Dynamical hypothesis tests and Decision Theory for Gibbs distributions

M. Denker, A. O. Lopes and S. R. C. Lopes

December 18, 2022

Abstract

We consider the problem of testing for two Gibbs probabilities μ_0 and μ_1 defined for a dynamical system (Ω, T). Due to the fact that in general full orbits are not observable or computable, one needs to restrict to subclasses of tests defined by a finite time series $h(x_0), h(x_1) = h(T(x_0)), ..., h(x_n) = h(T^n(x_0))$, $x_0 \in \Omega$, $n \geq 0$, where $h : \Omega \to \mathbb{R}$ denotes a suitable measurable function. We determine in each class the Neyman-Pearson tests, the minimax tests, and the Bayes solutions and show the asymptotic decay of their risk functions as $n \to \infty$. In the case of Ω being a symbolic space, for each $n \in \mathbb{N}$, these optimal tests rely on the information of the measures for cylinder sets of size n.

1 Introduction

We consider a compact metric space Ω with Borel σ-algebra and the dynamical action of an open and expanding transformation T on Ω which is topologically mixing.

Given a Hölder potential, i.e. a Hölder continuous function $A : \Omega \to \mathbb{R}$, the transfer operator $\mathcal{L}_A$ associated to A is the one acting on continuous functions $g \in C(\Omega)$ such that

$$[\mathcal{L}_A g](\omega) = \sum_{\{y \in \Omega \,|T(y)=\omega\}} g(y)\, e^{A(y)}. \tag{1.1}$$

Without loss of generality we may assume that for all $\omega \in \Omega$ the Jacobian

$$J = e^A \tag{1.2}$$

satisfies $J(\omega) > 0$ and $\sum_{\{y\,|T(y)=\omega\}} J(y) = 1$, for every $\omega \in \Omega$. It is well known that in this case the eigenmeasure m for the eigenvalue 1 of the dual operator $\mathcal{L}_A^*$ is T-invariant and is called a Gibbs measure.[1] Such Gibbs measures have finite Markov partitions $\gamma = \{\Gamma_1, ..., \Gamma_d\}$ for some $d \geq 2$.

We make the assumption throughout the paper that we are given two distinct Gibbs measures μ_0 and μ_1 on Ω which share a common Markov partition $\gamma = \{\Gamma_1, ..., \Gamma_d\}$ and that the available information on the orbits of points ω in Ω is given by the variables $X_n(\omega) = k \in \{1, ..., d\}$ if and only if $T^n(\omega) \in \Gamma_k$. In fact, this is not an essential restriction since such partitions can be obtained for all pairs of Gibbs measures. Their Jacobians will be denoted by J_0 and J_1, respectively, and we assume, without loss of generality, that both are strictly positive on their support.

Examples of open, expanding maps include hyperbolic rational functions, certain maps of the interval, expanding and differentiable maps on compact manifolds. The results of this paper also hold for invertible maps which admit Markov partitions like Axiom A diffeomorphisms, because we may restrict them to the forward orbit of points.

It follows from the assumption that we can and will restrict to the case when $\Omega = \{1, ..., d\}^{\mathbb{Z}_+}$ for some $d \geq 2$, since Markov partitions create almost surely one-to-one maps between the spaces. Ω is equipped with its Borel σ-algebra $\mathcal{F}$. In this setup the two measures μ_0 and μ_1 may be supported on different subspaces of finite type, but both are assumed here to be the same for simplicity.

We shall be using standard statistical terminology in the sequel as it is also explained in Section 6, the appendix. Notations, definition and facts of statistical nature used in this note are explained and stated there for the readers convenience.

The goal of the present note is to decide on μ_0 or μ_1 based on observed data. Loosely speaking, given a finite sample one has to decide between the hypothesis $H_0 \equiv \mu_0$ and the alternative $H_1 \equiv \mu_1$. The false alarm or type 1 error happens in case one announces H_1 when, in fact, H_0 is true (that is, the sample was originated by μ_0). The value $0 \leq \alpha \leq 1$ denotes the probability of a false alarm, which is called the test size or the significance level of the test. More formally, $\alpha :=$ (Prob. Decide $H_1\,|H_0$ is true).

The probability $\beta :=$ (Prob. Decide $H_1 \mid H_1$ is true) is called the power of the test. The value $1 - \beta$ is called the probability of type 2 error. When designing a test one would like to minimize type 1 and 2 errors under some

[1] Note that we abuse the terminology for a Jacobian here since taking $g = \mathbb{I}_C$ as the indicator function of a set C on which T acts invertible, then $\int_C e^{-A} dm = \int \mathcal{L}_A \mathbb{I}_C e^{-A} dm = m(TC)$. Hence the Jacobian of T is J^{-1}.

constraints.

Formally, we consider the *statistical experiment* $\mathcal{E} := (\Omega, \mathcal{F}, \mathcal{P})$, where $\mathcal{P} = \{\mu_0, \mu_1\}$. The objective is to make a decision about the *true* probability in $\mathcal{P}$ once a point in Ω is observed.

To do this we consider a the *test problem* which is specified by the subset $\mathcal{H}_0 = \{\mu_0, \mu_1\} \subset \mathcal{P}$, the hypothesis $H_0 \equiv \mu_0$ versus $H_1 \equiv \mu_1$, the decision space $D = \{0, 1\}$ and a loss function L to be set later (see Lemma 2.1 or the appendix. A test can be seen as a function

$$\varphi : \Omega \to [0, 1]$$

defined as

$$\varphi(\omega) = \delta(\omega, \{1\})$$

where $\delta \in \Delta$ is a decision function.

Since a point in the space Ω is in general not observable one needs to restrict to finite time series. Therefore we consider a dynamical setting where test problems $\mathcal{E}_n$ are defined for each $n \in \mathbb{N}$. We determine the best tests under Neyman-Pearson, minimax and Bayes distribution constraints and analyze the asymptotic behavior of their error properties, when $n \to \infty$.

We denote by S a set with $\mu_i(S) = i$ $(i = 0, 1)$, which exists since two distinct Gibbs measures are orthogonal. We denote by $E_m(g) = \int gdm$ the expectation of g with respect to the probability m. The first observation is well known see [11], page 201.

The Neyman-Pearson Lemma characterizes those tests which have maximal power subject to keeping a given significance level α. These are called Neyman-Pearson tests.

Theorem 1.1. *The test*

$$\phi^*(\omega) = \begin{cases} 1, & \text{if } \omega \in S \\ 0, & \text{if } \omega \notin S \end{cases}$$

is a Neyman-Pearson test at level $\alpha = 0$ *and is as well the minimax test and the Bayes solution for any risk function* $\phi \mapsto \mathcal{R}_\pi(\phi) = \pi_0 E_{\mu_0}(\phi) + \pi_1 E_{\mu_1}(1 - \phi)$, $\pi = (\pi_0, \pi_1)$ *(*ϕ *any randomized test) where* π *is the prior distribution on* $\{0, 1\}$.

All other Neyman-Pearson tests for this problem are inferior, so that full information on the orbit requires as well the knowledge of distinct supports of μ_0 and μ_1. So the problem arises to find a good computable test. This can be done using finite time series $X_0, X_1 = X_0 \circ T, ..., X_n = X_0 \circ T^n$ $(n \in \mathbb{N})$ where X_0 is the projection $\Omega \to \{1, ..., d\}$ onto the first coordinate.

We denote by $\mathcal{T}_n$ the collection of all tests which are measurable with respect to $X_0, ..., X_n$. This set can be described by the set of all tests for the test problem

$$\mathcal{E}_n = (\{1, ..., d\}^{n+1}, \mathcal{P}_n = \{\mu_i^n |\ i = 0, 1\}, H_0^n = \{\mu_0^n\}) \tag{1.3}$$

where μ_i^n $(i = 0, 1)$ denotes the marginal distribution of μ_i on cylinder sets c of length $n + 1$ which are defined as $c = [c_0, ..., c_n] = \{\omega \in \Omega |\ \omega_k = c_k\, (0 \leq k \leq n)\}$ $(1 \leq c_i \leq d$ for $0 \leq i \leq n)$.

Example 1.2. *In order to illustrate the foregoing setup, consider the unit interval $\Omega = [0, 1]$ together with the map $T(x) = 10 \cdot x \mod 1$. Let μ_0 denote the Lebesgue measure restricted on Ω and μ_1 the invariant measure associated to a potential $J : \Omega \to \mathbb{R}_+$ with $\sum_{T(y)=x} J(y) = 1$, for all $x \in \Omega$. The Markov partition is just $\gamma = \{[\frac{j}{10}, \frac{j+1}{10}]\ ; 0 \leq j \leq 9\}$. More precisely, the potential only needs to be Hölder continuous with respect to the sequence space metric in $\{0, 1, ..., 9\}^{\mathbb{Z}_+}$. The test problem then reads as follows: Given an observation $x \in [0, 1]$ by its decimal expansion $0.x_0\, x_1 \ldots x_n$ up to the $n + 1$st digit, test whether x is more likely to be a generic point for μ_0 or μ_1.*

This type of problem was recently studied in [15] and [12] using Birkhoff averages of the Jacobians to determine the classes of tests. Here we determine the Neyman-Pearson tests for the test problem $\mathcal{E}_n$ thus deriving the most powerful tests in the class $\mathcal{T}_n$. We also study the asymptotic behavior of these tests using large deviation theory and determine the minimax tests and Bayes solutions for the test problem $\mathcal{E}_n$ and show that these tests converge to the minimax test (Bayes solution) for the test problem $\mathcal{E}$ with exponentially fast decaying risk functions.

Comparing the setting of the present paper with the one in [12], we mention that in [12] (which likewise considers hypothesis tests) it also used LDP properties and a relation with the topological pressure. However, there the arguments are concerned just to rejected areas taking into account a loss function related to Jacobians, more precisely, $\log J_0 - \log J_1$. A similar expression like $f_i'(t) = \int (\log J_{i'} - \log J_i) dm_{i,t}$ in Theorem 3.1 was obtained. One of the main differences is that here we introduced the test $\phi_{n,\alpha}^*$ in Lemma 2.1, which takes into account the measure of cylinders. This is a different point of view, using a more basic information, and therefore, much more suitable for applications. Theorem 3.1 makes the connection of these two viewpoints.

The paper [15] has a quite different goal. It does not consider hypothesis tests or results on decision theory like here. [15] takes into account the Bayesian point of view, and considers a large class of loss functions (including some non additive expressions which were not our objective here). The prior probability on the set of parameters Θ (which does not have to be finite) in [15] covers a more general case, determining a more complex random source; the main issue there was to determine which Gibbs probability μ_{θ_0} (associated to a certain parameter $\theta_0 \in \Theta$) is responsible for the generation of the samples obtained from the random source. There it was used a LDP version for the non additive case.

In Section 2 we introduce for each value n the corresponding Neyman-Pearson test and we describe some basic properties. Section 3 considers asymptotic results, when $n \to \infty$, and large deviation estimates. In Section 4 we consider minimax tests and Bayes solutions. In Section 5 we present some classical results on large deviations for thermodynamic formalism (see [7], [22], [14], [17] and [10] for general references).

For results somehow related to Statistics in a dynamical setting we refer the reader to [9], [18], [21], [19], [20], [16], [13], [6], [16] and [5]. Classical results in Decision Theory can be found in [11], [24], [4] or [1]. Nice references in Thermodynamic Formalism are [23], [2] and [3].

2 Neyman-Pearson Tests

We keep the notation introduced in Section 1, in particular the notation for the Markov partition γ. For $n \geq 0$, we denote by γ_n ($n \in \mathbb{Z}_+ = \{0, 1, 2, ...\}$) the refinement of the partitions $T^{-j}\gamma$, with $j = 0, ..., n$. We also use the notation $S_n g = g + g \circ T + ... + g \circ T^n$ for a measurable map $g : \Omega \to \mathbb{R}$ and T_Γ^{-n-1} for the inverse mapping of $T^{n+1} : \Gamma \to T^{n+1}\Gamma$, where $\Gamma \in \gamma_n$. Finally, $\mathbb{I}_B$ stands for the characteristic function of the set B.

By the eigenvalue property (see [23]) of a Gibbs measure we have for $\Gamma \in \gamma_n$, $i = 0, 1$:

$$\mu_i(\Gamma) = \int \mathcal{L}^n_{\log J_i} \mathbb{I}_\Gamma d\mu_i = \int_{T^{n+1}\Gamma} \exp\{S_n \log J_i(T_\Gamma^{-n-1}(z))\} \mu_i(dz). \tag{2.1}$$

Lemma 2.1. *[see e.g. [11], p. 201] The Neyman-Pearson tests at level α are given by the formulas*

$$\phi^*_\alpha(\omega) = \begin{cases} 1, & \text{if } \omega \in S \\ \alpha, & \text{if } \omega \notin S \end{cases} \tag{2.2}$$

for the test problem $\mathcal{E}$ and - for the test problem $\mathcal{E}_n$ $(n \geq 0)$ - by

$$\phi^*_{n,\alpha}(\omega) = \begin{cases} 1 & \omega \in \Gamma \in \gamma_n; \quad \int_{T^{n+1}\Gamma} \exp\{S_n \log J_1(T_\Gamma^{-n-1}(z))\}\mu_1(dz) \\ & \qquad > c_{n,\alpha} \int_{T^{n+1}\Gamma} \exp\{S_n \log J_0(T_\Gamma^{-n-1}(z))\}\mu_0(dz) \\ 0 & \omega \in \Gamma \in \gamma_n; \quad \int_{T^{n+1}\Gamma} \exp\{S_n \log J_1(T_\Gamma^{-n-1}(z))\}\mu_1(dz) \\ & \qquad < c_{n,\alpha} \int_{T^{n+1}\Gamma} \exp\{S_n \log J_0(T_\Gamma^{-n-1}(z))\}\mu_0(dz) \\ \chi_{n,\alpha} & \omega \in \Gamma \in \gamma_n; \quad \int_{T^{n+1}\Gamma} \exp\{S_n \log J_1(T_\Gamma^{-n-1}(z))\}\mu_1(dz) \\ & \qquad = c_{n,\alpha} \int_{T^{n+1}\Gamma} \exp\{S_n \log J_0(T_\Gamma^{-n-1}(z))\}\mu_0(dz), \end{cases} \tag{2.3}$$

where $c_{n,\alpha} \in \mathbb{R}_+$ and $\chi_{n,\alpha} \in [0,1]$ are uniquely determined constants so that

$$\int \phi^*_{n,\alpha} d\mu_0^n = \alpha.$$

Proof. Taking the sum of the two measures involved for each of the test problems as their dominating measure and computing the densities we find for the test problem $\mathcal{E}$ the densities

$$\frac{d\mu_0}{d(\mu_0 + \mu_1)} = \mathbb{I}_{\Omega \setminus S}, \qquad \frac{d\mu_1}{d(\mu_0 + \mu_1)} = \mathbb{I}_S$$

and for the test problem $\mathcal{E}_n$ $(n \geq 0)$

$$\frac{d\mu_0^n}{d(\mu_0^n + \mu_1^n)}(\omega) = \mu_0(\Gamma), \qquad \frac{d\mu_1^n}{d(\mu_0^n + \mu_1^n)}(\omega) = \mu_1(\Gamma), \qquad \omega \in \Gamma \in \gamma_n,$$

Note that for $\Gamma \in \gamma_n$ the value $\mu_i(\Gamma)$, $i = 0, 1$, can be calculated by (2.1). The lemma follows from the Neyman-Pearson lemma, as formulated in [11], page 201, for example, which says that the Neyman-Pearson tests are defined by the quotients $\omega \mapsto \mu_1(\Gamma)/\mu_0(\Gamma)$, where $\omega \in \Gamma \in \gamma_n$. □

Remark: It follows from properties of the relative entropy of μ_0 and μ_1 (which are two distinct ergodic probabilities), that when n goes to infinity, the quotients of the integrals in each line of (2.3) will go to zero or infinity (see for instance [5]). The value $c_{n,\alpha}$ in some sense calibrate numerically these quotients. Therefore, the values 0 or 1, in the test defined by (2.3), will discriminate, when n is large, if the samples are being produced by the randomness of μ_0 or μ_1.

It follows immediately from the Neyman-Pearson lemma that these tests are optimal in the sense that the type 2 error $\int(1 - \phi)d\mu_1$ is minimal among all tests at level $\leq \alpha$. This is

Corollary 2.2. *The Neyman-Pearson tests defined in (2.2) and (2.3) are most powerful tests at their respective significance levels α.*

Let $L : \mathcal{P} \times \{0,1\} \to \mathbb{R}_+ = \{z \in \mathbb{R} | \ z \geq 0\}$ be a loss function and denote

$$\mathcal{R}(\mu, \phi) = \int L(\mu, t)\delta_\phi(\omega, dt))\mu(d\omega)$$

the associated risk function, where δ_ϕ denotes the decision function associated to the test ϕ, that is

$$\delta_\phi(\omega, \cdot) = \phi(\omega)\mathbb{I}_{\{0\}} + (1 - \phi(\omega))\mathbb{I}_{\{1\}}.$$

In the sequel we consider w.l.o.g. the Neyman-Pearson loss function for the simple test problem, that is

$$L(\mu, t) = \begin{cases} 1 & \text{if } \mu \in \mathcal{H}_0, t = 1 \text{ or } \mu \notin \mathcal{H}_0, t = 0 \\ 0 & \text{else.} \end{cases}$$

Recall that a test ϕ is called a minimax test if

$$\mathcal{R}(\phi) := \sup_{i \in \{0,1\}} \mathcal{R}(\mu_i, \phi) \leq \inf_{\phi'} \sup_{i \in \{0.1\}} \mathcal{R}(\mu_i, \phi') =: \mathcal{R}(\phi')$$

holds where $\mathcal{R}(\cdot, \phi')$ denotes the risk function of an arbitrary decision ϕ'. $\mathcal{R}(\phi)$ will be called the risk of the test (decision) ϕ.

Likewise a test ϕ is called a Bayes solution for the a priori distribution $\pi = (\pi_0, \pi_1)$ if

$$\mathcal{R}_\pi(\phi) := \int \mathcal{R}(\mu_i, \phi)\pi(di) \leq \int \mathcal{R}(\mu_i, \phi')\pi(di) =: \mathcal{R}_\pi(\phi')$$

holds for any test ϕ'. The Bayes risk of the test ϕ with respect to the a priori distribution π is $\mathcal{R}_\pi(\phi)$.

A well known consequence of Corollary 2.2 is

Proposition 2.3. *Let ϕ be a minimax test (or a Bayes solution with respect to the a priori distribution π, μ_i^n and $\mathcal{E}_n$). Then there exists a Neyman-Pearson test with the same risk function.*

Proof. Fix $n \in \mathbb{N}$. By definition

$$\sup_{i \in \{0,1\}} \mathcal{R}(\mu_i^n, \phi) \leq \sup_{i \in \{0,1\}} \mathcal{R}(\mu_i^n, \phi')$$

for all tests ϕ' of the test problem $\mathcal{E}_n$. Let $\alpha = E_{\mu_0^n}(\phi)$ denote the level of the test ϕ. Then $\phi_{n,\alpha}^*$ has level α as well and $E_{\mu_1^n}(1 - \phi_{n,\alpha}^*) \leq E_{\mu_1^n}(1 - \phi)$ so that

$$\mathcal{R}(\mu_\theta^n, \phi_{n,\alpha}^*) \leq \mathcal{R}(\mu_\theta^n, \phi), \qquad \theta \in \{0,1\}.$$

A similar argument works for the Bayes solution. $\square$

This implies the next proposition.

Proposition 2.4. *For each $n \geq 0$, there exists a minimax test and a Bayes solution to every a priori distribution π.*

Proof. The function

$$[0,1] \ni \alpha \mapsto \int \phi^*_{n\,\alpha} d\mu_i^n$$

is continuous for each $i = 0$ and $i = 1$. Indeed, if α increases also the corresponding $c_{n,\alpha}$ decreases, and if $c_{n,\alpha}$ is constant on some interval (α_0, α_1) the corresponding χ_α is inceasing. Thus

$$\int 1 - \phi^*_{n\alpha} g\mu_1 =$$

$$= \mu_1(\frac{d\mu_1}{d\mu_0 + \mu_1} < c_{n,\alpha}\frac{d\mu_0}{d\mu_0 + \mu_1}) + (1 - \chi_{n,\alpha})\mu_1(\frac{d\mu_1}{d\mu_0 + \mu_1} \leq c_{n,\alpha}\frac{d\mu_0}{d\mu_0 + \mu_1})$$

is decreasing and depends continuously on α. Therefore, the minimum of $\pi_0 \int \phi^*_{n,\alpha} d\mu_0 + \pi_1 \int (1 - \phi^*_{n,\alpha}) d\mu_1$ is attained, so it is a Bayes solution.

A similar argument works for the minimax test. □

3 Large deviation and Neyman-Pearson tests

We keep the notation from the last sections. Let $\mathcal{E}$ and $\mathcal{E}_n$ denote the test problems described in Section 1. We denote by J_i, $i = 0, 1$, the Jacobians corresponding, respectively, to μ_i, $i = 0, 1$ (cf. (1.2)). Accordingly, (1.1) will be taken with respect to these Jacobians. Furthermore, for each $n \geq 1$ and $0 \leq \alpha \leq 1$ the Neyman-Pearson test for the test problem $\mathcal{E}_n$ at level α is denoted by $\phi^*_{n,\alpha}$, see Lemma 2.1.

We shall use several facts from large deviation theory for Gibbs measures which are collected in an Appendix (Section 5).

Theorem 3.1. *The free energy functions*

$$f_i : \mathbb{R} \to \mathbb{R}, \quad (i \in \{0, 1\})$$

$$f_i(t) = \lim_{n\to\infty} \frac{1}{n} \log \int \exp\left\{ t \log \frac{\int_{T^n\Gamma} \exp\{S_n \log J_{i'}(T_\Gamma^{-n}(z))\}\mu_{i'}(dz)}{\int_{T^n\Gamma} \exp\{S_n \log J_i(T_\Gamma^{-n}(z))\}\mu_i(dz)} \right\} d\mu_i,$$

where $i' = i + 1 \mod 2$, *exist, are twice differentiable and satisfy*

$$\begin{aligned} f_i(t) &= P(\log J_i + t(\log J_{i'} - \log J_i)) \\ f_i'(t) &= \int (\log J_{i'} - \log J_i) dm_{i,t} \\ f_i''(t) &= \lim_{n\to\infty} \frac{1}{n} \int [S_n(\log J_{i'} - \log J_i - f_i'(t))]^2 dm_{i,t}, \end{aligned}$$

where for $i = 0, 1$ $m_{i,t}$ denotes the unique Gibbs measure for the potential $\log J_{i,t} = \log J_i + t(\log J_{i'} - \log J_i)$ *and where $P(\cdot)$ denotes the pressure function (its definition is recalled in the appendix).*

Proof. Let $i \in \{0, 1\}$ and let t be fixed. There exists a constant K such that for $\Gamma \in \gamma_n$, $n \geq 1$ and $i \in \{0, 1\}$

$$\begin{aligned} K^{-1} &\leq \frac{\mu_i(\Gamma)}{\exp\{-nP(\log J_i) + S_n \log J_i(z)\}} \leq K, \qquad z \in \Gamma, \\ K^{-1} &\leq \frac{m_{i,t}(\Gamma)}{\exp\{-nP(\log J_{i,t}) + S_n \log J_{i,t}(z)\}} \leq K, \qquad z \in \Gamma, \quad (3.1) \\ K^{-1} &\leq \exp\{S_n \log J_i(z) - S_n \log J_i(y)\} \leq K, \qquad z, y \in \Gamma. \end{aligned}$$

Writing

$$G_n^i(\omega) = \int_{T^n\Gamma} \exp\{S_n \log J_i(T_\Gamma^{-n} z)\} \mu_i(dz), \tag{3.2}$$

for $\omega \in \Gamma$ and using (3.1) it follows that

$$\left| \log G_n^{i'}(\omega) - \log G_n^i(\omega) - S_n \log \frac{J_{i'}}{J_i}(\omega) \right| \leq 2 \log[K] + \log \frac{\mu_{i'}(T^{n+1}\Gamma)}{\mu_i(T^{n+1}\Gamma)}$$

and so

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n} \log \int \exp\{t(\log G_n^{i'}(\omega) - \log G_n^i(\omega))\} \mu_i(d\omega) \\ &= \lim_{n\to\infty} \frac{1}{n} \log \int \exp(S_n(t \log \frac{J_{i'}}{J_i}))(\omega) \mu_i(\omega). \qquad (3.3) \end{aligned}$$

Considering (5.6) and (5.7), now apply relation (5.3) in the Appendix 5 to conclude that

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n} \log \int \exp\{t(\log G_n^{i'}(\omega) - \log G_n^i(\omega))\} \mu_i(d\omega) \\ &= P(\log J_i + t(\log J_{i'} - \log J_i)) - P(\log J_i). \end{aligned}$$

$P(\log J_i) = 0$, we arrive at

$$\begin{aligned} f_0(t) &= P(\log J_0 + t(\log J_1 - \log J_0)) \\ f_1(t) &= P(\log J_1 + t(\log J_0 - \log J_1)). \end{aligned}$$

The differentiability properties are well known for the pressure function (see equations 5.2, 5.4 and 5.5 in the Appendix 5). □

It is known that the ranges of the derivatives, restricted to $\mathbb{R}_+$, are $[f_i'(0), A_i]$ with

$$A_i = \lim_{n\to\infty} \text{ess sup} \frac{1}{n} S_n(\log J_{i'} - \log J_i),$$

the essential supremum is taken with respect to μ_i, where $i = \{0,1\}$, $i' = i+1 \mod 2$. Likewise the lower bounds for the ranges of the f_i''s on $\mathbb{R}$ are

$$\overline{A}_i = \lim_{n\to\infty} \text{ess inf} \frac{1}{n} S_n(\log J_{i'} - \log J_i).$$

Since both measures are supposed to be strictly positive on all cylinders, we have $A_i = -\overline{A}_{i'}$, $i = 0,1$, $i' = i+1 \mod 2$.

Lemma 3.2. *For $i \in \{0,1\}$ and $i' = i+1 \mod 2$, we have*

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{G_n^{i'}}{G_n^i} = f_i'(0) \qquad \mu_i \ a.s.$$

Proof. Note that

$$\frac{1}{n} \log \frac{G_n^{i'}}{G_n^i} = \frac{1}{n} \left(S_n(\log J_{i'} - \log J_i)\right) + o(n)$$

by the proof of Theorem 3.1, equation (3.3). Moreover,

$$f_i'(0) = \int (\log J_{i'} - \log J_i)\, d\mu_i$$

since μ_i is the equilibrium measure for the potential $\log J_i + t(\log J_{i'} - \log J_i)$ when $t = 0$. This proves the lemma by the ergodic theorem. $\square$

The large deviation property of the Neyman-Pearson tests can now be formulated in

Theorem 3.3. *For any $n \geq 1$, let $c_n \in \mathbb{R}_+$ be so that $c = \lim_{n\to\infty} \frac{1}{n} \log c_n$ exists and let ϕ^*_{n,α_n} denote a sequence of Neyman-Pearson tests for the test problem $\mathcal{E}_n$ with constants c_n given in (2.3).*

1. *The type 1 errors satisfy:*
   If $c = f_0'(t) \in (f_0'(0), A_0]$ then

$$\lim_{n\to\infty} \frac{1}{n} \log \int \phi^*_{n,\alpha_n} d\mu_0 = -t f_0'(t) + f_0(t). \tag{3.4}$$

If $c > A_0$ then

$$\lim_{n\to\infty} \frac{1}{n} \log \int \phi^*_{n,\alpha_n} d\mu_0 = -\infty. \tag{3.5}$$

If $c < f_0'(0)$ then

$$\lim_{n\to\infty} \frac{1}{n} \log \int \phi^*_{n,\alpha_n} d\mu_0 = 0. \tag{3.6}$$

.

2. *The type 2 errors satisfy*
   If $c = -f_1'(t) \in [-A_1, -f_1'(0))$, then

$$\lim_{n\to\infty} \frac{1}{n} \log \int (1 - \phi^*_{n,\alpha_n}) d\mu_1 = -t f_1'(t) + f_1(t). \tag{3.7}$$

If $c < -A_1$, then

$$\lim_{n\to\infty} \frac{1}{n} \log \int \phi^*_{n,\alpha_n} d\mu_0 = -\infty. \tag{3.8}$$

If $c \geq -f_1'(0)$, then

$$\lim_{n\to\infty} \frac{1}{n} \log \int \phi^*_{n,\alpha_n} d\mu_0 = 0. \tag{3.9}$$

Proof. 1. We show the first case (3.4). Using the notation in (3.2), for a suitable $\chi \in [0,1]$, chosen according to (2.3),

$$\begin{aligned}
\mu_0\left(\frac{G_n^1}{G_n^0} > c_n\right) &\leq \int \phi^*_{n,\alpha_n} d\mu_0 \\
&= \mu_0\left(\frac{G_n^1}{G_n^0} > c_n\right) + \chi\mu_0\left(\frac{G_n^1}{G_n^0} = c_n\right) \\
&\leq \mu_0\left(\frac{G_n^1}{G_n^0} \geq c_n\right)
\end{aligned}$$

By Markov's inequality for all $t > 0$

$$\begin{aligned}
\int \phi^*_{n,\alpha_n} d\mu_0 &\leq \mu_0\left(t \log \frac{G_n^1}{G_n^0} \geq t \log c_n\right) \\
&\leq \exp\{-t \log c_n + n f_{0,n}(t)\},
\end{aligned}$$

where

$$f_{0,n}(t) = \frac{1}{n} \log \int \exp\{t \log \frac{G_n^1}{G_n^0}\} d\mu_0.$$

Taking the infimum over $t > 0$ yields for n sufficiently large

$$\int \phi^*_{n,\alpha_n} d\mu_0 \le K_1 \exp\{-tnf_0'(t) + nf_0(t)\},$$

where t satisfies $f_0'(t) = \lim_{m\to\infty} \frac{1}{m} \log c_m$ and K_1 is some universal constant.

For the lower bound of (3.4) note that a Gibbs measure m with Jacobian J satisfies (see (3.1) and by T-invariance)

$$K^{-3} \le \frac{m([c_0, ..., c_{p+q-1}])}{m([c_0, ..., c_{p-1}]) \cdot m([c_p, ..., c_{p+q-1}])} \le K^3$$

for $p, q \ge 1$ and $[c_0, ..., c_{p+q-1}] \neq \emptyset$. Moreover, for a topologically mixing subshift of finite type there exists a constant $r \ge 1$ such that any cylinders $c, d \subset \Omega$ the set $c \cap T^{-r} d \neq \emptyset$. Since for $\omega \in \Gamma \in \gamma_n$ by (3.1)

$$\log \frac{G_n^1}{G_n^0}(\omega) - nf_0'(0) \ge K^2 \cdot \frac{\mu_1(T^n\Gamma)}{\mu_0(T^n\gamma)} \exp\{S_{n+1}[\log J_1 J_0^{-1} - f_0'(0)](\omega)\},$$

it follows that for n sufficiently large

$$\begin{aligned}\int \phi^*_{n,\alpha_n} d\mu_0 &\ge \mu_0\left(\log \frac{G_n^1}{G_n^0} - nf_0'(0) > \log c_n - nf_0'(0)\right) \\ &\ge \mu_0(S_{n+1}(\log J_1 J_0^{-1} - f_0'(0)) \ge \log c_n - nf_0'(0) + O(1)).\end{aligned}$$

Therefore the proof of Theorem 3.3 in [8] applies with minor adaptions as well for this case, proving the lower bound. In order to see this, note that the coordinate process of a Gibbs measure is ψ-mixing, so Theorem 3.3 in [8] is applicable here to partial sums above in view of (3.1). Alternatively, the arguments for its proof also work for cylinders. Moreover, one also can use [22].

Now we will show (3.5) and (3.6). If $c > A_0$ and $t > 0$, we have $\frac{d}{dt}[-tnc + nf_0(t)] = -nc + nf_0'(t) \le C < 0$, for some $C < 0$, so that the infimum is attained for $t \to \infty$.

If $c < f_0'(0)$ then by Lemma 3.2 $\log \frac{G_n^1}{G_n^0} - f_n'(0) \to 0 > c - f_0'(0)$, μ_0 a.s..

2. Using the notation in (3.1), for a suitable $\chi \in [0, 1]$

$$\begin{aligned} & \mu_1\left(\frac{G_n^1}{G_n^0} < c_n\right) \le \int 1 - \phi^*_{n,\alpha_n} d\mu_1 \\ = \quad & \mu_1\left(\frac{G_n^1}{G_n^0} < c_n\right) + (1-\chi)\mu_1\left(\frac{G_n^1}{G_n^0} = c_n\right) \le \mu_1\left(\frac{G_n^1}{G_n^0} \le c_n\right) \\ = \quad & \mu_1\left(\frac{G_n^0}{G_n^1} \ge \frac{1}{c_n}\right)\end{aligned}$$

Now this case is handled as case $i = 0$.

$\square$

4 Minimax tests and Bayes solutions

Here we prove the rate of convergence for the risk of the minimax tests and Bayes solutions in $\mathcal{E}_n$. We discuss the case of minimax tests first, the analogous arguments work for the Bayes solutions so that we only formulate those results.

We begin with

Lemma 4.1. *There exists a minimax test ψ_n^* in $\mathcal{E}_n$ that satisfies*

$$\int \psi_n^* d\mu_0 = 1 - \int \psi_n^* d\mu_1. \tag{4.1}$$

In particular, this test can be chosen to be a Neyman-Pearson test.

Proof. Let α denote the significance level of a minimax test $\psi_n^* \in \mathcal{T}_n$, where n is some fixed integer. Let $\beta = \int \psi_n^* d\mu_1$ be its power.

If $1 - \beta > \alpha$ then a Neyman-Pearson test $\phi_{n,\alpha'}^*$ at level $\alpha' \in [\alpha, 1-\beta]$ has at most a type 2 error of $1-\beta$, because it has a lower type 2 error than ψ_n^*. If for all such α' the Neyman-Pearson test has power β, then $\phi_{n,1-\beta}^*$ is a Neyman-Pearson test satisfying the requirements of the lemma. If the power is strictly larger than β for some α', then the test $\phi_{n,\alpha'}^*$ has a smaller risk than ψ_n^*, which is impossible. This proves the lemma if $1 - \beta > \alpha$.

If $1 - \beta = \alpha$ the assertion follows from the same argument as has been used in the proof of Proposition 2.3.

If $1 - \beta < \alpha$, a Neyman-Pearson test at level α has a power larger than or equal to β. This implies

$$\mathcal{R}(\phi_{n,\alpha}^*) = \alpha = \mathcal{R}(\psi_n^*)$$

and

$$\int \phi_{n,\alpha}^* d\mu_1 \geq \beta > 1 - \alpha.$$

Assume that

$$\int \phi_{n,\alpha}^* d\mu_0 = \alpha > 1 - \int \phi_{n,\alpha}^* d\mu_1.$$

Since the power of a Neyman-Pearson test is continuous, there is $\alpha' < \alpha$ such that

$$\int \phi_{n,\alpha'} d\mu_1 > \beta - (\beta - (1-\alpha)) = 1 - \alpha,$$

hence

$$\max\{\alpha', 1 - \int \phi^*_{n,\alpha'} d\mu_1\} = \mathcal{R}(\phi^*_{n,\alpha'}) < \alpha = \mathcal{R}(\psi^*_n),$$

a contradiction.

This finishes the proof. $\square$

Lemma 4.2. *For $i = 0, 1$ let $i' = i + 1 \mod 2$. Let*

$$F_i(t) = tf_i'(t) - f_i(t) = t \int (\log J_{i'} - \log J_i) dm_{i,t} - P(\log J_i + t(\log J_{i'} - \log J_i)),$$

$i = 0, 1$ and $i' = i + 1 \mod 2$, denote the (information) functions in (3.4) and (3.7), where $m_{i,t}$ denotes the unique equilibrium measure for the potential $\log J_i + t(\log J_{i'} - \log J_i)$. Then

1. *The functions F_i, $i = 0, 1$, are increasing on $(0, \infty)$.*
2. *$F_i(0) = 0$ for $i = 0, 1$.*
3. *For $0 \leq t \leq 1$ one has $m_{i,t} = m_{i',-t+1}$, $m_{i,0} = \mu_i$, $m_{1,1} = \mu_0$ and $m_{0,1} = \mu_1$.*
4. *For $0 \leq t \leq 1$ and $i = 0, 1$ one has $F_i(t) = F_{i'}(-t + 1) - 2t \int \log J_{i'} - \log J_i dm_{i',-t+1}$.*
5. *For $i = 0, 1$ one has $F_i(1) = \int (\log J_i - \log J_{i'}) d\mu_{i'} \geq 0$.*

Proof. 1. The derivative of F_i equals $F_i'(t) = tf_i''(t)$ which is positive on $\mathbb{R}_+$.

2.
$$F_i(0) = -f_i(0) = -P(\log J_i) = 0.$$

3. Note that $P(\log J_i + t(\log J_{i'} - \log J_i)) = P(\log J_{i'} + (t - 1)(\log J_{i'} - \log J_i))$ so that $m_{i,t} = m_{i',-t+1}$.

4. By 3. it follows that

$$\begin{aligned} F_i(t) &= t \int (\log J_{i'} - \log J_i)\, dm_{i,t} - P(\log J_i + t(\log J_{i'} - \log J_i)) \\ &= -t \int (\log J_i - \log J_{i'})\, dm_{i',-t+1} - P(\log J_{i'} + (-t + 1)(\log J_i - \log J_{i'})) \\ &= F_{i'}(-t + 1) - 2t \int (\log J_i - \log J_{i'}) dm_{i',-t+1}. \end{aligned}$$

5. This is obvious from 2., 3. and 4.: $F_i(1) = \int(\log J_{i'} - \log J_i)\, dm_{i,1} = \int(\log J_{i'} - \log J_i)\ d\mu_{i'}$. It follows from the variational principle and Rohklin's formula that

$$\int (\log J_{i'} - \log J_i)\, d\mu_{i'} = -[h_{\mu_{i'}}(T) + \int \log J_i\ d\mu_{i'}] \geq -P(\log J_i) = 0.$$

$\square$

Theorem 4.3. *Let ψ_n^* be a sequence of minimax tests in $\mathcal{E}_n$, $n \geq 1$. Then their risks*

$$\mathcal{R}(\psi_n^*) = \max\{\int \psi_n^* d\mu_0,\ 1 - \int \psi_n^* d\mu_1\}$$

satisfy

$$\lim_{n\to\infty} \frac{1}{n} \log \mathcal{R}(\psi_n^*) \leq \inf\{\max\{f_0(t) - tf_0'(t), f_1(s) - sf_1'(s)\}\} \tag{4.2}$$

where the infimum extends over all pairs (t, s) with $t \in (f_0'(0), A_0)$, $s \in (-A_0, -f_1'(0))$ and $f_0'(t) = -f_1'(s)$. More precisely,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathcal{R}(\psi_n^*) = f_0(t_0) - t_0 f_0'(t_0), \tag{4.3}$$

where t_0 is the solution of the equations

$$\begin{aligned} &f_0'(t) = -f_1'(s(t)) \\ &F_0(t_0) = \min\{F_0(t)|\ s(t)s'(t)f_1''(s(t)) - tf_0''(t) = 0\}. \end{aligned} \tag{4.4}$$

This solution is unique.

Proof. By Lemma 4.1 we may assume that the tests are Neyman-Pearson tests satisfying (4.1). Let $\phi_{n,\alpha_n}^* := \phi_n^*$ denote the Neyman-Pearson test for $\mathcal{E}_n$ with $\alpha_n = 1 - \int \phi_{n,\alpha_n}^* d\mu_1$. Let c_n denote the constant given by its definition as a Neyman-Pearson test.

We first show that

$$\overline{A}_0 \leq f_0'(0) \leq \liminf_{n\to\infty} \frac{1}{n} \log c_n \leq \limsup_{n\to\infty} \frac{1}{n} \log c_n \leq A_0.$$

1) If there is a subsequence $\frac{1}{n_k} \log c_{n_k} > A_0$, then for all k sufficiently large

$$\int \phi_{n_k}^* d\mu_0 = 0,$$

since μ_0-almost surely

$$\log G^1_{n_k} - \log G^0_{n_k} \leq \log K + n_k A_0 < \log c_{n_k}.$$

Likewise,

$$\int (1 - \phi^*_{n_k}) d\mu_1 = 1,$$

since $-A_0 = \overline{A}_1$ and

$$\log G^0_{n_k} - \log G^1_{n_k} \geq - \log K - n_k A_0 > - \log c_{n_k}.$$

2. If there is a subsequence $\frac{1}{n_k} \log c_{n_k} < f_0'(0)$, then for all k sufficiently large

$$\int \phi^*_{n_k} d\mu_0 = 1,$$

since μ_0-almost surely by Lemma 3.2

$$\log G^1_{n_k} - \log G^0_{n_k} \geq \log K + n_k f_0'(0) + o(n_k) > \log c_{n_k}.$$

Likewise,

$$\int (1 - \phi^*_{n_k}) d\mu_1 = 0,$$

since by the variational principle $f_0'(0) = \int \log J_0 - \log J_1 d\mu_0 = -[h_{\mu_0} + \int \log J_1 d\mu_0] < 0$. Then,

$$\log G^0_{n_k} - \log G^1_{n_k} \geq - \log K - \log c_{n_k} > -n_k f_0'(0) > 0,$$

and μ_1-a.s. by Lemma 3.2

$$\begin{aligned} &\frac{1}{n}(\log G^0_{n_k} - \log G^1_{n_k}) \to \int (\log J_0 - \log J_1) d\mu_1 \\ &= h_{\mu_1} + \int \log J_0 d\mu_1 < P(\log J_0) = 0. \end{aligned}$$

This is a contradiction.

It follows that the sequence c_n satisfies

$$-A_1 \leq f_0'(0) \leq \liminf_{n \to \infty} \frac{1}{n} \log c_n \leq \limsup_{n \to \infty} \frac{1}{n} \log c_n \leq A_0,$$

that is : c is contained in the image of the function f_0'. This also implies that $-c$ is contained in the interval $[-A_0, -f_0'(0)] \subset [\overline{A}_1, -\overline{A}_0] = [\overline{A}_1, A_1]$, which is the image of f_1'.

Assume first that $c = \lim_{n\to\infty} \frac{1}{n} \log c_n \in [f_0'(0), A_0]$ exists. Then there exists $t \geq 0$ with $c = f_0'(t)$. Moreover, $-c \in [-A_0, -f_0'(0)] \subset [\overline{A}_1, A_1]$ means that there is s with $f_1'(s) = -c$. It then follows that by Theorem 3.3

$$\lim_{n\to\infty} \frac{1}{n} \log \mathcal{R}(\phi_n^*) \leq \max\{-tf_0'(t) + f_0(t),\ -sf_1'(s) + f_1(s)\}.$$

By Lemma 4.1 we also must have that

$$-tf_0'(t) + f_0(t) = -sf_1'(s) + f_1(s)$$

and this value must be minimal. Since by Lemma 4.2 each of these functions is strictly decreasing, but the function $t \to s(t)$ defined by $f_0'(t) = -f_1'(s(t))$ is strictly increasing, it follows that the function $t \mapsto -s(t)f_1'(s(t)) + f_1(s(t))$ is increasing. This means that there is a unique t_0 with

$$-t_0 f_0'(t_0) + f_0(t_0) = -s(t_0)f_1(s(t_0)) + f_1(s(t_0))$$

and

$$f_0'(t_0) = f_1'(s(t_0)).$$

In particular, we must have that $c = \lim_{n\to\infty} \frac{1}{n} \log c_n$ exists because the functions f_i' are strictly increasing. $\square$

Bayes solutions can be handled much in the same way as the minimax test. Let $\pi = (\pi_0, \pi_1)$ be probability vector and let

$$\mathcal{R}_\pi(\phi) = \pi_0 \int \phi d\mu_0 + \pi_1 \int (1 - \phi) d\mu_1$$

to denote the Bayes risk for the Bayes distribution π of the test ϕ given the test problem $\mathcal{E}_n$.

Theorem 4.4. *Let $\pi = (\pi_0, \pi_1)$ be a Bayes prior distribution. Then, the Bayes solutions $\psi_{\pi,n}^*$ with respect to π for the test problem $\mathcal{E}_n$ have risks satisfying*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathcal{R}_\pi(\psi_{\pi,n}^*) \leq \inf\{\pi_0(-tf_0'(t) + f_0(t)) + \pi_1(-sf_1'(s) + f_1(s))\},$$

where the infimum extends over all pairs (s, t) so that $f_0'(t) = -f_1'(s)$.

More precisely, let t_π be chosen so that

$$-\pi_0(t_\pi f_0'(t_\pi) + f_0(t_\pi)) + \pi_1(s(t_\pi)f_0'(t_\pi) + f_1 s((t_\pi)))$$

and

$$f_0'(t_\pi) = f_1'(s(t_\pi)).$$

Then t_π is uniquely determined and satisfies

$$\lim_{n\to\infty} \frac{1}{n} \log \mathcal{R}_\pi(\psi_{\pi,n}^*) = 2\pi_0(-t_\pi f_0'(t_\pi) + f_0(t_\pi)).$$

5 Appendix on large deviation

Let $d \geq 2$ be an integer and $M = (m_{ij})_{1 \leq i,j \leq d}$ be an integral matrix with entries on $\{0, 1\}$. Gibbs states on mixing subshifts of finite type

$$\Omega = \{(\omega_n)_{n \geq 0} |\ 1 \leq \omega_n \leq d;\ m_{\omega_n, \omega_{n+1}} = 1\}$$

were introduced by Bowen in [3]. For a given Hölder continuous function $g : \Omega \to \mathbb{R}$ there exists a Gibbs measure μ_g such that

$$h_{\mu_g}(T) + \int g(\omega)\mu_g(d\omega) = \max\{h_m(T) + \int g(\omega)m(d\omega) |\ m \circ T^{-1} = m,\ m(\Omega = 1\}, \tag{5.1}$$

where $h_m(T)$ denotes the entropy of the invariant probability m, which by Rohklin's theorem satisfies $h_m(T) = -\int \log J dm$ where J^{-1} is the Jacobian of m (see the introduction for our use of the Jacobian). In particular, a Gibbs measure m for the potential $\log J$ satisfies $P(\log J) := h_m(T) + \int \log J dm = 0$ (Bowen's formula). The right hand side of equation (5.1) can be chosen as a definition of the pressure $P(g)$ for any continuous function g, hence $P : C(\Omega) \to \mathbb{R}$. It is well known that the function P is Gateaux differentiable in the sense that for Hölder continuous functions $g, h \in C(\Omega)$

$$\frac{d}{dt}P(g + th) = \int h(\omega)m_t(d\omega), \tag{5.2}$$

where m_t denotes the Gibbs measure for the function $g + th$. The free energy function f_h for a Hölder continuous function h with respect to the Gibbs measure μ_g exists and satisfies

$$f_h(t) = \lim_{n\to\infty} \frac{1}{n} \log \int \exp\{t \cdot S_n h\} d\mu_g = P(g + th) - P(g) \tag{5.3}$$

and hence is differentiable on its domain with first and second derivative

$$f_h'(t) = \int h dm_t, \tag{5.4}$$

m_t the equilibrium state for the potential $g + th$, and

$$f_h''(t) = \lim_{n\to\infty} \frac{1}{n} \int (S_n(h - \int h dm_t))^2 dm_t. \tag{5.5}$$

The domain of f_h is the real line, but the range of its derivative is a subinterval $(a, b) \subset \mathbb{R}$ defined by

$$\begin{aligned} a &= \lim_{n\to\infty} \text{ess inf} \frac{1}{n} S_n h \\ b &= \lim_{n\to\infty} \text{ess sup} \frac{1}{n} S_n h. \end{aligned}$$

We denote by $z \to I(z)$ the Legendre transform of the analytic function $t \to f_h(t)$. Then,

1. given an open interval $(a, b) \subset \mathbb{R}$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mu_g \left\{ y \in \Omega \,|\, \frac{1}{n} \sum_{i=0}^{n-1} h(T^i(y)) \in (a, b) \right\} \geq - \inf\{I(z) \,|\, z \in (a, b)\}. \tag{5.6}$$

2. given a closed interval $[a, b] \subset \mathbb{R}$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mu_g \left\{ y \in \Omega \,|\, \frac{1}{n} \sum_{i=0}^{n-1} h(T^i(y)) \in [a, b] \right\} \leq - \inf\{I(z) \,|\, z \in [a, b]\}. \tag{5.7}$$

For a proof see [7], [14] or [17].

6 Appendix on Statistical terminology and definitions

Here we collect basic facts and definition on statistical decision theory which are used in this note. It is included to make the paper self-contained for the readership in dynamical systems.

A *statistical experiment* is a triple $\mathcal{E} := (\Omega, \mathcal{F}, \mathcal{P})$ consisting of a measurable space $(\Omega, \mathcal{F})$ together with a family $\mathcal{P}$ of probability measures on $(\Omega, \mathcal{F})$. Here $\mathcal{F}$ denotes a σ-algebra on Ω. The objective is to make a decision about the *true* probability in $\mathcal{P}$ once a point in Ω is observed. For example, Ω may be chosen to be $\mathbb{R}^n$ and $\mathcal{P}$ may be chosen to be all Gaussian distributions which are the n-fold product measure of a one-dimensional normal distribution with expectation $\mu \in \mathbb{R}$ and variance 1. The objective may be to find the true μ.

Decisions are made with certain probabilities. Formally this is described by a measurable space $(D, \mathcal{D})$ (where $\mathcal{D}$ denotes the σ-algebra on D). It is called the space of decisions and a decision function is a stochastic kernel

$$\delta : \Omega \times \mathcal{D} \to [0, 1]$$

with the interpretation that a decision is in $d \in \mathcal{D}$ with probability $\delta(\omega, d)$ provided the observation is $\omega \in \Omega$. Such decisions (decision functions) in

$[0, 1]$ are called randomized decisions (decision functions) in contrary to the non-random case when the probability $\delta(\omega, d)$ is either 0 or 1. Let us denote the collection of all decision functions δ for a fixed statistical experiment by Δ.

It is common in statistics to value a decision using loss functions

$$L : \mathcal{P} \times D \to \mathbb{R}_+$$

which measures the *loss* of a decision $d \in D$ when P is "true". It is assumed that for all $P \in \mathcal{P}$ the function $d \to L(P, d)$ is measurable.

A *statistical problem* is then defined by $(\mathcal{E}, \Delta, L)$ where the triple is explained above. A *test problem* is a special statistical problem, specified by a subset $\mathcal{H}_0 \subset \mathcal{P}$, the hypothesis, the decision space $D = \{0, 1\}$ and a loss function L. L takes on the form

$$L(P, d) = \begin{cases} L_0 & \text{if } P \in \mathcal{H}_0 \text{ and } d = 1 \\ L_1 & \text{if } P \notin \mathcal{H}_0 \text{ and } d = 0 \\ 0 & \text{else} \end{cases}$$

The test problem is called *simple* if $\mathcal{H}_0$ and its complement consist of exactly one probability. This scenario is underlying the present article and we mostly assume that the loss function is of Neyman-Pearson type, that is $L_0 = L_1 = 1$.

We restrict the discussion to the special case of a simple test problem since it is the objective in this paper. A test is the function

$$\varphi : \Omega \to [0, 1]$$

defined as

$$\varphi(\omega) = \delta(\omega, \{1\})$$

where $\delta \in \Delta$ is a decision function. Randomized and non-randomized tests are those where δ has the corresponding property. The decision $d = 1$ then means that the observation suggests that the unknown distribution in $\mathcal{P}$ does not belong to $\mathcal{H}_0$ while $d = 0$ means that the unkown distribution belongs to $\mathcal{H}_0$. In the first case one rejects the hypothesis while in the second one does not reject the hypothesis.

Finally, tests φ for simple test problems are rated by their risk functions

$$R(\cdot, \delta) : \mathcal{P} \to [-\infty, \infty] \qquad \varphi = \delta(\cdot, \{1\})$$
$$R(P, \delta) = \int_\Omega \int_D L(P, \delta(\omega, t)\delta(\omega, dt)P(d\omega),$$

which amounts to the type 1 error (or significance level)

$$\alpha := R(P, \delta) = L_0 \int \varphi(\omega)P(d\omega),$$

for $P \in \mathcal{H}_0$ and to the type 2 error

$$1 - \beta := R(P, \delta) = L_1 \int 1 - \varphi(\omega) P(d\omega),$$

for $P \notin \mathcal{H}_0$. The value β is called the *power* of the test. One may assume that $L_0 = L_1$ when comparing tests.

The Neyman-Pearson Lemma characterizes those tests which have maximal power subject to keeping a given significance level α. It reads

Let $\mathcal{P} = \{\mu_0, \mu_1\}$, $\mathcal{H}_0 = \{\mu_0\}$ and μ be a dominating measure for μ_i, $i = 0, 1$, for example $\mu = \mu_0 + \mu_1$. Let f_i denote the densities of μ_i with respect to μ.

A Neyman-Pearson test φ is a test of the form

$$\varphi(\omega) = \begin{cases} 1 & \text{if } f_1(\omega) > C f_0(\omega) \\ 0 & \text{if } f_1(\omega) < C f_0(\omega) \\ \gamma(\omega) & \text{if } f_1(\omega) = C f_0(\omega) \end{cases}$$

where $\gamma(\omega) \in [0, 1]$ and $C \in [-\infty, \infty]$. In particular one may choose γ to be constant on $\{f_1 = C f_0\}$. Then we have the following facts:
1. A Neyman-Pearson test φ has maximal power among all tests ψ with

$$\int \psi d\mu_0 \leq \int \varphi d\mu_0.$$

2. Given $\alpha \in [0, 1]$ there is a Neyman-Pearson test φ satisfying

$$\int \varphi d\mu_0 = \alpha.$$

3. A test ψ at significance level α and with maximal power among all tests with significance level α is a.s. a Neyman-Pearson test.

References

[1] F. Abramovich and Y. Ritov, Statistical Theory, A Concise Introduction, CRC Press (2013).

[2] V. Baladi, Positive Transfer Operators and Decay of Correlations, World Scientific (2000).

[3] R. Bowen, Gibbs States and the Ergodic Theory of Anosov Diffeomorphisms, Lecture Notes in Math., **470**, Springer-Verlag, (1975).

[4] J. A. Bucklew, *Large Deviation Techniques in Decision, Simulation and Estimation.* New York: Wiley, (1990).

[5] J-R. Chazottes, R. Floriani and R. Lima, Relative entropy and identification of Gibbs measures in dynamical systems, J. Statist. Phys. 90 (1998) no. 3–4, 697–725.

[6] P. Collet, A. Galves and A. O. Lopes, Maximum likelihood and minimum entropy estimation of grammars, Random and Computational Dynamics, 3, pp-241–256 (1995).

[7] M. Denker, Large deviations and the pressure function. Transactions of the 11th Prague Conference on Information Theory, Statistical Decision Functions, Random Processes; Prague, 1990, 21–33. Academia Publ. House of the Czechoslovak Acad. of Science, 1992. = Mathematica Gottingensis **38**, (1988).

[8] M. Denker and Z. Kabluchko, An Erdös-Renyi law for mixing processes. Probability and Mathematical Statistics **27**, (2007), 139–149.

[9] M. Denker and G. Keller, Rigorous statistical procedures for data from dynamical systems. J. Stat. Physics **44** (1986), 67–93.

[10] R. S. Ellis, Entropy, Large Deviations, and Statistical Mechanics. Springer-Verlag New York, (1985).

[11] T.S. Ferguson, Mathematical Statistics. Academic Press New York and London, (1967).

[12] H. Ferreira, A. O. Lopes and S. R. C. Lopes, Decision theory and large deviations for dynamical hypothesies test: Neyman-Pearson, min-max and Baysian tests. Journal of Dynamics and Games. Volume 9, Number 2, April 2022 - pp 125-150

[13] C. Ji, Estimating functionals of one-dimensional Gibbs states, Probab. Th. Rel. Fields 82, (1989), 155–175.

[14] Y. Kifer, Large deviations in dynamical systems and stochastic processes. Trans. Amer. Math. Soc. 321 (1990), 505–524.

[15] A. O. Lopes, S. R. C. Lopes and P. Varandas, Bayes posterior convergence for loss functions via almost additive thermodynamic formalism, J. of Stat. Phys. 186:35 (2022)

[16] A. O. Lopes and J. K.Mengue, The generalized IFS Bayesian method and an associated variational principle covering the classical and dynamical cases, arXiv (2022)

[17] A. O. Lopes, Entropy, pressure and large deviation, cellular automata, dynamical systems and neural networks. In: Cellular Automata, Dynamical Systems and Neural Networks. Mathematics and Its Applications, **282**, E. Goles e S. Martinez (eds.), 79–146, Kluwer, Massachusets (1994).

[18] K. McGoff, S. Mukherjee and A. Nobel, Gibbs posterior convergence and thermodynamic formalism, Annals of Applied Probability. Ann. Appl. Probab. 32 (2022), no. 1, 461–496

[19] K. McGoff and A. B. Nobel, Empirical risk minimization and complexity of dynamical models. Ann. Statist. 48 (2020), no. 4, 2031–2054.

[20] K. McGoff, S. Mukherjee and N. Pillai, Statistical inference for dynamical systems: a review. Stat. Surv. 9 (2015), 209–252.

[21] A. B. Nobel, Hypothesis testing for families of ergodic processes, Bernoulli 12(2): 251-269 (April 2006).

[22] S. S. Orey, Large deviations in ergodic theory. In: Sem. on Stochastics. Proceedings 1984, pp 195–248, Birkhäuser (1986).

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

[24] V. K. Rohatgi and A. K. Md. Ehsanes Saleh, An Introduction to Probability and Statistics. Second Edition, Wiley New York (2001).