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

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

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 Darwiche and Ginsberg, and use it to analyze the FBST's value of evidence. This analysis helps us understand the FBST properties and interpretation. The definition of value of evidence against a sharp hypothesis, in the FBST setup, was motivated by applications of Bayesian statistical reasoning to legal matters where the sharp hypotheses were defendants statements, to be judged according to the Onus Probandi juridical principle.

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 papers, 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 (Darwiche and Ginsberg 1992) and (Darwiche 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 sections 4, 6 and 7; and, the ABC formalism is presented in sections 3 and 5.

2   The FBST Value of Evidence

Let $\theta \in \Theta \subseteq \mathcal{R}^p$ be a vector parameter of interest, and $L(\theta \,|\, x)$ be the likelihood associated to the observed data x, a standard statistical model. Under the Bayesian paradigm the posterior density, $p_x(\theta)$, is proportional to the product of the likelihood and a prior density,

$$p_x(\theta) \propto L(\theta \,|\, x)\, p(\theta).$$

The (null) hypothesis H states that the parameter lies in the null set, defined by inequality and equality constraints given by vector functions g and h in the parameter space,

$$\Theta_H = \{\theta \in \Theta \,|\, g(\theta) \leq \mathbf{0} \wedge h(\theta) = \mathbf{0}\}$$

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

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

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

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

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 Θ_H. When $r(\theta) \propto 1$, T_H is the Posterior's Highest Density Probability Set (HDPS) tangential to the null set Θ_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 $\text{Ev}(H)$, against the hypothesis.

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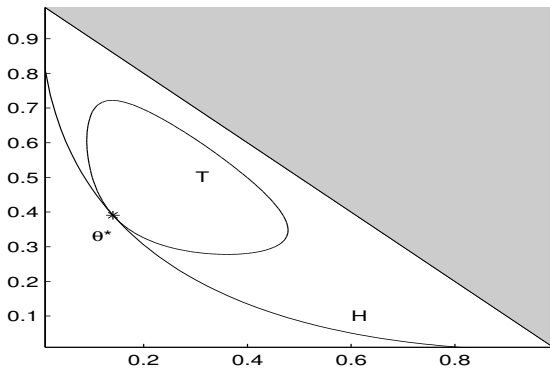

Figure 1: H-W: Hypothesis and Tangential Set

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:

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

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 the author's previous papers, including several examples of software certification and verification based on black-box simulation.

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 (Darwiche and Ginsberg 1992) and (Darwiche, 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 FBST setup.

The first concept in ABC is that of an abstract Support Function, Φ, which attributes abstract support values to statements in a universe $\mathcal{U}$, closed under disjunction, negation and conjunction. 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,

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

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 (Darwiche 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 $\neg(A \wedge B)$. 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.

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	$a \leq b$	Cl. Logic
$[0,1]$	$a + b$	0	1	$a \leq b$	Probablty
$[0,1]$	$\max(a,b)$	0	1	$a \leq b$	Possiblty
$\{0..\infty\}$	$\min(a,b)$	∞	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 Φ, however ABC defines the belief function

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

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 $\ddot{\Phi}(\mathcal{U})$. 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 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 sharp hypotheses were 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, $\Phi(H) = \overline{\text{Ev}}(H) = 1 - \text{Ev}(H)$, 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 a (measurable) subset Γ_H of the parameter space, i.e.,

$$\text{Ev}(H) = \int_{\Gamma_H} p_x(\theta) d\theta$$

If a parameter point $\theta \in \Theta$ is in the evidence set Γ_H we say that θ constitutes evidence against the hypothesis H. If θ is in the null Θ_H we say that θ is compatible with (or admissible, legal or valid by) hypothesis H.

R2, Relative Surprise: Whether a parameter point θ constitutes or not evidence against H depends only on the order in the parameter space established by the value of the posterior surprise relative to a given reference density, $s(\theta) = p_x(\theta)/r(\theta)$.

R3, No Self Incrimination: A parameter point compatible with an hypothesis can not constitute evidence against the same hypothesis, i.e.,

$$\Theta_H \cap \Gamma_H = \emptyset$$

R4, De Morgan's Law: A parameter point constitutes evidence against a composite hypothesis iff it constitutes evidence against all of its terms, i.e.,

$$\text{if } H = A \vee B \text{ then } \Gamma_H = \Gamma_A \cap \Gamma_B$$

R5, Most Favorable Interpretation: The evidence in favor of a composite hypothesis is the most favorable evidence in favor of its terms, i.e.,

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

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

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

$$\Theta_H = \{\theta \in \Theta \,|\, g(\theta) \leq \mathbf{0} \wedge h(\theta) = \mathbf{0}\},$$

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, see appendix.

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 information given by data 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 $\text{Ev}(H)$ to depend on the observed data only though the likelihood function is in the essence of the Likelihood Principle. This is enforced by R2.

Requirements R3, R4 and R5 try to capture the Onus Probandi principle, as it presented itself in the author's research and consulting practice, reported in previous and forthcomming papers. A simple example and its discussion is presented in section 6. 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."

R6 requires $\langle \overline{\text{Ev}}(H), \max \rangle$ to be a partial support structure, see appendix. (Darwiche 1993) gives a thorough analysis of why R6 establishes the minimal logical conditions for a support function. R7, R8 and R9 are standard desirable properties in statistical testing theory.

Invariance requirement R8 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, $\overline{\text{Ev}}(H)$, 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, chap. 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.

As pointed out to the author, instead of establishing some requirements on $\text{Ev}(H)$, we could consider only requirements given directly on the function $\lambda : \Gamma_H = \lambda(\Theta_H \,|\, s(\theta))$, like:

S1, Total Order: λ's range is totally ordered by $\subseteq$, that is, $\Gamma_A \subseteq \Gamma_B \ \vee \ \Gamma_B \subseteq \Gamma_A$

S2, Antimonotony: if $A \subseteq B$ then $\Gamma_B \subseteq \Gamma_A$

It is not hard to check that requirements R1 to R9 are fulfilled in the case of FBST, i.e. taking $\Gamma_H = T_H$, see appendix. Further interpretations of the FBST and its

partial support structure 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 Φ to a "posterior" support function Φ_A, after accepting a non-rejected statement A. (Darwiche and Ginsberg 1992) and (Darwiche 1993) define as Plausible Conditionalizations those given by a (partial) function,

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

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 $\oslash$ 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.,

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

It can be seen that axiom 6 is equivalent to

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

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 Φ and Ψ 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, A, the conditional support of A 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.,

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

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

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

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

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, $\overline{\mathrm{Ev}}(H)$, are computed using standard probability calculus on Θ 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, $\overline{\mathrm{Ev}}$, 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. In many legal applications with a composite sharp hypothesis, H, neither a probability distribution giving the probability measure of the null set, $\Pr(\Theta_H)$, nor the odds ratio $\Pr(\Theta_H)/\Pr(\Theta_{\overline{H}})$, 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 some tests based on Bayes factors, see (Good 1983).

As a subterfuge to obtain an artificial probability for a sharp (zero measure) Θ_H, many standard Bayesian

statistical tests use a particular parameterization of the hypothesis, and probability measures on the (sub-manifold representing the) sharp hypothesis derived from this parameterization, in conjunction with measures defined on Θ_H and/or a prior mass for sharp hypothesis, see (DeGroot 1970). Another device often used by standard statistical tests for sharp hypothesis is the sometimes cumbersome procedure of nuisance parameter elimination, see (Basu 1988), (Rubin 1984). The FBST does not need to follow the nuisance parameter 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 (Pereira and Stern 2001).

The FBST is based on the probability of the tangential set, and not directly on a probability of the null set. Therefore it can overcome several conceptual and practical difficulties of well known sharp hypothesis testing theories, both in frequentist and standard Bayesian Statistics, related to the direct or indirect use of $\Pr(\Theta_H)$, see (Basu 1988), (Good 1983), (Kempthorne 1980), and the author's previous papers.

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 again, this requirement contradicts many approaches in frequentist and standard Bayesian theories of hypothesis testing, that directly use a probabistic support structure.

(Darwiche 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

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

where $\widehat{\theta}$ is the unique maximizer of a smooth proper posterior density in the parameter space $\Theta = \mathcal{R}^p$, $\{\widehat{\theta}\} = \arg\max_{\theta \in \Theta} p_x(\theta)$. Assume a uniform reference, $r(\theta) \propto 1$. We have, $\overline{\mathrm{Ev}}(A) = \overline{\mathrm{Ev}}(B) = \overline{\mathrm{Ev}}(\neg B) = 1$ and $\overline{\mathrm{Ev}}(\neg A) = 0$. 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 a 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 θ, and claims that θ has been set to a value in a legal or valid null set, Θ_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 θ has been set to one, and only one value. Claiming that θ has been set at the most likely value, $\theta = \widehat{\theta}$, (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, $\Pr(\Theta_H)$. If Θ_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."

Further epistemological consequences of the FBST, as it departs from this tradition, and its capability of consistently (R9) handling really sharp hypothesis, are examined in the extended version of this paper.

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).

(Darwiche and Ginsberg 1992) remark that in several other uncertainty calculi, in particular multivalued logic calculi and generalizations of probability cal-

culus, 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.,

$$\Phi(\neg A) = f_\neg(\Phi(A))$$

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, chap. 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 ϕ 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 Θ, the uniform density or an uninformative prior that yields a proper posterior density $p_x(\theta)$. It is possible to use other reference densities, although doing so may impair the adherence to the Onus Probandi principle, or change its interpretation. For example, the use of a precautionary prior (or a precautionary evidence based on a convex contraction $\Gamma_H = C(T_H) \subseteq T_H$), may be justified in special circumstances, as examined in the extended version of this paper.

Appendix: Some Proofs for Section 4

We show that for the FBST, i.e. taking $\Gamma_H = T_H$, $\overline{\text{Ev}}(H)$ satisfies requirements R6 and R8.

Proof of Coherent Support (R6): For that, $\overline{\text{Ev}}$ 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

$$\sup_{\Theta_1 \cup \Theta_2} s(\theta) = \max\left(\sup_{\Theta_1} s(\theta), \sup_{\Theta_2} s(\theta)\right) \Rightarrow$$
$$\overline{\text{Ev}}(H_1 \vee H_2) = \max\left(\overline{\text{Ev}}(H_1), \overline{\text{Ev}}(H_2)\right)$$

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 $\overline{\text{Ev}}(H_1) = \overline{\text{Ev}}(H_3)$ 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 $\text{Ev}(H_3) \leq \text{Ev}(H_2) \leq \text{Ev}(H_1)$. The result follows.

To see that A4 holds, note that

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

To see that A5 holds, note that

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

Proof of invariance (R8): Consider a proper (bijective, integrable, and almost surely continuously differentiable) reparameterization $\omega = \phi(\theta)$. Under the reparameterization, the Jacobian, posterior, reference and surprise functions are:

$$\begin{aligned}
\widetilde{p}_x(\omega) &= p_x(\phi^{-1}(\omega))\,|J(\omega)| \\
\widetilde{r}(\omega) &= r(\phi^{-1}(\omega))\,|J(\omega)| \\
\widetilde{s}(\omega) &= \widetilde{p}_x(\omega)/\widetilde{r}(\omega) = p_x(\phi^{-1}(\omega))/r(\phi^{-1}(\omega))
\end{aligned}$$

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

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

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

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

$$\widetilde{\mathrm{Ev}}(H) = \int_{\widetilde{T}_H} \widetilde{p}_x(\omega) d\omega = \int_{T_H} p_x(\theta) d\theta = \mathrm{Ev}(H).$$

References

D.Basu (1988). Statistical Information and Likelihood. Edit J.K.Ghosh. *Lect. Notes in Statistics*, 45.

A.Y.Darwiche and M.L.Ginsberg (1992). A symbolic Generalization of Probability Theory. *AAAI-92, Tenth National Conference on Artificial Inteligence.*

A.Y.Darwiche (1993). A symbolic Generalization of Probability Theory. Ph.D. Thesis, Stanford Univ.

M.Evans (1997). Bayesian Inference Procedures Derived via the Concept of Relative Surprise. *Communications in Statistics*, 26, 1125–1143.

R.H.Gaskins (1992). *Burdens of Proof in Modern Discourse.* New Haven: Yale Univ. Press.

M.H.DeGroot (1970). *Optimal Statistical Decisions.* NY: McGraw-Hill.

I.J.Good (1983). *Good Thinking.* Univ. of Minesota.

I.Hacking (1965). *Logic of Statistical Inference.* Cambridge Univ. Press.

T.Z.Irony, M.Lauretto, C.A.B.Pereira, and J.M.Stern (2002). A Weibull Wearout Test: Full Bayesian Approach. In: Y.Hayakawa, T.Irony, M.Xie. *Systems and Bayesian Reliability*, 287–300. Singapore: World Scientific.

O.Kempthorne (1980). Foundations of Statistical Thinking and Reasoning. *Australian CSIRO-DMS Newsletter*, 68, 1–5; 69, 3–7.

D.A.Klein and G.C.Rota (1997). *Introduction to Geometric Probability.* Cambridge Univ. Press.

J.Kokott (1998). *The Burden of Proof in Comparative and International Human Rights Law.* The Hague: Kluwer.

B.O.Koopman (1940a). Axioms and Algebra of Intuitive Probability. *Annals of Mathematics*, 41, 269–292.

B.O.Koopman (1940b). Bases of Probability. *Bulletin of the Ammerican Mathematical Society*, 46, 763–774.

M.S.Lauretto, F. Nakano, C.O.Ribeiro, J.M.Stern (1998). REAL: Real Attribute Learning Algorithm for Strategic Market Operations. ISAS-SCI-98, 2, 315–321.

M.R.Madruga, L.G.Esteves and S.Wechsler (2001). On the Bayesianity of Pereira-Stern Tests. *Test*, 10, 291–299.

M.R.Madruga, C.A.B.Pereira and J.M.Stern (2003). Bayesian Evidence Test for Precise Hypotheses. *Journal of Statistical Planning and Inference.* In press, doi:10.1016/S0378-3758(02)00368-3.

C.A.B.Pereira and J.M.Stern (1999a). A Dynamic Software Certification and Verification Procedure. ISAS-SCI-99, 2, 426–435. Pereira and Stern (1999a and b) give the rationale for software certification and verification procedures adopted by the State of São Paulo, and later by Brazilian Federal Government, by ordinance SF-CAT-11/1996 and INDESP-104/1998.

C.A.B.Pereira and J.M.Stern (1999b). Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. *Entropy Journal*, 1, 69–80.

C.A.B.Pereira and J.M.Stern (2001). Model Selection: Full Bayesian Approach. *Environmetrics*, 12, 559–568.

C.A.B.Pereira and S.Wechsler (1993). On the Concept of *p-value*. *Brazilian Journal of Probability and Statistics*, 7, 159–177.

K.R.Popper (1989). *Conjectures and Refutations: The Growth of Scientific Knowledge.* London: Routledge.

R.Royall (1997). *Statistical Evidence: A Likelihood Paradigm.* London: Chapman & Hall.

D.B.Rubin (1984). Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician. *The Annals of Statistics*, 12, 1151–1172.

H.Rubin (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.

L.A.Santalo (1976). *Integral Geometry and Geometric Probability.* Encyclopedia of Mathematics and its Applications, 1. London: Addison-Wesley.

J.M.Stern and S.Zacks (2002). Testing the Independence of Poisson Variates under the Holgate Bivariate Distribution The Power of a new Evidence Test. *Statistical and Probability Letters*, 60, 313–320.

A.Zellner (1971). *An Introduction to Bayesian Inference in Econometrics.* NY: Wiley.