1**:** $\{X_n\}_{n\in\mathbb{N}}$ is described by μ_1 with Jacobian J_1.

The two measures μ_0 and μ_1 considered here are Hölder Gibbs probabilities and are, therefore, singular with respect to each other. As far as the authors know, this type of tests was not considered in the literature.

We want to decide which one of the two hypotheses is true from samples $x_i = \sigma^i(x_0)$, $i = 0, 1, 2, \cdots, n-1$, where $x_0 \in \Omega$ is chosen at random according to a given measure μ. In Sections 4 and subsection 4.1, we will choose to fix such μ as μ_1. We are interested on the Large Deviations properties for such type of tests.

One can announce H_1 when H_0 is true. This is called *false alarm* or *type I error*. The probability of false alarm is usually denoted by α, which is called *the test size*. Therefore, the value α denotes $\mathbb{P}(\text{Decide}\, H_1\, | H_0 \,\text{is true}) = \alpha$. We choose α such that $0 < \alpha < 1$.

On the other hand, one can announce H_0 when H_1 is true. This is called a *misspecification* or *type II error*. The probability of misspecification is usually denoted by $1-\beta$. The *detection rate* is the value $\beta \in (0, 1)$, which describes the probability $\mathbb{P}(\text{Decide}\, H_1 \mid H_1 \,\text{is true})$. The value β is called the *power of the test*. In general one hopefully would like to fix a value of β close to 1.

We do not know in advance which hypothesis H_0 or H_1 is more likely to happen (at least for non-Bayesian tests). For a fixed α we would like to choose a test that minimizes the total error probability, that is, we would like to maximize β.

We point out that the Bayesian point of view will be explored in Section 6.

The *test statistics* will be associated to samples and they are given in a log-likelihood form by

$$S_n(x) := \frac{1}{n}\sum_{i=0}^{n-1} \log\left(\frac{J_0(x_i)}{J_1\,(x_i)}\right). \tag{2.1}$$

A similar log likelihood test was considered in section VI in [9], but the author uses no dynamics.

We shall introduce a sequence u_n, $n \in \mathbb{N}$, which will be necessary for the test. The rejection region $\mathcal{R}_n$ is defined as

$$\mathcal{R}_n = \{x \in \Omega \,|\, S_n < u_n\}, \quad n \in \mathbb{N}. \tag{2.2}$$

We assume that

$$\lim_{n\to\infty} u_n = E. \tag{2.3}$$

In the dynamic sense, the important quantity is the limit value E and not the specific values u_n. Given a sample of size n, if $S_n < u_n$, we announce H_1, when H_0 is true, and if $S_n > u_n$ we announce H_0, when H_1 is true. If $S_n = u_n$, from an asymptotic perspective, the choice does not matter.

In all tests considered here, the main point is to find the optimal choice for E, the limit of u_n, for $n \in \mathbb{N}$) and its relationship with the asymptotic values of

$$\mu_1\,(S_n \geq u_n) \quad \text{and} \quad \mu_0\,(S_n \leq u_n)\,, \tag{2.4}$$

for large n.

For all tests, we shall consider large samples and we are interested in minimizing the exponential rate of the probability of a wrong decision. In this direction, it will be necessary to study Large Deviations properties first. The results on Large Deviations properties need in here are presented in Section 3.

For the Neyman-Pearson hypotheses test (see Section 4 and 4.1), the large samples will be taken according to μ_1 and we want to estimate how small is the probability $1-\beta_n$ of announcing H_0 when H_1 is true. In this case, we shall consider samples of the process S_n, for $n \in \mathbb{N}$, which will be produced with the random choice given by μ_1 and not by μ_0.

We denote by $\beta_n = \mu_1(\mathcal{R}_n)$ *the power of the test* at time n. We want to analyze the *misspecification probability, or type II error*, which will be denoted by $1-\beta_n$, from samples of size n.

The asymptotic values of the probabilities of $\mu_1(S_n > u_n)$ and $\mu_0(S_n \leq u_n)$, for $n \in \mathbb{N}$, are the essential information we shall consider. The main issue here is: $\mu_1\{x \,|\, S_n > u_n\}$ is associated with a wrong decision by announcing H_0 when H_1 is true. On the other hand $\mu_0\{x \,|\, S_n \leq u_n\}$ is associated with a wrong decision by announcing H_1 when H_0 is true.

The main result of Section 4 and subsection 4.1 is given by Theorem 4.1. We state this theorem below.

Theorem A. *The optimal choice for the value E, in the Neyman-Pearson hypotheses test, is given by*

$$E = \int \log J_0\, d\mu_0 - \int \log\, J_1\, \mathrm{d}\mu_0.$$

Moreover, the decay rate for minimizing the probability of wrong decisions will be of order $e^{-\,n\,(\int \log J_0\, d\mu_0 - \int \log\, J_1\, \mathrm{d}\mu_0\,)}$.

The value $\int (\log J_0 - \log J_1\,)\, d\mu_0$ is also known as *Kullback-Leibler divergence*. For some results on Kullback-Leibler divergence, we refer the reader [31], [18], [7] and [24].

In the other two tests (see Sections 5 and 6) we will also consider large samples S_n, but we have to compare the corresponding asymptotic laws according to μ_1 and also μ_0, in terms of the expression (2.4).

Section 5 will consider loss functions and the Min-Max hypotheses test. In this case, it will be natural to consider the *pressure* as a function of a real parameter $t \in \mathbb{R}$, more precisely, we shall need the function P_1 given by

$$t \to P_1(t) = P(\, t\,(\log J_0 - \log J_1) + \log J_1).$$

The main result in Section 5 is Theorem 5.1 that we state in here:

Theorem B. *In the Min-Max hypotheses test, the best choice of E will be $E = 0$. Moreover, the best decay rate for minimizing the probability of wrong decisions is given by $e^{n\,r}$, where r is the minimum of the pressure function P_1.*

In Section 6 a certain type of Bayesian hypotheses test will be studied. Hypothesis H_0 will have probability π_0 and hypothesis H_1 will have probability π_1, where $\pi_0 + \pi_1 = 1$. In this section, we shall consider rejections regions of the form

$$\mathcal{R}_{n,\lambda} = \left\{ x \in \Omega \,\middle|\, \frac{1}{n} \sum_{i=0}^{n-1} \log J_\lambda(x_i) < u_n \right\}, \quad \text{for} \quad n \in \mathbb{N}, \tag{2.5}$$

where

$$J_\lambda = \lambda J_1 + (1-\lambda) J_0, \quad \text{for} \quad \lambda \in [0,1]. \tag{2.6}$$

We shall estimate $\pi_1\, \mu_1(S_n > u_n)$ and $\pi_0\, \mu_0(S_n \leq u_n)$, for $n \in \mathbb{N}$. For this test, we shall exhibit the best value of E_λ that minimizes the probability of a wrong decision (see (6.21), (6.22)) and (6.24)), for each λ. We shall also find the best possible E_λ, producing the best decay rate, among all possible values of λ.

We will show a version of Chernoff's information in Section 6.

3 Preliminaries on Large Deviations Properties

In this section, we shall present the Large Deviations properties which will be necessary for the proof of our main results in Sections 4, 5, and 6.

We shall be interested in estimating

$$\mu_1\,(S_n > u_n) = \mu_1\,(S_n - u_n > 0) \quad \text{and} \quad \mu_0\,(S_n \leq u_n) = \mu_0\,(S_n - u_n \leq 0)\,, \tag{3.7}$$

where S_n is defined by (2.1) and $u_n \to E$, when n goes to infinity. We are interested in Large Deviations for $S_n - u_n$; that is, given an interval $(a,b) \subset \mathbb{R}$, we want to estimate, $\mu_j\{(S_n - u_n) \in (a,b)\}$, for $j = 0,1$. Intervals of the type $(-\infty, 0)$ and $(0, \infty)$ are particularly important.

It is a classical result (see, for instance, [29]), that

$$\int (\log J_0 - \log J_1\,)\, d\mu_0 \, > 0 \quad \text{and} \quad \int (\log J_1 - \log J_0\,) d\mu_1 > 0.$$

We need to estimate

$$\mu_j(S_n - u_n \in (a,b)) \, = \mathbb{P}_{\mu_j}\left(\frac{1}{n} \sum_{i=0}^{n-1} \left[\log\left(\frac{J_0(x_i)}{J_1\,(x_i)} \right) - u_n \right] \in (a,b) \right),$$

for $j = 0, 1$.

To get the correct large deviation rate, we need first to analyze the expression

$$\phi_n^j(t) := \frac{1}{n} \log \mathbb{E}_{\mu_j} \left\{ \exp \left[t \sum_{i=0}^{n-1} \left(\log \left(\frac{J_0(x_i)}{J_1(x_i)} \right) - u_n \right) \right] \right\}, \tag{3.8}$$

for each n and each real value t, where $\mathbb{E}_{\mu_j}$ denotes the expected values with respect to the probability μ_j, for $j = 0, 1$. Expression (3.8) is equivalent to

$$\phi_n^j(t) = \frac{1}{n} \log \left(\int e^{t \sum_{i=1}^{n} (\log J_0 - \log J_1)(\sigma^i(x))} \mathrm{d}\mu_j(x) \right) - t\, u_n, \tag{3.9}$$

for $j = 0, 1$.

It is known (see proposition 3.2 in [22] and theorem 3 in [26], or the references [27] and [25]) that

$$\lim_{n\to\infty} \frac{1}{n} \log \left(\int e^{t \sum_{i=1}^{n} (\log J_0 - \log J_1)(\sigma^i(x))} \mathrm{d}\mu_j(x) \right) = P(t(\log J_0 - \log J_1) + \log J_j). \tag{3.10}$$

Hence, from the expressions (3.8), (3.9) and (3.10), one has

$$\phi^j(t) := \lim_{n\to\infty} \phi_n^j(t) = P(t(\log J_0 - \log J_1) + \log J_j) - t\, E, \tag{3.11}$$

for $j = 0, 1$.

Denote by P_j the function

$$t \to P_j(t) := P(t(\log J_0 - \log J_1) + \log J_j), \tag{3.12}$$

for $j = 0, 1$. The function $t \to P_j(t)$ is convex, for $j = 0, 1$. Figure 3.1 shows the graphs of P_0 (in blue) and P_1 (in red), for the example in Section 7.

One can easily show that, for any $t \in \mathbb{R}$,

$$P_1(t) = P_0(t - 1). \tag{3.13}$$

The function

$$t \to P(t(\log J_0 - \log J_1) + \log J_j) - tE = P_j(t) - t\, E$$

is also convex, for $j = 0, 1$.

Moreover, $P_j(0) = P(0(\log J_0 - \log J_1) + \log J_j) = P(\log J_j) = 0$, for $j = 0, 1$. From chapter 4 in [29], note that

$$\frac{d}{dt} P_j(t)|_{t=0} = \frac{d}{dt} \left(P(t(\log J_0 - \log J_1) + \log J_j) \right|_{t=0} = \int (\log J_0 - \log J_1) d\mu_j, \tag{3.14}$$

for $j = 0, 1$.

Then, $\frac{d}{dt} P_0(t)|_{t=0} > 0$ and $\frac{d}{dt} P_1(t)|_{t=0} < 0$, if $\mu_1 \neq \mu_0$. Besides,

$$\frac{d}{dt} \left(P(t(\log J_0 - \log J_1) + \log J_j) \right|_t = \int (\log J_0 - \log J_1) d\mu_t^j, \tag{3.15}$$

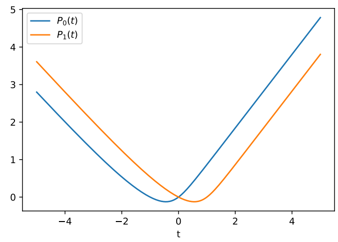

Figure 3.1: Graphs of P_0 (in blue) and P_1 (in orange) for the functions defined in (3.12). For these plots we use the data from the example in Section 7.

where μ_t^j is the equilibrium probability for $t(\log J_0 - \log J_1) + \log J_j$ (see [29]).

There exist values $c^+ > 0 > c^-$ defined by

$$c^- = \inf_{t\in\mathbb{R}} P_1'(t) \quad \text{and} \quad c^+ = \sup_{t\in\mathbb{R}} P_1'(t).$$

From expression (3.13) we also have

$$c^- = \inf_{t\in\mathbb{R}} P_0'(t) \quad \text{and} \quad c^+ = \sup_{t\in\mathbb{R}} P_0'(t).$$

The main interest here is thee following: for each value E, where $u_n \to E$, one wants to estimate the asymptotic values of $\mu_1 (S_n - u_n > 0)$ and $\mu_0 (S_n - u_n \le 0)$. The following proposition states the exact values for the deviation function (see [9], [26], [22] or [25]). Before addressing this issue, one observes the following two points:

- From expression (3.13), it holds

$$\begin{aligned} \frac{d}{dt} P_1(t)|_{t=1} &= \frac{d}{dt}(P(t(\log J_0 - \log J_1) + \log J_1)|_{t=1} \\ &= \frac{d}{dt} P_0(t)|_{t=0} = \int (\log J_0 - \log J_1)\, d\mu_0 > 0. \end{aligned} \tag{3.16}$$

- It is also true that

$$\frac{d}{dt} P_1(t)|_{t=0} = \frac{d}{dt} P_0(t)|_{t=-1}. \tag{3.17}$$

Proposition 3.1. *For a fixed value E, it is true that*

(i) If $E < \int (\log J_0 - \log J_1) d\mu_1$, then

$$\lim_{n\to\infty} \mu_1 (S_n - u_n > 0) = 1. \tag{3.18}$$

(ii) If $E > \int (\log J_0 - \log J_1) d\mu_0$, then

$$\lim_{n\to\infty} \mu_0 (S_n - u_n \le 0) = 1. \tag{3.19}$$

Proof: According to [26], [22] or [25], the deviation function I_j, for $(S_n - u_n)$, $n \in \mathbb{N}$, and for the measure μ_j, is

$$\begin{aligned} I_j(x) &= \sup_t \left[tx - \phi^j(t)\right] = \sup_t \left[t\,(x+E\,) - P(t(\log J_0 - \log J_1\,) + \log J_j\,)\right] \\ &= \sup_t \left[t\,(x+E\,) - P_j(t)\right], \end{aligned} \tag{3.20}$$

for a fixed value E and $j = 0, 1$.

That is,

$$\mu_j\,\{\,x \in \Omega \,|\, (S_n - u_n)(x) \in (a,b)\,\} \sim e^{-n\,\inf_{z\in(a,b)} I_j(z)}, \tag{3.21}$$

for $j = 0, 1$. The function $I_j(\cdot)$ is a real analytical one.

Given E, take

$$x = v_j = -E + \left(\int \log J_0\, d\mu_j - \int \log J_1 d\mu_j\right). \tag{3.22}$$

By using (3.20) and (3.14), the supremum is attained at $t = 0$. Then,

$$I_j(v_j) = 0, \quad \text{for} \quad j = 0, 1. \tag{3.23}$$

On the other hand, from (3.20) and (3.15), for $x = 0$, we get t_j^E, where

$$\begin{aligned} P_j'(t_j^E) &= \frac{d}{dt}(\,P(t(\log J_0 - \log J_1\,) + \log J_j\,)|_{t_j^E} = E \\ &= \int (\log J_0 - \log J_1\,) d\mu_{t_j^E}, \end{aligned} \tag{3.24}$$

and

$$\begin{aligned} I_j(0) &= t_j^E E - P(t_j^E(\log J_0 - \log J_1\,) + \log J_j\,) = t_j^E E - P_j(t_j^E) \\ &= t_j^E \left(\int (\log J_0 - \log J_1) d\mu_{t_j^E}\right) - \left[t_j^E \left(\int (\log J_0 - \log J_1) d\mu_{t_j^E}\right)\right. \\ &\quad \left. + \int \log J_j d\mu_{t_j^E} + h(\mu_{t_j^E})\right] = -\left[\int \log J_j d\mu_{t_j^E} + h(\mu_{t_j^E})\right] > 0, \end{aligned} \tag{3.25}$$

if $\mu_{t_j^E} \neq \mu_j$.

It follows from (3.13) that $t_0^E = t_1^E - 1$ and therefore, $P_0(t_0^E) = P_1(t_1^E)$. From this, follows that

$$I_1(0) = t_1^E E - P_1(t_1^E) \ \text{ and } \ I_0(0) = t_0^E E - P_0(t_0^E) = I_1(0) - E. \tag{3.26}$$

Item (i): From expression (3.21), with $(a,b) = (0,\infty)$, if $E < \int (\log J_0 - \log J_1\,) d\mu_1$, that is $v_1 > 0$, then

$$\lim_{n\to\infty} \mu_1\,(S_n - u_n > 0) = 1,$$

since $-\inf_{x>0} I_1(x) = 0$. Hence, expression (3.18) is true.

Now, if $v_1 < 0$, as $I_1(v_1) = 0$, from (3.21), with $(a,b) = (0,\infty)$,

$$\lim_{n\to\infty} \frac{1}{n} \log(\mu_1\,(S_n - u_n > 0)\,) = \lim_{n\to\infty} \frac{1}{n} \log(1 - \beta_n) = -\inf_{x>0} I_1(x) = -I_1(0) < 0. \tag{3.27}$$

Then, from (3.25), we get

$$1-\beta_n=\mu_1\,(S_n-u_n>0)\sim e^{-n\,\{\inf I_1(x)\,|\,x\geq 0\}}=e^{-n\,I_1(0)}\;\to 0.$$

This corresponds to

$$-E+\int(\log J_0-\log J_1\,)d\mu_1<0.$$

Note that if $E=0$, then $v_1<0$ and

$$\mu_1\,(S_n>0)\sim e^{-n\,I_1(0)},\tag{3.28}$$

where $I_1(0)>0$.

Item (ii): On the other hand, if $E>\int(\log J_0-\log J_1\,)\,d\mu_0$, that is, $v_0<0$, then

$$\lim_{n\to\infty}\mu_0\,(S_n-u_n\leq 0)=1,\tag{3.29}$$

since $-\inf_{x<0}I_0(x)=0$. Expression (3.19) is true by (3.21), with $(a,b)=(-\infty,0)$.

If $v_0>0$, as $I_0(v_0)=0$, from (3.21)

$$\lim_{n\to\infty}\frac{1}{n}\log(\mu_0\,(S_n-u_n\leq 0)\,)=-\inf_{x<0}I_0(x)=-I_0(0)<0.$$

Then, we obtain

$$\mu_0\,(S_n-u_n\leq 0)\sim e^{-n\,\{\inf I_0(x)\,|\,x\leq 0\}}=e^{-n\,I_0(0)}\to 0.$$

□

4 Neyman-Pearson Hypotheses Test

In this section, we shall deal with the dynamical Neyman-Pearson hypotheses test. Its optimality property will be shown in Section 4.1.

From the ergodicity of μ_0, we have

$$S_n\overset{a.s\,(\mu_0)}{\longrightarrow}\int\log\left(\frac{J_0}{J_1}\right)\mathrm{d}\mu_0=\int\log J_0\;d\mu_0\;\;-\int\log\;J_1\,\mathrm{d}\mu_0,\tag{4.1}$$

whenever hypothesis H_0 is true (that is, the samples are obtained from the measure μ_0). We point out that the right hand side of (4.1) is a relative entropy expression (see [8]).

Analyzing the asymptotic values of $\mu_0(S_n\leq u_n)$, associated to the *type I error* value $0<\alpha<1$, when $n\to\infty$, it follows that

$$u_n\longrightarrow\int\log J_0\,d\mu_0\;-\int\log\;J_1\,\mathrm{d}\mu_0>0.$$

This is true, otherwise, $\mu_0(S_n<u_n)$ would converge to 1 or 0.

We set in the present section a sequence u_n, such that

$$\lim_{n\to\infty}u_n=E=\int\log J_0\,d\mu_0\;-\int\log\;J_1\,\mathrm{d}\mu_0.\tag{4.2}$$

Later, in Section 4.1, we will show that such E is optimal.

The *probability of misspecification* is given by $\mathbb{P}(\text{Decide} H_0 \,|\, H_1 \text{ is true}) = 1 - \beta$. We will consider large samples of size n and we shall apply classical results of large deviations theory. In the Neyman-Pearson (NP, for short) hypotheses test we shall consider the *rejection region* given by (2.2). According to this test, at time n (obtained from samples of size n from measure μ_1) we want to estimate $1 - \beta_n$.

The sequence u_n should be consistent with the value α in the sense that it should preserve $\mathbb{P}(\text{Decide}\, H_1 \,|H_0 \text{ is true}) = \alpha$, that is, we define u_n by

$$\mu_0 \left(S_n < u_n\right) = \alpha. \tag{4.3}$$

As mentioned before

$$1 - \beta_n = \mu_1 \left(S_n \geq u_n\right) = \mu_1 \left(S_n - u_n \geq 0\right).$$

From expression (3.22), when $E = \int \log J_0 \, d\mu_0 - \int \log \, J_1 \, \mathrm{d}\mu_0$ (see expression (4.2)), we get

$$\begin{aligned} v_1 = - \, E \, + \left(\int \log J_0 \, d\mu_1 - \int \log J_1 d\mu_1 \right) = - \left(\int \log J_0 \, d\mu_0 \, - \int \log \, J_1 \, \mathrm{d}\mu_0 \right) \\ + \left(\int \log J_0 \, d\mu_1 - \int \log J_1 d\mu_1 \right) < 0. \end{aligned} \tag{4.4}$$

Then, $v_1 < 0$ and, from (3.12),

$$I_1(v_1) = 0. \tag{4.5}$$

When $E = \int \log J_0 \, d\mu_0 \, - \int \log \, J_1 \, \mathrm{d}\mu_0$, we get from (3.20) and (3.16), that $t_1^E = 1$. From expressions (3.27), (3.20), (3.21), (3.23) and (3.25) we get

$$\lim_{n \to \infty} \frac{1}{n} \log(\mu_1 \left(S_n - u_n > 0\right)) = -I_1(0) = -E = - \left(\int \log J_0 \;\, d\mu_0 - \int \log \, J_1 \;\, d\mu_0 \right) < 0. \tag{4.6}$$

The expression (4.6) shows that the Neyman-Pearson hypotheses test works well, since the *probability of misspecification* $1 - \beta_n$ goes exponentially fast to 0. The above expression can be seen as a version of Stein's Lemma (see, for instance, [8]).

4.1 The NP Optimality Property

In this section, we shall prove the optimality property for the Neyman-Pearson hypotheses test (NP, for short). One wonders if there is another alternative hypothesis that provides a smaller mean value error when the sample size n goes to infinity. We shall prove that the Neyman-Pearson hypotheses test is optimal in the sense of the largest power test. From the dynamical point of view, the limit value E is the most important issue and not the specific values u_n. We shall show, in the next theorem below, that $E = \int \log J_0 \, d\mu_0 \, - \int \log \, J_1 \, \mathrm{d}\mu_0$ is the best choice for E. For each possible value E we consider a sequence u_n such that the expression (2.3) holds. One may ask if there exist better values for E.

Theorem 4.1. *The optimal choice for the value E, in the Neyman-Pearson hypotheses test, is given by*

$$E = \int \log J_0 \, d\mu_0 \, - \int \log \, J_1 \, \mathrm{d}\mu_0.$$

Moreover, the decay rate for minimizing the probability of wrong decisions will be of order $e^{-n\,(\int \log J_0 \, d\mu_0 \, - \int \log \, J_1 \, \mathrm{d}\mu_0\,)}$.

Proof: We denote any other alternative hypotheses test by A. The sequence u_n is defined as

$$\mu_0\,(S_n < u_n) = \alpha,$$

which describes *the false alarm*. We choose α such that $0 < \alpha < 1$.

We are interested in optimizing the value of the probability of announcing H_1 when H_1 is true, that is, in optimizing the value $\beta_n = \mu_1(S_n - u_n < 0)$, $n \in \mathbb{N}$. The *rejection region* for the NP hypotheses test, at time n, is denoted by $\mathcal{R}_n^{NP}$ and defined as

$$\mathcal{R}_n^{NP} = \left\{ x \in \Omega | S_n(x) = \frac{1}{n} \sum_{i=0}^{n-1} \log \left(\frac{J_0(x_i)}{J_1\,(x_i)} \right) < u_n \right\}.$$

Similarly, the *rejection region* for another alternative hypotheses test A, at time n, will be denoted by $\mathcal{R}_n^{A}$, where

$$\mathcal{R}_n^{A} = \left\{ x \in \Omega | S_n(x) = \frac{1}{n} \sum_{i=0}^{n-1} \log \left(\frac{J_0(x_i)}{J_1\,(x_i)} \right) < \tilde{u}_n \right\},$$

for another sequence $\tilde{u}_n > 0$.

For a different alternative hypotheses test A corresponds a different choice of E, such that, $\tilde{u}_n \to E$. Assume that $\tilde{u}_n$ is such that

$$\tilde{\alpha} = \tilde{\alpha}_n = \mu_0\,(S_n < \tilde{u}_n) \leq \alpha = \alpha_n = \mu_0\,(S_n < u_n)\,, \tag{4.7}$$

for all n. This means

$$-\,\mu_0\,(S_n < u_n) + \mu_0\,(S_n < \tilde{u}_n) \leq 0. \tag{4.8}$$

Note that $\tilde{u}_n \leq u_n$, for all n. The inequality (4.8) means we are assuming that the alternative hypotheses test A has a smaller or equal false alarm probability. In other words, we do not want to increase the size of the test. Each choice of the sequence $\tilde{u}_n$, $n \in \mathbb{N}$, will be understood as an alternative possible hypotheses. For each alternative hypotheses test A, we obtain the associated value

$$\tilde{\beta}_n \,:=\, \mu_1(S_n < \tilde{u}_n),$$

for any $n \in \mathbb{N}$.

Notice that $\tilde{\beta}_n = \mu_1(\mathcal{R}_n^{A})$ and $\beta_n = \mu_1(\mathcal{R}_n^{NP})$. The optimality property means to compare the NP test with another test A based on the above sequence $\mathcal{R}_n^{A}$, $n \in \mathbb{N}$, using a sequence $\tilde{u}_n$ satisfying (4.7) and, such that,

$$G := \lim_{n\to\infty} \tilde{u}_n < \lim_{n\to\infty} u_n = \int \log J_0 \, d\mu_0 \, - \int \log \, J_1 \, \mathrm{d}\mu_0. \tag{4.9}$$

We shall assume $G > 0$. We point out that, for the alternative hypotheses test A, the associated value $\tilde{\beta}_n$ satisfies

$$\mu_1(\{S_n - \tilde{u}_n < 0\}) = \mu_1(\mathcal{R}_n^A) = \tilde{\beta}_n.$$

We want to show that

$$\tilde{\beta}_n < \beta_n = \mu_1(\mathcal{R}_n^{NP}) = \mu_1(\{S_n - u_n < 0\}),$$

for large n. More precisely, we want to show that

$$\lim_{n\to\infty} \frac{1-\beta_n}{1-\tilde{\beta}_n} = 0. \tag{4.10}$$

The expression (4.10) will guarantee that the NP test is exponentially better than the other alternative test A.

It is known that

$$1 - \beta_n \sim e^{-n\, I_1(0)} = e^{-n(\int \log J_0 d\mu_0 - \int \log\, J_1\, d\mu_0)}.$$

Now we will show that, for large n,

$$\mu_1(\mathcal{R}_n^{NP}) = \beta_n = \mu_1(S_n < u_n) \geq \tilde{\beta}_n = \mu_1(S_n < \tilde{u}_n) = \mu_1(\mathcal{R}_n^A), \tag{4.11}$$

or, equivalently, that

$$1 - \beta_n = \mu_1(S_n \geq u_n) \leq 1 - \tilde{\beta}_n = \mu_1(S_n > \tilde{u}_n). \tag{4.12}$$

Note that

$$\mu_1(S_n > \tilde{u}_n) = \mu_1(S_n - u_n > \tilde{u}_n - u_n). \tag{4.13}$$

As $u_n \to (\int \log J_0\, d\mu_0 - \int \log\, J_1\, \mathrm{d}\mu_0)$ and $\tilde{u}_n \to G$, from expressions (4.13), (3.20)-(3.21) and (3.25), we get the following limit

$$\lim_{n\to\infty} \frac{1}{n} \log(\, \mu_1(S_n > \tilde{u}_n)\,) = -\inf \left\{ I_1(x) \,|\, x \geq G - \left(\int \log J_0\, d\mu_0 - \int \log\, J_1\, \mathrm{d}\mu_0 \right) \right\}.$$

From the expression (4.9), we obtain

$$G_1 := G - \left(\int \log J_0\, d\mu_0 - \int \log\, J_1\, \mathrm{d}\mu_0 \right) < 0. \tag{4.14}$$

Observe that $v_1 < G_1$. Indeed, as $G > 0$, from the expression (4.14), one has

$$\begin{aligned} v_1 &= \left(\int \log J_0\, d\mu_1 - \int \log J_1 d\mu_1 \right) - \left(\int \log J_0\, d\mu_0 - \int \log\, J_1\, \mathrm{d}\mu_0 \right) \\ &< G - \left(\int \log J_0\, d\mu_0 - \int \log\, J_1\, \mathrm{d}\mu_0 \right). \end{aligned}$$

Then, from the expression (4.14), we obtain

$$\begin{aligned} I(G_1) = I\left(G - \left(\int \log J_0\, d\mu_0 - \int \log\, J_1\, \mathrm{d}\mu_0 \right) \right) &< \int \log J_0\, d\mu_0 - \int \log\, J_1\, d\mu_0 \\ &= I_1(0). \end{aligned}$$

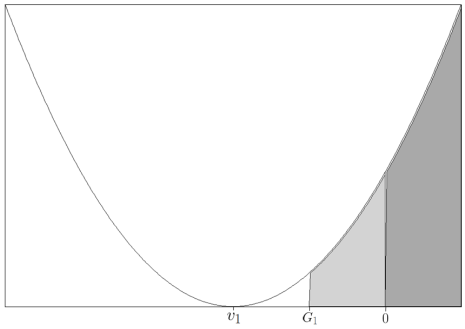

Figure 4.1: Large deviation rate function $I_1(\cdot)$ at points: v_1, $G_1 = G - \left(\int \log J_0 \, d\mu_0 - \int \log\ J_1\ d\mu_0\right)$ and zero.

Therefore,

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \log(\, \mu_1(S_n > \tilde{u}_n)\,) &= -I_1(G_1) > -I_1(0) \\
&= -\left(\int \log J_0 \, d\mu_0 - \int \log\ J_1\ d\mu_0\right) \\
&= \lim_{n\to\infty} \frac{1}{n} \log(\, \mu_1(S_n > u_n)\,).
\end{aligned}$$

This proves expressions (4.12) and (4.10).

□

Figure 4.1 shows the large deviation rate function $I_1(\cdot)$ at points v_1, G_1 and zero, where the point G_1 is given by (4.14).

5 The Min-Max Hypotheses Test

In this section, we shall present the Min-Max Hypotheses text in the dynamical sense. Once more, for S_n, $n \in \mathbb{N}$, given by (2.1), $\mu_1\{x \,|\, S_n > u_n\}$ is associated to a wrong decision by announcing H_0 when H_1 is true while $\mu_0\{x \,|\, S_n \leq u_n\}$ is associated to a wrong decision by announcing H_1 when H_0 is true. For each value E we consider a sequence u_n, such that, (2.3) holds. In the same way as before, the limit value E is more important than the specific values u_n.

In this section we shall consider large deviation properties for both $\mu_1\{x \,|\, S_n > u_n\}$ and $\mu_0\{x \,|\, S_n < u_n\}$. For the Min-Max hypotheses test, we need loss functions for a false alarm. This is a classical ingredient in Hypotheses Tests (see [8] or [30]).

Here we consider the case when the loss functions for false alarm for H_0 and H_1 are constants, respectively, given by y_0 and y_1. The main question here is once again what is the best value for E? The idea behind the use of loss functions is wrong decisions can have a cost. We are interested in finding some optimality property in this setting. From the large deviation properties for this setting, for each choice of limit value E we shall obtain $C_1(E) = C_1 \geq 0$ and $C_0(E) = C_0 \geq 0$, such that,

$$\mu_1\{x \,|\, S_n > u_n\} \sim\ e^{-C_1\, n} \ \text{ and } \ \mu_0\{x \,|\, S_n \leq u_n\} \sim\ e^{-C_0\, n}.$$

In the Min-Max hypotheses test, we have to compare the asymptotic values of the maximum of

$$y_1\,\mu_1\{x\,|\,S_n > u_n\} \sim y_1\,e^{-C_1\,n} \;\text{ and }\; y_0\,\mu_0\{x\,|\,S_n \leq u_n\} \sim y_0\,e^{-C_0\,n}, \tag{5.1}$$

that shall take into account the loss functions y_1 and y_0, for each value E. Finally, we shall consider the minimum $\tilde{E}$ among all possible values of E, that is, the minimum of the function $E \to \max\{C_0(E), C_1(E)\}$. This means that in the Min-Max hypotheses test we are interested in minimizing the maximal cost of wrong decisions by taking either H_0 or H_1. Later, we will show that $\tilde{E} = 0$.

Given an interval $(a,b) \subset \mathbb{R}$ we will be interested simultaneously in Large Deviation properties for $S_n - u_n$, where $u_n \to E$. Then, we have to estimate both $\mu_1\{(S_n - u_n) \in (a,b)\}$ and $\mu_0\{(S_n - u_n) \in (a,b)\}$. This problem was addressed in Section 3. The main properties we shall need in the Min-Max hypotheses test are related to the deviation functions I_1 and I_0 values. From Section 3, for each value E, we obtain the corresponding values t_1^E and t_0^E, such that

$$\frac{d}{dt}P_1(t_1^E) = E \;\text{ and }\; \frac{d}{dt}P_0(t_0^E) = E.$$

From the expression (3.13), we obtain $t_0^E = t_1^E - 1$. And, from expression (3.25), we get

$$I_j(0) = t_j^E\,E - P_j(t_j^E),$$

for $j = 0,1$. Recall, from expression (3.26), that $I_0(0) = I_1(0) - E$.

Denote by $c^+ > 0$, the limit of the derivative

$$\frac{d}{dt}P_1(t)|_t = \frac{d}{dt}(\,P(t(\log J_0 - \log J_1\,) + \log J_1\,)|_t,$$

when $t \to \infty$. The value c^+ is the maximal value of the ergodic optimization for the potential $\log J_0 - \log J_1$ (see [2] and also Section 7). When $E \to c^+$, we get that

$$P_1(t_1^E) - E = P_1(t_1^E) - \frac{d}{dt}P_1(t)|_{t=t_1^E} \to \infty,$$

since $t_1^E \to \infty$. On the other hand, denote by $c^- < 0$, the limit of the derivative

$$\frac{d}{dt}P_0(t)|_t = \frac{d}{dt}(\,P(t(\log J_0 - \log J_1\,) + \log J_0\,)|_t,$$

when $t \to -\infty$. The value c^- is the maximal value of the ergodic optimization for the potential $\log J_1 - \log J_0$ (see [2] and also Section 7). When $E \to c^-$, we get that

$$P_0(t_0^E) - E = P_0(t_0^E) - \frac{d}{dt}P_0(t)|_{t=t_0^E} \to \infty,$$

since $t_0^E \to -\infty$.

Remark 5.1. Observe that both $I_0(0) \geq 0$ and $I_1(0) \geq 0$ **depend on** E. Only the values of E, such that, $I_1(0) = t_1^E\,E - P_1(t_1^E) \geq 0$ and $I_0(0) = t_0^E\,E - P_0(t_0^E) \geq 0$, are relevant for the Min-Max hypotheses test analysis. Indeed, if one of the two options do not happen, then

$$y_1\,\mu_1\{x\,|\,S_n > u_n\} \sim y_1\,e^{-C_1\,n} \;\text{ or }\; y_0\,\mu_0\{x\,|\,S_n \leq u_n\} \sim y_0\,e^{-C_0\,n}$$

will be large and this value of E should be discarded in the search for the optimal $\tilde{E}$.

Recall that from expressions (3.18) and (3.19), if $v_0 < 0$, then $I_0(0) = 0$ and, if $v_1 > 0$, then $I_1(0) = 0$.

If $C_0 > C_1$, then the dominant part of the maximum of (5.1) is $y_1\, e^{-C_1 n}$. On the other hand, if $C_1 > C_0$, then the dominant part of the maximum of the same expression (5.1) is $y_0\, e^{-C_0 n}$. The specific values of y_0 and y_1 are irrelevant for this test and we just have to look for the **minimum value** $\tilde{E}$ of the function

$$E \to r(E) := \max\{\, \inf\{I_0(x) \,|\, x \leq 0\}, \inf\{I_1(x) \,|\, x \geq 0\}\, \},$$

but only for values E such that both $I_0(0) > 0$ and $I_1(0) > 0$.

From the expression (3.26), we have that $I_0(0) = t_0^E E - P_0(t_0^E) = I_1(0) - E$. Then, we just have to find the minimum value $\tilde{E}$ of the function

$$E \to \inf\{I_1(0), I_0(0)\} = \inf\{I_1(0), I_1(0) - E\}, \tag{5.2}$$

for values E, such that both $I_0(0) > 0$ and $I_1(0) > 0$.

From Section 3, we have that $\mu_1\{x \,|\, S_n > u_n\} \sim e^{-n\, \inf\{I_1(x)\,|\,x\geq 0\}}$, which depends on each value of E through the limit (2.3). These values $\mu_1\{x \,|\, S_n > u_n\}$ will be maximum when $\inf\{I_1(x) \,|\, x \geq 0\}$ is minimum. In the same way, according to Section 3, for each value of E, we have that $\mu_0\{x \,|\, S_n > u_n\} \sim e^{-n\, \inf\{I_0(x)\,|\,x\leq 0\}}$. These values $\mu_0\{x \,|\, S_n \leq u_n\}$ will be maximum when $\inf\{I_0(x) \,|\, x \leq 0\}$ is minimum.

In the search for the optimal $\tilde{E}$, we consider several different cases according to the position of E in the set (c^-, c^+).

- **Case 1:** $c^- < E < \int (\log J_0 - \log J_1\,)\, d\mu_1 < 0 < \int (\log J_0 - \log J_1\,)\, d\mu_0 < c^+$. In this situation, $\inf\{I_0(x) \,|\, x \leq 0\} = E - P_0(t_0^E)$ and $\inf\{I_1(x) \,|\, x \geq 0\} = 0$. Hence, $v_1 > 0$ and $v_0 > 0$. Therefore, such values of E should be discarded according to Remark 5.1.

- **Case 2:** $c^+ < \int (\log J_0 - \log J_1\,)\, d\mu_1 < 0 < \int (\log J_0 - \log J_1\,)\, d\mu_0 < E < c^+$. In this situation, $\inf\{I_0(x) \,|\, x \leq 0\} = 0$ and $\inf\{I_1(x) \,|\, x \geq 0\} = E - P_1(t_1^E)$. Hence, $v_1 < 0$ and $v_0 < 0$. Therefore, such values of E should be discarded according to Remark 5.1.

- **Case 3:** $\int (\log J_0 - \log J_1\,)\, d\mu_1 \leq E \leq \int (\log J_0 - \log J_1\,)\, d\mu_0$. As r is a continuous function it follows from the above that there exists a minimum $\tilde{E}$ for the function described by (5.2) restricted to this interval of values of E. This corresponds to $v_1 < 0$ and $v_0 > 0$.

Observe that, in **Case 3**, t_1^E range in an increasing monotonous way from 0 to 1. From Section 3 we obtain $\inf\{I_0(x) \,|\, x \leq 0\} = t_0^E\, E - P_0(t_0^E)$ and $\inf\{I_1(x) \,|\, x \geq 0\} = t_1^E\, E - P_1(t_1^E)$. When $\int (\log J_0 - \log J_1\,)\, d\mu_1 \leq E < 0$, from (5.2) we obtain

$$r(E) = t_1^E\, E \,-\, P_1(t_1^E) - E = (t_1^E - 1)\, E - P_1(t_1^E), \tag{5.3}$$

and, when $\int (\log J_0 - \log J_1\,)\, d\mu_0 \geq E > 0$, from (5.2) we obtain

$$r(E) = t_1^E\, E \,-\, P_1(t_1^E). \tag{5.4}$$

We shall analyze the following two functions: for $\int(\log J_0 - \log J_1)\, d\mu_1 \leq E \leq \int(\log J_0 - \log J_1)\, d\mu_0$

$$E \to P_1(t_1^E) - t_1^E\, E + E = P_1(t_1^E) - (t_1^E - 1)P_1'(t_1^E) \tag{5.5}$$

and

$$E \to P_1(t_1^E) - t_1^E\, E = P_1(t_1^E) - t_1^E\, P_1'(t_1^E). \tag{5.6}$$

Observe that (5.6) is a monotonous decreasing function. Indeed,

$$\frac{d}{dE}[\, P_1(t_1^E) - t_1^E\, P_1'(t_1^E)\,] = -\,(t_1^E)\,(t_1^E)'\, P_1''(t_1^E) < 0,$$

since $(t_1^E)' > 0$, $t_1^E > 0$ and the result follows from the convexity. On the other hand, expression (5.5) is a monotonous increasing function since

$$\frac{d}{dE}[\, P_1(t_1^E) - t_1^E\, P_1'(t_1^E) + P_1'(t_1^E)\,] = (1 - t_1^E)\,(t_1^E)'\, P_1''(t_1^E) > 0,$$

since $(1 - t_1^E)$, $(t_1^E)' > 0$ and the result follows from the convexity.

When $E = \int(\log J_0 - \log J_1)\, d\mu_1$, which is a negative value, we have $t_1^E = 0$ and $P_1(t_1^E) - t_1^E E + E = P_1(0) - 0E + E = E$. Hence, expression (5.5) is equal to $\int(\log J_0 - \log J_1)\, d\mu_1 < 0$. On the other hand, when $E = \int(\log J_0 - \log J_1)\, d\mu_0$, which is a positive value, we have $t_1^E = 1$ and $P_1(t_1^E) = 0$. Hence, $P_1(t_1^E) - E + E = -E + E = 0$. This describes the values of the function given by (5.5) on the interval $\int(\log J_0 - \log J_1)\, d\mu_1 < E < \int(\log J_0 - \log J_1)\, d\mu_0$. Note that when $E = 0$ the two functions (5.6) and (5.5) coincide.

Now, we shall analyze the function (5.6). At the point $E = \int(\log J_0 - \log J_1)\, d\mu_1 < 0$, we get that $t_1^E = 0$, and (5.6) is equal to 0. Moreover, when $E = \int(\log J_0 - \log J_1)\, d\mu_0 > 0$, then, $t_1^E = 1$, and the function (5.6) attains the value

$$-\int(\log J_0 - \log J_1)\, d\mu_0 < 0.$$

Therefore, there exists a unique point E where the two functions (5.6) and (5.5) are equal. This is the value $\tilde{E} = 0$, and then, t_0^1 is such that $P_1'(t_0^1) = 0$. Therefore,

$$r(\tilde{E}) = r(0) = P_1(t_E^1) - t_0^1\, E = P_1(t_0^1),$$

which is the **minimum value of the function** P_1.

It follows from the above and from expression (5.2) that the minimum of $r(E)$ on the interval $\int(\log J_0 - \log J_1)\, d\mu_1 < E < \int(\log J_0 - \log J_1)\, d\mu_0$ is attained when $\tilde{E} = 0$. The point $\tilde{E}$ satisfies $P_1'(t_{\tilde{E}}^1) = P_1'(t_0^1) = 0$.

In this case, the Min-Max solution for $\tilde{E}$ satisfies

$$\min\max\{\, y_1\, \mu_1\{x \mid S_n > u_n\},\, y_0\, \mu_0\{x \mid S_n \leq u_n\}\ \} \sim e^{P_1(t_0^1)\, n}, \tag{5.7}$$

giving us the best rate.

We emphasize that the values of the two loss functions were irrelevant.

6 A Bayesian Hypotheses Test and a Chernoff's Information Version

In this section, we are interested in finding a version of Chernoff's information for the setting of Thermodynamic Formalism (see theorem 11.9.1 in [8]).

Given two Hölder Jacobians J_0 and J_1, for a given parameter $0 \leq \lambda \leq 1$, consider their convex combination given by

$$J_\lambda = \lambda J_1 + (1-\lambda) J_0. \tag{6.1}$$

We point out that J_λ is also a Hölder Jacobian and our setting has a different nature than the one mentioned in section 10 of [8]. Making an analogy, for being consistent with [8], we should consider a convex combination of the logarithm of the Jacobians J_0 and J_1; instead, we did not use the logarithm function. In [8] the probabilities are on finite sets.

We denote by μ_λ the Hölder Gibbs probability associated with $\log J_\lambda$. For $\lambda \in [0,1]$ and $n \in \mathbb{N}$, set $S_n^\lambda(x)$ as

$$S_n^\lambda(x) := \frac{1}{n} \sum_{i=0}^{n-1} \log\left(J_\lambda(x_i)\right).$$

The *rejections regions* are of the form

$$\mathcal{R}_{n,\lambda} = \left\{ x \in \Omega \,|\, \frac{1}{n} \sum_{i=0}^{n-1} \log J_\lambda(x_i) < u_n \right\}, \quad \text{for} \quad n \in \mathbb{N}. \tag{6.2}$$

The *a priori probability* of H_0 is given by π_0 and the *a priori probability* of H_1 is given by $\pi_1 = 1 - \pi_0$. We shall consider for the Bayesian hypotheses test a sequence $u_n \to E$, $n \in \mathbb{N}$, in the same way as in (2.3). The expression

$$\pi_1 \; \mu_1\{x \,|\, S_n^\lambda > u_n\}$$

represents the mean value probability of a wrong decision by announcing H_0 when H_1 is true while the expression

$$\pi_0 \, \mu_0\{x \,|\, S_n^\lambda \leq u_n\},$$

represents the mean value probability of a wrong decision by announcing H_1 when H_0 is true.

In the Bayes hypotheses test, for each value of $\lambda \in [0,1]$, we want to minimize the average total mean probability. We want to choose u_n, $n \in \mathbb{N}$, this means to choose E, that asymptotically minimizes

$$\pi_0 \, \mu_0\{x \,|\, S_n^\lambda \leq u_n\} + \pi_1 \, \mu_1\{x \,|\, S_n^\lambda > u_n\}, \quad \text{as} \quad n \to \infty. \tag{6.3}$$

We shall denote by E_λ the best value of E, for each value λ. We will show later the explicit expression for such E_λ. We can also ask: among the different values of λ which one determines the best E_λ, in the sense of getting the best rate? We shall denote by $\tilde{E}, \tilde{\lambda}$ the optimal value, among all possible values of E and λ, for the asymptotic (6.3) minimizer. In our reasoning, we want to find the best choice of $\tilde{\lambda}$ for which a best choice of $\tilde{E}$ is possible.

In the same way as before, it will follow that, for each choice of λ and limit value E, we shall obtain $C_1(E,\lambda) = C_1 \geq 0$ and $C_0(E,\lambda) = C_0 \geq 0$, such that,

$$\mu_1\{x \,|\, S_n^\lambda > u_n\} \sim e^{-C_1 n} \quad \text{and} \quad \mu_0\{x \,|\, S_n^\lambda \leq u_n\} \sim e^{-C_0 n}. \tag{6.4}$$

From the expression (6.4), for each E and λ we get that the asymptotic of (6.3) is of order

$$e^{-\min\{C_0(E,\lambda), C_1(E,\lambda)\}\, n}. \tag{6.5}$$

When $C_0(E,\lambda) = 0$ or $C_1(E,\lambda) = 0$ we do not get the optimal values of E and λ for the asymptotic in (6.3). Such values of E and λ should be discarded. We will show that for the optimal solution it is required that $C_0(E,\lambda) = C_1(E,\lambda)$. This optimal solution is called *Chernoff information* (we refer the reader to the end of the proof of theorem 11.9.1 in [8], which considers a different setting).

The optimal choice of E and λ will be described by expressions (6.21), (6.22) and (6.24) at the end of this section. The optimal value $C_0(E,\lambda)$ will have a relative entropy expression given by (6.22).

We shall be interested in estimating

$$\mu_1\left(S_n^\lambda > u_n\right) = \mu_1\left(S_n^\lambda - u_n > 0\right), \quad \text{and also} \quad \mu_0\left(S_n^\lambda \leq u_n\right) = \mu_0\left(S_n^\lambda - u_n \leq 0\right),$$

where $u_n \to E$. This requires to estimate

$$\mu_j(\,(\,S_n^\lambda - u_n\,) \in (a,b)\,) \;= \mathbb{P}_{\mu_j}\left(\frac{1}{n}\sum_{i=0}^{n-1}\left[\log\left(J_\lambda(x_i)\right) - u_n\right] \in (a,b)\right),$$

for $j = 0,1$.

In order to get the correct large deviation rate we need first to analyze the following expression

$$\phi_{n,\lambda}^j(t) := \frac{1}{n}\log\left(\int e^{t\sum_{i=1}^n \log J_\lambda(\sigma^i(x))} \mathrm{d}\mu_j(x)\right) - t\,u_n,$$

for each n, λ and real value t.

For $j = 0,1$, $\lambda \in [0,1]$ and $t \in \mathbb{R}$,

$$\lim_{n\to\infty}\frac{1}{n}\log\left(\int e^{t\sum_{i=1}^n \log J_\lambda(\sigma^i(x))} \mathrm{d}\mu_j(x)\right) = P(t\,\log J_\lambda + \log J_j\,).$$

Then, for $j = 0,1$, $\lambda \in [0,1]$ and $t \in \mathbb{R}$, denote

$$\phi_\lambda^j(t) := \lim_{n\to\infty}\phi_{n,\lambda}^j(t) = P(t\,\log J_\lambda + \log J_j\,) - t\,E.$$

We denote by $P_{j,\lambda}$, for $j = 0,1$ and $\lambda \in [0,1]$, the function

$$t \to P_{j,\lambda}(t) = P(t\,\log J_\lambda + \log J_j\,),$$

which is convex and also monotone decreasing in t. Moreover, $P_{j,\lambda}(0) = P(0\,\log J_\lambda + \log J_j\,) = 0$, for $j = 0,1$ and for $\lambda \in [0,1]$.

Note that

$$\frac{d}{dt}P_{j,\lambda}(t)|_{t=0} = \frac{d}{dt}\,P(t\log J_\lambda + \log J_j\,)|_{t=0} = \int \log J_\lambda d\mu_j. \tag{6.6}$$

Furthermore, for $t \in \mathbb{R}$,

$$\frac{d}{dt}\, P(t\, \log J_\lambda + \log J_j\,)|_t = \int \log J_\lambda d\mu_j^{t,\lambda} < 0, \tag{6.7}$$

where $\mu_j^{t,\lambda}$ is the equilibrium probability for $t \log J_\lambda + \log J_j$.

The deviation function I_j^λ for $(S_n^\lambda - u_n)$, $n \in \mathbb{N}$ and for μ_j, $j = 0, 1$, is

$$\begin{aligned} I_j^\lambda(x) &= \sup_t \left[tx - \phi_\lambda^j(t)\right] = \sup_t \left[t\,(x + E\,) - P(t\, \log J_\lambda + \log J_j\,)\right] \\ &= \sup_t \left[t\,(x + E\,) - P_{j,\lambda}(t)\right]. \end{aligned} \tag{6.8}$$

If

$$x = v_j^\lambda = v_j^{E,\lambda} = -E\, + \int \log J_\lambda\, d\mu_j,$$

then, $t = 0$ and,

$$I_j^\lambda(v_j^{E,\lambda}) = 0. \tag{6.9}$$

The suitable values $v_j^{E,\lambda}$ are the ones such that $v_1^{E,\lambda} < 0$ and $v_0^{E,\lambda} > 0$. For each fixed λ, this will require that

$$E \geq \int \log J_\lambda\, d\mu_1 \quad \text{and} \quad E \leq \int \log J_\lambda\, d\mu_0.$$

We will show there exist values λ such that it is possible to find a non-trivial interval for E. We just have to find values λ, such that

$$\int \log J_\lambda\, d\mu_0 > \int \log J_\lambda\, d\mu_1. \tag{6.10}$$

We claim that there are values of λ such that the expression (6.10) holds. Indeed,

$$\lambda = 0 \Rightarrow \int \log J_0\, d\mu_0 - \int \log J_0\, d\mu_1 > 0, \text{ while } \lambda = 1 \Rightarrow \int \log J_1\, d\mu_0 - \int \log J_1\, d\mu_1 < 0.$$

There exists a value λ such that

$$\int \log J_\lambda\, d\mu_0 - \int \log J_\lambda\, d\mu_1 = 0, \tag{6.11}$$

that is, there exists a value λ such that

$$\int \log(\lambda J_1 + (1-\lambda) J_0) d\mu_0 = \int \log(\lambda J_1 + (1-\lambda) J_0) d\mu_1.$$

In fact, consider the functions

$$g_j(\lambda) = \int \log(\lambda J_1 + (1-\lambda) J_0) d\mu_j,$$

for $j = 0, 1$.

Note that $g_0(0) > g_1(0)$ and $g_0(1) < g_1(1)$. From this fact, the claim follows. The functions g_j, for $j = 0, 1$, are concave and g_0 is *a decreasing function* while g_1 is *an increasing* one. Besides, the point where the two graphs coincide is unique.

Therefore, *there exists a value* λ_s, such that for $0 \leq \lambda < \lambda_s$, a non trivial interval of suitable parameters E exists and it holds that

$$0 > \int \log J_\lambda \, d\mu_0 > E > \int \log J_\lambda \, d\mu_1. \tag{6.12}$$

In this case, for such parameters E, we have $v_1^{E,\lambda} < 0$ and $v_0^{E,\lambda} > 0$. From now on we assume that E is in the interval described by the expression (6.12).

When $x = 0$, for $\lambda \in [0, \lambda_s]$, for $j = 0, 1$, we get $t_j^{E,\lambda} \in \mathbb{R}$, for which

$$P'_{j,\lambda}(t_j^{E,\lambda}) = \frac{d}{dt} P(t \log J_\lambda + \log J_j)|_{t_j^{E,\lambda}} = E = \int \log J_\lambda d\mu_{t_j^{E,\lambda}}, \tag{6.13}$$

where $\mu_{t_j^{E,\lambda}}$ is the equilibrium probability for $t_j^{E,\lambda} \log J_\lambda + \log J_j$, where E satisfies (6.12).

From the convexity argument and expression (6.7), for fixed λ and for $j = 0, 1$, the value $t_j^{E,\lambda}$ is monotonous increasing on E. That is, for fixed λ and for $j = 0, 1$, the function $E \to t_j^{E,\lambda}$ satisfies

$$\frac{d}{dE} t_j^{E,\lambda} > 0. \tag{6.14}$$

Denote $I_j^{E,\lambda}(0)$ by

$$\begin{aligned} I_j^{E,\lambda}(0) \, : &= t_j^{E,\lambda} E - P(t_j^{E,\lambda} \log J_\lambda + \log J_j) = t_j^{E,\lambda} E - P_{j,\lambda}(t_j^{E,\lambda}) \\ &- \left[\int \log J_j d\mu_{t_j^{E,\lambda}} + h(\mu_{t_j^{E,\lambda}}) \right] = - \left[\int \log J_j d\mu_{t_j^{E,\lambda}} - \int \log J_{j,\lambda,E}(\mu_{t_j^{E,\lambda}}) \right] > 0, \end{aligned} \tag{6.15}$$

where $J_{j,\lambda,E}$ is the Jacobian of the invariant probability $\mu_{t_j^{E,\lambda}}$ which is, by it turns, the equilibrium probability for $t_j^{E,\lambda} \log J_\lambda + \log J_j$, for $j = 0, 1$ and $\lambda \in [0, \lambda_s]$.

Using expressions (6.15) and (6.8), when $u_n \to E$, one can rewrite them both, as mentioned in (6.3), by

$$\mu_j \{x \,|\, S_n^\lambda > u_n\} \sim e^{-I_j^{E,\lambda}(0) n}, \tag{6.16}$$

for $j = 0, 1$. Hence, in the notation of (6.4), we get $C_j(\lambda, E) = I_j^{E,\lambda}(0)$, for $j = 0, 1$.

Furthermore, for fixed λ and $j = 0, 1$,

$$\frac{d}{dE} \left[t_j^{E,\lambda} \, P'_{j,\lambda}(t_j^{E,\lambda}) \; - \; P_{j,\lambda}(t_j^{E,\lambda}) \right] \; = \; (t_j^{E,\lambda}) \, (t_j^{E,\lambda})' \, P''_{j,\lambda}(t_j^{E,\lambda}). \tag{6.17}$$

Since

$$P(0 \, \log J_\lambda + \log J_0) = 0,$$

$$\frac{d}{dt} P_{0,\lambda}(t)|_{t=0} = \int \log J_\lambda d\mu_0 \quad \text{and} \quad 0 > \int \log J_\lambda \, d\mu_0 > E.$$

From the pressure convexity, we get $t_0^{E,\lambda} < 0$.

From expressions (6.15), (6.17) and (6.14), it follows that we get

$$E \; \to \; I_0^{E,\lambda}(0) \quad \text{decreases with} \quad E, \tag{6.18}$$

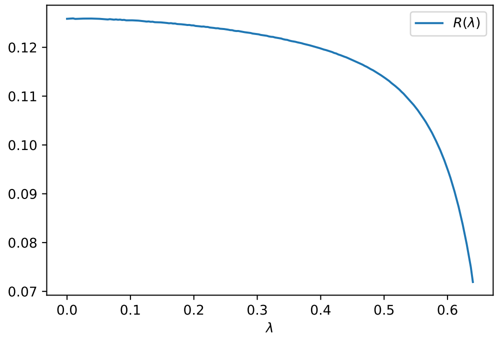

Figure 6.1: Graph of the function $R(\lambda) = I_0^{E_\lambda,\lambda}(0)$, when $0 \leq \lambda \leq \lambda_s$, using the stochastic matrix $\mathcal{P}_j$, for $j = 0, 1$, from the example in Section 7.

for each fixed λ. As $\int \log J_\lambda \, d\mu_1 < E$, we obtain $t_1^{E,\lambda} > 0$. From this property, one can show, in a similar way, that

$$E \to I_1^{E,\lambda}(0) \quad \text{increases with} \quad E, \tag{6.19}$$

for each fixed λ.

For fixed $0 \leq \lambda \leq \lambda_s$, consider the functions

$$E \in \left[\int \log J_\lambda d\mu_1, \int \log J_\lambda d\mu_0 \right] \to y_a(E) = I_0^{E,\lambda}(0)$$

and

$$E \in \left[\int \log J_\lambda d\mu_1, \int \log J_\lambda d\mu_0 \right] \to y_b(E) = I_1^{E,\lambda}(0).$$

As $y_a(\int \log J_\lambda d\mu_0) = 0$, it follows, from the decreasing monotonicity (see expression (6.18)), that $y_a(\int \log J_\lambda d\mu_1) > 0$. In fact, $\int \log J_\lambda d\mu_0 - \int \log J_\lambda d\mu_1 > 0$.

Moreover, as $y_b(\int \log J_\lambda d\mu_1) = 0$, from the increasing monotonicity (see expression (6.19)), we get $y_b(\int \log J_\lambda d\mu_0) > 0$. Hence, *for each* λ, $0 \leq \lambda \leq \lambda_s$, *there exists a point* E_λ, such that $I_1^{E_\lambda,\lambda}(0) = I_0^{E_\lambda,\lambda}(0)$. The value E_λ *determines the best rate for the parameter* λ.

Note also that this point E_λ belongs to the interval $[\int \log J_\lambda d\mu_1, \int \log J_\lambda d\mu_0]$. Furthermore, if $\lambda_p < \lambda_q \leq \lambda_s$, then, as g_0 is a decreasing function and g_1 is an increasing one, we have that $[\int \log J_{\lambda_q} d\mu_1, \int \log J_{\lambda_q} d\mu_0] \subset [\int \log J_{\lambda_p} d\mu_1, \int \log J_{\lambda_p} d\mu_0]$. Moreover, $E_{\lambda_s} = \int \log J_{\lambda_s} d\mu_1 = \int \log J_{\lambda_s} d\mu_0$.

Note that, for any $0 \leq \lambda < \lambda_s$, the value

$$\int \log J_{\lambda_s} \, d\, \mu_{t_1^{E_{\lambda_s},\lambda_s}} \in \left[\int \log J_\lambda d\mu_1, \int \log J_\lambda d\mu_0 \right].$$

From the expression (6.5), property (6.16) and the fact that $I_0^{E,\lambda}(0)$ decreases with E (while $I_1^{E,\lambda}(0)$ increases with E), for each fixed λ, we get the best value E is when $E = E_\lambda$ (see definition above). That is, when

$$t_0^{E_\lambda,\lambda} E_\lambda - P_{0,\lambda}(t_0^{E_\lambda,\lambda}) = I_0^{E_\lambda,\lambda}(0) = I_1^{E_\lambda,\lambda}(0) = t_1^{E_\lambda,\lambda} E_\lambda - P_{1,\lambda}(t_1^{E_\lambda,\lambda}). \tag{6.20}$$

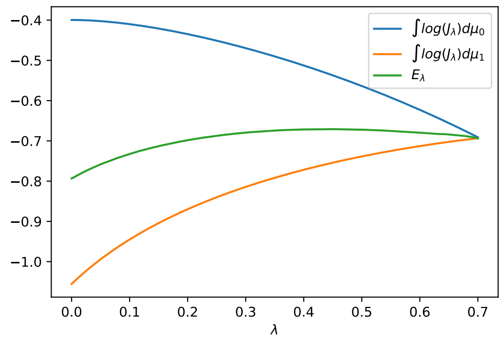

Figure 6.2: Graphs of the functions $\lambda \to \int \log J_\lambda d\mu_0$ (in blue) and $\lambda \to \int \log J_\lambda d\mu_1$ (in orange) together with the graph of the values E_λ (in green), as a function of λ, when $0 \leq \lambda \leq \lambda_s$. The stochastic matrix $\mathcal{P}_j$, for $j = 0, 1$, is from the example in Section 7.

Then, we need to find the value $\tilde{\lambda}$, which maximizes $\lambda \to I_1^{E_\lambda,\lambda}(0)$, simultaneously with (6.20) holds, among all λ, such that $\lambda \in [0, \lambda_s]$.

Note that when $\lambda = \lambda_s$, we have both $I_0^{E_\lambda,\lambda}(0) = 0 = I_1^{E_\lambda,\lambda}(0)$, which is not a good choice.

The function $\lambda \to I_0^{E_\lambda,\lambda}(0)$ is monotonous decreasing on $\lambda \in [0, \lambda_s]$. Therefore, the largest $I_0^{E_\lambda,\lambda}(0)$ occurs when $\lambda = 0$. This means that we need to take $J_\lambda = J_0$.

The value E_0 belongs to the interval $[\int \log J_0 d\mu_1, \int \log J_0 d\mu_0]$ and, by definition, $I_0^{E_0,0}(0) = I_1^{E_0,0}(0)$.

The value E_0 is determined (see expression (6.13)) by the equation

$$P'_{0,0}(t_0^{E_0,0}) = E_0 = P'_{1,0}(t_1^{E_0,0}). \tag{6.21}$$

From expression (6.15), the corresponding value of $I_0^{E_0,0}(0)$ is given by

$$I_0^{E_0,0}(0) = t_0^{E_0,0} E_0 - P_{0,0}(t_0^{E_0,0}) = -\left[\int \log J_0 \, d\mu_{t_0^{E_0,0}} - \int \log J_{0,0,E_0} \, d\,\mu_{t_0^{E_0,0}}\right] > 0. \tag{6.22}$$

Expression (6.22) is a relative entropy.

Finally, the best choice for the hypotheses test, under thermodynamic formalism sense, will be when the rejection region is of the form

$$\mathcal{R}_{n,0} = \left\{ x \in \Omega \,\middle|\, \frac{1}{n} \sum_{i=0}^{n-1} \log J_0(x_i) < u_n \right\}, \quad n \in \mathbb{N}, \tag{6.23}$$

with $u_n \to E_0$, and E_0 satisfies (6.21).

In this case,

$$\pi_0\, \mu_0\{x \,|\, S_n^0 \leq u_n\} + \pi_1\, \mu_1\{x \,|\, S_n^0 > u_n\} \;\sim e^{-\,I_0^{E_0,0}(0)\, n} \tag{6.24}$$

will describe the **best possible rate among λ for minimizing the probability of a wrong decision**.

7 An Example

In this section, we present an example for the Min-Max Hypotheses test. Recall that, given the two by two line stochastic matrix $\mathcal{P}$, the value of the Jacobian J on the cylinder $\overline{i\,j}$ has the constant value $\frac{\pi_i\, p_{ij}}{\pi_j} = p_{ji}$, where $\pi = (\pi_1, \pi_2)$ is the initial stationary vector for $\mathcal{P}$.

We consider the case where $\mathcal{P}_0$ and $\mathcal{P}_1$ are described by the following two column stochastic matrices

$$\mathcal{P}_0 := \begin{pmatrix} \frac{1}{4} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{2} \end{pmatrix} \quad \text{and} \quad \mathcal{P}_1 := \begin{pmatrix} \frac{2}{3} & \frac{1}{5} \\ \frac{1}{3} & \frac{4}{5} \end{pmatrix}.$$

In this case, the best rate is described by (5.7) as it was explained in Section 5. We shall present now some explicit values for this example.

Using techniques given in [17] one can show that the maximizing probability (see [2]) for $\log J_0 - \log J_1$ is an orbit of period two. More precisely, $m(\log J_0 - \log J_1) = \frac{1}{2}\log\left(\frac{45}{8}\right)$ which is a value close to $c^+ \sim 0.8636$. For $m(\log J_1 - \log J_0)$, which is realized by a orbit of period 1, we get $c^- \sim 0.9808$.

In fact, we are interested in finding the image of the functions P_0' and P_1', that is, in finding c^+ and c^-. We will show the domain of the Legendre transform for P_1, which is the same as for P_0.

Define $K = \log(J_0/J_1)$. It is true that $\lim_{t\to+\infty} P_0'(t) = \lim_{t\to+\infty} P_1'(t)$, and this limit value (see [2]) is given by $m(K)$, as long as exists a function u, the so-called a *calibrated subaction*, such that

$$\max_y [K(yx) + u(yx)] = m(K) + u(x),$$

for all $x \in \Omega$. We claim that this equation is satisfied when

$$m(K) = \frac{1}{2}\log\left(\frac{45}{8}\right).$$

We refer the reader to the reference [2] for the max-plus algebra's properties in Ergodic Optimization. The proof of the claim will be as follows. Define the matrix, with $\varepsilon = -\infty$, by

$$W := \begin{pmatrix} K_{11} & \varepsilon & K_{21} & \varepsilon \\ K_{11} & \varepsilon & K_{21} & \varepsilon \\ \varepsilon & K_{12} & \varepsilon & K_{22} \\ \varepsilon & K_{12} & \varepsilon & K_{22} \end{pmatrix}.$$

Now,

$$m(K) = \bigoplus_{n=1}^{4} \frac{Tr_\oplus(W^{\otimes n})}{n}$$

is simply the maximum cyclic mean in the directed graph which has transition costs W_{ij} from node i to node j. Here, we denote $Tr_\oplus$ the max-plus trace and $W^{\otimes n}$ is the n-th max-plus power of W. It is easy to see that the maximal cyclic mean in such graph is given by the mean $\frac{W_{23}+W_{32}}{2} = \frac{1}{2}\log\left(\frac{45}{8}\right)$. In fact, from this the following matrix

$$u = \begin{pmatrix} K_{21} + K_{12} - 2m(K) & 0 \\ K_{12} - m(K) & K_{12} - m(K) \end{pmatrix}$$

is a calibrated sub-action, that is, the calibrated subaction is determined by the matrix u. Hence, we conclude that $\lim_{t\to+\infty} P'_{i=1,2}(t) = m(K) = \frac{1}{2}\log\left(\frac{45}{8}\right)$.

We can also compute $\lim_{t\to-\infty} P'_{i=1,2}(t) = -m(-K)$ by following the same procedure but now with the matrix $-W$ (here we only change the sign of the finite terms in the matrix W). In this way, we consider

$$m(-K) = \log\left(\frac{8}{3}\right) \quad \text{and} \quad v = \begin{pmatrix} 0 & 0 \\ -K_{12} - m(-K) & -K_{12} - m(-K) \end{pmatrix},$$

which satisfies $\max_{Ty=x}\{-K(y) + v(y)\} = m(-K) + v(x)$. This means that

$$\lim_{t\to-\infty} P'_{i=1,2}(t) = -\log\left(\frac{8}{3}\right).$$

This concludes the example. $\diamondsuit$

The method described in this section can be adapted to other cases.

Acknowledgments

H.H. Ferreira was supported by CAPES-Brazil. A.O.Lopes and S.R.C. Lopes' researchs were partially supported by CNPq-Brazil.

References

[1] F. Abramovich and Y. Ritov, Statistical Theory: A Concise Introduction, Chapman-Hall, 2013.

[2] A. Baraviera, R. Leplaideur and A. O. Lopes, *Ergodic Optimization, zero temperature and the Max-Plus algebra*. 23° Colóquio Brasileiro de Matemática, IMPA, Rio de Janeiro, 2013.

[3] R. R. Bahadur. Large deviations of the maximum likelihood estimate in the Markov chain case. In J. S. Rostag, M. H. Rizvi and D. Siegmund, editors, Recent advances in statistics, 273-283. Academic Press, Boston, 1983.

[4] T. Benoist, V. Jaksic , Y. Pautrat, C.-A. Pillet, On Entropy Production of Repeated Quantum Measurements I. General Theory. *Commun. Math. Phys.*, Vol. 357, 77–123, 2018.

[5] L. D. Broemeling, *Bayesian Inference for Stochastic Processes*, CRC Press, 2018.

[6] D. Bohle, A. Marynych and M. Meiners, A Fundamental Problem of Hypotesis testing with finite e-commerce, arXiv, 2020.

[7] J-R Chazottes and D Gabrielli, Large deviations for empirical entropies of g-measures. *Nonlinearity*, Vol. 18, 2545–2563, 2005.

[8] T. Cover and J. Thomas, *Elements of Information Theory*, second edition. New York: Wiley Press, 2006

[9] J. A. Buckle, *Large Deviation Techniques in Decision, Simulation and Estimation.* New York: Wiley, 1990.

[10] A. Caticha, Entropic Physics: Lectures on Probability, Entropy and Statistical Physics, arXiv, 2021.

[11] G.B. Cybis, S. R. C. Lopes and H.P. Pinheiro, Power of the Likelihood Ratio Test for Models of DNA Base Substitution. *Journal of Applied Statistics*, Vol. 38, 2723–2737, 2011.

[12] M. Denker and W. Woyczynski. *Introductory Statistics and Random Phenomena: Uncertainty, Complexity and Chaotic Behavior in Engineering and Science*. New York: Birkhäuser, 2012.

[13] M. Denker, Basics of Thermodynamics, Lecture Notes - Penn State Univ., 2011.

[14] R. Dakovic, M. Denker and M. Gordin, Circular unitary ensembles: parametric models and their asymptotic maximum likelihood estimates. *Journal of Mathematical Sciences*, Vol. 219(5), 714–730, 2016.

[15] A. Dembo and O. Zeitouni, *Large Deviation Techniques and Applications*. New York: Springer Verlag, 2010.

[16] A.C.D. van Enter, A. O. Lopes, S. R. C Lopes and J. K. Mengue, How to get the Bayesian a posteriori probability from an a priori probability via Thermodynamic Formalism for plans; the connection to Disordered Systems. To appear, 2020.

[17] H.H. Ferreira, A.O. Lopes and E.R. Oliveira, An iteration process for approximating subactions. To appear in "Modeling, Dynamics, Optimization and Bioeconomics IV" Editors: Alberto Pinto and David Zilberman, Springer Proceedings in Mathematics and Statistics, Springer Verlag, 2020.

[18] V. Girardin and P. Regnault P, Escort distributions minimizing the KullbackLeibler divergence for a large deviations principle and tests of entropy level. *Ann Inst Stat Math.*, Vol. 68, 439-468, 2016.

[19] V. Girardin, L. Lhote and P. Regnault, Different Closed-Form Expressions for Generalized Entropy Rates of Markov Chains. *Methodology and Computing in Applied Probability*, Vol. 21, 1431–1452, 2019.

[20] S. Ihara, *Information Theory for continuous systems*. World Scientific, 1993.

[21] M.J. Karling, S.R.C. Lopes and R.M. de Souza, A Bayesian Approach for Estimating the Parameters of an α-Stable Distribution. Journal of Statistical Computation and Simulation, available online at https://doi.org/10.1080/00949655.2020.1865958, 2020.

[22] Y. Kifer, Large Deviations in Dynamical Systems and Stochastic processes, *TAMS*, Vol. 321(2), 505–524, 1990.

[23] A.O. Lopes, S.R.C. Lopes and P. Varandas, Bayes posterior convergence for loss functions via almost additive Thermodynamic Formalism, arXiv, 2020.

[24] A.O. Lopes and J.K. Mengue, On information gain, Kullback-Leibler divergence, entropy production and the involution kernel, arXiv, 2020.

[25] A.O. Lopes, Entropy, Pressure and Large Deviation, *Cellular Automata, Dynamical Systems and Neural Networks*, E. Goles e S. Martinez (eds.), Kluwer, Massachusets, 79–146, 1994.

[26] A. O. Lopes, Entropy and Large Deviation, *NonLinearity*, Vol. 3(2), 527–546, 1990.

[27] A.O. Lopes, *Thermodynamic Formalism, Maximizing Probabilities and Large Deviations*, Preprint - UFRGS.

[28] K. McGoff, S. Mukherjee and A. Nobel, Gibbs posterior convergence and Thermodynamic formalism, arXiv, 2019.

[29] W. Parry and M. Pollicott. Zeta functions and the periodic orbit structure of hyperbolic dynamics. *Astérisque*, Vol. 187-188, 1990.

[30] V. K. Rohatgi, *An Introduction to Probability Theory and Mathematical Statistics*. New York: Wiley, 1976.

[31] T. Sagawa, Entropy, divergence and majorization in classical and quantum theory, arXiv, 2020.

[32] Y. Suhov and M. Kelbert, Probability and Statistics by Example. II, Cambridge: Cambridge University Press, 2014.

[33] W. von der Linden, V. Dose and U. von Toussaint. *Bayesian Probability Theory Applications in the Physical Sciences.* Cambridge: Cambridge University Press, 2014.