

Article

## A Bayesian Decision-Theoretic Approach to Logically-Consistent Hypothesis Testing

Gustavo Miranda da Silva <sup>1</sup>, Luis Gustavo Esteves <sup>1,\*</sup>, Victor Fossaluza <sup>1</sup>, Rafael Izbicki <sup>2</sup> and Sergio Wechsler <sup>1</sup>

<sup>1</sup> Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 05508-090, Brazil; E-Mails: gmirandasilva@gmail.com(G.M.S.); victor.ime@gmail.com (V.F.); sw@ime.usp.br (S.W.)

<sup>2</sup> Department of Statistics, Federal University of São Carlos, São Carlos, 13565-905, Brazil; E-Mail: rafaelizbicki@gmail.com

\* Author to whom correspondence should be addressed; E-Mail: lesteves@ime.usp.br; Tel.: +55-11-3091-6324.

Academic Editor: Kevin H. Knuth

Received: 27 May 2015 / Accepted: 9 September 2015 / Published: 24 September 2015

---

**Abstract:** This work addresses an important issue regarding the performance of simultaneous test procedures: the construction of multiple tests that at the same time are optimal from a statistical perspective and that also yield logically-consistent results that are easy to communicate to practitioners of statistical methods. For instance, if hypothesis  $A$  implies hypothesis  $B$ , is it possible to create optimal testing procedures that reject  $A$  whenever they reject  $B$ ? Unfortunately, several standard testing procedures fail in having such logical consistency. Although this has been deeply investigated under a frequentist perspective, the literature lacks analyses under a Bayesian paradigm. In this work, we contribute to the discussion by investigating three rational relationships under a Bayesian decision-theoretic standpoint: coherence, invertibility and union consonance. We characterize and illustrate through simple examples optimal Bayes tests that fulfill each of these requisites separately. We also explore how far one can go by putting these requirements together. We show that although fairly intuitive tests satisfy both coherence and invertibility, no Bayesian testing scheme meets the desiderata as a whole, strengthening the understanding that logical consistency cannot be combined with statistical optimality in general. Finally, we associate Bayesian hypothesis testing with Bayes point estimation procedures. We prove the performance of logically-consistent hypothesis testing by means of a Bayes point estimator to be optimal only under very restrictive conditions.

**Keywords:** Bayes tests; decision theory; logical consistency; loss functions; multiple hypothesis testing

---

## 1. Introduction

One could (...) argue that ‘power is not everything’. In particular for multiple test procedures one can formulate additional requirements, such as, for example, that the decision patterns should be logical, conceivable to other persons, and, as far as possible, simple to communicate to non-statisticians.

—G. Hommel and F. Bretz [1]

Multiple hypothesis testing, a formal quantitative method that consists of testing several hypotheses simultaneously [2], has gained considerable ground in the last few decades with the aim of drawing conclusions from data in scientific experiments regarding unknown quantities of interest. Most of the development of multiple hypothesis testing has been focused on the construction of test procedures satisfying statistical optimality criteria, such as the minimization of posterior expected loss functions or the control of various error rates. These advances are detailed, for instance, in [2], [3] (p. 7), [4] and the references therein. However, another important issue concerning multiple hypothesis testing, namely the construction of simultaneous tests that yield coherent results easier to communicate to practitioners of statistical methods, has not been so deeply investigated yet, especially under the Bayesian paradigm. As a matter of fact, most traditional multiple hypothesis testing schemes do not combine statistical optimality with logical consistency. For example, [5] (p. 250) presents a situation regarding the parameter,  $\theta$ , of a single exponential random variable,  $X$ , in which uniformly most powerful (UMP) tests of level 0.05 for the one-sided hypothesis  $H_0^{(1)} : \theta \leq 1$  and the two-sided hypothesis  $H_0^{(2)} : \theta \leq 1 \cup \theta \geq 2$ , say  $\varphi_1$  and  $\varphi_2$ , respectively, lead to puzzling decisions. In fact, for the sample outcome  $X = 0.7$ , the test  $\varphi_2$  rejects  $H_0^{(2)}$ , and because  $H_0^{(1)}$  implies  $H_0^{(2)}$ , one may decide to reject  $H_0^{(1)}$ , as well. On the other hand, the test  $\varphi_1$  does not reject  $H_0^{(1)}$ , a fact that makes a practitioner confused given these conflicting results. In this example, an inconsistency related to nested hypotheses named coherence [6] takes place. Frequently, other logical relationships one may expect from the conclusions drawn from multiple hypothesis testing, such as consonance [6] and compatibility [7], are not met either.

Although several of these properties have been deeply investigated under a frequentist hypothesis-testing framework, Bayesian literature lacks such analyses. In this work, we contribute to this discussion by examining three rational requirements in simultaneous tests under a Bayesian decision-theoretic perspective. In short, we characterize the families of loss functions that induce multiple Bayesian tests that satisfy partially such desiderata. In Section 2, we review and illustrate the concept of a testing scheme (TS), a mathematical object that assigns to each statistical hypothesis of interest a test function. In Section 3, we formalize three consistency relations one may find important to hold in simultaneous tests: coherence, union consonance and invertibility. In Section 4, we provide necessary and sufficient conditions on loss functions to ensure Bayesian tests to meet each desideratum separately, whatsoever the prior distribution for the relevant parameters is. In Section 5, we prove, under

quite general conditions, the impossibility of creating multiple tests under a Bayesian decision-theoretic framework that fulfill the triplet of requisites simultaneously with respect to all prior distributions. We also explore the connection between logically-consistent Bayes tests and Bayes point estimation procedures. Final remarks and suggestions for future inquiries are presented in Section 6. All theorems are proven in the Appendix.

## 2. Testing Schemes

We start by formulating the mathematical setup for multiple Bayesian tests. For the remainder of the manuscript, the parameter space is denoted by  $\Theta$  and the sample space by  $\mathcal{X}$ . Furthermore,  $\sigma(\Theta)$  and  $\sigma(\mathcal{X})$  represent  $\sigma$ -fields of subsets of  $\Theta$  and  $\mathcal{X}$ , respectively. We consider the Bayesian statistical model  $(\mathcal{X} \times \Theta, \sigma(\mathcal{X} \times \Theta), \mathbf{P})$ . The  $\mathbf{P}$ -marginal distribution of  $\theta$ , namely the prior distribution for  $\theta$ , is denoted by  $\pi$ , while  $\pi_x(\cdot)$  represents the posterior distribution for  $\theta$  given  $X = x$ ,  $x \in \mathcal{X}$ . Moreover,  $P(\cdot|\theta)$  stands for the conditional distribution of the observable  $X$  given  $\theta$ , and  $L_x(\theta)$  represents the likelihood function at the point  $\theta \in \Theta$  generated by the sample observation  $x \in \mathcal{X}$ . Finally, let  $\Psi$  be the set of all test functions, that is the set of all  $\{0, 1\}$ -valued measurable functions defined on  $\mathcal{X}$ . As usual, “1” denotes the decision of rejecting the null hypothesis and “0” the decision of not rejecting or accepting it.

Next, we review the definition of a TS, a mathematical device that formally describes the idea that to each hypothesis of interest it is assigned a test function. Although the specification of the hypotheses of interest most of the times depends on the scientific problem under consideration, here, we assume that a decision-maker has to assign a test to each element of  $\sigma(\Theta)$ . This assumption not only enables us to precisely define the relevant consistency properties, but it also allows multiple Bayesian testing based on posterior probabilities of the hypotheses (a deeper discussion on this issue may be found in [3] (p. 5) and [8]).

**Definition 1.** (Testing scheme (TS)) *Let the  $\sigma$ -field of subsets of the parameter space  $\sigma(\Theta)$  be the set of hypotheses to be tested. Moreover, let  $\Psi$  be the set of all test functions defined on  $\mathcal{X}$ . A TS is a function  $\varphi : \sigma(\Theta) \rightarrow \Psi$  that assigns to each hypothesis  $A \in \sigma(\Theta)$  the test  $\varphi_A \in \Psi$  for testing  $A$ .*

Thus, for  $A \in \sigma(\Theta)$  and  $x \in \mathcal{X}$ ,  $\varphi_A(x) = 1$  represents the decision of rejecting the hypothesis  $A$  when the datum  $x$  is observed. Similarly,  $\varphi_A(x) = 0$  represents the decision of not rejecting  $A$ . We now present examples of testing schemes.

**Example 1.** (Tests based on posterior probabilities) *Assume  $\Theta = \mathbb{R}^d$  and  $\sigma(\Theta) = \mathcal{B}(\mathbb{R}^d)$ , the Borelians of  $\mathbb{R}^d$ . Let  $\pi$  be the prior probability distribution for  $\theta$ . For each  $A \in \sigma(\Theta)$ , let  $\varphi_A : \mathcal{X} \rightarrow \{0, 1\}$  be defined by:*

$$\varphi_A(x) = \mathbb{I} \left( \pi_x(A) < \frac{1}{2} \right),$$

where  $\pi_x(\cdot)$  is the posterior distribution of  $\theta$ , given  $x$ . This is the TS that assigns to each hypothesis  $A \in \mathcal{B}(\mathbb{R}^d)$  the test that rejects it when its posterior probability is smaller than  $1/2$ .

Recall that, under a Bayesian decision-theoretic perspective, a hypothesis testing for the hypothesis  $\Theta_0 \subseteq \Theta$  [5] (p. 214) is a decision problem in which the action space is  $\{0, 1\}$  and the loss function  $L : \{0, 1\} \times \Theta \rightarrow \mathbb{R}$  satisfies:

$$L(1, \theta) \geq L(0, \theta) \text{ for } \theta \in \Theta_0 \text{ and } L(1, \theta) \leq L(0, \theta) \text{ for } \theta \in \Theta_0^c, \tag{1}$$

that is,  $L$  is such that the wrong decision ought to be assigned a loss at least as large as that assigned to a correct decision (many authors consider strict inequalities in Equation (1)). We call such a loss function a (strict) hypothesis testing loss function.

A solution of this decision problem, named a Bayes test, is a test function  $\varphi^* \in \Psi$  derived, for each sample point  $x \in \mathcal{X}$ , by minimizing the expectation of the loss function  $L$  over  $\{0, 1\}$  with respect to the posterior distribution. That is, for each  $x \in \mathcal{X}$ ,

$$\varphi^*(x) = 1 \iff \mathbb{E}[L(1, \theta)|X = x] < \mathbb{E}[L(0, \theta)|X = x],$$

where  $\mathbb{E}[L(d, \theta)|X = x] = \int_{\Theta} L(d, \theta) d\pi_x(\theta)$ ,  $d \in \{0, 1\}$ . In the case of the equality of the posterior expectations, both zero and one are optimal decisions, and either of them can be chosen as  $\varphi^*(x)$ .

When dealing with multiple tests, one can use the above procedure for each hypothesis of interest. Hence, one can derive a Bayes test for each null hypothesis  $A \in \sigma(\Theta)$  considering a specified loss function  $L_A : \{0, 1\} \times \Theta \rightarrow \mathbb{R}$  satisfying Equation (1). This is formally described in the following definition.

**Definition 2.** (TS generated by a family of loss functions) *Let  $(\mathcal{X} \times \Theta, \sigma(\mathcal{X} \times \Theta), \mathbf{P})$  be a Bayesian statistical model. Let  $(L_A)_{A \in \sigma(\Theta)}$  be a family of hypothesis testing loss functions, where  $L_A : \{0, 1\} \times \Theta \rightarrow \mathbb{R}$  is the loss function for testing  $A \in \sigma(\Theta)$ . A TS generated by the family of loss functions  $(L_A)_{A \in \sigma(\Theta)}$  is any TS  $\varphi$  defined over  $\sigma(\Theta)$ , such that,  $\forall A \in \sigma(\Theta)$ ,  $\varphi_A$  is a Bayes test for hypothesis  $A$  with respect to  $\pi$  considering the loss  $L_A$ .*

The following example illustrates this concept.

**Example 2.** (Tests based on posterior probabilities) *Assume the same scenario as Example 1 and that  $(L_A)_{A \in \sigma(\Theta)}$  is a family of loss functions, such that  $\forall A \in \sigma(\Theta)$  and  $\forall \theta \in \Theta$ ,*

$$L_A(0, \theta) = \mathbb{I}(\theta \notin A) \text{ and } L_A(1, \theta) = \mathbb{I}(\theta \in A),$$

*that is,  $L_A$  is the 0–1 loss for  $A$  ([5] (p. 215)). The testing scheme introduced in Example 1 is a TS generated by the family of 0–1 loss functions.*

The next example shows a TS of Bayesian tests motivated by different epistemological considerations (see [9,10] for details), the full Bayesian significance tests (FBST).

**Example 3.** (FBST testing scheme) *Let  $\Theta = \mathbb{R}^d$ ,  $\sigma(\Theta) = \mathcal{B}(\mathbb{R}^d)$  and  $f(\cdot)$  be the prior probability density function (pdf) for  $\theta$ . Suppose that, for each  $x \in \mathcal{X}$ , there exists  $f(\cdot|x)$ , the pdf of the posterior distribution of  $\theta$ , given  $x$ . For each hypothesis  $A \in \sigma(\Theta)$ , let:*

$$T_x^A = \left\{ \theta \in \Theta : f(\theta|x) > \sup_{\theta \in A} f(\theta|x) \right\}$$

be the set tangent to the null hypothesis, and let  $ev_x(A) = 1 - \pi_x(T_x^A)$  be the Pereira–Stern evidence value for  $A$  (see [11] for a geometric motivation). One can define a TS  $\varphi$  by:

$$\varphi_A(x) = \mathbb{I}(ev_x(A) \leq c), \forall A \in \sigma(\Theta) \text{ and } \forall x \in \mathcal{X},$$

in which  $c \in [0, 1]$  is fixed. In other words, one does not reject the null hypothesis when its evidence is larger than  $c$ .

We end this section by defining a TS generated by a point estimation procedure, an intuitive concept that plays an important role in characterizing logically-consistent simultaneous tests.

**Definition 3.** (TS generated by a point estimation procedure) Let  $\delta : \mathcal{X} \rightarrow \Theta$  be a point estimator for  $\theta$  ([5] (p. 296)). The TS generated by  $\delta$  is defined by:

$$\varphi_A(x) = \mathbb{I}(\delta(x) \notin A).$$

Hence, the TS generated by the point estimator  $\delta$  rejects hypothesis  $A$  after observing  $x$  if, and only if, the point estimate for  $\theta$ ,  $\delta(x)$ , is not in  $A$ .

**Example 4.** (TS generated by a point estimation procedure) Let  $\Theta = \mathbb{R}$ ,  $\sigma(\Theta) = \mathcal{P}(\Theta)$  and  $X_1, \dots, X_n | \theta$  i.i.d.  $N(\theta, 1)$ . The TS generated by the sample mean,  $\bar{X}$ , rejects  $A \in \sigma(\Theta)$  when  $x$  is observed if  $\bar{x} \notin A$ .

### 3. The Desiderata

In this section, we review three properties one may expect from simultaneous test procedures: coherence, invertibility and union consonance.

#### 3.1. Coherence

When a hypothesis is tested by a significance test and is not rejected, it is generally agreed that all hypotheses implied by that hypothesis (its “components”) must also be considered as non-rejected.

—K. R. Gabriel [6]

The first property concerns nested hypotheses and was originally defined by [6]. It states that if hypothesis  $H_0^{(1)}$  implies hypothesis  $H_0^{(2)}$ , that is  $H_0^{(1)} \subseteq H_0^{(2)}$ , then the rejection of  $H_0^{(2)}$  implies the rejection of  $H_0^{(1)}$ . In the context of TSs, we have the following definition.

**Definition 4.** (Coherence) A testing scheme  $\varphi$  is coherent if:

$$\forall A, B \in \sigma(\Theta), A \subseteq B \Rightarrow \varphi_A \geq \varphi_B, \text{ i.e., } \forall x \in \mathcal{X}, \varphi_A(x) \geq \varphi_B(x).$$

In other words, if after observing  $x$ , a hypothesis is rejected, any hypothesis that implies it has to be rejected, as well.

The testing schemes introduced in Examples 1, 3 and 4 are coherent. Indeed, in Example 1, coherence is a consequence of the monotonicity of probability measures, while in Example 3, it follows from the fact that if  $A \subseteq B$ , then  $T_x^B \subseteq T_x^A$  and, therefore,  $ev_x(A) \leq ev_x(B)$ . In Example 4, coherence is immediate. On the other hand, testing schemes based on UMP tests or generalized likelihood ratio tests with a common fixed level of significance are not coherent in general. Neither are TSs generated by some families of loss functions (see Section 4). Next, we illustrate that even test procedures based on  $p$ -values or Bayes factors may be incoherent.

**Example 5.** Suppose that in a case-control study, one measures the genotype in a certain locus for each individual of a sample. Results are shown in Table 1. These numbers were taken from a study presented by [12] that had the aim of verifying the hypothesis that subunits of the gene  $GABA_A$  contribute to a condition known as methamphetamine use disorder. Here, the set of all possible genotypes is  $\mathbb{G} = \{AA, AB, BB\}$ . Let  $\gamma = (\gamma_{AA}, \gamma_{AB}, \gamma_{BB})$ , where  $\gamma_i$  is the probability that an individual from the case group has genotype  $i$ . Similarly, let  $\pi = (\pi_{AA}, \pi_{AB}, \pi_{BB})$ , where  $\pi_i$  is the probability that an individual of control group has genotype  $i$ .

In this context, two hypotheses are of interest: the hypothesis that the genotypic proportions are the same in both groups,  $H_0^G : \gamma = \pi$ , and the hypothesis that the allelic proportions are the same in both groups  $H_0^A : \gamma_{AA} + \frac{1}{2}\gamma_{AB} = \pi_{AA} + \frac{1}{2}\pi_{AB}$ . The  $p$ -values obtained using chi-square tests for these hypotheses are, respectively, 0.152 and 0.069. Hence, at the level of significance  $\alpha = 10\%$ , the TS given by chi-square tests rejects  $H_0^A$ , but does not reject  $H_0^G$ . That is, the TS leads a practitioner to believe that the allelic proportions are different in both groups, but it does not suggest any difference between the genotypic proportions. This is absurd! If the allelic proportions are not the same in both groups, the genotypic proportions cannot be the same either. Indeed, if the latter were the same, then  $\gamma_i = \pi_i, \forall i \in \mathbb{G}$ , and hence,  $\theta \in H_0^A$ . This example is further discussed in [8,13].

**Table 1.** Genotypic sample frequencies.

	AA	AB	BB	Total
Case	55	83	50	188
Control	24	42	39	105

Several other (in)coherent testing schemes are explored by [8,14].

Coherence is by far the most emphasized logical requisite for simultaneous test procedures in the literature. It is often regarded as a sensible property by both theorists and practitioners of statistical methods who perceive a hypothesis test as a two-fold (accept/reject) decision problem. On the other hand, adherents to evidence-based approaches to hypothesis testing [15] do not see the need for coherence. Under the frequentist approach to hypothesis testing, the construction of coherent procedures is closely associated with the so-called closure methods [16,17]. Many results on coherent classical tests are shown in [6,17], among others. On the other hand, coherence has not been deeply investigated from a Bayesian standpoint yet, except for [18], who relate coherence with admissibility and Bayesian optimality in certain situations of finitely many hypotheses of interest. In Section 4, we provide a characterization of coherent testing schemes under a decision-theoretic framework.



### 3.2. Invertibility

There is a duality between hypotheses and alternatives which is not respected in most of the classical hypothesis-testing literature. (...) suppose that we decide to switch the names of alternative and hypothesis, so that  $\Omega_H$  becomes  $\Omega_A$ , and vice versa. Then we can switch tests from  $\phi$  to  $\psi = 1 - \phi$  and the “actions” accept and reject become switched.

—M. J. Schervish [5] (p. 216)

The duality mentioned in the quotation above is formally described in the next definition.

**Definition 5.** (Invertibility) *A testing scheme  $\varphi$  satisfies invertibility if:*

$$\forall A \in \sigma(\Theta), \varphi_{A^c} = 1 - \varphi_A.$$

*In other words, it is irrelevant to decision-making which hypothesis is labeled as null and which is labeled as alternative.*

Unlike coherence, there is no consensus among statisticians on how reasonable invertibility is. While it is supported by many decision-theorists, invertibility is usually discredited by advocates of the frequentist theory owing to the difference between the interpretations of “not reject a hypothesis” and “accept a hypothesis” under various epistemological viewpoints (the reader is referred to [7] for a discussion on this distinction). As a matter of fact, invertibility can also be seen, from a logic perspective, as a version of the law of the excluded middle, which itself represents a gap between schools of logic ([19] (p. 32)). In spite of the controversies on invertibility, it seems to be beyond any argument the fact that the absence of invertibility in multiple tests may lead a decision-maker to be puzzled by senseless conclusions, such as the simultaneous rejections of both a hypothesis and its alternative. The following example illustrates this point.

**Example 6.** *Suppose that  $X|\theta \sim \text{Normal}(\theta, 1)$ , and consider that the parameter space is  $\Theta = \{-3, 3\}$ . Assume one wants to test the following null hypotheses:*

$$H_0^A: \theta = 3 \quad \text{and} \quad H_0^B: \theta = -3$$

*The Neyman–Pearson tests for these hypotheses have the following critical regions, at the level 5%, respectively:*

$$\{x \in \mathbb{R} : x < 1.35\} \quad \text{and} \quad \{x \in \mathbb{R} : x > -1.35\}.$$

*Hence, if we observe  $x = -0.5$ , we reject both  $H_0^A$  and  $H_0^B$ , even though  $H_0^A \cup H_0^B = \Theta$ !*

The testing schemes of Examples 2 and 4 satisfy invertibility. In Example 4, it is straightforward to verify this. In Example 2, it follows essentially from the equivalence  $\pi_x(A) < 1/2 \Leftrightarrow \pi_x(A^c) > 1/2$ . If  $\pi_x(A) \neq 1/2$  for each sample  $x$  and for all  $A \in \sigma(\Theta)$ , the unique TS generated by the 0–1 loss functions satisfies invertibility. Otherwise, there is a testing scheme generated by such losses that is still in line with this property. Indeed, for any  $A \in \sigma(\Theta)$  and  $x_0 \in \mathcal{X}$ , such that  $\pi_{x_0}(A) = 1/2$ , the decision of rejecting  $A$  (not rejecting  $A^c$ ) after observing  $x_0$  has the same expected loss as the decision of not

rejecting (rejecting) it. Thus, among all testing schemes generated by the 0–1 loss functions, which are all equivalent from a decision-theoretic point of view ([20] (p. 123)), a decision-maker can always choose a TS  $\varphi'$ , such that  $\varphi'_{A^c}(x_0) = 1 - \varphi'_A(x_0)$  for all  $A \in \sigma(\Theta)$ , and  $x_0 \in \mathcal{X}$ , such that  $\pi_{x_0}(A) = 1/2$ . Such a TS  $\varphi'$  meets invertibility.

### 3.3. Consonance

... a test for  $(\cup_{i \in I} H_i)^c$  versus  $(\cup_{i \in I} H_i)$  may result in rejection which then indicates that at least one of the hypotheses  $H_i, i \in I$ , may be true.

—H. Finner and K. Strassburger [21]

The third property concerns two hypotheses, say  $A$  and  $B$ , and their union,  $A \cup B$ . It is motivated by the fact that in many cases, it seems reasonable that a testing scheme that retains the union of these hypotheses should also retain at least one of them. This idea is generalized in Definition 6.

**Definition 6.** (Union Consonance) A TS  $\varphi$  satisfies the finite (countable) union consonance if for all finite (countable) set of indices  $I$ ,

$$\forall \{A_i\}_{i \in I} \subseteq \sigma(\Theta), \varphi_{\cup_{i \in I} A_i} \geq \min\{\varphi_{A_i}\}_{i \in I}.$$

In other words, if we retain the union of the hypotheses  $\cup_{i \in I} A_i$ , we should not reject at least one of the  $A_i$ 's.

There are several testing schemes that meet union consonance. For instance, TSs generated by point estimation procedures, TSs of Aitchison's confidence-region tests [22] and FBST TSs (under quite general conditions; see [8]) satisfy both finite and countable union consonance.

Although union consonance may not be considered as appealing as coherence for simultaneous test procedures, it was hinted at in a few relevant works. For instance, the interpretation given by [21] on the final joint decisions derived from partial decisions implicitly suggests that union consonance is reasonable: they suggest one should consider  $B := \cup_{A: \varphi_A(x)=1} A$  to be the set of all parameter values rejected by the simultaneous procedure at hand when  $x$  is observed. Under this reading, it seems natural to expect that  $\varphi_B(x) = 1$ , which is exactly what the union consonance principle states. As a matter of fact, the general partitioning principle proposed by these authors satisfies union consonance. It should also be mentioned that union consonance, together with coherence, plays a key role in the possibilistic abstract belief calculus [23]. In addition, an evidence-based approach detailed in [24] satisfies both consonance and invertibility.

We end this section by stating a result derived from putting these logical requirements together.

**Theorem 1.** Let  $\Theta$  be a countable parameter space and  $\sigma(\Theta) = \mathcal{P}(\Theta)$ . Let  $\varphi$  be a testing scheme defined on  $\sigma(\Theta)$ . The TS  $\varphi$  satisfies coherence, invertibility and countable union consonance if, and only if, there is a point estimator  $\delta : \mathcal{X} \rightarrow \Theta$ , such that  $\varphi$  is generated by  $\delta$ .

Theorem 1 is also valid for finite union consonance with the obvious adaptation.



#### 4. A Bayesian Look at Each Desideratum

In the previous section, we provided several examples of testing schemes satisfying some of the logical properties reviewed therein. In particular, a testing scheme generated by the family of 0–1 loss functions (Example 2) was shown to fulfill both coherence and invertibility. However, not all families of loss functions generate a TS meeting any of these requisites, as is shown in the examples below.

**Example 7.** Suppose that  $X|\theta \sim \text{Bernoulli}(\theta)$  and that one is interested in testing the null hypotheses:

$$H_0^A : \theta \leq 0.4 \text{ and } H_0^B : \theta \leq 0.5.$$

Furthermore, assume  $\theta \sim \text{Uniform}(0, 1)$  a priori and that he uses the loss functions from Table 2 to perform the tests.

Thus, Bayes tests for testing  $H_0^A$  and  $H_0^B$  are, respectively,

$$\varphi_A(x) = \mathbb{I}(\mathbf{P}(\theta \leq 0.4|x) \leq 1/7) \text{ and } \varphi_B(x) = \mathbb{I}(\mathbf{P}(\theta \leq 0.5|x) \leq 1/3).$$

As  $\theta|x \sim \text{Beta}(2, 1)$  if  $x = 1$  is observed, then  $\mathbf{P}(\theta \leq 0.4|x) = 0.16$  and  $\mathbf{P}(\theta \leq 0.5|x) = 0.25$ , so that one does not reject  $H_0^A$ , but rejects  $H_0^B$ . Since  $H_0^A \subset H_0^B$ , we conclude that coherence does not hold.

**Table 2.** Loss functions for tests of Example 7.

State of Nature			State of Nature		
Decision	$\theta \in H_0^A$	$\theta \notin H_0^A$	Decision	$\theta \in H_0^B$	$\theta \notin H_0^B$
0	0	1	0	0	1
1	6	0	1	2	0

Intuitively, incoherence takes place because the loss of falsely rejecting  $H_0^A$  is three-times as large as the loss of falsely rejecting  $H_0^B$ , while the corresponding errors of Type II are of the same magnitude. Hence, these loss functions reveal that the decision-maker is more reluctant to reject  $H_0^A$  than to reject  $H_0^B$  in such a way that he only needs little evidence to accept  $H_0^A$  (posterior probability greater than 1/7) when compared to the amount of evidence needed to accept  $H_0^B$  (posterior probability greater than 1/3). Thus, it is not surprising at all that in this case, the tests do not cohere for some priors.

**Example 8.** In the setup of Example 7, suppose one also needs to test the null hypothesis  $H_0^{B^c} : \theta > 0.5$  by taking into account the loss function in Table 3.

The Bayes test for  $H_0^{B^c}$  is then to reject it if  $\mathbf{P}(\theta > 0.5|x) < 4/5$ . For  $x = 1$ ,  $\mathbf{P}(\theta > 0.5|x) = 0.75$ , and consequently,  $H_0^{B^c}$  is rejected. As both  $H_0^B$  and  $H_0^{B^c}$  are rejected when  $x = 1$  is observed, these tests do not satisfy invertibility.

**Table 3.** Loss function for Example 8.

State of Nature		
Decision	$\theta \in H_0^{B^c}$	$\theta \notin H_0^{B^c}$
0	0	4
1	1	0

The absence of invertibility is somewhat expected here, because the degree to which the decision-maker believes an incorrect decision of choosing  $H_0^B$  to be more serious than an incorrect decision of choosing  $H_0^{B^c}$  is not the same whether  $H_0^B$  is regarded as the “null” or the “alternative” hypothesis. More precisely, while the decision-maker assigns a loss to the error of Type I that is the double of the one assigned to the error of Type II when testing the null hypothesis  $H_0^B$ , he evaluates the loss of falsely accepting  $H_0^{B^c}$  to be four-times (not twice!) as large as that of falsely rejecting it when  $H_0^{B^c}$  is the null hypothesis.

The examples we have examined so far give rise to the question: from a decision-theoretic perspective, what conditions must be imposed on a family of loss functions so that the resultant Bayesian testing scheme meets coherence (invertibility)? Next, we offer a solution to this question. We first give a definition in order to simplify the statement of the main results of this section.

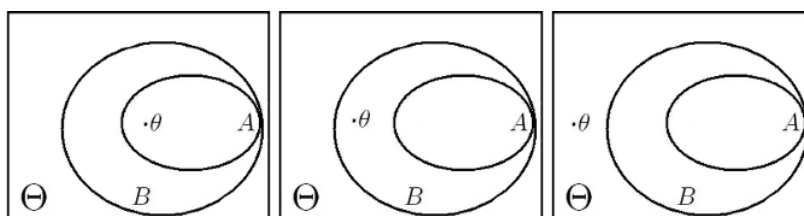
**Definition 7.** (Relative loss) Let  $L_A$  be a loss function for testing the hypothesis  $A \in \sigma(\Theta)$ . The function  $\Delta_A : \Theta \rightarrow \mathbb{R}$  defined by:

$$\Delta_A(\theta) = \begin{cases} L_A(1, \theta) - L_A(0, \theta), & \text{if } \theta \in A \\ L_A(0, \theta) - L_A(1, \theta), & \text{if } \theta \notin A \end{cases}$$

is named the relative loss of  $L_A$  for testing  $A$ .

In short, the relative loss measures the difference between losses of taking the wrong and the correct decisions. Thus, the relative loss of any hypothesis testing loss function is always non-negative.

A careful examination of Example 7 hints that in order to obtain coherent tests, the “larger” (the “smaller”) the null hypothesis of interest is, the more cautious about falsely rejecting (accepting) it the decision-maker ought to be. This can be quantified as follows: for hypotheses  $A$  and  $B$ , such that  $A \subseteq B$  and with corresponding hypothesis testing loss functions  $L_A$  and  $L_B$ , if  $\theta_1 \in A$ , then  $\Delta_B(\theta_1)$  should be at least as large as  $\Delta_A(\theta_1)$ . Similarly, if  $\theta_2 \in B^c$ , then  $\Delta_B(\theta_2)$  should be at most  $\Delta_A(\theta_2)$ . Such conditions are also appealing, since it seems reasonable that greater relative losses should be assigned to greater “distances” between the parameter and the wrong decision. For instance, if  $\theta \in A$  (and consequently,  $\theta \in B$ ), the rougher error of rejecting  $B$  should be penalized more heavily than the error of rejecting  $A$ ; Figure 1 enlightens this idea.



**Figure 1.** Interpretation of sensible relative losses: rougher errors of decisions should be assigned larger relative losses.

These conditions, namely:

$$\Delta_A(\theta_1) \leq \Delta_B(\theta_1), \forall \theta_1 \in A \quad \text{and} \quad \Delta_A(\theta_2) \geq \Delta_B(\theta_2), \forall \theta_2 \in B^c,$$

are sufficient for coherence. As a matter of fact, Theorem 2 states that the weaker condition:

$$\Delta_A(\theta_1)\Delta_B(\theta_2) \leq \Delta_A(\theta_2)\Delta_B(\theta_1), \forall \theta_1 \in A, \forall \theta_2 \in B^c,$$

is necessary and sufficient for a family of hypothesis testing loss functions to induce a coherent testing scheme with respect to each prior distribution for  $\theta$ . Henceforward, we assume that  $\mathbb{E}(L_A(d, \theta)|x) < \infty$ , for all  $A \in \sigma(\Theta)$ ,  $d \in \{0, 1\}$  and  $x \in \mathcal{X}$ .

**Theorem 2.** *Let  $(L_A)_{A \in \sigma(\Theta)}$  be a family of hypothesis testing loss functions. Suppose that for all  $\theta_1, \theta_2 \in \Theta$ , there is  $x \in \mathcal{X}$ , such that  $L_x(\theta_1), L_x(\theta_2) > 0$ . Then, for all prior distributions  $\pi$  for  $\theta$ , there exists a testing scheme generated by  $(L_A)_{A \in \sigma(\Theta)}$  with respect to  $\pi$  that is coherent if, and only if,  $(L_A)_{A \in \sigma(\Theta)}$  is such that for all  $A, B \in \sigma(\Theta)$  with  $A \subseteq B$ :*

$$\Delta_A(\theta_1)\Delta_B(\theta_2) \leq \Delta_A(\theta_2)\Delta_B(\theta_1), \forall \theta_1 \in A, \forall \theta_2 \in B^c. \tag{2}$$

Notice that the “if” part of Theorem 2 still holds for families of hypothesis testing loss functions that depend also on the sample. Theorem 2 characterizes, under certain conditions, all families of loss functions that induce coherent tests, no matter what the decision-maker’s opinion (prior) on the unknown parameter is. Although the result of Theorem 2 is not properly normative, any Bayesian decision-maker can make use of it to prevent himself from drawing incoherent conclusions from multiple hypothesis testing by checking whether his personal losses satisfy the condition in Equation (2).

Many simple families of loss functions generate coherent tests, as we illustrate in Examples 9 and 10.

**Example 9.** *Consider, for each  $A \in \sigma(\Theta)$ , the loss function  $L_A$  in Table 4 to test the null hypothesis  $A$ , in which  $\lambda : \sigma(\Theta) \rightarrow \mathbb{R}_+$  is any finite measure, such that  $\lambda(\Theta) > 0$ . This family of loss functions satisfies the condition in Equation (2) for coherence as for all  $A, B \in \sigma(\Theta)$ , such that  $A \subseteq B$ , and for all  $\theta_1 \in A$  and  $\theta_2 \in B^c$ ,  $\Delta_A(\theta_1) = \lambda(A)$ ,  $\Delta_B(\theta_2) = \lambda(B^c)$ ,  $\Delta_A(\theta_2) = \lambda(A^c)$  and  $\Delta_B(\theta_1) = \lambda(B)$ .*

**Table 4.** Loss function  $L_A$  for testing A.

	State of Nature	
Decision	$\theta \in A$	$\theta \notin A$
0	0	$\lambda(A^c)$
1	$\lambda(A)$	0

As a matter of fact, if for each  $A \in \sigma(\Theta)$ ,  $L_A$  is a 0 – 1 –  $c_A$  loss function ([5] (p. 215)), with  $0 < c_A \leq c_B$  if  $A \subseteq B$ , then the family  $(L_A)_{A \in \sigma(\Theta)}$  will induce a coherent TS for each prior for  $\theta$ .

**Example 10.** *Assume  $\Theta$  is equipped with a distance, say  $d$ . Define, for each  $A \in \sigma(\Theta)$  the loss function  $L_A$  for testing A by:*

$$L_A(0, \theta) = d^*(\theta, A) \text{ and } L_A(1, \theta) = d^*(\theta, A^c),$$

where  $d^*(\theta, A) = \inf_{a \in A} d(\theta, a)$  is the distance between  $\theta \in \Theta$  and  $A$ . For  $A, B \in \sigma(\Theta)$ , such that  $A \subseteq B$ , and for  $\theta_1 \in A$  and  $\theta_2 \in B^c$ ,  $\Delta_A(\theta_1) = d^*(\theta_1, A^c)$ ,  $\Delta_B(\theta_2) = d^*(\theta_2, B)$ ,  $\Delta_A(\theta_2) = d^*(\theta_2, A)$  and  $\Delta_B(\theta_1) = d^*(\theta_1, B^c)$ . These values satisfy Equation (2) from Theorem 2. Hence, families of loss functions based on distances as the above generate Bayesian coherent tests.

Next, we characterize Bayesian tests with respect to invertibility. In order to obtain TSs that meet invertibility, it seems reasonable that when the null and alternative hypotheses are switched, the relative losses ought to remain the same. That is to say, when testing the null hypothesis  $A$ , the relative loss at each point  $\theta \in \Theta$ ,  $\Delta_A(\theta)$ , should be equal to the relative loss  $\Delta_{A^c}(\theta)$  when  $A^c$  is the null hypothesis instead. This condition is sufficient, but not necessary for a family of loss functions to induce tests fulfilling this logical requisite with respect to all prior distributions. In Theorem 3, however, we provide necessary and sufficient conditions for invertibility.

**Theorem 3.** *Let  $(L_A)_{A \in \sigma(\Theta)}$  be a family of hypothesis testing loss functions. Suppose that for all  $\theta_1, \theta_2 \in \Theta$ , there is  $x \in \mathcal{X}$ , such that  $L_x(\theta_1), L_x(\theta_2) > 0$ . Then, for all prior distributions  $\pi$  for  $\theta$ , there exists a testing scheme generated by  $(L_A)_{A \in \sigma(\Theta)}$  with respect to  $\pi$  that satisfies invertibility if, and only if,  $(L_A)_{A \in \sigma(\Theta)}$  is such that for all  $A \in \sigma(\Theta)$ :*

$$\Delta_A(\theta_1)\Delta_{A^c}(\theta_2) = \Delta_{A^c}(\theta_1)\Delta_A(\theta_2), \quad \forall \theta_1 \in A, \theta_2 \in A^c. \quad (3)$$

Condition Equation (3) is equivalent (for strict hypothesis testing loss functions) to impose, for each  $A \in \sigma(\Theta)$ , that the function  $\frac{\Delta_A(\cdot)}{\Delta_{A^c}(\cdot)}$  to be constant over  $\Theta$ . We should mention that the “if” part of Theorem 3 still holds for hypothesis testing loss functions satisfying (Equation (3)) that also depend on the sample  $x$ .

The families of loss functions introduced in Examples 9 and 10 satisfy (Equation (3)). Thus, such families of losses ensure the construction of simultaneous Bayes tests that are in conformity with both coherence and invertibility for all prior distributions on  $\sigma(\Theta)$ . Thus, if one believes these (two) logical requirements to be of primary importance in multiple hypothesis testing, he can make use of any of these families of loss functions to perform tests satisfactorily. Other simple loss functions also lead to TSs that meet invertibility: for instance, any family of 0–1– $c$  loss functions for which  $c_{A^c} = 1/c_A$  for all  $A \in \sigma(\Theta)$  leads to invertible TSs.

We end this section by examining union consonance under a decision-theoretic point of view. From Definition 6, it appears that a necessary condition for the derivation of consonant tests is that “smaller” (“larger”) null hypotheses ought to be assigned greater losses for false rejection (acceptance). More precisely, for  $A, B \in \sigma(\Theta)$ , if  $\theta_1 \in A \cup B$ , then it seems that either  $\Delta_{A \cup B}(\theta_1) \leq \Delta_A(\theta_1)$  or  $\Delta_{A \cup B}(\theta_1) \leq \Delta_B(\theta_1)$  should hold. If  $\theta_2 \in (A \cup B)^c$ , then it is reasonable that either  $\Delta_{A \cup B}(\theta_2) \geq \Delta_A(\theta_2)$  or  $\Delta_{A \cup B}(\theta_2) \geq \Delta_B(\theta_2)$ . The next theorem shows that this is nearly the case. However, it is still unknown whether sufficient conditions for union consonance are determinable.

**Theorem 4.** *Let  $(L_A)_{A \in \sigma(\Theta)}$  be a family of hypothesis testing loss functions. Suppose that for all  $\theta_1, \theta_2 \in \Theta$ , there is  $x \in \mathcal{X}$ , such that  $L_x(\theta_1), L_x(\theta_2) > 0$ . If for all prior distribution  $\pi$  for  $\theta$ , there exists a testing scheme generated by  $(L_A)_{A \in \sigma(\Theta)}$  with respect to  $\pi$  that satisfies finite union consonance, then  $(L_A)_{A \in \sigma(\Theta)}$  is such that for all  $A, B \in \sigma(\Theta)$  and for all  $\theta_1 \in A \cup B, \theta_2 \in (A \cup B)^c$ ,*

$$\text{either } \Delta_{A \cup B}(\theta_1)\Delta_A(\theta_2) \leq \Delta_{A \cup B}(\theta_2)\Delta_A(\theta_1) \text{ or } \Delta_{A \cup B}(\theta_1)\Delta_B(\theta_2) \leq \Delta_{A \cup B}(\theta_2)\Delta_B(\theta_1).$$

## 5. Putting the Desiderata Together

In Section 4, we showed that there are infinitely many families of loss functions that induce, for each prior distribution for  $\theta$ , a TS that satisfies both coherence and invertibility (Examples 9 and 10). However, requiring the three logical consistency properties we presented to hold simultaneously with respect to all priors is too restrictive: under mild conditions, no TS constructed under a Bayesian decision-theoretic approach to hypothesis testing fulfills this, as stated in the next theorem.

**Theorem 5.** *Assume that  $\Theta$  and  $\sigma(\Theta)$  are such that  $|\Theta| \geq 3$  and that there is a partition of  $\Theta$  composed of three nonempty measurable sets. Assume also that for all triplet  $\theta_1, \theta_2, \theta_3 \in \Theta$ , there is  $x \in \mathcal{X}$ , such that  $L_x(\theta_i) > 0$  for  $i = 1, 2, 3$ . Then, there is no family of strict hypothesis testing loss functions that induces, for each prior distribution for  $\theta$ , a testing scheme satisfying coherence, invertibility and finite union consonance.*

Theorem 5 states that Bayesian optimality (based on standard loss functions that do not depend on the sample) cannot be combined with complete logical consistency. This fact can lead one to wonder whether such properties are indeed sensible in multiple hypothesis testing. The following result shows us that the desiderata are in fact reasonable in the sense that a TS meeting these requirements does correspond to the optimal tests of some Bayesian decision-makers. We return to this point in the concluding remarks.

**Theorem 6.** *Let  $\Theta$  be a countable (finite) parameter space,  $\sigma(\Theta) = \mathcal{P}(\Theta)$ , and  $\mathcal{X}$  be a countable sample space. Let  $\varphi$  be a testing scheme that satisfies coherence, invertibility and countable (finite) union consonance. Then, there exist a probability measure  $\mu$  over  $\mathcal{P}(\Theta \times \mathcal{X})$  and a family of strict hypothesis testing loss functions  $(L_A)_{A \in \sigma(\Theta)}$ , such that  $\varphi$  is generated by  $(L_A)_{A \in \sigma(\Theta)}$  with respect to the  $\mu$ -marginal distribution of  $\theta$ .*

We end this section by associating logically-consistent Bayesian hypothesis testing with Bayes point estimation procedures in case both  $\Theta$  and  $\mathcal{X}$  are finite. This relationship is characterized in Theorem 7.

**Theorem 7.** *Let  $\Theta$  and  $\mathcal{X}$  be finite sets and  $\sigma(\Theta) = \mathcal{P}(\Theta)$ . Let  $\varphi$  be the testing scheme generated by the point estimator  $\delta : \mathcal{X} \rightarrow \Theta$ . Suppose that for all  $x \in \mathcal{X}$ ,  $L_x(\delta(x)) > 0$ .*

- (a) *If there exist a probability measure  $\pi : \sigma(\Theta) \rightarrow [0, 1]$  for  $\theta$ , with  $\pi(\delta(x)) > 0$  for all  $x \in \mathcal{X}$ , and a loss function  $L : \Theta \times \Theta \rightarrow \mathbb{R}_+$ , satisfying  $L(\theta, \theta) = 0$  and  $L(d, \theta) > 0$  for  $d \neq \theta$ , such that  $\delta$  is a Bayes estimator for  $\theta$  generated by  $L$  with respect to  $\pi$ , then there is a family of hypothesis testing loss functions  $(L_A)_{A \in \sigma(\Theta)}$ ,  $L_A : \{0, 1\} \times (\Theta \times \mathcal{X}) \rightarrow \mathbb{R}_+$  for each  $A \in \sigma(\Theta)$ , such that  $\varphi$  is generated by  $(L_A)_{A \in \sigma(\Theta)}$  with respect to  $\pi$ .*
- (b) *If there exist a probability measure  $\pi : \sigma(\Theta) \rightarrow [0, 1]$  for  $\theta$ , with  $\pi(\delta(x)) > 0$  for all  $x \in \mathcal{X}$ , and a family of strict hypothesis testing loss functions  $(L_A)_{A \in \sigma(\Theta)}$ ,  $L_A : \{0, 1\} \times \Theta \rightarrow \mathbb{R}_+$  for each  $A \in \sigma(\Theta)$ , such that  $\varphi$  is generated by  $(L_A)_{A \in \sigma(\Theta)}$  with respect to  $\pi$ , then there is a loss function  $L : \Theta \times \Theta \rightarrow \mathbb{R}_+$ , with  $L(\theta, \theta) = 0$  and  $L(d, \theta) > 0$  for  $d \neq \theta$ , such that  $\delta$  is a Bayes estimator for  $\theta$  generated by  $L$  with respect to  $\pi$ .*

Theorem 7 ensures that multiple Bayesian tests that fulfill the desiderata cannot be separated from Bayes point estimation procedures. One may find in Theorem 7, Part (a), a decision-theoretic justification

for performing simultaneous tests by means of a Bayes point estimator. However, the optimality of such tests is derived under very restrictive conditions, as the underlying loss functions depend both on the sample and on a point estimator. This fact reinforces that one can reconcile statistical optimality and logical consistency in multiple tests only in very particular cases. We should also emphasize that, under the conditions of Part (a), if, in addition,  $\pi(\theta) > 0$  for all  $\theta \in \Theta$ , then, for all  $A \in \sigma(\Theta)$ ,  $\varphi_A$  is an admissible test for  $A$  with regard to  $L_A$  (the standard proof of this result developed for losses that do not depend on the sample also works here). The second part of Theorem 7 states that if a Bayesian testing scheme meets coherence, invertibility and finite union consonance, then the point estimator that generates it cannot be devoid of optimality: it must be a Bayes estimator for specific loss functions. Example 11 illustrates the first part of this theorem.

**Example 11.** Assume that  $\Theta = \{\theta_1, \theta_2, \dots, \theta_k\}$  and  $\mathcal{X}$  is finite. Assume also that there is a maximum likelihood estimator (MLE) for  $\theta$ ,  $\delta_{ML} : \mathcal{X} \rightarrow \Theta$ , such that  $L_x(\delta_{ML}(x)) > 0$ , for all  $x \in \mathcal{X}$ . Then, the testing scheme generated by  $\delta_{ML}$  is a TS of Bayes tests. Indeed, when  $\Theta$  is finite, an MLE for  $\theta$  is a Bayes estimator generated by the loss function  $L(d, \theta) = \mathbb{I}(d \neq \theta)$ ,  $d, \theta \in \Theta$ , with respect to the uniform prior over  $\Theta$  (that is,  $\delta_{ML}(x)$  corresponds to a mode of the posterior distribution  $\pi_x$ , for each  $x \in \mathcal{X}$ ). Consequently (recall that  $|\Theta| = k$ ),  $\pi_x(\delta_{ML}(x)) \geq 1/k$  and  $\mathbb{E}[L(\delta_{ML}(x), \theta)|x] = 1 - \pi_x(\delta_{ML}(x))$ , for each  $x \in \mathcal{X}$ . Thus,

$$\max_{x \in \mathcal{X}} \frac{E[L(\delta_{ML}(x), \theta)|x]}{\pi_x(\delta_{ML}(x))} = \max_{x \in \mathcal{X}} \frac{1 - \pi_x(\delta_{ML}(x))}{\pi_x(\delta_{ML}(x))} \leq \frac{1 - \frac{1}{k}}{\frac{1}{k}} = k - 1,$$

as  $g : (0, 1] \rightarrow \mathbb{R}_+$  given by  $g(t) = (1 - t)/t$  is strictly decreasing.

By Theorem 7, it follows that the TS generated by the MLE  $\delta_{ML}$  is a Bayesian TS generated by (for instance) the family of loss functions  $(L_A)_{A \in \sigma(\Theta)}$  given, for each  $A \in \sigma(\Theta)$ , by  $L_A(1, (\theta, x)) = 0$  and  $L_A(0, (\theta, x)) = 1$ , for  $\theta \in A^c$ , and  $L_A(0, (\theta, x)) = 0$  and  $L_A(1, (\theta, x)) = k\mathbb{I}_A(\delta_{ML}(x)) + (1/k)\mathbb{I}_{A^c}(\delta_{ML}(x))$ , for  $\theta \in A$ .

It is worth mentioning that the development of Theorem 7(a) and Example 11 is in a sense related to the optimality of least relative surprise estimators under prior-based loss functions [24] (Section 2).

## 6. Conclusions

While several studies on frequentist multiple tests deal with the question of seeking for a balance between statistical optimality and logical consistency, this issue has not been addressed yet under a decision-theoretic standpoint. For this reason, in this work, we examine simultaneous Bayesian hypothesis testing with respect to three rational properties: coherence, invertibility and union consonance. Briefly, we characterize the families of loss functions that yield Bayes tests meeting each of these requisites separately, whatever the prior distribution for the relevant parameter is. These results not only shed some light on when each of these relationships may be considered to be sensible for a given scientific problem, but they also serve as a guide for a Bayesian decision-maker aiming at performing tests in line with the requirement he finds more important. In particular, this can be done through the usage of the loss functions described in the paper.



We also explore how far one can go by putting these properties together. We provide examples of fairly intuitive loss functions that induce testing schemes satisfying both coherence and invertibility, no matter what one's prior opinion on the parameter is. On the other hand, we prove that no family of reasonable loss functions generates Bayes tests that respect the logical properties as a whole with respect to all priors, although any testing scheme meeting the desiderata corresponds to the optimal tests of several Bayesian decision-makers.

Finally, we discuss the relationship between logically-consistent Bayesian hypothesis testing and Bayes point estimations procedures when both the parameter space and the sample space are finite. We conclude that the point estimator generating a testing scheme fulfilling the rational properties is inevitably and unavoidably a Bayes estimator for certain loss functions. Furthermore, performing logically-consistent procedures by means of a Bayes estimator is one's best approach towards multiple hypothesis testing only under very restrictive conditions in which the underlying loss functions depend not only on the decision to be made and the parameter as usual, but also on the observed sample. See [24–26] for some examples of such loss functions. That is, a more complex framework is needed to combine Bayesian optimality with logical consistency. This fact and the impossibility result of Theorem 5 corroborate the thesis that full rationality and statistical optimality rarely can be combined in simultaneous tests. In practice, this suggests that when testing hypotheses at once, a practitioner may abandon in part the desiderata so as to preserve statistical optimality. This is further discussed in [8].

Several issues remain open, among which we mention three. First, the extent to which the results derived in this work can be generalized to infinite (continuous) parameter spaces is an important problem from both theoretical and practical aspects. Furthermore, the consideration of different decision-theoretic approaches to hypothesis testing, such as the “agnostic” tests with three-fold action spaces proposed by [27], may bring new insight into which logical properties may be expected, not only in the current, but also in alternative frameworks. In epistemological terms, one may be concerned with the question of whether multiple hypothesis testing is the most adequate way to draw inferences about a parameter of interest from data given the incompatibility between full logical consistency and the achievement of statistical optimality. As a matter of fact, many Bayesians regard the whole posterior distribution as the most complete inference one can make about the unknown parameter. These analyses may contribute to better decision-making.

## Acknowledgments

The authors are thankful for Carlos Alberto de Bragança Pereira, José Carlos Simon de Miranda, José Galvão Leite, Julio Michael Stern, Marcelo Esteban Coniglio, Márcio Alves Diniz and Paulo Cilas Marques Filho for fruitful discussions and important comments and suggestions, which improved the manuscript. We are also grateful to the referees for all of the detailed comments that helped improve the paper. This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (2009/03385-5, 2014/25302-2) Brazil and Conselho Nacional de Pesquisa e Desenvolvimento Científico e Tecnológico (131982/2009-5) Brazil.

**Author Contributions**

The manuscript has come to fruition by the substantial contributions of all authors from conceiving the idea of examining Bayes tests with respect to logical consistency to obtaining the main theorems and providing several examples. All authors have also been involved in either writing the article or carefully revising it. All authors have read and approved the submitted version of the paper.

**Conflicts of Interest**

The authors declare no conflict of interest.

**Appendix**

**A. Proof of Theorem 1**

That a testing scheme generated by a point estimation procedure  $\delta$  satisfies the desiderata follows from Theorem 4.3 from [8] and the fact that for all  $x \in \mathcal{X}$  and all countable partition  $(A_n)_{n \geq 1}$  of  $\Theta$ , there is a unique  $i^* \in \mathbb{N}^*$ , such that  $\delta(x) \in A_{i^*}$  and, consequently,  $\sum_{i=1}^{\infty} [1 - \mathbb{I}(\delta(x) \notin A_i)] = 1$ . For the converse, Theorem 4.3 from [8] implies that  $\forall x \in \mathcal{X}, \exists! \theta_0 = \theta_0(x) \in \Theta$ , such that  $\varphi_{\{\theta_0\}}(x) = 0$ . Thus, for  $A \in \sigma(\Theta)$ ,  $\theta_0 \in A \Rightarrow \{\theta_0(x)\} \subseteq A$  and, as coherence holds,  $\varphi_A(x) = 0$ . On the other hand,  $\theta_0 \notin A \Rightarrow \{\theta_0(x)\} \subseteq A^c$ . Coherence and invertibility yield  $\varphi_A(x) = 1$ . Hence, for each  $A \in \sigma(\Theta)$ ,  $\varphi_A(x) = 1 \Leftrightarrow \theta(x) \notin A$ . We conclude the proof by defining  $\delta : \mathcal{X} \rightarrow \Theta$  by  $\delta(x) = \theta_0(x)$ .

**B. Proof of Theorem 2**

First, we prove the necessary condition by the contrapositive. Thus, let us suppose there are  $A, B \in \sigma(\Theta)$  with  $A \subseteq B$  and  $\theta_1 \in A$  and  $\theta_2 \in B^c$ , such that:

$$\Delta_A(\theta_1)\Delta_B(\theta_2) > \Delta_A(\theta_2)\Delta_B(\theta_1),$$

which implies that  $\Delta_A(\theta_1) > 0$  and  $\Delta_B(\theta_2) > 0$ .

Adding  $\Delta_A(\theta_2)\Delta_B(\theta_2)$  to both sides of the inequality above, straightforward manipulations yield:

$$0 \leq \frac{\Delta_A(\theta_2)}{\Delta_A(\theta_2) + \Delta_A(\theta_1)} < \frac{\Delta_B(\theta_2)}{\Delta_B(\theta_2) + \Delta_B(\theta_1)} \leq 1.$$

Thus, there is  $\alpha_0 \in (0, 1)$ , such that:

$$\frac{\Delta_A(\theta_2)}{\Delta_A(\theta_2) + \Delta_A(\theta_1)} < \alpha_0 < \frac{\Delta_B(\theta_2)}{\Delta_B(\theta_2) + \Delta_B(\theta_1)}. \tag{4}$$

Furthermore, there is  $x' \in \mathcal{X}$ , such that  $L_{x'}(\theta_1), L_{x'}(\theta_2) > 0$ . Considering the prior distribution  $\pi^*$  for  $\theta$  given by:

$$\pi^*(\theta_1) = \frac{\alpha_0 L_{x'}(\theta_2)}{\alpha_0 L_{x'}(\theta_2) + (1 - \alpha_0) L_{x'}(\theta_1)} \text{ and } \pi^*(\theta_2) = 1 - \pi^*(\theta_1),$$

the corresponding posterior distribution given  $x'$  is  $\pi_{x'}^*(\theta_1) = \alpha_0$  and  $\pi_{x'}^*(\theta_2) = 1 - \alpha_0$ . Let  $\varphi^*$  be any TS generated by  $(L_A)_{A \in \sigma(\Theta)}$  with respect to  $\pi^*$ . Thus,

$$\varphi_A^*(x') = 0, \text{ if } \alpha_0 > \frac{\Delta_A(\theta_2)}{\Delta_A(\theta_2) + \Delta_A(\theta_1)} \text{ and } \varphi_B^*(x') = 0, \text{ if } \alpha_0 > \frac{\Delta_B(\theta_2)}{\Delta_B(\theta_2) + \Delta_B(\theta_1)}.$$

From Equation (4), we have  $\varphi_A^*(x') = 0$  and  $\varphi_B^*(x') = 1$ . Therefore, there is a prior distribution  $\pi^*$  for  $\theta$  with respect to which any TS generated by  $(L_A)_{A \in \sigma(\Theta)}$  is not coherent.

We now prove the “if” part. We suppose that the family  $(L_A)_{A \in \sigma(\Theta)}$  satisfies the condition that for all  $A, B \in \sigma(\Theta)$  with  $A \subseteq B$ ,  $\Delta_A(\theta_1)\Delta_B(\theta_2) \leq \Delta_A(\theta_2)\Delta_B(\theta_1)$ ,  $\forall \theta_1 \in A, \forall \theta_2 \in B^c$ . Integrating (with respect to  $\theta_1$ ) over  $A$  with respect to any probability measure  $P$ , we obtain:

$$\Delta_B(\theta_2) \int_A \Delta_A(\theta_1)dP(\theta_1) \leq \Delta_A(\theta_2) \int_A \Delta_B(\theta_1)dP(\theta_1), \forall \theta_2 \in B^c.$$

Similarly, integration (with respect to  $\theta_2$ ) over  $B^c$  with respect to the same measure  $P$  yields:

$$\int_{B^c} \Delta_B(\theta_2)dP(\theta_2) \int_A -\Delta_A(\theta_1)dP(\theta_1) \geq \int_{B^c} \Delta_A(\theta_2)dP(\theta_2) \int_A -\Delta_B(\theta_1)dP(\theta_1). \tag{5}$$

Now, let  $\varphi$  be a testing scheme generated by the family  $(L_A)_{A \in \sigma(\Theta)}$ . For  $A, B \in \sigma(\Theta)$  with  $A \subseteq B$  and  $x \in \mathcal{X}$ ,

$$\varphi_A(x) = 0 \Rightarrow \int_{\Theta} [L_A(0, \theta) - L_A(1, \theta)]d\pi_x(\theta) \leq 0,$$

where  $\pi_x(\cdot)$  denotes the posterior distribution of  $\theta$  given  $X = x$ . Thus,

$$\int_A -\Delta_A(\theta)d\pi_x(\theta) + \int_{A^c \cap B} \Delta_A(\theta)d\pi_x(\theta) + \int_{B^c} \Delta_A(\theta)d\pi_x(\theta) \leq 0.$$

Multiplying the last inequality by  $\int_{B^c} \Delta_B(\theta)d\pi_x(\theta) \geq 0$ , we get:

$$\int_{B^c} \Delta_B(\theta)d\pi_x(\theta) \int_A -\Delta_A(\theta)d\pi_x(\theta) + \int_{B^c} \Delta_B(\theta)d\pi_x(\theta) \int_{A^c \cap B} \Delta_A(\theta)d\pi_x(\theta) + \int_{B^c} \Delta_B(\theta)d\pi_x(\theta) \int_{B^c} \Delta_A(\theta)d\pi_x(\theta) \leq 0.$$

From inequality Equation (5), it follows that:

$$\int_A -\Delta_B(\theta)d\pi_x(\theta) \int_{B^c} \Delta_A(\theta)d\pi_x(\theta) + \int_{B^c} \Delta_B(\theta)d\pi_x(\theta) \int_{A^c \cap B} \Delta_A(\theta)d\pi_x(\theta) + \int_{B^c} \Delta_B(\theta)d\pi_x(\theta) \int_{B^c} \Delta_A(\theta)d\pi_x(\theta) \leq 0.$$

As  $\int_{B^c} \Delta_B(\theta)d\pi_x(\theta) \int_{A^c \cap B} \Delta_A(\theta)d\pi_x(\theta) \geq 0$  and  $\int_{A^c \cap B} -\Delta_B(\theta)d\pi_x(\theta) \int_{B^c} \Delta_A(\theta)d\pi_x(\theta) \leq 0$ , we have that:

$$\int_A -\Delta_B(\theta)d\pi_x(\theta) \int_{B^c} \Delta_A(\theta)d\pi_x(\theta) + \int_{A^c \cap B} -\Delta_B(\theta)d\pi_x(\theta) \int_{B^c} \Delta_A(\theta)d\pi_x(\theta) + \int_{B^c} \Delta_B(\theta)d\pi_x(\theta) \int_{B^c} \Delta_A(\theta)d\pi_x(\theta) \leq 0,$$

and, consequently,

$$\int_{B^c} \Delta_A(\theta)d\pi_x(\theta) \left\{ \int_A -\Delta_B(\theta)d\pi_x(\theta) + \int_{A^c \cap B} -\Delta_B(\theta)d\pi_x(\theta) + \int_{B^c} \Delta_B(\theta)d\pi_x(\theta) \right\} \leq 0.$$

Finally,

$$\int_{\Theta} [L_B(0, \theta) - L_B(1, \theta)]d\pi_x(\theta) \leq 0.$$

If  $\int_{\Theta} [L_B(0, \theta) - L_B(1, \theta)]d\pi_x(\theta) < 0$ , then  $\varphi_B(x) = 0$ . If this integral is equal to zero, then both zero and one are optimal solutions, and we can choose the decision zero as  $\varphi_B(x)$  in order to ensure that  $\varphi_B(x) \leq \varphi_A(x)$ . Hence, with respect to each prior  $\pi$ , there is a TS generated by  $(L_A)_{A \in \sigma(\Theta)}$  that is coherent.

### C. Proof of Theorem 3

The proof is analogous to that of Theorem 2. First, we prove the necessary condition by the contrapositive. Suppose that there are  $A \in \sigma(\Theta)$  and  $\theta_1 \in A$  and  $\theta_2 \in A^c$ , such that:

$$\Delta_A(\theta_1)\Delta_{A^c}(\theta_2) \neq \Delta_{A^c}(\theta_1)\Delta_A(\theta_2).$$

Assume  $\Delta_A(\theta_1)\Delta_{A^c}(\theta_2) < \Delta_{A^c}(\theta_1)\Delta_A(\theta_2)$  (the other case is developed in the same way), which implies that  $\Delta_{A^c}(\theta_1) > 0$  and  $\Delta_A(\theta_2) > 0$ . Adding  $\Delta_{A^c}(\theta_2)\Delta_A(\theta_2)$  to both sides of the inequality, we easily obtain that:

$$0 \leq \frac{\Delta_{A^c}(\theta_2)}{\Delta_{A^c}(\theta_1) + \Delta_{A^c}(\theta_2)} < \frac{\Delta_A(\theta_2)}{\Delta_A(\theta_1) + \Delta_A(\theta_2)} \leq 1.$$

Thus, there is  $\alpha_0 \in (0, 1)$ , such that:

$$0 \leq \frac{\Delta_{A^c}(\theta_2)}{\Delta_{A^c}(\theta_1) + \Delta_{A^c}(\theta_2)} < \alpha_0 < \frac{\Delta_A(\theta_2)}{\Delta_A(\theta_1) + \Delta_A(\theta_2)} \leq 1. \tag{6}$$

In addition, there is  $x' \in \mathcal{X}$ , such that  $L_{x'}(\theta_1), L_{x'}(\theta_2) > 0$ . For the prior distribution  $\pi^*$  for  $\theta$  given by:

$$\pi^*(\theta_1) = \frac{\alpha_0 L_{x'}(\theta_2)}{\alpha_0 L_{x'}(\theta_2) + (1 - \alpha_0)L_{x'}(\theta_1)} \text{ and } \pi^*(\theta_2) = 1 - \pi^*(\theta_1),$$

the posterior distribution given  $x'$  is  $\pi_{x'}^*(\theta_1) = \alpha_0$  and  $\pi_{x'}^*(\theta_2) = 1 - \alpha_0$ . Let  $\varphi^*$  be any TS generated by  $(L_A)_{A \in \sigma(\Theta)}$  with respect to  $\pi^*$ . Thus,

$$\varphi_A^*(x') = 0, \text{ if } \alpha_0 > \frac{\Delta_A(\theta_2)}{\Delta_A(\theta_1) + \Delta_A(\theta_2)} \text{ and } \varphi_{A^c}^*(x') = 0, \text{ if } \alpha_0 < \frac{\Delta_{A^c}(\theta_2)}{\Delta_{A^c}(\theta_1) + \Delta_{A^c}(\theta_2)}.$$

From Equation (6), we have  $\varphi_A^*(x') = 1$  and  $\varphi_{A^c}^*(x') = 1$ . Therefore, there is a prior distribution  $\pi^*$  for  $\theta$  with respect to which any TS generated by  $(L_A)_{A \in \sigma(\Theta)}$  does not meet invertibility.

Now, we prove the sufficiency. Suppose that for all  $A \in \sigma(\Theta)$ :

$$\Delta_A(\theta_1)\Delta_{A^c}(\theta_2) = \Delta_{A^c}(\theta_1)\Delta_A(\theta_2), \forall \theta_1 \in A, \theta_2 \in A^c.$$

Integrating (with respect to  $\theta_2$ ) over the set  $A^c$  with respect to any probability measure  $P$  defined on  $\sigma(\Theta)$ , we have:

$$\int_{A^c} \Delta_A(\theta_2)\Delta_{A^c}(\theta_1)dP(\theta_2) = \int_{A^c} \Delta_A(\theta_1)\Delta_{A^c}(\theta_2)dP(\theta_2), \text{ for all } \theta_1 \in A.$$

Similarly, integrating (with respect to  $\theta_1$ ) over  $A$ , we get:

$$\int_A \Delta_{A^c}(\theta_1)dP(\theta_1) \int_{A^c} \Delta_A(\theta_2)dP(\theta_2) = \int_A \Delta_A(\theta_1)dP(\theta_1) \int_{A^c} \Delta_{A^c}(\theta_2)dP(\theta_2). \tag{7}$$

Let  $\varphi$  be a TS generated by  $(L_A)_{A \in \sigma(\Theta)}$ . If  $\varphi_A(x) = 0$ , then:

$$\int_{\Theta} [L_A(0, \theta) - L_A(1, \theta)]d\pi_x(\theta) = \int_A -\Delta_A(\theta)d\pi_x(\theta) + \int_{A^c} \Delta_A(\theta)d\pi_x(\theta) \leq 0.$$

Multiplying both sides by  $\int_A \Delta_{A^c}(\theta)d\pi_x(\theta) \geq 0$ , we get:

$$\int_A \Delta_{A^c}(\theta)d\pi_x(\theta) \int_A -\Delta_A(\theta)d\pi_x(\theta) + \int_A \Delta_{A^c}(\theta)d\pi_x(\theta) \int_{A^c} \Delta_A(\theta)d\pi_x(\theta) \leq 0.$$

From Equation (7), it follows that:

$$\int_A -\Delta_A(\theta)d\pi_x(\theta) \left\{ \int_A \Delta_{A^c}(\theta)d\pi_x(\theta) + \int_{A^c} -\Delta_{A^c}(\theta)d\pi_x(\theta) \right\} \leq 0.$$

Thus,

$$\int_A \Delta_{A^c}(\theta)d\pi_x(\theta) + \int_{A^c} -\Delta_{A^c}(\theta)d\pi_x(\theta) \geq 0,$$

since  $\int_A -\Delta_A(\theta)d\pi_x(\theta) \leq 0$ . In this way,

$$\int_{\Theta} [L_{A^c}(0, \theta) - L_{A^c}(1, \theta)]d\pi_x(\theta) \geq 0.$$

If  $\int_{\Theta} [L_{A^c}(0, \theta) - L_{A^c}(1, \theta)]d\pi_x(\theta) > 0$ , then  $\varphi_{A^c}(x) = 1$ . If the integral is zero, then we can choose  $\varphi_{A^c}(x) = 1$ , so as to obtain  $\varphi_{A^c}(x) = 1 - \varphi_A(x)$ . Similarly, we prove that if  $\varphi_A(x) = 1$ , then there is a Bayes test for  $A^c$ ,  $\varphi_{A^c}$ , generated by  $L_{A^c}$ , such that  $\varphi_{A^c}(x) = 0$ . Consequently, there is a TS generated by  $(L_A)_{A \in \sigma(\Theta)}$  that satisfies invertibility.

**D. Proof of Theorem 4**

Suppose that there are  $A, B \in \sigma(\Theta)$ ,  $\theta_1 \in A \cup B$  and  $\theta_2 \in (A \cup B)^c$  such that both:

$$\Delta_{A \cup B}(\theta_1)\Delta_A(\theta_2) > \Delta_{A \cup B}(\theta_2)\Delta_A(\theta_1) \text{ and } \Delta_{A \cup B}(\theta_1)\Delta_B(\theta_2) > \Delta_{A \cup B}(\theta_2)\Delta_B(\theta_1)$$

hold, from which it follows that  $\Delta_{A \cup B}(\theta_1) > 0$ ,  $\Delta_A(\theta_2) > 0$  and  $\Delta_B(\theta_2) > 0$ . Proceeding as in the previous proofs, we obtain that:

$$0 \leq \frac{\Delta_{A \cup B}(\theta_2)}{\Delta_{A \cup B}(\theta_1) + \Delta_{A \cup B}(\theta_2)} < \min \left\{ \frac{\Delta_A(\theta_2)}{\Delta_A(\theta_1) + \Delta_A(\theta_2)}, \frac{\Delta_B(\theta_2)}{\Delta_B(\theta_1) + \Delta_B(\theta_2)} \right\} \leq 1.$$

Thus, there is  $\alpha_0 \in (0, 1)$  such that:

$$0 \leq \frac{\Delta_{A \cup B}(\theta_2)}{\Delta_{A \cup B}(\theta_1) + \Delta_{A \cup B}(\theta_2)} < \alpha_0 < \min \left\{ \frac{\Delta_A(\theta_2)}{\Delta_A(\theta_1) + \Delta_A(\theta_2)}, \frac{\Delta_B(\theta_2)}{\Delta_B(\theta_1) + \Delta_B(\theta_2)} \right\} \leq 1.$$

In addition, there is  $x' \in \mathcal{X}$  such that  $L_{x'}(\theta_1), L_{x'}(\theta_2) > 0$ . For the prior distribution  $\pi^*$  for  $\theta$  given by:

$$\pi^*(\theta_1) = \frac{\alpha_0 L_{x'}(\theta_2)}{\alpha_0 L_{x'}(\theta_2) + (1 - \alpha_0) L_{x'}(\theta_1)} \text{ and } \pi^*(\theta_2) = 1 - \pi^*(\theta_1),$$

the posterior distribution is  $\pi_{x'}^*(\theta_1) = \alpha_0$  and  $\pi_{x'}^*(\theta_2) = 1 - \alpha_0$ . Let  $\varphi^*$  be any TS generated by  $(L_A)_{A \in \sigma(\Theta)}$  with respect to  $\pi^*$ . Next, we consider three cases:

(i) if  $\theta_1 \in A \cap B$ , then:

$$\varphi_C^*(x') = 0 \text{ , if } \alpha_0 > \frac{\Delta_C(\theta_2)}{\Delta_C(\theta_1) + \Delta_C(\theta_2)} \text{ ,}$$

for any  $C \in \{A, B, A \cup B\}$ . Thus, we have  $\varphi_A^*(x') = 1$ ,  $\varphi_B^*(x') = 1$  and  $\varphi_{A \cup B}^*(x') = 0$ ;

(ii) if  $\theta \notin A$ , then:

$$\varphi_B^*(x') = 0 \text{ , if } \alpha_0 > \frac{\Delta_B(\theta_2)}{\Delta_B(\theta_1) + \Delta_B(\theta_2)} \text{ and } \varphi_{A \cup B}^*(x') = 0 \text{ , if } \alpha_0 > \frac{\Delta_{A \cup B}(\theta_2)}{\Delta_{A \cup B}(\theta_1) + \Delta_{A \cup B}(\theta_2)} \text{ ,}$$

and:

$$\int_{\Theta} [L_A(0, \theta) - L_A(1, \theta)] d\pi_{x'}(\theta) = \Delta_A(\theta_1)\alpha_0 + \Delta_A(\theta_2)(1 - \alpha_0) > 0.$$

Thus,  $\varphi_A^*(x') = 1$ ,  $\varphi_B^*(x') = 1$  and  $\varphi_{A \cup B}^*(x') = 0$ ;

(iii) if  $\theta \notin B$ , a development similar to that of Case (ii) yields the same results:  $\varphi_A^*(x') = 1$ ,  $\varphi_B^*(x') = 1$  and  $\varphi_{A \cup B}^*(x') = 0$ .

Therefore, in any case, there is a prior distribution  $\pi^*$  for  $\theta$  with respect to which no TS generated by  $(L_A)_{A \in \sigma(\Theta)}$  meets finite union consonance, concluding the proof.

### E. Proof of Theorem 5

The proof of Theorem 5 consists of verifying the inexistence of such a family of loss functions that generates Bayes tests satisfying the desiderata with respect to all priors concentrated on three points in  $\Theta$  (of course, there will not be such a family satisfying these requisites with respect to all priors over  $\sigma(\Theta)$ ).

Let  $\{A_1, A_2, A_3\}$  be a measurable partition of  $\Theta$  and  $\theta_1, \theta_2, \theta_3 \in \Theta$ , such that  $\theta_i \in A_i, i = 1, 2, 3$ . First, notice that for all  $x \in \mathcal{X}$ , such that  $L_x(\theta_i) > 0$  for  $i = 1, 2, 3$ , there is a one-to-one correspondence between prior and posterior distributions concentrated on  $\{\theta_1, \theta_2, \theta_3\}$ . Indeed, for all  $(\alpha_1, \alpha_2, \alpha_3) \in \mathcal{A} = \{(a, b, c) \in \mathbb{R}_+^3 : a + b + c = 1\}$  and  $x \in \mathcal{X}$ , such that  $L_x(\theta_i) > 0$  for  $i = 1, 2, 3$ , there is a unique prior distribution for  $\theta$ ,  $\pi$ , such that the corresponding posterior distribution given  $x$ ,  $\pi_x$ , satisfies  $\pi_x(\theta_i) = \alpha_i, i = 1, 2, 3$ , namely:

$$\pi(\theta_i) = \frac{\frac{\alpha_i}{L_x(\theta_i)}}{\frac{\alpha_1}{L_x(\theta_1)} + \frac{\alpha_2}{L_x(\theta_2)} + \frac{\alpha_3}{L_x(\theta_3)}} \text{ , } i = 1, 2, 3.$$

Henceforth, we will refer to the above posterior by  $(\alpha_1, \alpha_2, \alpha_3)$  for short. Let  $(L_A)_{A \in \sigma(\Theta)}$  be any family of strict hypothesis testing loss functions. For each  $(\alpha_1, \alpha_2, \alpha_3) \in \mathcal{A}$ , the difference between the posterior risk of accepting  $H_0^{(i)} : \theta \in A_i$  and that of rejecting it is given by:

$$\int_{\Theta} [L_{A_i}(0, \theta) - L_{A_i}(1, \theta)] d\pi_x(\theta) = \Delta_{A_i}(\theta_1)\alpha_1 + \Delta_{A_i}(\theta_2)\alpha_2 + \Delta_{A_i}(\theta_3)\alpha_3,$$

where  $\Delta_{A_i}(\theta_j) = L_{A_i}(0, \theta_j) - L_{A_i}(1, \theta_j)$  (note that  $\Delta_{A_i}(\theta_j) > 0$ , if  $i \neq j$ , while  $\Delta_{A_i}(\theta_i) < 0$ ). In order to evaluate the tests for the hypotheses  $H_0^{(1)}, H_0^{(2)}$  and  $H_0^{(3)}$  with respect to all posterior distributions concentrated on  $\{\theta_1, \theta_2, \theta_3\}$ , we consider the transformation  $T : \mathcal{A} \rightarrow \mathbb{R}^3$  defined by:

$$T(\alpha_1, \alpha_2, \alpha_3) = \left( \int_{\Theta} \Delta_{A_1}(\theta) d\pi_x(\theta), \int_{\Theta} \Delta_{A_2}(\theta) d\pi_x(\theta), \int_{\Theta} \Delta_{A_3}(\theta) d\pi_x(\theta) \right) \text{ ,}$$



where  $\int_{\Theta} \Delta_{A_i}(\theta) d\pi_x(\theta) = \Delta_{A_i}(\theta_1)\alpha_1 + \Delta_{A_i}(\theta_2)\alpha_2 + \Delta_{A_i}(\theta_3)\alpha_3$ . Thus,  $T$  assigns to each posterior  $(\alpha_1, \alpha_2, \alpha_3) \in \mathcal{A}$  the differences between the risks of accepting  $H_0^{(i)}$  and of rejecting it,  $i = 1, 2, 3$ . It is easy to verify that  $\mathcal{B} = T(\mathcal{A}) = \{T(\alpha_1, \alpha_2, \alpha_3) : (\alpha_1, \alpha_2, \alpha_3) \in \mathcal{A}\}$  is a convex set. Indeed,  $\mathcal{B}$  is a triangle (see Figure 2) with vertices  $P_1 = T(1, 0, 0) = (\Delta_{A_1}(\theta_1), \Delta_{A_2}(\theta_1), \Delta_{A_3}(\theta_1))$ ,  $P_2 = T(0, 1, 0) = (\Delta_{A_1}(\theta_2), \Delta_{A_2}(\theta_2), \Delta_{A_3}(\theta_2))$  and  $P_3 = T(0, 0, 1) = (\Delta_{A_1}(\theta_3), \Delta_{A_2}(\theta_3), \Delta_{A_3}(\theta_3))$  (these points are not aligned owing to the restrictions on the quantities  $\Delta_{A_i}(\theta_j)$ , [14]).

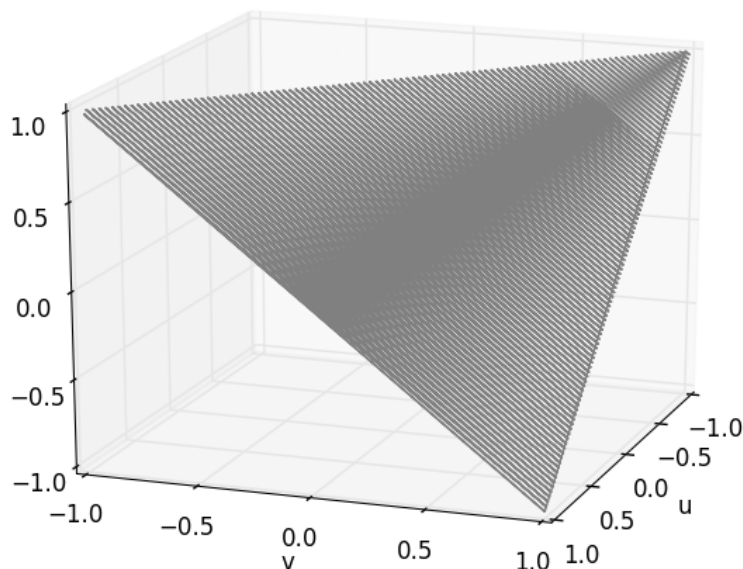


Figure 2. Set  $\mathcal{B}$ .

Now, we turn to the main argument of the proof. By Theorem 4.3 from [8], it is necessary for a Bayesian testing scheme to satisfy the logical requirements with respect to all priors over  $\sigma(\Theta)$  that exactly one of the  $A_i$ 's is accepted for each vector of probabilities  $(\alpha_1, \alpha_2, \alpha_3)$ . Geometrically, such a necessary condition is equivalent to the triangle  $\mathcal{B}$  to be contained in the union of the octants that comprise the triplets with only one negative coordinate, namely  $\mathbb{R}_- \times \mathbb{R}_+ \times \mathbb{R}_+$ ,  $\mathbb{R}_+ \times \mathbb{R}_- \times \mathbb{R}_+$  and  $\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_-$ . However, this is impossible. To verify this fact, we consider three cases (Figure 3 illustrates the projection of  $\mathcal{B}$  over the plane  $\mathbf{w} = \{(u, v, 0) : u, v \in \mathbb{R}\}$  in each of these cases):

- (i) if  $\Delta_{A_1}(\theta_1)\Delta_{A_2}(\theta_2) > \Delta_{A_1}(\theta_2)\Delta_{A_2}(\theta_1)$ , then the projection of the line segment joining  $P_1$  and  $P_2$  over the plane  $\mathbf{w}$  intersects the (third) quadrant  $\mathbb{R}_- \times \mathbb{R}_- \times \{0\}$  (see the first graphic in Figure 3). Thus, there is  $\gamma \in (0, 1)$ , such that  $\gamma\Delta_{A_i}(\theta_1) + (1 - \gamma)\Delta_{A_i}(\theta_2) < 0, i = 1, 2$ . As  $\gamma P_1 + (1 - \gamma)P_2 \in \mathcal{B}$ , there is a posterior  $(\alpha_1, \alpha_2, \alpha_3)$  concentrated on  $\{\theta_1, \theta_2, \theta_3\}$  with respect to which any TS generated by  $(L_A)_{A \in \sigma(\Theta)}$  does not reject both  $A_1$  and  $A_2$  and, therefore, does not respect coherence, invertibility and finite union consonance;
- (ii) if  $\Delta_{A_1}(\theta_1)\Delta_{A_2}(\theta_2) = \Delta_{A_1}(\theta_2)\Delta_{A_2}(\theta_1)$ , then the projection of the line segment joining  $P_1$  and  $P_2$  over  $\mathbf{w}$  intersects the origin  $(0, 0, 0)$  (see the second graphic in Figure 3). Thus, there is  $t_0 > 0$ , such that the point  $P_0 = (0, 0, t_0) \in \mathcal{B}$ . Considering now the line segment joining  $P_0$  and  $P_3$ , it is easily seen that for any  $\gamma \in (\frac{-\Delta_{A_3}(\theta_3)}{t_0 - \Delta_{A_3}(\theta_3)}, 1)$ ,  $\gamma t_0 + (1 - \gamma)\Delta_{A_1}(\theta_3) > 0, \gamma t_0 + (1 - \gamma)\Delta_{A_2}(\theta_3) > 0$  and  $\gamma t_0 + (1 - \gamma)\Delta_{A_3}(\theta_3) > 0$ . As  $\gamma P_0 + (1 - \gamma)P_3 \in \mathcal{B}$ , there is a posterior distribution with

respect to which any TS generated by  $(L_A)_{A \in \sigma(\Theta)}$  rejects  $A_1, A_2$  and  $A_3$  and, therefore, does not satisfy the logical consistency properties all together;

- (iii) if  $\Delta_{A_1}(\theta_1)\Delta_{A_2}(\theta_2) < \Delta_{A_1}(\theta_2)\Delta_{A_2}(\theta_1)$ , then the projection of the above-mentioned segment over  $\mathbf{w}$  intersects the (first) quadrant  $\mathbb{R}_+ \times \mathbb{R}_+ \times \{0\}$  (third graphic in Figure 3). Thus, there is  $\gamma \in (0, 1)$  such that  $\gamma\Delta_{A_i}(\theta_1) + (1 - \gamma)\Delta_{A_i}(\theta_2) > 0, i = 1, 2$ . As  $\gamma P_1 + (1 - \gamma)P_2 \in \mathcal{B}$ , there is a posterior  $(\alpha_1, \alpha_2, \alpha_3)$  concentrated on  $\{\theta_1, \theta_2, \theta_3\}$  with respect to which any TS generated by  $(L_A)_{A \in \sigma(\Theta)}$  rejects  $A_1, A_2$  and  $A_3$  and, consequently, does not meet the desiderata.

From (i)–(iii), the result follows.

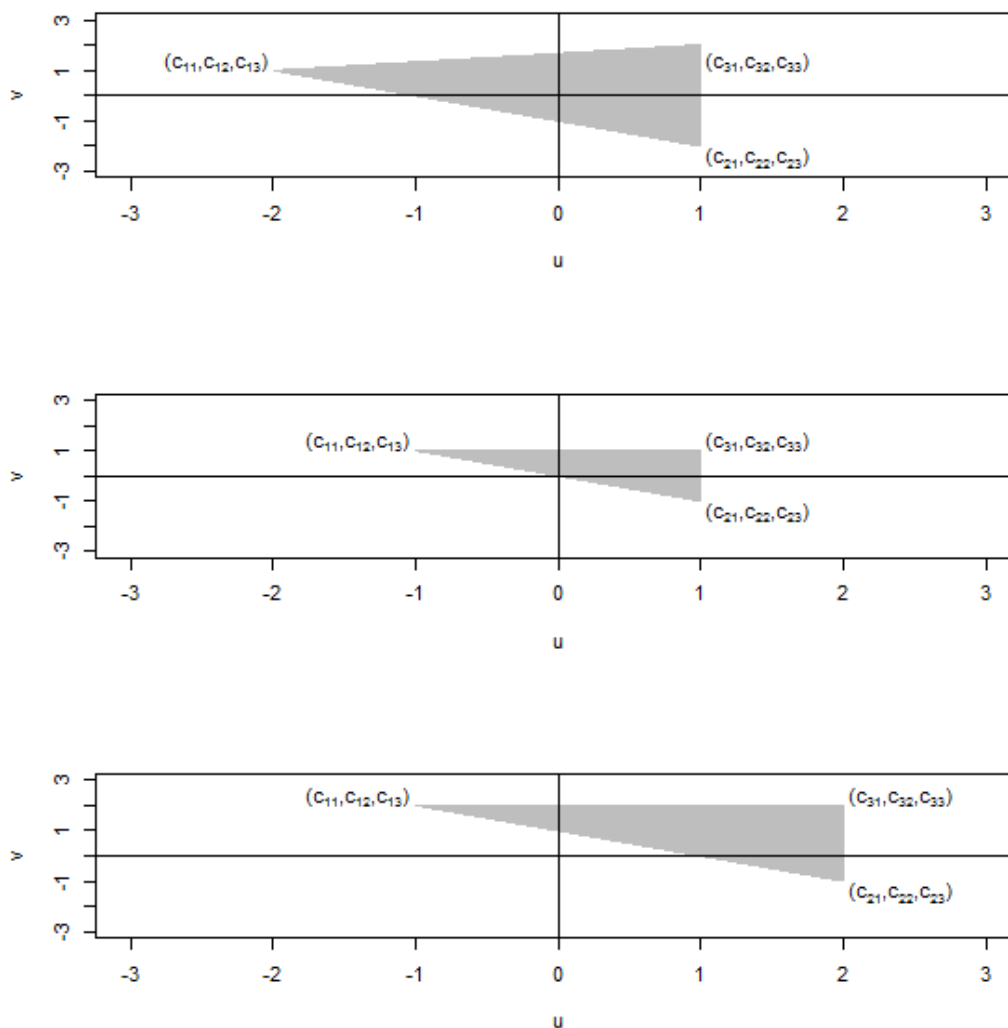


Figure 3. Projection of  $\mathcal{B}$  in  $u \times v$ .

### F. Proof of Theorem 6

Let  $\varphi$  be a TS satisfying coherence, invertibility and countable union consonance. From Theorem 1, there is a unique point estimator  $\delta : \mathcal{X} \rightarrow \Theta$ , such that for all  $A \in \sigma(\Theta)$  and  $x \in \mathcal{X}, \varphi_A(x) = \mathbb{I}(\delta(x) \notin A)$ . For each  $x \in \mathcal{X}$ , define  $\mu_x : \sigma(\Theta) \rightarrow \mathbb{R}_+$  by:

$$\mu_x(A) = 1 - \varphi_A(x) = \mathbb{I}(\delta(x) \in A),$$

that is,  $\mu_x$  is the probability measure degenerate at the point  $\delta(x)$  [14]. Furthermore, let  $\mu_0$  be any probability measure defined on  $\mathcal{P}(\mathcal{X})$ . Defining  $\mu : \mathcal{P}(\Theta \times \mathcal{X}) \rightarrow \mathbb{R}_+$  by:

$$\mu(B) = \sum_{(\theta,x) \in B} \mu_0(\{x\})\mu_x(\{\theta\}), \quad B \in \mathcal{P}(\Theta \times \mathcal{X}),$$

it is immediate that  $\mu$  is a probability measure and that  $\mu_x$  is the conditional distribution of  $\theta$  given  $X = x$ , for each  $x \in \mathcal{X}$ . Next, let  $(L_A)_{A \in \sigma(\Theta)}$  be any family of strict hypothesis testing loss functions. Let  $\varphi^*$  be a testing scheme generated by  $(L_A)_{A \in \sigma(\Theta)}$  with respect to the  $\mu$ -marginal distribution of  $\theta$ . Let us verify that  $\varphi^*$  coincides with  $\varphi$ . Indeed, for  $x \in \mathcal{X}$  and  $A \in \sigma(\Theta)$ , we have:

$$\begin{aligned} \varphi_A(x) = 0 &\Rightarrow \mu_x(A) = 1 \Rightarrow \\ \Rightarrow \sum_{\theta \in \Theta} [L_A(0, \theta) - L_A(1, \theta)]\mu_x(\theta) &= \sum_{\theta \in A} [L_A(0, \theta) - L_A(1, \theta)]\mu_x(\theta) < 0 \Rightarrow \varphi_A^*(x) = 0. \end{aligned}$$

Similarly,

$$\begin{aligned} \varphi_A(x) = 1 &\Rightarrow \mu_x(A) = 0 \Rightarrow \\ \Rightarrow \sum_{\theta \in \Theta} [L_A(0, \theta) - L_A(1, \theta)]\mu_x(\theta) &= \sum_{\theta \in A^c} [L_A(0, \theta) - L_A(1, \theta)]\mu_x(\theta) > 0 \Rightarrow \varphi_A^*(x) = 1, \end{aligned}$$

concluding the proof. It should be emphasized that there are many other probability measures over  $\mathcal{P}(\Theta \times \mathcal{X})$  and families of strict hypothesis testing loss functions that yield the result of Theorem 6. For instance, considering for each  $x \in \mathcal{X}$ , a conditional probability measure  $\mu'_x$ , such that  $\mu'_x(\delta(x)) > 1/2$  and  $\mu'_x(\theta) > 0$ , for all  $\theta \in \Theta$ , together with the family of 0–1 loss functions, one will obtain a Bayesian TS that coincides with  $\varphi$ , as well (see [14] for the details).

**G. Proof of Theorem 7**

To prove Part (a), we define a family of loss functions that generates Bayesian testing schemes satisfying both coherence and invertibility with respect to all prior distributions for  $\theta$ , which implies, by Theorem 3.1 from [28], that, for each sample point, at most one hypothesis of each partition of  $\Theta$  is not rejected. Next, we prove that, for each  $x \in \mathcal{X}$ , there is a singleton that is not rejected with respect to the prior  $\pi$ . Combining these facts, we prove that, for each sample point, exactly one hypothesis of each partition of  $\Theta$  is accepted, which is equivalent (Theorem 4.3 from [8]) to asserting that the TS generated by that family of losses with respect to  $\pi$  meets the desiderata.

Thus, for  $A \in \sigma(\Theta)$ , let  $L_A : \{0, 1\} \times (\Theta \times \mathcal{X}) \rightarrow \mathbb{R}_+$  be given, for  $\theta \in A^c$  and  $x \in \mathcal{X}$ , by  $L_A(1, (\theta, x)) = 0$  and:

$$L_A(0, (\theta, x)) = \min \left\{ \min \left\{ L(d, \theta); \frac{1}{L(d, \theta)} \right\} I_A(\delta(x)) + \max \left\{ L(d, \theta); \frac{1}{L(d, \theta)} \right\} I_{A^c}(\delta(x)) : d \in A \right\},$$

and, for  $\theta \in A$  and  $x \in \mathcal{X}$ , by  $L_A(0, (\theta, x)) = 0$  and:

$$L_A(1, (\theta, x)) = \min \left\{ \frac{1}{C} \min \left\{ L(d, \theta); \frac{1}{L(d, \theta)} \right\} I_{A^c}(\delta(x)) + C \max \left\{ L(d, \theta); \frac{1}{L(d, \theta)} \right\} I_A(\delta(x)) : d \in A^c \right\},$$

where  $C > 1$  is any constant greater than  $\max \left\{ \frac{E[L(\delta(x), \theta)|x]}{\pi_x(\delta(x))} : x \in \mathcal{X} \right\}$ .

These hypothesis testing loss functions do not penalize correct decisions. They also reflect the decision-maker’s tendency to not reject the hypotheses that comprise the best estimate for  $\theta$ ,  $\delta(x)$ .

For instance, if  $\theta \in A$  and one decides to reject  $A$  on the basis of the sample  $x$ , the loss of falsely rejecting  $A$  is:

$$\begin{cases} C \min \left\{ \max \left\{ L(d, \theta); \frac{1}{L(d, \theta)} \right\} : d \in A^c \right\}, & \text{if } \delta(x) \in A \\ \frac{1}{C} \min \left\{ \min \left\{ L(d, \theta); \frac{1}{L(d, \theta)} \right\} : d \in A^c \right\}, & \text{otherwise.} \end{cases}$$

If  $C$  is large enough, of course the decision-maker will be more reluctant to reject  $A$  in case of  $\delta(x) \in A$  than to reject it if not.

This family of loss functions satisfies the condition in Equation (2) of Theorem 2. In fact, for all  $A, B \in \sigma(\Theta)$ , with  $A \subseteq B$ , for all  $\theta_1 \in A, \theta_2 \in B^c$  and  $x \in \mathcal{X}$ , we have:

(i) if  $\delta(x) \in A$ , then:

$$\frac{\Delta_A(\theta_1)}{\Delta_B(\theta_1)} = \frac{C \min \left\{ \max \left\{ L(d, \theta_1); \frac{1}{L(d, \theta_1)} \right\} : d \in A^c \right\}}{C \min \left\{ \max \left\{ L(d, \theta_1); \frac{1}{L(d, \theta_1)} \right\} : d \in B^c \right\}} \leq 1 \leq \frac{\min \left\{ \min \left\{ L(d, \theta_2); \frac{1}{L(d, \theta_2)} \right\} : d \in A \right\}}{\min \left\{ \min \left\{ L(d, \theta_2); \frac{1}{L(d, \theta_2)} \right\} : d \in B \right\}} = \frac{\Delta_A(\theta_2)}{\Delta_B(\theta_2)}.$$

(ii) if  $\delta(x) \in B \cap A^c$  (recall  $C \geq 1$ ), it follows that:

$$\frac{\Delta_A(\theta_1)}{\Delta_B(\theta_1)} = \frac{\frac{1}{C} \min \left\{ \min \left\{ L(d, \theta_1); \frac{1}{L(d, \theta_1)} \right\} : d \in A^c \right\}}{C \min \left\{ \max \left\{ L(d, \theta_1); \frac{1}{L(d, \theta_1)} \right\} : d \in B^c \right\}} \leq 1 \leq \frac{\min \left\{ \max \left\{ L(d, \theta_2); \frac{1}{L(d, \theta_2)} \right\} : d \in A \right\}}{\min \left\{ \min \left\{ L(d, \theta_2); \frac{1}{L(d, \theta_2)} \right\} : d \in B \right\}} = \frac{\Delta_A(\theta_2)}{\Delta_B(\theta_2)}.$$

(iii) if  $\delta(x) \in B^c$ ,

$$\frac{\Delta_A(\theta_1)}{\Delta_B(\theta_1)} = \frac{\frac{1}{C} \min \left\{ \min \left\{ L(d, \theta_1); \frac{1}{L(d, \theta_1)} \right\} : d \in A^c \right\}}{\frac{1}{C} \min \left\{ \min \left\{ L(d, \theta_1); \frac{1}{L(d, \theta_1)} \right\} : d \in B^c \right\}} \leq 1 \leq \frac{\min \left\{ \max \left\{ L(d, \theta_2); \frac{1}{L(d, \theta_2)} \right\} : d \in A \right\}}{\min \left\{ \max \left\{ L(d, \theta_2); \frac{1}{L(d, \theta_2)} \right\} : d \in B \right\}} = \frac{\Delta_A(\theta_2)}{\Delta_B(\theta_2)}.$$

Therefore,  $(L_A)_{A \in \sigma(\Theta)}$  generates coherent testing schemes with respect to all prior distribution for  $\theta$ , if  $C \geq 1$ . Furthermore, for all  $A \in \sigma(\Theta)$ ,  $\theta_1 \in A$  and  $\theta_2 \in A^c$ , we have, if  $\delta(x) \in A$ , that:

$$\frac{\Delta_A(\theta_2)}{\Delta_{A^c}(\theta_2)} = \frac{\min \left\{ \min \left\{ L(d, \theta_2); \frac{1}{L(d, \theta_2)} \right\} : d \in A \right\}}{\frac{1}{C} \min \left\{ \min \left\{ L(d, \theta_2); \frac{1}{L(d, \theta_2)} \right\} : d \in A \right\}} = \frac{1}{C} = C = \frac{C \min \left\{ \max \left\{ L(d, \theta_1); \frac{1}{L(d, \theta_1)} \right\} : d \in A^c \right\}}{\min \left\{ \max \left\{ L(d, \theta_1); \frac{1}{L(d, \theta_1)} \right\} : d \in A^c \right\}} = \frac{\Delta_A(\theta_1)}{\Delta_{A^c}(\theta_1)}.$$

Analogously, we prove that the condition in Equation (3) of Theorem 3 is fulfilled if  $\delta(x) \notin A$ . Thus, there are testing schemes generated by  $(L_A)_{A \in \sigma(\Theta)}$  that respect invertibility with respect to all priors. Finally, let us prove that a TS  $\varphi$  generated by  $(L_A)_{A \in \sigma(\Theta)}$  is such that  $\varphi_{\{\delta(x)\}}(x) = 0$ , for all  $x \in \mathcal{X}$ . Indeed,

$$\begin{aligned} & \sum_{\theta \in \Theta} L_{\{\delta(x)\}}(0, (\theta, x)) \pi_x(\theta) = \sum_{\theta \neq \delta(x)} L_{\{\delta(x)\}}(0, (\theta, x)) \pi_x(\theta) \\ & = \sum_{\theta \neq \delta(x)} \min \left\{ L(\delta(x), \theta); \frac{1}{L(\delta(x), \theta)} \right\} \pi_x(\theta) \leq \sum_{\theta \neq \delta(x)} L(\delta(x), \theta) \pi_x(\theta) \end{aligned} \tag{8}$$

and:

$$\begin{aligned} & \sum_{\theta \in \Theta} L_{\{\delta(x)\}}(1, (\theta, x)) \pi_x(\theta) = L_{\{\delta(x)\}}(1, (\delta(x), x)) \pi_x(\delta(x)) \\ & = C \min \left\{ \max \left\{ L(d, \delta(x)); \frac{1}{L(d, \delta(x))} \right\} : d \neq \delta(x) \right\} \pi_x(\delta(x)) \\ & > \min \left\{ \max \left\{ L(d, \delta(x)); \frac{1}{L(d, \delta(x))} \right\} : d \neq \delta(x) \right\} \sum_{\theta \neq \delta(x)} L(\delta(x), \theta) \pi_x(\theta), \end{aligned} \tag{9}$$

since  $C > \max \left\{ \frac{E[L(\delta(x), \theta) | x]}{\pi_x(\delta(x))} : x \in \mathcal{X} \right\} \geq \frac{\sum_{\theta \neq \delta(x_0)} L(\delta(x_0), \theta) \pi_{x_0}(\theta)}{\pi_{x_0}(\delta(x_0))}$ , for any  $x_0 \in \mathcal{X}$ .

From Equations (8) and (9), it follows that:

$$\sum_{\theta \in \Theta} L_{\{\delta(x)\}}(1, (\theta, x)) \pi_x(\theta) > \min \left\{ \max \left\{ L(d, \delta(x)); \frac{1}{L(d, \delta(x))} \right\} : d \neq \delta(x) \right\} \sum_{\theta \in \Theta} L_{\{\delta(x)\}}(0, (\theta, x)) \pi_x(\theta).$$

Therefore,  $\sum_{\theta \in \Theta} L_{\{\delta(x)\}}(1, (\theta, x)) \pi_x(\theta) > \sum_{\theta \in \Theta} L_{\{\delta(x)\}}(0, (\theta, x)) \pi_x(\theta)$  and, consequently,  $\varphi_{\{\delta(x)\}}(x) = 0$ , concluding the proof of Part (a).

For Part (b), suppose  $\varphi$  is generated by  $(L_A)_{A \in \sigma(\Theta)}$  with respect to  $\pi$ . From Theorem 4.3 from [8] and Theorem 1, it follows that for all  $x \in \mathcal{X}$ ,  $\varphi_{\{\delta(x)\}}(x) = 0$  and  $\varphi_{\{d\}}(x) = 1$ , for all  $d \neq \delta(x)$ . Thus,

$$\sum_{\theta \in \Theta} [L_{\{\delta(x)\}}(0, \theta) - L_{\{\delta(x)\}}(1, \theta)] \pi_x(\theta) \leq 0 \leq \sum_{\theta \in \Theta} [L_{\{d\}}(0, \theta) - L_{\{d\}}(1, \theta)] \pi_x(\theta),$$

for  $d \neq \delta(x)$ , where  $\pi_x$  is the posterior distribution for  $\theta$  given  $x$ . Defining  $L : \Theta \times \Theta \rightarrow \mathbb{R}_+$  by:

$$\begin{aligned} L(d, \theta) &= [L_{\{d\}}(0, \theta) - L_{\{d\}}(1, \theta)] - \min\{L_{\{d'\}}(0, \theta) - L_{\{d'\}}(1, \theta) : d' \in \Theta\} \\ &= [L_{\{d\}}(0, \theta) - L_{\{d\}}(1, \theta)] - [L_{\{\theta\}}(0, \theta) - L_{\{\theta\}}(1, \theta)], \end{aligned}$$

it follows that:

$$\sum_{\theta \in \Theta} L(\delta(x), \theta) \pi_x(\theta) \leq \sum_{\theta \in \Theta} L(d, \theta) \pi_x(\theta),$$

for  $d \neq \delta(x)$ , for each  $x \in \mathcal{X}$ . Therefore,  $\delta$  is a Bayes estimator for  $\theta$  generated by  $L$  with respect to  $\pi$ . Notice that  $L$  (essentially) assigns to the estimate  $d$  for the parameter the difference between the loss of not rejecting the hypothesis  $\{d\}$  and that of rejecting it when the state of nature is  $\theta$ . It seems reasonable that the greater the “distance” between  $d$  and  $\theta$ , the greater this difference (and, consequently,  $L(d, \theta)$ ) should be.

**References**

1. Hommel, G.; Bretz, F. Aesthetics and power considerations in multiple testing—A contradiction? *Biom. J.* **2008**, *20*, 657–666.
2. Shaffer, J.P. Multiple hypothesis testing. *Ann. Rev. Psychol.* **1995**, *46*, 561–584.
3. Hochberg, Y.; Tamhane, A.C. *Multiple Comparison Procedures*; Wiley: New York, NY, USA, 1987.
4. Farcomeni, A. A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion. *Stat. Methods Med. Res.* **2008**, *17*, 347–388.
5. Schervish, M.J. *Theory of Statistics*; Springer: New York, NY, USA, 1997.
6. Gabriel, K.R. Simultaneous test procedures—Some theory of multiple comparisons. *Ann. Math. Stat.* **1969**, *41*, 224–250.
7. Lehmann, E.L. A theory of some multiple decision problems, II. *Ann. Math. Stat.* **1957**, *28*, 547–572.
8. Izbicki, R.; Esteves, L.G. Logical consistency in Simultaneous Statistical Test Procedures. *Log. J. IGPL* **2015**, doi:10.1093/jigpal/jzv027.

9. Pereira, C.A.B.; Stern, J.M.; Wechsler, S. Can a significance test be genuinely Bayesian? *Bayesian Anal.* **2008**, *3*, 79–100.
10. Stern, J.M. Constructive verification, empirical induction and fallibilist deduction: A threefold contrast. *Information* **2001**, *2*, 635–650.
11. Pereira, C.A.B.; Stern, J.M. Evidence and credibility: Full Bayesian significance test for precise hypotheses. *Entropy* **1999**, *1*, 99–110.
12. Lin, S.K.; Chen, C.K.; Ball, D.; Liu, H.C.; Loh, E.W. Gender-specific contribution of the GABAA subunit genes on 5q33 in methamphetamine use disorder. *Pharm. J.* **2003**, *3*, 349–355.
13. Izbicki, R.; Fossaluzza, V.; Hounie, A.G.; Nakano, E.Y.; Pereira, C.A. Testing allele homogeneity: The problem of nested hypotheses. *BMC Genet.* **2012**, *13*, doi:10.1186/1471-2156-13-103.
14. Silva, G.M. Propriedades Lógicas de Classes de Testes de Hipóteses. Ph.D. Thesis, University of São Paulo, São Paulo, Brazil, 2014. (in Portuguese)
15. Evans, M. *Measuring Statistical Evidence Using Relative Belief*; Chapman & Hall/CRC: London, UK, 2015.
16. Marcus, R.; Eric, P.; Gabriel, K.R. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* **1976**, *63*, 655–660.
17. Sonnemann, E. General solutions to multiple testing problems. *Biom. J.* **2008**, *50*, 641–656.
18. Lavine, M.; Schervish, M.J. Bayes Factors: What they are and what they are not. *Am. Stat.* **1999**, *53*, 119–122.
19. Kneale, W.; Kneale, M. *The Development of Logic*; Oxford University Press: Oxford, UK, 1962.
20. DeGroot, M.H. *Optimal Statistical Decisions*; McGraw-Hill: New York, NY, USA, 1970.
21. Finner, H.; Strassburger, K. The partitioning principle: A powerful tool in multiple decision theory. *Ann. Stat.* **2002**, *30*, 1194–1213.
22. Aitchison, J. Confidence-region tests. *J. R. Stat. Soc. Ser. B* **1964**, *26*, 462–476.
23. Darwiche, A.Y.; Ginsberg, M.L. A Symbolic Generalization of Probability Theory. In Proceedings of the Tenth National Conference on Artificial Intelligence, AAAI-92, San Jose, CA, USA, 12–16 July 1992
24. Evans, M.; Jang, G.H. Inferences from Prior-Based Loss Functions. **2011**, arXiv:1104.3258.
25. Berger, J. O. In defense of the likelihood principle: axiomatics and coherency. *Bayesian Stat.* **1985**, *2*, 33–66.
26. Madruga, M.R.; Esteves, L.G.; Wechsler, S. On the bayesianity of pereira-stern tests. *Test* **2001**, *10*, 291–299.
27. Ripley, B.D. *Pattern Recognition and Neural Networks*; Cambridge University Press: Cambridge, UK, 1996.
28. Izbicki, R. Classes de Testes de Hipóteses. Ph.D. Thesis, University of São Paulo, São Paulo, Brazil, 2010. (in Portuguese)