Bayesian Epistemic Values: Focus on Surprise, Measure Probability!

Julio Michael Stern, *IME, University of São Paulo, Brazil.*
E-mail: jstern@ime.usp.br

Carlos A. de Bragança Pereira, *IME, University of São Paulo, Brazil.*
E-mail: cpereira@ime.usp.br

Abstract

The e-value or epistemic value, $\mathrm{ev}(H)$, measures the statistical significance of H, a hypothesis about the parameter θ of a Bayesian model. The e-value is obtained by a probability-possibility transformation of the model's posterior measure, $p(\theta)$, and can, in turn, be used to define the FBST or Full Bayesian Significance Test. This article investigates the relation of this novel approach to more standard probability-possibility transformations. In particular, we show how and why the e-value focus on or conforms with $s(\theta) = p(\theta)/r(\theta)$, the model's surprise function relative to the reference density $r(\theta)$, while it keeps itself consistent with the model's posterior probability measure. In addition, we investigate traditional objections raised in decision theoretic Bayesian statistics against measures of significance engendered by probability-possibility transformations.

Keywords: Bayesian models; Belief calculi transformations; Complex hypotheses; Epistemic values; Possibilistic and probabilistic reasoning; Significance tests; Surprise function; Truth function.

Shackle [51] interpreted the possibility of an event as the absence of surprise felt when it occurs... The possibilistic interpretation of histograms may be carried out in several ways, at first glance, at least... It is supposed that the events [are] sufficient in order to equalize frequencies and probabilities.
Didier Dubois and Henri Prade [19, p.177].

1 Introduction

This paper compares the concept of statistical significance of sharp (precise) hypotheses in three schools of statistical thinking, namely: (1) Frequentist or classical statistics; (2) Decision theoretic Bayesian statistics; (3) Bayesian statistics based on the epistemological framework of cognitive constructivism. Closely related topics are discussed in Stern [61].

The first goal of this article is to clarify some formal aspects and emphasize the role played by possibility theory and probability-possibility transformations. The basic reference for possibility theory used in this article is Dubois and Prade [19]. This pioneering article is direct, intuitive and concise, covering however all the pertinent concepts.

The second goal of this article is to explain some traditional objections raised in decision theoretic Bayesian statistics against measures of statistical significance engendered by probability-possibility transformations. We identify and analyze four

major (perceived) obstacles, namely: (1) Technical difficulties in obtaining invariant procedures. (2) Endorsement of nuisance parameter elimination procedures. (3) Decision theoretic rejection of an optimal point (constrained maximum-a-posteriori or maximum-likelihood) as a legitimate representative of a composite hypothesis. (4) Traditional understandings of significance tests as coverage (or not) of a point hypothesis by a credibility interval of prescribed size.

In order to fulfill the aforementioned goals, this paper is organized as follows: Section 2 presents a short review and establishes the language concerning statistical models. Section 3 reviews the basic concepts of possibility theory and, strictly within this framework, defines the epistemic value of hypotheses H, $\mathrm{ev}(H)$, that, in turn, is used to define the FBST, the Full Bayesian Significance Test. The FBST is a novel theory of statistical significance developed within the epistemological framework of cognitive constructivism. For a general review of the FBST stressing the connections with logic and other aspects relevant to the scope of this article, see Borges and Stern [8]. Section 4 explains the role played in the FBST by the reference density and the surprise function, and how they are used to achieve the fundamental property of invariance in the resulting measure of significance. In contrast, this section reviews Box and Tiao [9] arguments for forfeiting general invariance in significance tests. Section 5 presents the frequentist p-value and discusses its pseudo-possibilitic characteristics. Section 5 also discusses the traditional rationale for nuisance parameter elimination procedures. Section 6 presents Bayes factors, the probability all-the-way decision theoretic solution for hypothesis test. Section 6 also reviews some decision theoretic arguments against the use of logic rules of possibilistic compositionality in statistical procedures. Section 7 discusses Lindley's method, an approach for testing hypotheses based on their coverage (or not) by credibility intervals of prescribed size. Lindley's method is a probabilistic-possibilistic compromise approach that can be regarded as a precursor for the FBST. Section 7 also discusses traditional requirements on credibility regions used by Lindley's method, like topological connectivity, or on the underlying probability distribution, like unimodality and monotonicity. Section 8 presents our final remarks and directions for further research.

2 Bayesian and Frequentist Statistical Models

A standard model of (parametric) Bayesian statistics concerns an observed (vector) random variable, x, that has a *sampling* distribution with a specified functional form, $p(x \mid \theta)$, indexed by the (vector) parameter θ. This same functional form, regarded as a function of the free variable θ with a fixed argument x, is the model's *likelihood* function.

In *frequentist* or classical statistics, θ should be taken as a 'fixed but unknown quantity'. Hence, in the frequentist framework, one is allowed to use probability calculus in the sample space, but strictly forbidden to do so in the parameter space, that is, x is to be considered as a random variable, while θ is not to be regarded as random in any way. Further consequences of the frequentist prohibition of probabilistic statements on parameters are examined in Section 6.

In the Bayesian context, the parameter θ is regarded as a latent (non-observed) random variable. Accordingly, the same formalism, namely, probability as an abstract belief calculus, is used to express credibility or (un)certainty in both the sample and

the parameter space. Consequently, the joint probability distribution, $p(x, \theta)$, should summarize all the information available in a statistical model, see Wechsler et al. [63]. Following the rules of probability calculus, the model's joint distribution of x and θ can be factorized either as the likelihood function of the parameter given the observation times the *prior* distribution on θ, or as the *posterior* density of the parameter times the observation's marginal density,

$$p(x, \theta) = p(x \,|\, \theta)p(\theta) = p(\theta \,|\, x)p(x) \ .$$

The *prior* probability distribution $p_0(\theta)$ represents the initial information available about the parameter. In this setting, a *predictive* distribution for the observed random variable, x, is represented by a mixture (or superposition) of stochastic processes, all of them with the functional form of the sampling distribution, according to the prior mixing (or weights) distribution,

$$p(x) = \int_\Theta p(x \,|\, \theta)p_0(\theta)d\theta \ .$$

If we now observe a single event, x, it follows from the factorizations of the joint distribution above that the *posterior* probability distribution of θ, representing the available information about the parameter after the observation, is given by

$$p_1(\theta) \propto p(x \,|\, \theta)p_0(\theta) \ .$$

The subscript n in $p_n(\theta)$ counts the number of observed events. In order to replace the 'proportional to' symbol, $\propto$, by an equality, $=$, it is necessary to divide the right hand site by the normalization constant, $c_1 = \int_\Theta p(x \,|\, \theta)p_0(\theta)d\theta$.

This is *Bayes rule* - the basic learning mechanism of Bayesian statistics, giving the (inverse) probability of the parameter given the data. Computing normalization constants is often difficult or cumbersome. Hence, especially in large models, it is customary to work with unnormalized densities or *potentials* as long as possible, computing normalization constants only at the very end (if ever). It is interesting to observe that the joint distribution function, taken with fixed x and free argument θ, is a potential for the posterior distribution.

Bayesian learning is a recursive process, where the posterior distribution after a learning step becomes the prior distribution for the next step. Assuming that the observations are c.i.i.d., conditionally (on the parameter) independent and identically distributed, the posterior distribution after n observations, $x^{(1)}, \ldots, x^{(n)}$, becomes,

$$p_n(\theta) \ \propto \ p(x^{(n)} \,|\, \theta)p_{n-1}(\theta) \ \propto \ \prod_{i=1}^{n} p(x^{(i)} \,|\, \theta)p_0(\theta) \ .$$

Whenever possible, it is very convenient to use a *conjugate prior*, that is, a mixing distribution whose functional form is invariant by the Bayes operation in the statistical model at hand, see Castillo et al. [1] and [12, Sec.13.8].

The 'beginnings and the endings' of the Bayesian learning process also need further discussion, that is, we should present some rationale for choosing the prior distribution used to start the learning process, and some convergence theorems for the posterior as the number observations increases. In order to do so, we must access and measure the information content of a (posterior) distribution. Stern [60] gives a short review explaining why and how the concept of entropy is the key that unlocks the mysteries related to the problems at hand.

Statistical Hypotheses

A statistical hypothesis, H, (sometimes called the null hypothesis) states that the parameter θ of a statistical model lies in the *hypothesis set*, Θ_H. For the sake of simplicity, we may, from now on, use a relaxed notation, writing H instead of Θ_H. The hypothesis set is usually defined by inequality and equality constraints given by vector functions, $g = [g_1(\theta), g_2(\theta), \ldots, g_l(\theta)]'$ and $h = [h_1(\theta), h_2(\theta), \ldots, h_k(\theta)]'$, in the parameter space,

$$H : \theta \in \Theta_H \ , \quad \Theta_H = \{\theta \in \Theta \,|\, g(\theta) \leq 0, h(\theta) = 0\} \ .$$

Throughout this paper we assume these vector constraints are regular and non-degenerate, that is, that they nowhere imply a singularity in (the algebraic submanifold) $H \subseteq \Theta$. Hence, the dimension of the hypothesis is specified by the number of equality constraints. We use the following notation: $t = \dim(\Theta)$, is the dimension of the parameter space; k, the codimension of H, counts the scalar equations constraining H; $\dim(H) = h = t - k$, the dimension of H, counts the degrees of freedom of a particle moving on H. We are particularly interested in *sharp* or *precise* hypotheses, i.e., those in which $\dim(H) < \dim(\Theta)$, that is, $k > 0$. A point-hypothesis has dimension zero, that is, the hypothesis set is a singleton, $H = \{\theta^0\}$. Throughout this article we assume, whenever necessary, appropriate topological and analytical regularity conditions, like continuity, differentiability and the existence of unconstrained or constrained maximal arguments.

3 ev(H) - A Probability-Possibility Transformation

The first goal of this section is to define $\mathrm{ev}(H)$, the e-value or *epistemic value* of an hypothesis $H \subseteq \Theta$, given a Bayesian statistical model as described in the last section, with posterior density $p_n(\theta)$ and reference density $r(\theta)$. For a few interesting applications illustrating the use of e-values and the FBST to practical problems, see Diniz et al. [16], Lauretto et al. [32], Irony et al. [25], Johnson et al. [26], Pereira and Stern [43], Pereira et al. [44], Rifo and Torres [47] and Rodrigues [48].

The *surprise function*, $s(\theta)$, indicates the change of the posterior probability density, $p_n(\theta)$, relative to a reference density, $r(\theta)$, representing an initial situation of minimum-information, see Section 4. The 'hat' and 'star' superscripts indicate unconstrained and constrained maximal arguments and supremal surprise values, as follows:

$$s(\theta) = \frac{p_n(\theta)}{r(\theta)} \ , \quad \begin{array}{ll} \widehat{s} = \sup_{\theta\in\Theta} s(\theta) \ , & \widehat{\theta} = \arg\max_{\theta\in\Theta} s(\theta) \ , \\ s^* = \sup_{\theta\in H} s(\theta) \ , & \theta^* = \arg\max_{\theta\in H} s(\theta) \ . \end{array}$$

The (closed, lower-level) *v-cut* of function $s(\theta)$, $T(v)$, its complement, the *highest surprise function set* (HSFS) above level v, $\overline{T}(v)$, and its *rim* (aka level-v set), $M(v)$, are defined as

$$T(v) = \{\theta \in \Theta \,|\, s(\theta) \leq v\} \ , \quad M(v) = \{\theta \in \Theta \,|\, s(\theta) = v\} \ , \quad \overline{T}(v) = \Theta - T(v) \ .$$

The statistical model's *truth function*, $W(v)$, is the cumulative probability function up to surprise level v. The complement and the derivative of $W(v)$ are also defined as

follows. If $W(v)$ is discontinuous, $m(v)$ must be interpreted as a generalized function in the sense of Schwartz [50].

$$W(v) = \int_{T(v)} p_n(\theta)\, d\theta \ , \quad \overline{W}(v) = 1 - W(v) \ , \quad m(v) = \frac{d}{dv} W(v) \ .$$

Several important properties of $W(v)$ follow directly from the *nesting* property exhibited by the v-cuts that, in turn, give the integration range defining the truth function, see Klir and Folger [31, Ch.4], and Zadeh [67],

$$u \leq v \Rightarrow T(u) \subseteq T(v) \Rightarrow W(u) \leq W(v) \ .$$

Finally, the e-value and its complement for an hypothesis $H \subseteq \Theta$, are defined as follows.

$$\mathrm{ev}(H) = W(v^*) \ , \quad \overline{\mathrm{ev}}(H) = 1 - \mathrm{ev}(H) \ .$$

As defined, the e-value is a *set function*. However, for the sake of simplicity, we may use a relaxed notation, writing $\mathrm{ev}(\{\theta^0\}) = \mathrm{ev}(\theta^0)$ as a *point function* for singleton arguments, that is, in the case of a *point hypothesis* $H = \{\theta^0\}$.

The e-value of an hypothesis H is based on the most favorable case, $\mathrm{ev}(H) = \mathrm{ev}(\theta^*)$, a property that characterizes $\mathrm{ev}(H)$ as a *possibilistic* abstract belief calculus, see Darwiche [14], Darwiche and Ginsberg [15], and Borges and Stern [8]. Using the nesting property of v-cuts, it is easy to establish that $\mathrm{ev}(H)$ also has the desired properties of *consistency* with its underlying probability measure and *conformity* (to be similarly shaped) with its underlying surprise function, that is,

$$\text{Consistency:} \quad \mathrm{ev}(H) \geq p_n(H) \ , \quad \forall\ H \subseteq \Theta \ ;$$

$$\text{Conformity:} \quad \mathrm{ev}(\theta) \geq \mathrm{ev}(\tau) \ \Leftrightarrow \ s(\theta) \geq s(\tau) \ , \quad \forall\ \theta, \tau \in \Theta \ .$$

A *plausibility measure*, $\mathrm{Pl}(H)$, is defined by its *basic probability assignment*, $m : 2^\Theta \mapsto [0,1]$, such that $\int_{S \subseteq \Theta} m(S) = 1$. The *focal elements* of m are the subsets of the universe with non-zero basic probability assignment, $\mathcal{F} = \{E \subseteq \Theta \,|\, m(E) > 0\}$. Finally, the plausibility of $H \subseteq \Theta$, $\mathrm{Pl}(H)$, is defined as

$$\mathrm{Pl}(H) = \int_{E \in \mathcal{F} \,|\, E \cap H \neq \emptyset} m(E) \ .$$

Hence, $\mathrm{ev}(H)$ can be characterized as a plausibility function having v-cuts of the surprise function as focal elements, $\mathcal{F} = \{T(v), 0 \leq v \leq \widehat{v}\}$, while the basic probability density assigned to $T(v)$ is obtained integrating the posterior probability density over its rim, $m(v) = \int_{M(v)} p_n(\theta) d\theta$.

A plausibility function defines its dual *belief* function by the relations

$$\mathrm{Bel}(\overline{H}) = \int_{E \in \mathcal{F} \,|\, E \subseteq \overline{H}} m(E) \ = \ 1 - \mathrm{Pl}(H) \ .$$

Epistemic Onus Probandi

This subsection makes further comments on the interpretation of plausibility and (dis)belief functions in the context of law and epistemology.

Dubois and Prade [19, p.12] (adapting the mathematical notation to the conventions used in this article) give the following insights about the meaning of the belief and plausibility functions:

> *The mass $m(E_i)$ can be interpreted as a global allocation of probability to the whole set of elementary events making up E_i, without specifying how this mass is distributed over the elementary events themselves... In this situation, the probability of an event A will be imprecise, that is, will be contained in the interval $[Bel(A), Pl(A)]$.*
>
> *The events E_i are called* focal elements*, Shafer [52], and may be used to model imprecise observations. In this situation, the probability of an event A will be mprecise, that is, will be contained in the interval $[Bel(A), Pl(A)]$.*
>
> *$Bel(A)$ is calculated by considering all focal elements which make the occurrence of A necessary (i.e. which imply A). $Pl(A)$ is obtained by considering all the focal elements which make the occurrence of A possible.*

That is, *only* the focal elements completely contained in $\overline{H}$ contribute to its credibility, while *all* the elements that intersect H contribute to its plausibility.

In the legal context, the observations in the last paragraph validate the interpretation of $\overline{\text{ev}}(H) = \text{Bel}(\overline{H})$ as the belief that H is a *misstatement*, that is, the belief that someone stating H is guilty of lying. In the legal context, valid accusations must conform to two basic juridical principles known as *onus probandi* and *in dubito pro reo*, for further details analysis see Stern [54, 55]. The natural epistemological foundation of frequentist statistics is Popperian falsificationism, where significance measures are used to falsify a theory in a 'scientific tribunal'. For a detailed analysis of the scientific tribunal metaphor and its role in statistical analysis, see Stern [61].

The Standard Probability-Possibility Transformation

Dubois and Prade [19, p.178] define (for discrete variables) the *standard* possibility measure, $\pi(H)$, that coincides (generalizing the same definition for continuous variables) with $\text{ev}(H)$ in the trivial case of a uniform (possibly improper) reference density, $r(\theta) = 1$. In this case, $s(\theta) = p_n(\theta)$ and the HSFSs are ordinary *highest probability density sets*, or HPDSs. Considering that the standard probability-possibility transformation has been extensively used for a long time in the areas of artificial intelligence and computer science, it would be natural to consider the intuitive suggestion of using $\pi(H)$ as a measure of *statistical significance* for the hypothesis H. It turns out, however, that $\pi(H)$ is not appropriate for this task.

The e-value, $\text{ev}(H)$, is more flexible than the standard possibility measure, $\pi(H)$, for it allows the use of a non-trivial reference density, $r(\theta)$, and, therefore, a surprise functions, $s(\theta)$, that has a 'different shape' than the underlying probability density, $p_n(\theta)$. Hence, $\text{ev}(H)$ can be kept consistent with $p_n(H)$, while the same $\text{ev}(\theta)$ conforms with $s(\theta)$. Better said, the possibility measure $\text{ev}(H)$ has focal elements that are defined by the level sets of the surprise function, while its basic probability assignment is obtained integrating the underlying probability density. The following section will examine why and how this added flexibility can be used to obtain a possibility measure that can be used as an invariant measure of statistical significance.

4 Invariance and Reference Geometry

Invariance is a key property of well-defined significance measures, see Stern [60]. An invariant measure is independent of the (regular) parameterizations being used to describe the statistical model. For example, an invariant significant measure cannot depend on the particular coordinate system being used as a reference frame in the parameter space, or the particular algebraic form of the equations being used to describe the hypothesis set.

In the FBST, the role of the reference density, $r(\theta)$, is to make $\mathrm{ev}(H)$ explicitly invariant under suitable transformations of the coordinate system. The natural choice of reference density is an uninformative prior, interpreted as a representation of no, minimum or low information in the parameter space, or the limit prior for no observations, or the neutral ground state for the Bayesian operation. Standard (possibly improper) uninformative priors include the uniform, Jeffreys' and maximum entropy densities, see Stern [60] for a detailed discussion.

Invariance, as used in statistics, is a metric concept. The reference density can be interpreted as induced by the information metric in the parameter space, $dl^2 = d\theta' G(\theta) d\theta$. Jeffreys' invariant prior is given by $p(\theta) = \sqrt{\det G(\theta)}$, see Amari [1], Amari et al. [2] and Stern [60] for further interpretations. For a formal proof of the e-value invariance, see Borges and Stern [8, p.405-406].

The operator used to expand a point-wise possibility measure to a set measure is maximization (or, more technically, the supremum). The $\max_H$ operator is essentially independent on the algebraic description of the hypothesis set. Hence, $\mathrm{ev}(H)$ is invariant by alternative parameterizations of the hypothesis set.

The maximization operator is used in several procedures of classical or frequentist statistics. Decision theoretic Bayesian statistical procedures, however, favor averaging or integration operations. These preferences are not fortuitous, but deeply grounded in the epistemological foundations of these well established frameworks, for an extensive investigation see Stern [61].

Resigning Invariance?

This sub-section analyses the non-invariance of HPDSs and some implications of this non-invariance in Bayesian procedures using these sets, like Lindley's method, discussed in Section 7. HPDSs are not invariant objects, as acknowledged in Box and Tiao [9, p.1469]:

> *Effect of transformations: Let $\varphi = \phi(\theta)$ be, say, a one to one transformation of parameters θ to φ. But it is clear from their definition that HPD regions in θ will not in general transform into HPD regions in φ.*

Nuisance parameter elimination procedures are an important technique used in both frequentist and decision theoretic Bayesian statistic, as discussed in Sections 5, 6 and 7. These procedures are based on reparameterization maneuvers that are tailor-made for a given statistical model and hypothesis. Hence, any non-invariance relative to choice of coordinates in the parameter space contaminates nuisance parameter elimination procedures, This state of affairs is fully acknowledged at Box and Tiao [9, p.1475-1477]:

> *Suppose in general we have k parameters $\theta = [\theta_1, \ldots, \theta_k]$. We shall define*

> *$(k-1)$ non-redundant comparisons as $(k-1)$ independent functions $\phi_i = f_i(\theta)$, $k = 1 \ldots (k-1)$, which are all equal to zero if and only if $\theta_1 = \ldots = \theta_k$. There is clearly a very wide range of choices of functions of this kind. Since HPD regions, like the confidence regions, are not invariant under non-linear transformations, some thought must be given as to how we parameterize such comparisons.*
>
> *In particular, we may be interested to discover if $\varphi_0 = 0$ is so included [in a $(1-\alpha)$ HPD region]. The point $\varphi_0 = 0$ corresponds to the situation where $\theta_1 = \ldots = \theta_k$, and is often of special concern in comparing location of distributions.*

Box and Tiao [9, p.1470] advocate the use of statistical methods based on HPDSs, see section 7. However, they understate the importance of invariance, downplaying the issue altogether. After all, any unfeasible property should be regarded as an inessential characteristic:

> *It seems that we cannot hope for invariance for a genuine measure of credibility. It needs to be remembered that invariance under transformations and virtues are not synonymous. For problems which should not be invariant under transformation, a search for invariance serves only to guarantee inappropriate solutions.*

As far as we know, Box and Tiao [9] option to forfeit invariance while using measures of significance based on credible regions remained consensual in mainstream Bayesian analysis. However, we beg to strongly disagree with Box and Tiao final conclusion about the importance of invariance properties in statistical procedures. For an in-depth discussion of this issue, see Stern [60].

5 Frequentist p-values and their Deconstruction

The general idea of a p-value is to compute the probability that, repeating a random experiment under a given statistical model and a given hypothesis H, one would obtain an observation that is more extreme, that is, more unlikely than the one that was actually observed.

In section 2 it was stated that, in the frequentist conceptual framework, the (vector) parameter θ of the statistical model is considered as a 'fixed but unknown' quantity. Moreover, under the frequentist paradigm θ must not be regarded as a random variable. Furthermore, the language of probability is strictly forbidden for any direct description or manipulation of the existing uncertainty about θ. In this setting, it is easy to argue that the *maximum-likelihood* or ML estimator, $\widehat{\theta}$ is the best choice for fixing the parameters, if they are free. Likewise, the constrained ML, θ^*, is arguably the best choice for fixing the parameters under the constraints established by a given hypothesis H.

Under these conditions, the predictive distribution $p_n(x \,|\, \theta^*)$ can be used to induce an order in the sample space, so that the so far vague idea of a p-value becomes a well-defined concept, see Pereira and Wechsler [45]. $\overline{\text{pv}}(H)$, the complement of the p-value of hypothesis H, is defined as follows:

$$\overline{\text{pv}}(H) = \int_{C(y_1,\ldots,y_n)} \prod_{i=1}^{n} p(x_i \,|\, \theta^*)\, dx^n \;, \quad \theta^* = \arg\max_{\theta \in H} \prod_{i=1}^{n} p(y_i \,|\, \theta) \;,$$

$$C(y_1,\ldots,y_n)=\left\{x_1,\ldots,x_n \mid \prod_{i=1}^{n} p(x_i\,|\,\theta^*)\leq \prod_{i=1}^{n} p(y_i\,|\,\theta^*)\right\}\ .$$

Pseudo-possibilistic nature of p-values

$\mathrm{pv}\,(H)=1-\overline{\mathrm{pv}}(H)$ measures the probability of 'extreme' events, extreme according to the order in the sample space engendered the by the optimal (most likely) parameter θ^*. Hence, $\mathrm{pv}\,(H)$ considers the probability of worst case outcomes under the best case parameters. In this sense, p-values have a "possibilistic" appearance.

However, although the hypothesis H is stated in the parameter space, $H\subseteq\Theta$, $\mathrm{pv}\,(H)$ 'shifts the problem' computing a probability in the sample space. Hence, $\mathrm{pv}\,(H)$ only superficially resembles the basic conceptual framework for probability-possibility transformations studied in Section 3.

Construction of Practical p-values.

As last defined, $\overline{\mathrm{pv}}(H)$ may be extremely difficult to compute. Please note that, in the equation defining a p-value, y_i and x_i, $i=1\ldots n$, stand for, respectively, the observed and a possible new data bank. Each data bank is coded in matrix form, $x_{i,j}, i=1\ldots n, j=1\ldots m$, consisting of n singular observations, $x_{i,\bullet}$, each of these coded as a vector of dimension m. Hence, the dimension of the integration space, $m*n$, increases linearly with the number of observations, demanding an exponential (in n) computing time. Moreover, the geometry of the cut-region $C(y_1,\ldots,y_n)$ is specified by non-linear constraints that may be hard to handle either analytically or numerically. Therefore, it should not come as a surprise that the equation defining a general p-value is seldomly directly used in practice. Instead, practical implementations make use of several approximation techniques, reparameterization maneuvers, dimensionality reduction strategies and pre-compiled algorithms that greatly simplify the subsequent computational procedures. In the sequel we put in perspective some of these methods and how they relate and interrelate in practical implementations.

We hope that this deconstructive analysis will help us to understand why p-values are such a successful tool and, later on in this paper, also to understand why e-values can match or supersede them in practice. An important source of inspiration for the development of the e-value was a challenge made by Oscar Kempthorne to the second author, asking for a Bayesian measure of significance able to compete with the frequentist p-values, including the case of sharp hypotheses. Section 7 will revisit this issue, explaining how, in the authors' view, the e-value is a worthy answer to Kempthorne's challenge.

(a) Asymptotic approximations: Under mild regularity conditions that are commonly assumed in the practice of statistical practice, and for large data banks, $n\to\infty$, it is reasonable to approximate the behavior of certain random variables of interest by well-known and convenient to manipulate statistical distributions. The central limit theorem, stating the convergence of (vector) means to (multivariate) Normal random variables provides the best known of such approximations.

(b) Sufficient statistics: It would be very useful to reduce $[x_1,\ldots,x_n]$, the entire sample, to a compact or condensed statistic $S(x_1,\ldots,x_n)$. Typically, although the pre-image of the map $S(\ )$ is a space of dimension $m*n$, growing linearly with the

sample size, its image is a space of fixed dimension. Perhaps the best known example is the reduction of sample of n real scalar values to their mean and variance, $S(x_1, \ldots, x_n) = [m, s^2]$, where $m = (1/n)\sum x_i$ and $s^2 = (1/n)\sum(x_i - \mu)^2$. In general, a lot of relevant information would be lost by such a reduction. Under special appropriate conditions, for a given map S in a specific statistical model, no relevant information is lost. In these circumstances S is called a *sufficient statistics*, see Kempthorne and Folks [30]. In many practical applications, (approximate) sufficient statistics are obtained in conjunction with asymptotic approximations.

(c) Numerical algorithms for Gaussian and related distributions: For the multivariate Normal, Chi-2, central and non-central F and T, and other related distributions, many useful computations can be handled with the aid of precompiled algorithms, tables, or even analog devices, see Pickett [46]. We believe that these simple but powerful techniques offer at least a partial explanation for the extraordinary success of frequentist statistics in the times predating personal computers. In section 7 we discuss how these techniques can be successfully applied also in the context of Bayesian statistics.

(d) Nuisance parameter elimination: In mainstream mathematical statistics, it is customary, if by all means feasible, to reparameterize a model using new coordinates $[\delta, \lambda]$, $\dim(\delta) = k$, $\dim(\lambda) = h$, so that, the *parameters of interest,* δ, are completely specified by the hypothesis, while the *nuisance parameters,* λ are free. Such a $\Theta = \Delta \times \Lambda$ decomposition can then be followed by a *nuisance parameter elimination* procedure, that is, a mapping or 'projection', $\mathcal{D}(\Theta) = \Delta$, that reduces the original composite hypotheses to the point-hypothesis $\mathcal{D}(H) = \{\delta^0\}$.

Basu and Ghosh [5] give an extensive list of at least 10 categories of procedures for achieving this goal, like using max_λ or $\int\ d\lambda$, the maximization or integration operators, in order to obtain a projected profile likelihood or marginal posterior function, $p(\delta \mid x)$, see also Pereira and Lindley [42]. Maximization operations and profile likelihoods are especially important in the frequentist framework, while integration operations and marginal posteriors are especially important in the decision theoretic Bayesian framework, as discussed in the next section.

We can now begin to contrast the characteristics of p-values and e-values, a contrast analysis that will, in the following sections, be extended to decision theoretic Bayesian methods. The FBST performs probability calculations in the parameter space and, in so doing, falls within the Bayesian framework. However, the FBST does not follow the nuisance parameters elimination paradigm, working in the original parameter space, in its full dimension. In this respect the FBST breaks away from both the frequentist and the decision theoretic Bayesian tradition. Furthermore, the FBST is defined by a first optimization step, a possibilistic operation on the surprise, followed by a second posterior probability integration step. The combination of these two steps looks alien in any of the two traditional frameworks for statistical theory. Moreover, with the currently available numerical optimization and integration algorithms, the FBST has little need to rely on calculation or computational procedures based on asymptotic approximation techniques. The consequences of all these departures from and compromises with different means and methods developed by the two mainstream statistical schools is further explored in the next sections.

6 Bayes Factors - Probability as a Tool for all Trades

This section presents Bayes factors, the probability all-the-way decision theoretic solution for hypothesis test. Given two alternative simple hypotheses (or models), H_1 and H_2, and an observed data (matrix), $X = [x_1, x_2 \ldots, x_n]$, the *Bayes factor* $B_{1,2}$ transforms the *prior odds ratio* to the *posterior odds ratio*, that is,

$$\frac{\Pr(H_1 \,|\, X)}{\Pr(H_2 \,|\, X)} \;=\; \frac{\Pr(X \,|\, H_1)}{\Pr(X \,|\, H_2)} \frac{\Pr(H_1)}{\Pr(H_2)} \;=\; B_{1,2} \frac{\Pr(H_1)}{\Pr(H_2)} \;.$$

This transformation is a direct consequence of Bayes rule,

$$\Pr(H_k \,|\, X) = \frac{\Pr(X \,|\, H_k)\Pr(H_k)}{\Pr(X \,|\, H_1)\Pr(H_1) + \Pr(X \,|\, H_2)\Pr(H_2)} \;.$$

The calculation of Bayes factors for parameterized hypotheses if further explained in Kass [29, p.776]:

> *In the simplest case, when the two hypothesis are single distributions with no free parameters (the case of simple versus simple testing), $B_{1,2}$ is the likelihood ratio. In other cases, when there are unknown parameters under either or both of the hypotheses, the Bayes factor is still given by [the formula above], and. in a sense, it continues to have the form of a likelihood ratio. Then, however, the densities $p(X \,|\, H_k)$ are obtained by* integrating *(not maximizing) over the parameter space* [where g_k is the prior probability of the (vector) parameter θ_k under hypothesis H_k],
>
> $$B_{1,2} \;=\; \frac{\int_{H_1} p(X \,|\, H_1, \theta_1) g_1(\theta_1 \,|\, H_1) d\theta_1}{\int_{H_2} p(X \,|\, H_2, \theta_2) g_2(\theta_2 \,|\, H_2) d\theta_2} \;.$$

Building such a weighted or average likelihood ratio is the traditional course of action within the decision theoretic Bayesian paradigm driven by the betting-utility or scientific casino metaphors, as discussed at length in Stern [61]. A strong point of this approach is that a single abstract belief calculus, namely probability, is used to handle all computations involving uncertain quantities. This approach is very effective in several cases involving non-sharp hypothesis, but runs into serious difficulties in the case of sharp hypotheses representing lower dimensional sub-models of a given model.

Lindley's paradox and related symptoms indicate that naive ad-hoc assignments of non-zero measures to null-Lebesgue measure sets is problematic. There is a vast and ever increasing literature proposing ever more sophisticated case-specific prior measures specifically designed for this purpose. However, the very need of such convoluted solutions seems to indicate that it may be worthwhile to develop different approaches for testing sharp hypotheses.

There is also a strong line of thought in the decision theoretic Bayesian school arguing that sharp hypotheses make little sense in this framework, and that the technical difficulties alluded in the last paragraph further confirm and justify regarding sharp hypotheses as ill-posed statements. Although tempting from a theoretical perspective internal to the decision theoretic framework, this position cannot withstand

the strong demand by the scientific community for adequate procedures for testing sharp hypotheses. Sharp hypotheses are naturally stated in many areas of research, specially so in the exact sciences, as discussed in Stern [56-61].

Decision Theoretic Rejection of Possibilistic Logic

The optimization step of the FBST, $\theta^* = \arg\max_H p_n(\theta)$, can be interpreted as 'representing' H by its best case. In section 5 we argued that this choice for the true value of a fixed but unknown parameter was a very reasonable under the frequentist framework. This choice follows a well established principles of possibilistic logic, as explained in Darwiche [14], Darwiche and Ginsberg [15], and Borges and Stern [8]. From the conceptual considerations in the present section, however, one can understand that the same optimization step lies completely outside the traditional decision theoretic Bayesian framework. Denis Lindley [36] explicitly warns us that taking this path is, in his view, a very unwise course of action:

> *Several proposals have been made as how ev(H) might be calculated. Two are p-values and posterior probabilities, to which you added a third. Suppose we look at ev(H) as an abstract concept and ask ourselves what properties it should have. For example, if A and B are two hypotheses... I find it compelling that ev(H) should satisfy the assumptions SP_1 to SP_5 in chapter 6 of DeGroot's book... Accepting this five assumptions, DeGroot proves that ev(H) must obey all the rules of probability, in particular that ev$(A$ or $B) =$ ev$(A) +$ ev$(B) -$ ev$(A$ and $B)$. As you clearly point out... your form of ev(H) does not satisfy this rule, but rather ev$(A$ or $B) = \max[$ev$(A),$ ev$(B)]$.*
>
> *Separate from the axiomatic approach, many people, including myself, have objected to the concept of possibility on the grounds that with it ev$(A$ or $B)$ can be calculated from ev(A) and ev(B) without any consideration of the relationship between A and B... Probability requires three numbers adequately to describe the relationship between two hypotheses; possibility uses only two and, for that reason, is often thought to be inadequate.*

John Skilling [53], give us even stronger warnings against following a renegade possibilistic logic that lead us into temptation of accepting the optimization heresy:

> *Take maximum likelihood. The difficulty there is that the single point of maximum may be highly atypical of the measure being assessed... By using the single point θ^* to represent the entire hypothesis H, the FBST fails to escape the inadequacy of representing a set with a point.... I can't accept that.*

7 Lindley's Method

The treatment given to sharp hypotheses by frequentist p-values is far less problematic than that offered by Bayes factors, at least for the aspects previously discussed that directly concern the properties of zero-measure sets. Optimization on algebraic sub-manifolds defined by equality and inequality constraints is the standard problem of mathematical programming, requiring only good optimization algorithms, see for example Andreani et al. [3]. In contrast, computing non-zero integrals over proper algebraic sub-manifolds requires several ad-hoc choices and definition of tailor-made

objects (like artificial priors and other oximora) for the statistical model at hand. Furthermore, all this effort may be of no avail resulting in statistical procedures that are problematic in the case of sharp hypotheses.

This state of affairs stimulated research on compromise methods allowing for the joint use of 'probabilistic' integral operators and 'possibilistic' maximization operators within a framework that is fundamentally Bayesian insofar as it allows and makes use of probability measures in the parameter space. Several of such compromise methods have been developed. However, many of these hybrid possibilistic-probabilistic procedures (not the FBST) lack a solid theoretical and epistemological foundation. Anyway, when presenting any of the pre-existing methods for handling sharp hypothesis, many Bayesian authors give an explicit *caveat emptor* to the users, warning them that such procedures are pragmatic tools that should be used with great care, see for example Williams [65, p.234].

> *Bayesian significance of sharp hypothesis: a plea for sanity: ...It astonishes me therefore that some Bayesian now assign non-zero prior probability that a sharp hypothesis is exactly true to obtain results which seem to support strongly null hypotheses which frequentists would very definitely reject. (Of course, it is blindingly obvious that such results must follow).*

Bayesian p-values look at the *frequency of extreme samples under average parameters* located at the hypothesis, that is, they measure the frequentist probability of observing a data bank more extreme than the one actually observed, under a parameter θ distributed over H according to a convenient posterior probability $g_n(\theta)$ on H. However interesting, Bayesian p-values and other compromise solution that are not directly linked with the ideas leading to the FBST, have to be analyzed in future articles.

This section discusses Lindley's method, an approach for testing a point hypotheses $H : \theta = \theta^0$ based on the coverage (or not) of θ^0 by a credibility interval of prescribed size, see Lee [33, Sec4.3, p.123]. Hence, this is an approach that looks at *extreme epistemic intervals covering a given point-hypotheses.* Hald [24] gives a detailed account of the use of credibility and confidence intervals in parametric statistics since the times of Laplace and Gauss; see also Barnett[4, Sec.5.5]. In Lindley's method, HPDSs are at the center-stage of all computational procedures and, in this sense, it could be considered as a direct precursor of the FBST. Denis Lindley himself regards such methods only as practical procedures for testing point-hypotheses, having however serious reservations about its underlying theoretical foundations. We introduce this method by the words of some authors who used it in important applications:

> *The normal distribution has the remarkable property that equivalent statements can be made with either X or Θ as the relevant space supporting the probability distributions... A Bayesian interpretation of the common F-test is then available by rephrasing the sampling-theory notion that a null value is significant if the confidence interval does not include it, confidence being replaced by credible... Essentially in rejecting the null value we are saying that it has not got high posterior probability (density) in comparison with other values.*
>
> *Although these ideas enable orthodox practice to be interpreted in probability terms, it does not follow that the practice is to be adopted... We now turn from*

> *sampling-theory concepts to an honest Bayesian analysis of a decision problem (and hence of an associated inference problem).* D.V.Lindley [35, p.18,19].

> *In normal linear models, HPD regions are always connected regions, or intervals, and these provide interval-based inferences and tests of hypotheses through the use of posterior normal, T and F distributions. West and Harrison [64, Sec.17.3.5, p.643].*
>
> *A time plot of the estimates [values], with an indication of the associated uncertainty as measured by the posterior variance, provides a clear and useful visual indication of the contribution of the [scalar parameter of interest].* West and Harrison [64, Sec.8.6.7, p.256,257].

> *Testing of a Sharp Null Hypothesis Through Credible Intervals:*
> *Some Bayesians are in favor of testing, say, $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ by accepting H_0 if θ_0 belongs to a chosen credible set. This is similar to the relation between confidence intervals and classical testing, except that there the tests are inverted to get confidence intervals. This must be thought of as a very informal way of testing. If one really believes that the sharp null is a well-formulated theory and deserves to be tested, one would surely want to attach a posterior probability to it. That is not possible in this approach.*
>
> *Because the inference based on credible intervals often has good frequency properties, a test based on them also is similar to a classical test. This is in sharp contrast with inference based on Bayes factors or posterior odds.* Ghosh et al. [23, Sec.2.7.3-4, p.48-50].

Other influential authors in the field of Bayesian statistics are not willing to be so complaisant about heterodox transgressions relative to decision theoretic doctrine, see for example Berger and Delampady [7, Sec1.2, p.319; Sec.4.3, p.328]:

> Opinion 3: Testing is Somewhat Irrelevant; One Should Concentrate on Confidence Sets, Testing from Them if Necessary. *This opinion is wrong, because it ignores the supposed special nature of θ_0. A point can be outside a 95% confidence set an, yet not be so strongly contradicted by the data. Only by calculating a Bayes factor (or related conditional measures) can one judge how well the data supports a distinguished point θ_0...*
>
> *The Bayes factor communicates the evidence in the data against θ_0, and [a credible region] C the magnitude of the possible discrepancy.*

As already stated, in Lindley's method, HPDSs are at the center-stage of all computational procedures and, in this sense, it could be considered as a precursor of the FBST. Nevertheless, Lindley's method and the FBST are in essence very different approaches to the concept of hypothesis significance. The following list of contrasting points should make the distinction clear. This list also has the purpose of exposing some of the obstacles that, in the authors' view, have prevented further research in this area, at least within the historical path taken in the development of mainstream decision theoretical Bayesian statistics.

(a) Covering-interval interpretation: The original covering-interval interpretation of Lindley's method is often translated into a strong topological requirement of having simply-connected HPDSs. These topological requirements, in turn, render

uni-modality and monotonicity sufficiency conditions for the underlying posterior density or its marginals at the statistical model in study.

(b) Nuisance parameter elimination: Lindley's method can be coherently extended to handle a composite hypothesis H by reduction to a point-hypothesis $\mathcal{D}(H) = \{\delta^0\}$ through an acceptable nuisance parameter elimination procedure. As discussed in Sections 4 and 5, in the decision theoretic Bayesian framework, acceptable procedures for that purpose may be based on a marginalization or other integration operation, but never on a possibilistic optimization operation.

(c) Non-invariance: As discussed in Section 4, Lindley's method is essentially non-invariant, a consecrated reason to question the theoretical foundations of any method based on HPDSs.

(d) Loss function: Before the formulation of an appropriate loss function by Madruga et al. [37], test procedures for sharp hypotheses based on general HPDSs had not been duly derived within the decision theoretic framework, even though closely related loss functions are discussed in O'Hagan [40, Sec.2.5, p.54-59]. The lack of such a foundation was a steady source of complaints concerning the heterodox character of such procedures.

(e) Ontological status of sharp hypotheses: Even after the work of Madruga et al. [37], from the decision theoretic epistemological perspective, sharp hypotheses are perceived to be ill-posed problems. In contrast, the epistemological framework of cognitive constructivism fully supports sharp hypotheses, see Stern [56-61].

It can be argued that, within the decision theoretic framework, points (c), (d) and (e) made Lindley's method and related ideas only barely acceptable as pragmatic procedures that *must be thought of as a very informal way of testing.* Moreover, points (a), (b) and (c) seem to have restricted the scope of interest for practical applications of Lindley's method almost exclusively to Gaussian linear models, as in Box and Tiao [10] or West and Harrison [64]. Nevertheless, the authors recently became aware of the work of Sanjib Basu [6], where the author departs from the strict interval coverage interpretation, and breaks away from the standard simple-connectivity requirement for HPDSs. In so doing, even if non-invariant and limited to point-hypotheses, his work should be regarded as an even closer precursor to the FBST.

Paradoxically (from the cognitive constructivism plus FBST perspective), but also very interestingly (from an historical perspective), the paradigm of interval-based inference seems to have percolated (backwards) from (sometimes very old sources in) the literature of mathematical statistics to that of possibility and fuzzy sets theory. For recent developments on interval-based possibilistic inference, see Castineira et al [13] and Salicone [49]. The next quotation, from Dubois et al. [18, Sec,3,p.282], explains the authors' motivation for this course of investigation even if, as acknowledged by the authors in Remark 3.2, p.282,283, consequent definitions only make sense for unimodal and monotonic distributions.

> *A closed form expression of the possibility distribution induced by confidence intervals around the mode $\widehat{x}$ is obtained for unimodal continuous probability densities strictly increasing on the left and decreasing on the right of $\widehat{x}$...*
> *In the paper, we use the terminology 'confidence interval' for reliable interval substitutes to probability distributions. It does not correspond to the traditional*

terminology...
Our notion of confidence interval is much closer to Fishers fiducial interval.

8 Final Remarks and Future Research

Epistemic vs. Predictive Significance Tests and their Performances

In the quotation of Ghosh et al. [23] at the last section, the authors praise the good frequency properties of Bayesian inference based on credible intervals. Leonard and Hsu [34, p.142-143] make similar remarks. These good frequency properties can explain the FBST outstanding performance at comparative benchmarks based on frequencies of type-I and type-II errors, used to gauge the FBST performance at several already published applications. Nevertheless, Skilling [53] sees these measures of performance with great suspicion. Contradictory opinions about this issue are common in the Bayesian statistics literature: Frequency properties of covering intervals are sometimes exalted as a wonderful feature, sometimes dismissed as an irrelevant aspect.

We believe that contradictory appraisals of these frequency properties and different opinions about performance measures are deeply entangled with two very distinct concepts of statistical significance, namely, the notion of predictive power of a theory vs. the idea of epistemic verification of the same theory. The last paragraph in the quotation of Berger and Delampady [7] at the last section has already hinted at making this same distinction. We also believe that these two concepts, although intertwined, require essentialy different tests of significance that ought to be accompanied by compatible, and hence also different, measures of performance. This issue should benefit from extensive further research.

Alternative Probability-Possibility Transformations

In Dubois and Prade [19, p.177,180] the authors define two alternative probability-possibility transformations specially suited for continuous densities, namely,

$$\kappa(\varphi) = \int_{\Theta} \min\left[p\left(\theta\right) p\left(\varphi\right)\right] p(\theta) d\theta \;; \quad \xi(\varphi) = \frac{p(\varphi)}{\widehat{p}} \;,\; \widehat{p} = \sup_{\Theta} p(\theta) \;.$$

These alternative transformations have several interesting interpretations. Under reasonable regularity conditions, both transformations are easily extended to possibility measures if computed at the argument $\theta^* = \arg\max_H p(\theta)$.

For the first alternative transformation, the identity $\pi(\widehat{\theta}) = \kappa(\widehat{\theta}) = 1$ can be interpreted as two alternative ways of calculating the total volume under $p(\theta)$, integrating over vertical or horizontal 'slices'. The equivalence of this two ways of computing the total probability is a consequence of Fubini theorem that, in turn, can be interpreted by Cavalieri principle, expressing a notion that predates the formalization of calculus by either Newton or Leibniz, see Fubini [22] and Palmieri [41].

Analytically, the alternative transformation $\kappa(\varphi)$ can be easier to handle than the standard possibility transformation, $\pi(\varphi)$, because the (necessarily) discontinuous integrand, $\mathbf{1}\left[p(\theta) \leq p(\varphi)\right] p(\theta)$, is replaced the (possibly) continuous function $\min\left[p(\theta), p(\varphi)\right]$. Furthermore, all the results stated for one of these measures can be immediately translated to the other for, taking $v = p(\theta)$, we get $\delta(\theta) = \kappa(\theta) - \pi(\theta) =$

$v * \mu(\overline{T}(v))$, where μ stands for the appropriate Lebesgue measure.

The second alternative measure, $\xi(H)$, is even easier to handle analytically, but it also departs even farther away from the standard measure. Several other alternative probability-possibility transformations have been proposed in the literature, see for example Dubois and Prade [19], Dubois et al. [18, 21], Castineira et al. [13], Dhar [17], Jumarie [27, 28], Mauris et al. [39], Salicone [49], Yamada [62] and Wonneberger [66]. In future research, we intend to study generalizations of the e-value based on some of these alternative probability-possibility transformations.

Acknowledgments

The authors are grateful for the support of the Department of Applied Mathematics and the Department of Statistics of the Institute of Mathematics and Statistics of the University of São Paulo, FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo, and CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico (grants PQ-306318-2008-3 and PQ-302046-2009-7).

Most of the material in this article was presented at the following conferences: EBL-2011, the XVI Brazilian Logic Conference, held on May 9-13 at LNCC - Laboratório Nacional de Computação Cienífica, Petrópolis, Brazil; COBAL-2011, the III Latin American Meeting on Bayesian Statistics, held on October 23–27 at UFRO—Universidad de La Frontera, Pucón, Araucanía, Chile; MBR-2012, Model-Based Reasoning in Science and Technology, held on June 21–23 at Sestri Levante, Genoa, Italy. Finally, a condensed version in Portuguese language is available at the proceedings of CBSF-2012, the II Brazilian Congress on Fuzzy Systems, held on November 06–09 at Natal, Rio Grande do Norte, Brazil. We are grateful to the organizers, reviewers and participants of all these conferences for their support, comments and suggestions, that helped us to improve early versions of this manuscript.

References

[1] S.I.Amari (2007). *Methods of Information Geometry.* Providence, RI: AMS.

[2] S.I.Amari, O.E.Barndorff-Nielsen, R.E.Kass, S.L.Lauritzen, C.R.Rao (1987). *Differential Geometry in Statistical Inference.* IMS Lecture Notes Monograph, v.10. Hayward, CA: IMS.

[3] R.Andreani, E.Birgin, J.Martinez, M.Schuverdt (2007). On Augmented Lagrangian Methods with General Lower-level Constraints. *SIAM J.on Optimization,* 18, 1286-1309.

[4] V.Barnett (1999). *Comparative Statistical Inference.* Chichester John Wiley.

[5] D.Basu (1988). Statistical Information and Likelihood. Edited by J.K.Ghosh. *Lect. Notes in Statistics*, 45.

[6] S.Basu (1996). A New Look at Bayesian Point Null Hypothesis Testing. *Sankhya,* 58,A,2, 292-310.

[7] J.O.Berger, M.Delampady (1987). Testing Precise Hypothesis. Statistical Science, 2,3,317-335. Comments of D.R.Cox, p.335-336, M.Eaton, p.337-338; A.Zellner, p.339-341; M.J.Bayarry, p.342-344; G.Casella and R.L.Berger, p.344-347; J.B.Kadane 347-348; and Rejoinder of J.O.Berger and M.Delampady p.348-352.

[8] W.Borges, J.M.Stern (2007). The Rules of Logic Composition for the Bayesian Epistemic e-values. *Logic J.of the IGPL*, 15, 5-6, 401-420.

[9] G.E.P.Box, G.C.Tiao (1965). Multiparameter Problems From a Bayesian Point of View. *Ann. Math. Statist.,* 36,5,1468-1482.

[10] G.E.P.Box, G.C.,Tiao (1973). *Bayesian Inference in Statistical Analysis.* Reading, MA: Addison-Wesley.

[11] E.Castillo, A.Cobo J.A.Gutierrez, R.E.Pruneda (1998). *Functional Networks with Applications: A Neural-Based Paradigm.* NY: Springer

[12] E.Castillo, A.Iglesias, R.Ruíz-Cobo (2005). *Functional Equations in Applied Sciences.*

[13] E.Castineira, S.Cubillo, E.Trillas (2007). On the Coherence Between Probability and Possibility Measures. *Information Theories & Applications,* 14, 303-310.

[14] A.Y.Darwiche (1993). *A Symbolic Generalization of Probability Theory.* Ph.D. Thesis, Stanford Univ.

[15] A.Y.Darwiche, M.L.Ginsberg (1992). A Symbolic Generalization of Probability Theory. AAAI-92. 10-th Conf. American Association for Artificial Intelligence.

[16] M.Diniz, C.A.B.Pereira, J.M.Stern (2011). Unit Roots: Bayesian Significance Test. *Communications in Statistics - Theory and Methods,* 40,23,4200-4213.

[17] M.Dhar (2012). Probability- Possibility Transformations: A Brief Revisit. To appear in *Annals of Fuzzy Mathematics and Informatics.*

[18] D.Dubois, L.Foulloy, G.Mauris, H.Prade (2004) Probability-Possibility Transformations, Triangular Fuzzy Sets, and Probabilistic Inequalities. *Reliable Computing,* 10, 273-297.

[19] D.Dudois H.Prade (1982). On Several Representations of an Uncertain Body of Evidence. p. 167-181 in M.M.Gupta, E.Sanchez (eds.) *Fuzzy Information and Decision Processes,* North-Holland.

[20] D.Dudois H.Prade (1988). *Possibility Theory: An Approach to Computerized Processing of Uncertainty.* NY: Plenum.

[21] D.Dubois, H.Prade, S.Sandri, (1993). On possibility-probability transformations. In *Fuzzy Logic,* R. Lowen, M. Roubens, eds. p.103-112.

[22] G. Fubini (1958). Sugli integrali multipli. *Opere scelte,* 2,243-249. Cremonese.

[23] J.K.Ghosh, M.Delampady, T.Samanta (2006). *An Introduction to Bayesian Analysis: Theory and Methods.* NY: Springer.

[24] A.Hald (2007). *A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713-1935.* NY: Springer. Ch.8 Credibility and Confidence Intervals by Laplace and Gauss.

[25] 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, edit. Systems and Bayesian Reliability, 287–300. *Quality, Reliability & Engineering Statistics*, 5, Singapore: World Scientific.

[26] R.Johnson, D.Chakrabarty, E.O'Sullivan, S.Raychaudhury (2009). Comparing X-ray and Dynamical Mass Profiles in The Early-Type Galaxy NGC 4636. *The Astrophysical Journal,* 706.

[27] G.Jumarie (1995a). Possibility, probability and relative information: a unified approach via geometric programming. *Kybernetes,* 24,1,18-33.

[28] G.Jumarie (1995b). Further results on possibility-probability conversion via relative information and informational invariance. *Cybernetics and Systems,* 26,1,111-128.

[29] R.E.Kass, A.E.Raftery (1995). Bayes Factors. J.of the American Statistical Association, 90, 430, 773-795.

[30] O.Kempthorne, L.Folks (1971). *Probability, Statistics and Data Analysis.* Ames: Iowa State Univ. Press.

[31] G.J.Klir, T.A.Folger (1988). *Fuzzy Sets, Uncertainty and Information.* NY: Prentice Hall.

[32] M.Lauretto, C.A.B.Pereira, J.M.Stern, S.Zacks (2003). Full Bayesian Significance Test Applied to Multivariate Normal Structure Models. *Brazilian J.of Probability and Statistics,* 17, 147-168.

[33] P.M.Lee (2004). *Bayesian Statistics.* Chichester: Wiley.

[34] T.Leonard, j.s.j.Hsu (1999). *Bayesian Methods: An Analysis for Statisticians and Interdisciplinary researchers.* Sec.3.2.E, p.109, (Bayesian Intervals) and Sec.3.2.F, p.109-110, (Investigating Hypotheses).

[35] D.V.Lindley (1972). *Bayesian Statistics, A Review.* Montpelier,VM: SIAM.

[36] D.V.Lindley (2006). Personal letter to Prof. Carlos Pereira concerning the FBST. `www.ime.usp.br/∼jstern/miscellanea/citacoes/Lindley06p1.pdf, Lindley06p2.pdf`

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

[38] M.R.Madruga, C.A.B.Pereira, J.M.Stern (2003). Bayesian Evidence Test for Precise Hypotheses. *J.of Statistical Planning and Inference,* 117, 185–198.

[39] G.Mauris, V.Lasserre, L.Foulloy (1997). A Simple Probability-Possibility Transformation for Measurement Error Representation: A truncated triangular transformation. World Congress of International Fuzzy Systems Assoc., IFSA, Prague, Czech Republic, 476-481,

[40] A.O'Hagan (1994). *Bayesian Inference.* NY: Halsted Press.

[41] P.Palmieri (2009). Superposition: on Cavalieri's Practice of Mathematics. *Archive for History of Exact Sciences,* 63,5, 471-495.

[42] C.A.B.Pereira, D.V.Lindley (1987). Examples Questioning the use of Partial Likelihood. *The Statistician*, 36, 15–20.

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

[44] C.A.B.Pereira, J.M.Stern, S.Wechsler (2008). Can a Significance Test be Genuinely Bayesian? *Bayesian Anal.* 3,1,79-100.

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

[46] Pickett Inc. (1965). N525 Stat-Rule, A Multi-Purpose Sliderule for General and Statistical Use (Instruction manual). Santa Barbara, CA, USA.

[47] L.L.R.Rifo, S.Torres (2009). Full Bayesian Analysis for a Class of Jump-Diffusion Models. *Communications in Statistics - Theory and Methods,* 38, 1262-1271.

[48] J.Rodrigues (2006). Full Bayesian Significance Test for Zero-Inflated Distributions. *Communications in Statistics - Theory and Methods,* 35, 299-307.

[49] S.Salicone (2007). *Measurement Uncertainty: An Approach via the Mathematical Theory of Evidence.* NY: Springer.

[50] L.Schwartz (1966). *Mathematics for the Physical Sciences.* NY: Addison-Wesley.

[51] G.L.S. Shackle (1961). *Decision Order and Time in Human Affairs.* Cambridge Univ.Press.

[52] G.Shafer (1975). *A Mathematical Theory of Evidence.* Princeton Univ.Press.

[53] J.Skilling (2011). Open review for paper 1146, MaxEnt 2010.

[54] J.M.Stern (2003). Significance Tests, Belief Calculi, and Burden of Proof in Legal and Scientific Discourse. Laptec-2003, *Frontiers in Artificial Intelligence and its Applications,* 101, 139–147.

[55] J.M.Stern (2004). Paraconsistent Sensitivity Analysis for Bayesian Significance Tests. SBIA'04, *Lecture Notes Artificial Intelligence,* 3171, 134–143.

[56] J.M.Stern (2007a). Cognitive Constructivism, Eigen-Solutions, and Sharp Statistical Hypotheses. *Cybernetics and Human Knowing*, 14, 1, 9-36.

[57] J.M.Stern (2007b). Language and the Self-Reference Paradox. *Cybernetics and Human Knowing*, 14, 4, 71-92.

[58] J.M.Stern (2008a). Decoupling, Sparsity, Randomization, and Objective Bayesian Inference. *Cybernetics and Human Knowing*, 15, 2, 49-68.

[59] J.M.Stern (2008b). *Cognitive Constructivism and the Epistemic Significance of Sharp Statistical Hypotheses.* Tutorial book for MaxEnt 2008, The 28th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering. July 6-11 of 2008, Boracéia, São Paulo, Brazil.

[60] J.M.Stern (2011a). Symmetry, Invariance and Ontology in Physics and Statistics. *Symmetry,* 3, 3, 611-635.

[61] J.M.Stern (2011b). Constructive Verification, Empirical Induction, and Falibilist Deduction: A Threefold Contrast. *Information,* 2, 635-650.

[62] K.Yamada (2001) Probability-Possibility Transformation Based on Evidence Theory. Joint 9th IFSA World Congress and 20th NAFIPS International Conference, v.1, 70-75.

[63] S.Wechsler, C.A.B.Pereira, P.C.Marques (2008). Birnbaum's Theorem Redux. *AIP Conference Proceedings,* 1073, 96-100.

[64] M.West, J.Harrison (1997). *Bayesian Forecasting and Dynamic Models,* 2nd.ed. NY: Springer.

[65] D.Williams (2001) *Weighing the Odds.* Cambridge Univ. Press.

[66] S.Wonneberger (1994). Generalization of an invertible mapping between probability and possibility. *Fuzzy Sets and Systems,* 64, 229-240.

[67] L.A.Zadeh (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems,* 1,1,3-28.

Received Submitted December 22, 2011; Rev.2 November 16, 2012.