

DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Relatório Técnico

RT-MAC-2002-07

*SIGNIFICANCE TESTS, BELIEF CALCULI, AND BURDEN  
OF PROOF IN LEGAL AND SCIENTIFIC DISCOURSE*

Julio Michael Stern

Setembro de 2002

# Significance Tests, Belief Calculi, and Burden of Proof in Legal and Scientific Discourse

Julio Michael Stern  
BIOINFO and Computer Science Department  
University of Sao Paulo, Brazil

September 27, 2002

## Abstract

We review the definition of the Full Bayesian Significance Test (FBST), and summarize its main statistical and epistemological characteristics. We review also the Abstract Belief Calculus (ABC) of Derwiche and Ginsberg, and use it to analyze the FBST's value of evidence. This analysis helps us understand the FBST properties, specially its consistency with the Onus Probandi principle, and why it overcomes several deficiencies of well known sharp hypothesis testing theories, both in frequentist and standard Bayesian Statistics.

KEY WORDS: Abstract Belief Calculus; Bayesian statistics; Common-sense reasoning, Evidence; Onus Probandi; Significance test; Uncertainty.  
E-MAIL: [jsfern@ime.usp.br](mailto:jsfern@ime.usp.br)

## 1 Introduction and Summary

The Full Bayesian Significance Test (FBST), first presented in Pereira and Stern (1999) as a coherent Bayesian significance test for sharp hypothesis, is a test based on a value of evidence concept, whose definition was originally motivated by practical, juridical and epistemological requirements. These requirements, however, even though mentioned in the author's previous articles, have never been formally analyzed. This analysis, which we pursue in sections 4, 6 and 7 below with the aid of the Abstract Belief Calculus (ABC) formalism, as defined in Derwiche and Ginsberg (1992) and Derwiche (1993), constitutes the main objective of the present article.

For clarity and completeness: the FBST is defined in section 2; the value of evidence concept and its motivating requirements are presented in section 4; and, the ABC formalism is presented in sections 3 and 5.

## 2 The FBST Value of Evidence

Let  $X_1, \dots, X_n$  be random variables having joint density  $f(x_1, \dots, x_n; \theta)$ , indexed by a vector parameter  $\theta$  in a parameter space  $\Theta \subseteq \mathcal{R}^p$ . The (null) hypothesis  $H$  states that the parameter lies in the null set,

$$\Theta_H = \{\theta \in \Theta \mid g(\theta) \leq \gamma \wedge h(\theta) = \zeta\}$$

We are particularly interested in sharp (precise) hypotheses, i.e., those in which  $\dim(\Theta_H) < \dim(\Theta)$ .

Let  $L(\theta; x_1, \dots, x_n)$  denote the likelihood function of  $\theta$  on  $\Theta$ . Let  $p(\theta)$  be a prior density, and  $r(\theta)$  a reference density on  $\Theta$ . We denote by  $p_n(\theta)$  the posterior density,

$$p_n(\theta) \propto L(\theta; x_1, \dots, x_n)p(\theta).$$

The FBST value of evidence against the hypothesis,  $Ev(H)$ , is defined by

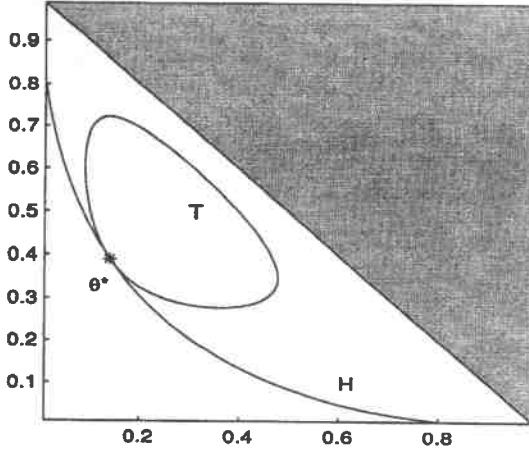
$$\begin{aligned} Ev(H) &= \int_{T_H} p_n(\theta) d\theta, \text{ where} \\ T_H &= \{\theta \in \Theta \mid s(\theta) > s_H\} \\ s_H &= \sup_{\theta \in \Theta_H} s(\theta) \\ s(\theta) &= \left( \frac{p_n(\theta)}{r(\theta)} \right) \end{aligned}$$

The function  $s(\theta)$  is known as the posterior surprise relative to the reference density  $r(\theta)$ . The surprise function was used, among other statisticians, by Good (1983), Evans (1997) and Royall (1997). Its role is to make  $Ev(H)$  explicitly invariant under suitable transformations on the coordinate system of the parameter space, see appendix A.

The tangential set  $T_H$  is a Highest Relative Surprise Set (HRSS). It contains the points of the parameter space with higher surprise, relative to the reference density, than any point in the null set  $\Theta_H$ . When  $r(\theta) \propto 1$ ,  $T_H$  is the Posterior's Highest Density Probability Set (HDPS) tangential to the null set  $\Theta_H$ .

The posterior probability of  $T_H$  gives an indication of inconsistency between the posterior and the hypothesis: "Small" values of  $Ev(H)$  indicate that the hypothesis traverses high density regions, providing weak evidence against the hypothesis. On the other hand, if the posterior probability of  $T_H$  is "large", the null set is in a region of low posterior density, and the data provides strong evidence, large  $Ev(H)$ , against the hypothesis.

The value of evidence, defined above, has a simple and intuitive geometric characterization. Figure 1 shows the null set  $\Theta_H$ , the tangential HRSS  $T_H$ , and the point of constrained maximum,  $\theta^*$ , for testing Hardy-Weinberg equilibrium law in a population genetics problem, as discussed in Pereira and Stern (1999).



In this biological application  $n$  is the sample size,  $x_1$  and  $x_3$  are the two homozygote sample counts and  $x_2 = n - x_1 - x_3$  is heterozygote sample count.  $\theta = [\theta_1, \theta_2, \theta_3]$  is the parameter vector. The posterior and reference densities for this trinomial model, the parameter space and the null set are:

$$\begin{aligned} p_n(\theta | x) &\propto \theta_1^{x_1} \theta_2^{x_2} \theta_3^{x_3}, \quad r(\theta) \propto 1 \\ \Theta &= \{\theta \geq 0 \mid \theta_1 + \theta_2 + \theta_3 = 1\} \\ \Theta_H &= \{\theta \in \Theta \mid \theta_3 = (1 - \sqrt{\theta_1})^2\} \end{aligned}$$

Several other applications of the FBST, details of its numerical implementation, suggestive remarks on its epistemological implications, and an extensive list of references can be found in Irony et al. (2002), Lauretto et al. (2002), Madruga et al. (2001), Madruga et al. (2002) and Pereira and Stern (1999,2001).

### 3 Abstract Belief Calculus

The FBST was originally motivated by some requirements on what constitutes a valid value of evidence against a hypothetical statement. Under appropriate circumstances, these requirements are commonsense in juridical reasoning. These requirements will be precisely stated in the next section, using the ABC formalism presented below.

Abstract Belief Calculus (ABC) is defined in Derwiche and Ginsberg (1992) and Derwiche (1993) as a symbolic generalization of Probability calculus. ABC is a powerful tool. Besides being capable of handling both numerical and symbolic beliefs, it also sets the foundations for computational algorithms for abstract belief propagation. ABC also unifies a number of concrete uncertainty calculi proposed in the literature. It is in this particular context that we will use ABC to analyze the value of evidence concept in the Full Bayesian Significance Testing setup.

The first concept in ABC is that of an abstract Support Function,  $\Phi$ , which attributes abstract support values to statements in a universe  $\mathcal{U}$ . We use conventional set theory notation to denote the range of statements support values by  $\Phi(\mathcal{U})$ . Axioms A1 to A5, below, impose coherence conditions on support states.

A1: Under any support function, equivalent statements must have the same support value, i.e.,

$$(A \Leftrightarrow B) \Rightarrow \Phi(A) = \Phi(B)$$

A2: There exists a Support Summation,

$$\oplus : \Phi(\mathcal{U}) \times \Phi(\mathcal{U}) \mapsto \Phi(\mathcal{U})$$

such that, under any support function, the support value of the disjunction of any two logically disjoint statements is a function of their individual support values.

$$\neg(A \wedge B) \Rightarrow \Phi(A \vee B) = \Phi(A) \oplus \Phi(B)$$

A3: Under any support function, if statement A implies statement B, which, in turn, implies statement C, and statements A and C have the same support value, then all three statements have the same support value, i.e.,

$$((A \Rightarrow B \Rightarrow C) \wedge (\Phi(A) = \Phi(C))) \Rightarrow \Phi(B) = \Phi(A)$$

A4: Under any support function, false statements have zero support value, i.e.,

$$A \text{ false} \Rightarrow \Phi(A) = 0$$

A5: Under any support function, tautological statements have full support value

$$A \text{ true} \Rightarrow \Phi(A) = 1$$

It can be shown, see Derwiche (1993), that under Axioms A1 to A5 the support summation is a partial function defined for each  $a, b \in \Phi(\mathcal{U})$  which are support values of logically disjoint statements. More precisely, for each  $a, b \in \Phi(\mathcal{U})$  such that there are statements  $A, B \in \mathcal{U}$  for which  $a = \Phi(A)$ ,  $b = \Phi(B)$  and  $\neg(A \vee B)$ . Moreover, support summation has the following algebraic properties:

X0: Symmetry,

$$a \oplus b = b \oplus a$$

X1: Transitivity,

$$(a \oplus b) \oplus c = a \oplus (b \oplus c)$$

X2: Convexity,

$$\text{if } a \oplus b \oplus c = a \text{ then } a \oplus b = a$$

X3: There is a unique element  $0 \in \Phi(\mathcal{U})$  such that

$$\forall a \in \Phi(\mathcal{U}), a \oplus 0 = a$$

X4: There is a unique element  $1 \in \Phi(\mathcal{U})$  such that  $1 \neq 0$  and

$$\forall a \in \Phi(\mathcal{U}), \exists b \in \Phi(\mathcal{U}) \mid a \oplus b = 1$$

The pair support function and support summation,  $\langle \Phi, \oplus \rangle$  is called a Partial Support Structure. Partial support structures for some uncertainty calculi, namely, classical logic, probability calculus, possibility calculus, and disbelief calculus, are given in table 1. In table 1 all support functions have numeric values, however a support function could also give symbolic (non-numeric) values.

Table 1: Examples of partial support structures

$\Phi(\mathcal{U})$	$a \oplus b$	0	1	$a \preceq b$	Calculus
$\{0, 1\}$	$\max(a, b)$	0	1	$a \leq b$	Classical Logic
$[0, 1]$	$a + b$	0	1	$a \leq b$	Probability
$[0, 1]$	$\max(a, b)$	0	1	$a \leq b$	Possibility
$\{0 \dots \infty\}$	$\min(a, b)$	$\infty$	0	$b \leq a$	Disbelief

The support value of a statement does not determine, in general, the support value of its negation. For any support function  $\Phi$ , however ABC defines the belief function

$$\tilde{\Phi}(A) = \langle \Phi(A), \Phi(\neg A) \rangle$$

for which the belief value of a statement does determine the belief value of its negation.

The partial support structures can also be used to define partial orders on  $\Phi(\mathcal{U})$  and on  $\tilde{\Phi}(\mathcal{U})$ . The symbol  $\preceq$  is used for the support order, and the symbol  $\sqsubseteq$  is used for the belief order.

$$a \preceq b \Leftrightarrow \exists c \mid a \oplus c = b$$

$$\langle a, b \rangle \sqsubseteq \langle c, d \rangle \Leftrightarrow a \preceq c \text{ and } d \preceq b$$

The extreme, minimal and maximal, states of support and belief, with respect to these orders are, respectively, 0 and 1 for the support order, and  $\langle 0, 1 \rangle$  and  $\langle 1, 0 \rangle$  for the belief order. Statements with minimal and maximal belief are said to be, respectively, Rejected and Accepted.

## 4 Value of Evidence and Onus Probandi

The definition of value of evidence against a hypothesis, in the Full Bayesian Significance Testing setup, was motivated by applications of Bayesian statistical reasoning to legal matters where the hypotheses were, possibly sharp, defendants statements, to be judged according to the Onus Probandi principle, Pereira and Stern (1999). In this setup, our interpretation of the Onus Probandi principle in the Bayesian statistics context establishes some basic requirements for the

support value,  $\Phi(H) = \overline{\text{Ev}}(H) \leftarrow 1 - \text{Ev}(H)$ , of a hypothesis,  $H : \theta \in \Theta_H \subseteq \Theta$ . Namely:

**R1 Value of Evidence as a Probability:** The value of evidence against a hypothesis,  $H$ , must be defined by a posterior probability on the parameter space, i.e.,

$$\text{Ev}(H) = \int_{T_H} p_n(\theta) d\theta$$

**R2 Relative Surprise:** Whether a parameter point  $\theta$  belongs or not to  $T_H$  depends only on the value of its posterior surprise, relative to a reference density  $r(\theta)$ ,  $s(\theta) = p_n(\theta)/r(\theta)$ , in comparison to the range  $s(\Theta_H)$ .

**R3 No Self Incrimination:**

$$\Theta_H \cap T_H = \emptyset$$

**R4 De Morgan's Law:**

$$\text{if } H = A \vee B \text{ then } T_H = T_A \cap T_B$$

**R5 Most Favorable Interpretation:**

$$\text{if } H = A \vee B \text{ then } \overline{\text{Ev}}(H) = \max(\overline{\text{Ev}}(A), \overline{\text{Ev}}(B))$$

**R6 Coherent Support:**  $(\overline{\text{Ev}}, \max)$  must be a partial support structure.

**R7 Continuity:** If the posterior density  $p_n(\theta)$  and the constraints defining the null set,

$$\Theta_H = \{\theta \in \Theta \mid g(\theta) \leq \gamma \wedge h(\theta) = \zeta\},$$

are smooth (continuous, differentiable, etc.) functions on its arguments, then so is  $\text{Ev}(H)$ .

**R8 Invariance:**  $\text{Ev}(H)$  must be invariant under bijective smooth reparameterizations, i.e. transformations of the parameter space coordinate system.

**R9 Consistency:**  $\text{Ev}(H)$  must be a consistent acceptance/rejection indicator for the hypothesis being tested, in the sense that  $\text{Ev}(H)$  converges to 0 or 1, according to whether  $H$  is true or false, as the sample size increases.

Defining value of evidence by means of a probability measure is common to most statistical theories of significance. In frequentist statistics, for example, a p-value is defined as the probability that, under the hypothesis, a sample point is at least as "extreme" as the observed data. This is a probability on the sample space. The concept of p-value also requires an order in the sample space to define how extreme a point is. For a critical analysis of p-values, see Kempthorne (1980) and Pereira and Wechsler (1993).

In Bayesian statistics, a value of evidence is usually defined as a probability on the parameter space, as required in R1. According to Basu (1988), Good

(1983) and other statisticians, requiring  $Ev(H)$  to depend on the observed data only though the likelihood function is in the essence of the Likelihood Principle. This is enforced by R1 and R2.

Onus Probandi is a basic principle of legal reasoning, also known as Burden of Proof, see Gaskins (1992), Kokott (1998). It also manifests itself in accounting through the Safe Harbor Liability Rule. The principle can be stated as:

"There is no liability as long as there is a reasonable basis for belief, effectively placing the burden of proof (Onus Probandi) on the plaintiff, who, in a lawsuit, must prove false a defendant's misstatement, without making any assumption not explicitly stated by the defendant, or tacitly implied by an existing law or regulatory requirement."

The Most Favorable Interpretation principle, which, depending on the context, is also known as Benefit of the Doubt, In Dubito Pro Reo, or Presumption of Innocence, is a consequence of the Onus Probandi principle, and requires the court to consider the evidence in the light of what is most favorable to the defendant, see *Ruta v. Breckenridge-Remy Co.* (1982).

"Moreover, the party against whom the motion is directed is entitled to have the trial court construe the evidence in support of its claim as truthful, giving it its most favorable interpretation, as well as having the benefit of all reasonable inferences drawn from that evidence."

To be consistent with our interpretation of the Onus Probandi principle in the Full Bayesian Significance Testing context, requirements R3, R4, and R5 must be imposed to the value of evidence. A concrete example and its discussion is presented in section 6.

R6 requires  $(\overline{Ev}(H), \max)$  to be a partial support structure. For that,  $\overline{Ev}$  must satisfy Axioms A1 to A5, which can be easily checked:

To see that A2 holds, note that if  $H_1 : \theta \in \Theta_1 \subseteq \Theta$  and  $H_2 : \theta \in \Theta_2 \subseteq \Theta$  then

$$\sup_{\Theta_1 \cup \Theta_2} s(\theta) = \max \left( \sup_{\Theta_1} s(\theta), \sup_{\Theta_2} s(\theta) \right)$$

and therefore

$$\overline{Ev}(H_1 \vee H_2) = \max(\overline{Ev}(H_1), \overline{Ev}(H_2))$$

Notice that the assumption  $\neg(H_1 \wedge H_2)$ , i.e.,  $\Theta_1 \cap \Theta_2 = \emptyset$ , was not necessary to prove the validity of A2.

To see that A3 holds, note that if we have  $\overline{Ev}(H_1) = \overline{Ev}(H_3)$  and

$$H_1 : \theta \in \Theta_1 \subseteq \Theta, H_2 : \theta \in \Theta_2 \subseteq \Theta, H_3 : \theta \in \Theta_3 \subseteq \Theta,$$

then  $\Theta_1 \subseteq \Theta_2 \subseteq \Theta_3$  implies  $s_{H1} \leq s_{H2} \leq s_{H3}$ , which, in turn implies

$$T_3 \subseteq T_2 \subseteq T_1$$

Consequently  $Ev(H_3) \leq Ev(H_2) \leq Ev(H_1)$ . The result follows.

To see that A4 holds, note that

$$\overline{Ev}(\emptyset) = 1 - Ev(\emptyset) = 1 - \int_{\Theta} p_n(\theta) d\theta = 1 - 1 = 0$$



To see that A5 holds, note that

$$\overline{\text{Ev}}(\Theta) = 1 - \text{Ev}(\Theta) = 1 - \int_0 p_n(\theta) d\theta = 1 - 0 = 1$$

The proof of R8 is given in appendix A. Invariance means that two observers, each one using a different measuring system (rules, clocks, etc.), will have different measurement figures, but they must agree on the support value for an hypothesis correctly translated into each system.

Consider for example Einstein's (sharp) hypothesis, H:

"The speed of any light wave, measured in any inertial frame, is a constant."

The support value of H,  $\overline{\text{Ev}}(H)$ , given by Michelson and Morley experimental data,  $\text{Ev}(H)$ , must be the same, whether they used the English or the International Metric System (186K miles per second or 300K kilometers per second).

Consistency requirement R9 is a corollary of the convergence theory of posterior distributions, see DeGroot (1970, chapter 10). Consistency means that, as we keep making more and more observations,  $\overline{\text{Ev}}(H)$  must converge to 0 or 1, according to whether the hypothesis is false or truth.

Other interpretations of the partial support structure  $(\overline{\text{Ev}}, \max)$  are given in sections 6 and 7. Before that, however, we shall introduce additional facts on the ABC formalism.

## 5 Conditionalization

ABC's formalism also establishes a set of axioms for Conditionalization, i.e. on how to update a support function  $\Phi$  to a "posterior" support function  $\Phi_A$ , after accepting a non-rejected statement A. Derwiche and Ginsberg (1992) and Derwiche (1993) define as Plausible Conditionalizations those given by a (partial) function,

$$\odot : \Phi(\mathcal{U}) \times \Phi(\mathcal{U}) \mapsto \Phi(\mathcal{U})$$

attending Axioms A6 to A11 bellow. For ease of writing we will refer to  $\Phi(B)$  and  $\Phi_A(B)$ , respectively, as the unconditional support value of B and the conditional support value of B given (the acceptance of) A. The function  $\odot$  is called Support Scaling.

A6: The conditional support value of B given  $A \vee B$  is a function of the unconditional support values of B and  $A \vee B$ , i.e.,

$$\Phi_{A \vee B}(B) = \Phi(B) \odot \Phi(A \vee B)$$

It can be seen that axiom 6 is equivalent to

$$\Phi_A(B) = \Phi(A \wedge B) \odot \Phi(A)$$

A7: Accepting a non-rejected statement retains all accepted statements, i.e.,

$$(\Phi(A) \neq 0 \wedge \Phi(B) = 0) \Rightarrow \Phi_A(B) = 0$$

A8: Accepting an accepted statement leads to no change in the conditional support function, i.e.,

$$\Phi(A) = 1 \Rightarrow \Phi_A = \Phi$$

A9: When  $A \vee B$  is equally supported by two support functions, conditioning on  $A \vee B$  in either case does not introduce equality or order between the unconditional supports of  $A$ , i.e., if  $\Phi$  and  $\Psi$  are support functions and  $\Phi(A \vee B) = \Psi(A \vee B)$ , then

$$\Phi_{A \vee B}(A) \preceq (=) \Psi_{A \vee B}(A) \Rightarrow \Phi(A) \preceq (=) \Psi(A)$$

A10: After accepting the logical consequences of a statement, its conditional support either increases or does not change, i.e.,

$$\Phi(A \vee B) \neq 0 \Rightarrow \Phi(A) \preceq \Phi_{A \vee B}(A)$$

A11: If the conditional support of  $A$  given  $C$  equals its conditional support given  $B \wedge C$ , then the conditional support of  $B$  given  $C$  equals its conditional support given  $A \wedge C$ , i.e.,

$$(\Phi(A \wedge B \wedge C) \neq 0 \wedge \Phi_C(A) = \Phi_{B \wedge C}(A)) \Rightarrow \Phi_C(B) = \Phi_{A \wedge C}(B)$$

$\langle \Phi(\mathcal{U}), \oplus, \odot \rangle$  is called a Support Structure. For the examples in table 1, the scaling functions are:

$$\Phi_A(B) = \min(\Phi(A \wedge B), \Phi(A))$$

for classical logic;

$$\Phi_A(B) = \frac{\Phi(A \wedge B)}{\Phi(A)}$$

for probability and possibility calculus; and

$$\Phi_A(B) = \Phi(A \wedge B) - \Phi(A)$$

for disbelief calculus.

## 6 Coexistent Belief Calculi

A critical interpretation of FBST's value of evidence, in the context set by the previous sections, can help us elucidate the benefits and some apparent paradoxes of using the FBST in statistical testing.

In the FBST, the support values,  $\text{Ev}(H)$ , are computed using standard probability calculus on  $\Theta$  which has an intrinsic conditionalization operator. The computed evidences, on the other hand, form a possibilistic partial support structure, the evidence calculus. It is impossible however to define a scaling function for the evidence calculus that is compatible with the FBST's support,

$\overline{Ev}$ , as it is defined. Therefore, two belief calculi are in simultaneous use in the Full Bayesian Testing setup: probability and evidence calculus.

Most standard (frequentist or Bayesian) theories of statistical testing try to use a single belief calculus. Namely: probability calculus. In order to do so they try to use the probability of the null set as a support value for the hypothesis. This can also take an indirect form, such as integrating a utility or loss function. This approach is the source of endless problems, specially in the case of sharp hypothesis.

In many legal applications with a composite hypothesis,  $H = A \vee B$ , neither a probability distribution giving the probabilities  $Pr(A)$  and  $Pr(B)$ , nor the relative probability  $Pr(A)/Pr(B)$ , is explicitly stated by the defendant, or tacitly implied by an existing law or regulatory requirement. According to requirement R2, if no such probabilities are given, then no such probabilities can be used. This statement contradicts many practices of standard Bayesian approaches to hypothesis testing, including all tests based on Bayes factors, see DeGroot (1970), Zelter (1971).

As a subterfuge to obtain artificial probabilities when they are not explicitly stated, many standard Bayesian statistical tests use a particular parameterization of the hypothesis, and probably measures on the (sub-manifold representing the) sharp hypothesis derived from this parameterization, in conjunction with measures defined on  $\Theta$  and/or a prior mass for sharp hypothesis, see DeGroot (1970). Even the improvement suggested in Irony and Pereira (1995) uses such a strategy. Requirement R2 admits no such adhoceries.

Let us examine some aspects of the partial support structure of FBST's evidence. The most favorable interpretation requirement implies that the evidence calculus should have a possibilistic rather than a probabilistic partial support structure. Once gain, this requirement contradicts many approaches in frequentist and standard Bayesian theories of hypothesis testing, that directly use a probabilistic support structure.

Derwiche (1993) makes some interesting remarks concerning support and belief orders. Namely:

1. If two statements are equally believed, then they are equally supported; but not the converse.
2. Rejected statements are always minimally supported, and accepted statements are always maximally supported. But although minimally supported statements are rejected, maximally supported sentences are not necessarily accepted.
3. A statement and its negation may be maximally supported at the same time, while neither of them may be accepted.

Consider, as an illustrative example, the hypotheses

$$A : \theta \in \Theta \text{ and } B : \theta \in \{\hat{\theta}\}$$

where  $\hat{\theta}$  is the unique point at which a smooth proper posterior density attains its maximum in the parameter space  $\Theta = \mathcal{R}^p$ ,  $\{\hat{\theta}\} = \arg \max_{\theta \in \Theta} p_n(\theta)$ . Assume

the uniform reference,  $r(\theta) \propto 1$ . We have,  $\overline{\text{Ev}}(A) = \overline{\text{Ev}}(B) = \overline{\text{Ev}}(\neg B) = 1$  and  $\overline{\text{Ev}}(\neg A) = 0$ . So both A and B have full support, but A is accepted, while B is not.

This example, or variations of it, were given to the author as either an example of how a support function should work in the juridical context, or as an FBST paradox, in the context of traditional statistical tests of significance.

In the juridical context, the interpretation is as follows: A defendant describes a system (machine, software, genetic code etc.) by a parameter  $\theta$ , and claims that  $\theta$  has been set to a value in a legal or valid null set,  $\Theta_H$ . The parameter can not be observed directly, but we can observe a random variable whose distribution is a function  $f(x; \theta)$ . The parameter  $\theta$  has been set to one, and only one value. Claiming that  $\theta$  has been set at the most likely value,  $\theta = \hat{\theta}$ , (given  $n$  observed outcomes) must give the defendant's claim full support, for being absolutely vague, i.e., claiming only that  $\theta \in \Theta$ , cannot put him in a better position.

In most traditional statistical tests of significance,  $\Phi(\Theta_H)$  is a probability measure of the null set,  $\mu(\Theta_H)$ . If  $\Theta_H$  is a singleton in  $\mathcal{R}^p$ , with a smooth posterior, then it should have null support. Indeed, the refutation of any sharp hypothesis is a price many philosophers, see Popper (1989), and most statisticians are ready to pay, as explicitly stated by I.J. Good:

"If by the truth of Newtonian mechanics we mean that it is approximately true in some appropriate well defined sense we could obtain strong evidence that it is true; but if we mean by its truth that it is exactly true then it has already been refuted. ... Very often the statistician doesn't bother to make it quite clear whether his null hypothesis is intended to be sharp or only approximately sharp. ... It is hardly surprising then that many Fisherians (and Popperians) say that - you can't get (much) evidence in favor of the null hypothesis but can only refute it."

Epistemological consequences of the FBST, as it departs from this tradition, and its capability of consistently (R9) handling really sharp hypothesis, are examined in forthcoming papers.

## 7 Final Remarks

In order to discuss concepts such as: testing a hypothesis (acceptance/rejection) at a certain level; test power; and optimal levels, the FBST theory must be further developed. This is done in Laureto et al. (2002), Stern and Zacks (2002) and other forthcoming papers. For an alternative view of the FBST, in the context of decision theory, see Madruga et al. (2001) and Rubin (1987).

Derwiche and Ginsberg (1992) remark that in several other uncertainty calculi, in particular multivalued logic calculi and generalizations of probability calculus, at least one of the following axioms hold:

A13: The support value of a conjunction is a function of the support values of its factors, i.e.,

$$\Phi(A \wedge B) = f_{\wedge}(\Phi(A), \Phi(B))$$

A14: The support value of a statement's negation is a function of the statement's support value, i.e.,

$$\Phi(\neg A) = f_{-}(\Phi(A))$$

Neither axiom A13 nor axiom A14 can be imposed to the FBST's evidence partial support structure. Once again, trying to impose one or both of these axioms can be viewed as the source of many problems in traditional theories of hypothesis testing, both in the frequentist and standard Bayesian approaches, see Good(1983), Hacking (1965), Koopman (1940a,b).

The literal interpretation of the Onus Probandi principle suggests taking the, possibly improper, uniform density as the reference density, in the "natural" parameter space. In the Bayesian context, this is usually the parameter space where the scientist accesses his/her prior. We can generalize the procedure using other reference densities. For example, we may use as reference density the uninformative prior (also known as neutral or reference prior), if one is available. This possibility is suggested by the paper of Evans (1997), in conjunction with Jeffreys' rules to obtain uninformative priors, Zellner (1971, appendix to chapter 2).

One of Jeffreys' rules to obtain an uninformative prior is to define a transformation  $\omega = \phi(\theta)$  of the parameter space so that, in the new coordinate system, the uniform uninformative prior in  $\mathcal{R}^p$  is "natural". According to this perspective, using the uninformative prior as reference density is equivalent to specify a transformation  $\phi$  of the parameter space, so that, in the transformed parameter space, the uninformative prior is uniform. We also observe that, in  $\mathcal{R}^p$ , the uniform measure and the evidence computed fixing the uniform reference are both invariant under proper linear transformations, see Klein (1997) and Santalo (1976).

In order to be consistent with the Onus Probandi principle, applications of the FBST generally use as reference density on  $\Theta$ , the uniform density or an uninformative prior that yields a proper posterior density  $p_n(\theta)$ . It is possible to use other reference densities, although doing so may impair the adherence to the Onus Probandi principle, or change its interpretation.

Finally, we must stress that the FBST departs from another major statistical paradigm. Namely: Nuisance Parameter Elimination.

Consider the situation where the hypothesis constraint,  $H : h(\theta) = h(\delta) = 0$ ,  $\theta = [\delta, \lambda]$  is not a function of some of the parameters,  $\lambda$ . This situation is described in Basu (1988):

"If the inference problem at hand relates only to  $\delta$ , and if information gained on  $\lambda$  is of no direct relevance to the problem, then we classify  $\lambda$  as the Nuisance Parameter. The big question in statistics is: How can we eliminate the nuisance parameter from the argument?"

Basu (1988) lists at least 10 categories of procedures to achieve this goal, including the ones using the  $\max_{\lambda}$  or  $\int d\lambda$  operators in order to obtain a projected profile or marginal posterior density,  $f(\delta|x)$ . The very word Nuisance, used by Fisher, reveals how strongly this paradigm is encrusted in statistical

thinking. Compare the etymological derivation of the radical Nuisance, related to (An)noyance, in contrast to Nonsense, which is far more natural in the context:

Nuisance < O.F. Nuire < V.L. Nocere (harm), Necare (kill).

Noyance < O.F. Enoyer, Sp. Enojar < V.L. in-Odiare (hate).

Nonsense < F. non Sens (meaningless) < L. Sentire (feel), Ger. Sinn (meaning).

The FBST does not follow the nuisance parameters elimination paradigm. In fact, staying in the original parameter space, in its full dimension, explains the “Intrinsic Regularization” property of the FBST, when it is used for model selection, see Pereira and Stern (2001).

## Acknowledgements

The author has benefited greatly from comments from the Bayesian group at University of Sao Paulo, including Luis Esteves, Marcelo Lauretto, Fabio Nakano, Sergio Wechsler and, most specially, Wagner Borges and Carlos Alberto de Bragança Pereira.

## Appendix A: Proof of Invariance

Consider a proper (bijective, integrable, and almost surely continuously differentiable) reparameterization  $\omega = \phi(\theta)$ . We denote the Jacobian of the reparameterization by

$$J(\omega) = \left[ \frac{\partial}{\partial \omega} \phi^{-1}(\omega) \right] = \left[ \frac{\partial \theta}{\partial \omega} \right] = \begin{bmatrix} \frac{\partial \theta_1}{\partial \omega_1} & \cdots & \frac{\partial \theta_1}{\partial \omega_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \theta_n}{\partial \omega_1} & \cdots & \frac{\partial \theta_n}{\partial \omega_n} \end{bmatrix}$$

Under the reparameterization, the posterior, reference and surprise functions are

$$\begin{aligned} \tilde{p}_n(\omega) &= p_n(\phi^{-1}(\omega)) |J(\omega)| \\ \tilde{r}(\omega) &= r(\phi^{-1}(\omega)) |J(\omega)| \\ \tilde{s}(\omega) &= \tilde{p}_n(\omega) / \tilde{r}(\omega) = p_n(\phi^{-1}(\omega)) / r(\phi^{-1}(\omega)) \end{aligned}$$

Let  $\Omega_H = \phi(\Theta_H)$ . It follows that

$$\tilde{s}_H = \sup_{\omega \in \Omega_H} \tilde{s}(\omega) = \sup_{\theta \in \Theta_H} s(\theta) = s_H$$

hence,  $T_H \mapsto \phi(T_H) = \tilde{T}_H$ , and

$$\widetilde{\text{Ev}}(H) = \int_{\tilde{T}_H} \tilde{p}_n(\omega) d\omega = \int_{T_H} p_n(\theta) d\theta = \text{Ev}(H) \quad , \quad \text{Q.E.D.}$$

## References

- Basu, D. (1988). *Statistical Information and Likelihood*. Edited by J.K. Ghosh. *Lect. Notes in Statistics*, 45.
- Darwiche, A.Y., Ginsberg, M.L. (1992). A symbolic Generalization of Probability Theory. *AAAI-92, Proceedings of the Tenth National Conference on Artificial Intelligence*.
- Darwiche, A.Y. (1993). A symbolic Generalization of Probability Theory. Ph.D. Thesis, Stanford Univ.
- Evans, M. (1997). Bayesian Inference Procedures Derived via the Concept of Relative Surprise. *Communications in Statistics*, 26, 1125-1143.
- Gaskins, R.H. (1992). *Burdens of Proof in Modern Discourse*. New Haven: Yale Univ. Press.
- DeGroot, M.H. (1970). *Optimal Statistical Decisions*. NY: McGraw-Hill.
- Good, I.J. (1983). *Good Thinking*. Univ. of Minnesota.
- Hacking, I. (1965). *Logic of Statistical Inference*. Cambridge Univ. Press.
- Irony, T.Z., Lauretto, M., Pereira, C.A.B. and Stern, J.M. (2002). A Weibull Wearout Test: Full Bayesian Approach. In: Hayakawa, Y., Irony, T., Xie, M. (Eds). *Systems and Bayesian Reliability*. pp 287-300. Singapore: World Scientific.
- Irony, T.Z., Pereira, C.A.B. (1995). Bayesian Hypothesis test: Using surface integrals to distribute prior information among hypotheses. *Resenhas*, 2, 27-46.
- Kempthorne, O. (1980). Foundations of Statistical Thinking and Reasoning. *Australian CSIRO-DMS Newsletter*, 68, 1-5; 69, 3-7.
- Klein, D.A. and Rota, G.C. (1997). *Introduction to Geometric Probability*. Cambridge Univ. Press.
- Kokott, J. (1998). *The Burden of Proof in Comparative and International Human Rights Law*. The Hague: Kluwer.
- Koopman, B.O. (1940a). The Axioms and Algebra of Intuitive Probability. *Annals of Mathematics*, 41, 269-292.
- Koopman, B.O. (1940b). The Bases of Probability. *Bulletin of the American mathematical Society*, 46, 763-774.
- Lauretto, M., Pereira, C.A.B. and Stern, J.M. (2002). Comparing Parameters of Two Bivariate Normal Distributions Using the Full Bayesian Significance Test. *Seventh Valencia International Meeting on Bayesian Statistics*.
- Madruza, M.R., Esteves, L.G. and Wechsler, S. (2001). On the Bayesianity of Pereira-Stern Tests. *Test*, 10, 291-299.
- Madruza, M.R., Pereira, C.A.B., Stern, J.M., (2002). Bayesian Evidence Test for Precise Hypotheses. *Journal of Statistical Planning and Inference*. In press.
- Pereira, C.A.B. and Stern, J.M. (1999). Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. *Entropy Journal*, 1, 69-80.
- Pereira, C.A.B. and Stern, J.M. (2001). Model Selection: Full Bayesian Approach. *Environmetrics*, 12, (6), 559-568.

- Pereira, C.A.B. and Wechsler, S. (1993). On the Concept of  $p$ -value. *Braz J Prob Statist*, 7, 159–177.
- Popper, K.R. (1989). *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Routledge.
- Royall, R. (1997). *Statistical Evidence: A Likelihood Paradigm*. London: Chapman & Hall.
- Rubin, H. (1987). A Weak System of Axioms for “Rational” Behaviour and the Non-Separability of Utility from Prior. *Statistics and Decisions*, 5, 47–58.
- Ruta v. Breckenridge-Remy Co., USA, 1982.
- Santalo, L.A. (1976). *Integral Geometry and Geometric Probability*. Encyclopedia of Mathematics and its Applications, 1. London: Addison-Wesley.
- Stern, J.S. and Zacks, S. (2002). Testing the Independence of Poisson Variates under the Holgate Bivariate Distribution The Power of a new Evidence Test. To appear, *Statistical and Probability Letters*.
- Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. NY: Wiley.



## RELATÓRIOS TÉCNICOS

### DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Instituto de Matemática e Estatística da USP

A listagem contendo os relatórios técnicos anteriores a 1998 poderá ser consultada ou solicitada à Secretaria do Departamento, pessoalmente, por carta ou e-mail (mac@ime.usp.br).

Carlos Alberto de Bragança Pereira, Fabio Nakano e Julio Michael Stern

*A DYNAMIC SOFTWARE CERTIFICATION AND VERIFICATION PROCEDURE*

RT-MAC-9901, março 1999, 21pp.

Carlos E. Ferreira e Dilma M. Silva

*BCC DA USP: UM NOVO CURSO PARA OS DESAFIOS DO NOVO MILÊNIO*

RT-MAC-9902, abril 1999, 12pp.

Ronaldo Fumio Hashimoto and Junior Barrera

*A SIMPLE ALGORITHM FOR DECOMPOSING CONVEX STRUCTURING ELEMENTS*

RT-MAC-9903, abril 1999, 24 pp.

Jorge Euler, Maria do Carmo Noronha e Dilma Menezes da Silva

*ESTUDO DE CASO: DESEMPENHO DEFICIENTE DO SISTEMA OPERACIONAL LINUX PARA CARGA MISTA DE APLICAÇÕES.*

RT-MAC-9904, maio 1999, 27 pp.

Carlos Humes Junior e Paulo José da Silva e Silva

*AN INEXACT CLASSICAL PROXIMAL POINT ALGORITHM VIEWED AS DESCENT METHOD IN THE OPTIMIZATION CASE*

RT-MAC-9905, maio 1999, pp.

Carlos Humes Junior and Paulo José da Silva e Silva

*STRICT CONVEX REGULARIZATIONS, PROXIMAL POINTS AND AUGMENTED LAGRANGIANS*

RT-MAC-9906, maio 1999, 21 pp.

Ronaldo Fumio Hashimoto, Junior Barrera, Carlos Eduardo Ferreira

*A COMBINATORIAL OPTIMIZATION TECHNIQUE FOR THE SEQUENTIAL DECOMPOSITION OF EROSIONS AND DILATIONS*

RT-MAC-9907, maio 1999, 30 pp.

Carlos Humes Junior and Marcelo Queiroz

*ON THE PROJECTED PAIRWISE MULTICOMMODITY FLOW POLYHEDRON*

RT-MAC-9908, maio 1999, 18 pp.

Carlos Humes Junior and Marcelo Queiroz

*TWO HEURISTICS FOR THE CONTINUOUS CAPACITY AND FLOW ASSIGNMENT  
GLOBAL OPTIMIZATION*

RT-MAC-9909, maio 1999, 32 pp.

Carlos Humes Junior and Paulo José da Silva e Silva

*AN INEXACT CLASSICAL PROXIMAL POINT ALGORITHM VIEWED AS A  
DESCENT METHOD IN THE OPTIMIZATION CASE*

RT-MAC-9910, julho 1999, 13 pp.

Markus Endler and Dilma M. Silva and Kunio Okuda

*A RELIABLE CONNECTIONLESS PROTOCOL FOR MOBILE CLIENTS*

RT-MAC-9911, setembro 1999, 17 pp.

David Robertson, Fábio S. Corrêa da Silva, Jaume Agustí and Wamberto W. Vasconcelos

*A LIGHTWEIGHT CAPABILITY COMMUNICATION MECHANISM*

RT-MAC-9912, novembro 1999, 14 pp.

Flávio S. Corrêa da Silva, Jaume Agustí, Roberto Cássio de Araújo and Ana Cristina V.  
de Melo

*KNOWLEDGE SHARING BETWEEN A PROBABILISTIC LOGIC AND BAYESIAN  
BELIEF NETWORKS*

RT-MAC-9913, novembro 1999, 13 pp.

Ronaldo F. Hashimoto, Junior Barrera and Edward R. Dougherty

*FINDING SOLUTIONS FOR THE DILATION FACTORIZATION EQUATION*

RT-MAC-9914, novembro 1999, 20 pp.

Marcelo Finger and Wamberto Vasconcelos

*SHARING RESOURCE-SENSITIVE KNOWLEDGE USING COMBINATOR LOGICS*

RT- MAC-2000-01, março 2000, 13pp.

Marcos Alves e Markus Endler

*PARTICIONAMENTO TRANSPARENTE DE AMBIENTES VIRTUAIS DISTRIBUÍDOS*

RT- MAC-2000-02, abril 2000, 21pp.

Paulo Silva, Marcelo Queiroz and Carlos Humes Junior

*A NOTE ON "STABILITY OF CLEARING OPEN LOOP POLICIES IN  
MANUFACTURING SYSTEMS"*

RT- MAC-2000-03, abril 2000, 12 pp.

Carlos Alberto de Bragança Pereira and Julio Michael Stern  
*FULL BAYESIAN SIGNIFICANCE TEST: THE BEHRENS-FISHER AND  
COEFFICIENTS OF VARIATION PROBLEMS*

RT-MAC-2000-04, agosto 2000, 20 pp.

Telba Zalkind Irony, Marcelo Lauretto, Carlos Alberto de Bragança Pereira and Julio Michael Stern

*A WEIBULL WEAROUT TEST: FULL BAYESIAN APPROACH*

RT-MAC-2000-05, agosto 2000, 18 pp.

Carlos Alberto de Bragança Pereira and Julio Michael Stern  
*INTRINSIC REGULARIZATION IN MODEL SELECTION USING THE FULL  
BAYESIAN SIGNIFICANCE TEST*

RT-MAC-2000-06, outubro 2000, 18 pp.

Douglas Moreto and Markus Endler  
*EVALUATING COMPOSITE EVENTS USING SHARED TREES*

RT-MAC-2001-01, janeiro 2001, 26 pp.

Vera Nagamura and Markus Endler  
*COORDINATING MOBILE AGENTS THROUGH THE BROADCAST CHANNEL*

RT-MAC-2001-02, janeiro 2001, 21 pp.

Júlio Michael Stern  
*THE FULLY BAYESIAN SIGNIFICANCE TEST FOR THE COVARIANCE PROBLEM*

RT-MAC-2001-03, fevereiro 2001, 15 pp.

Marcelo Finger and Renata Wassermann  
*TABLEAUX FOR APPROXIMATE REASONING*

RT-MAC-2001-04, março 2001, 22 pp.

Julio Michael Stern  
*FULL BAYESIAN SIGNIFICANCE TESTS FOR MULTIVARIATE NORMAL  
STRUCTURE MODELS*

RT-MAC-2001-05, junho 2001, 20 pp.

Paulo Sérgio Naddeo Dias Lopes and Hernán Astudillo  
*VIEWPOINTS IN REQUIREMENTS ENGINEERING*

RT-MAC-2001-06, julho 2001, 19 pp.

Fabio Kon  
*O SOFTWARE ABERTO E A QUESTÃO SOCIAL*

RT-MAC-2001-07, setembro 2001, 15 pp.

Isabel Cristina Italiano, João Eduardo Ferreira and Osvaldo Kotaro Takai  
*ASPECTOS CONCEITUAIS EM DATA WAREHOUSE*  
RT – MAC-2001-08, setembro 2001, 65 pp.

Marcelo Queiroz , Carlos Humes Junior and Joaquim Júdice  
*ON FINDING GLOBAL OPTIMA FOR THE HINGE FITTING PROBLEM*  
RT- MAC –2001-09, novembro 2001, 39 pp.

Marcelo Queiroz , Joaquim Júdice and Carlos Humes Junior  
*THE SYMMETRIC EIGENVALUE COMPLEMENTARITY PROBLEM*  
RT- MAC-2001-10, novembro 2001, 33 pp.

Marcelo Finger, and Fernando Antonio Mac Cracken Cezar  
*BANCO DE DADOS OBSOLEScentes E UMA PROPOSTA DE IMPLEMENTAÇÃO.*  
RT- MAC - 2001-11- novembro 2001, 90 pp.

Flávio Soares Correa da Silva  
*TOWARDS A LOGIC OF PERISHABLE PROPOSITIONS*  
RT- MAC- 2001-12 - novembro 2001, 15 pp.

Alan M. Durham  
*O DESENVOLVIMENTO DE UM INTERPRETADOR ORIENTADO A OBJETOS PARA ENSINO DE LINGUAGENS*  
RT-MAC-2001-13 – dezembro 2001, 21 pp.

Alan M. Durham  
*A CONNECTIONLESS PROTOCOL FOR MOBILE AGENTS*  
RT-MAC-2001-14 – dezembro 2001, 12 pp.

Eugênio Akihiro Nassu e Marcelo Finger  
*O SIGNIFICADO DE “AQUI” EM SISTEMAS TRANSACIONAIS MÓVEIS*  
RT-MAC-2001-15 – dezembro 2001, 22 pp.

Carlos Humes Junior, Paulo J. S. Silva e Benar F. Svaiter  
*SOME INEXACT HYBRID PROXIMAL AUGMENTED LAGRANGIAN ALGORITHMS*  
RT-MAC-2002-01 – Janeiro 2002, 17 pp.

Roberto Speicys Cardoso e Fabio Kon  
*APLICAÇÃO DE AGENTES MÓVEIS EM AMBIENTES DE COMPUTAÇÃO UBÍQUA.*  
RT-MAC-2002-02 – Fevereiro 2002, 26 pp.

Julio Stern and Zacks

*TESTING THE INDEPENDENCE OF POISSON VARIATES UNDER THE HOLGATE  
BIVARIATE DISTRIBUTION: THE POWER OF A NEW EVIDENCE TEST.*

RT- MAC – 2002-03 – Abril 2002, 18 pp.

E. N. Cáceres, S. W. Song and J. L. Szwarcfiter

*A PARALLEL ALGORITHM FOR TRANSITIVE CLOSURE*

RT-MAC – 2002-04 – Abril 2002, 11 pp.

Regina S. Burachik, Suzana Scheimberg, and Paulo J. S. Silva

*A NOTE ON THE EXISTENCE OF ZEROES OF CONVEXLY REGULARIZED SUMS  
OF MAXIMAL MONOTONE OPERATORS*

RT- MAC 2002-05 – Maio 2002, 14 pp.

C.E.R. Alves, E.N. Cáceres, F. Dehne and S. W. Song

*A Parameterized Parallel Algorithm for Efficient Biological Sequence Comparison*

RT-MAC-2002-06 – Agosto 2002, 11pp.

Julio Michael Stern

*Significance Tests, Belief Calculi, and Burden of Proof in Legal and Scientific Discourse*

RT- MAC – 2002-07 – Setembro 2002, 20pp.