The e-value: a fully Bayesian significance measure for precise statistical hypotheses and its research program

C. A. B. Pereira & J. M. Stern

São Paulo Journal of Mathematical Sciences

ISSN 1982-6907

São Paulo J. Math. Sci.
DOI 10.1007/s40863-020-00171-7

Springer

São Paulo Journal of Mathematical Sciences
https://doi.org/10.1007/s40863-020-00171-7

SPECIAL ISSUE COMMEMORATING THE GOLDEN JUBILEE OF THE INSTITUTE OF MATHEMATICS AND STATISTICS OF THE UNIVERSITY OF SÃO PAULO

The *e*-value: a fully Bayesian significance measure for precise statistical hypotheses and its research program

C. A. B. Pereira[1] · **J. M. Stern**[2]

Abstract

This article gives a survey of the *e*-value, a statistical significance measure a.k.a. the evidence rendered by observational data, X, in support of a statistical hypothesis, H, or, the other way around, the epistemic value of H given X. The *e*-value and the accompanying FBST, the Full Bayesian Significance Test, constitute the core of a research program that was started at IME-USP, is being developed by over 20 researchers worldwide, and has, so far, been referenced by over 200 publications. The *e*-value and the FBST comply with the best principles of Bayesian inference, including the likelihood principle, complete invariance, asymptotic consistency, etc. Furthermore, they exhibit powerful logic or algebraic properties in situations where one needs to compare or compose distinct hypotheses that can be formulated either in the same or in different statistical models. Moreover, they effortlessly accommodate the case of sharp or precise hypotheses, a situation where alternative methods often require ad hoc and convoluted procedures. Finally, the FBST has outstanding robustness and reliability characteristics, outperforming traditional tests of hypotheses in many practical applications of statistical modeling and operations research.

Keywords Bayesian inference · Hypothesis testing · Foundations of statistics

Communicated by Luiz Renato G. Fontes.

C. A. B. Pereira and J. M. Stern have contributed equally to this work.

✉ J. M. Stern
jmstern@hotmail.com; jstern@ime.usp.br

C. A. B. Pereira
cpereira@ime.usp.br

[1] Institute of Mathematics, Federal University of Mato Grosso do Sul, Avenida Senador Filinto Müller, 1555, Campo Grande 79074-460, Brazil

[2] Institute of Mathematics and Statistics, University of São Paulo, Rua do Matão, 1010, São Paulo 05508-900, Brazil

Published online: 05 June 2020

Mathematics Subject Classification 62F15 · 62F03 · 62A01

1 Introduction

The Full Bayesian Significance Test (FBST) is a novel statistical test of hypothesis published in 1999 by both authors [34] and further extended in Ref. [37, 86]. This solution is anchored by a novel measure of *statistical significance* known as the *e*-value, ev($H \mid X$), a.k.a. the *evidence value* provided by observational data X in support of the statistical hypothesis H or, the other way around, the *epistemic value* of hypothesis H given the observational data X. The *e*-value, its theoretical properties and its applications have been a topic of research for the Bayesian Group at USP, the University of São Paulo, for the last 20 years, including collaborators working at UNICAMP, the State University of Campinas, UFSCar, the Federal University of São Carlos, and other universities in Brazil and around the world. The bibliographic references list a selection of contributions to the FBST research program and its applications.

The FBST was specially designed to provide a significance measure to *sharp* or *precise* statistical hypothesis, namely, hypotheses consisting of a zero-volume (or zero Lebesgue measure) subset of the parameter space. Furthermore the e-value has many necessary or desirable properties for a statistical support function, such as:

(i) Give an intuitive and simple measure of significance for the hypothesis in test, ideally, a *probability* defined directly in the original or *natural parameter space*.
(ii) Have an intrinsically geometric definition, independent of any non-geometric aspect, like the particular parameterization of the (manifold representing the) null hypothesis being tested, or the particular coordinate system chosen for the parameter space, in short, be defined as an *invariant* procedure.
(iii) Give a measure of significance that is smooth, i.e. *continuous and differentiable*, on the hypothesis parameters and sample statistics, under appropriate regularity conditions for the model.
(iv) Obey the *likelihood principle* , i.e., the information gathered from observations should be represented by, and only by, the likelihood function, [13, 96, 143].
(v) Require *no ad hoc artifice* like assigning a positive prior probability to zero measure sets, or setting an arbitrary initial belief ratio between hypotheses.
(vi) Be a *possibilistic* support function, where the support of a logical disjunction is the maximum support among the support of the disjuncts, see [133].
(vii) Be able to provide a *consistent* test for a given sharp hypothesis.
(viii) Be able to provide *compositionality* operations in complex models.
(ix) Be an *exact* procedure, i.e., make no use of "large sample" asymptotic approximations when computing the *e*-value.
(x) Allow the incorporation of previous experience or expert's opinion via (subjective) *prior distributions*.

The objective of the next two sections is to recall standard nomenclature and provide a short survey of the FBST theoretical framework, summarizing the most important statistical properties of its statistical significance measure, the *e*-value; these introductory sections follow closely the tutorial [122, appendix A], see also [37].

2 Bayesian 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 function, 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, 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. In frequentist statistics, θ should be taken as a "fixed but unknown quantity", and neither probability nor any other belief calculus may be used to directly represent or handle the uncertain knowledge about the parameter.

In the Bayesian context, the parameter θ is regarded as a latent (non-observed) random variable. Hence, the same formalism used to express (un)certainty or belief, namely, probability theory, is used in both the sample and the parameter space. Accordingly, the joint probability distribution, $p(x, \theta)$ should summarize all the information available in a statistical model. 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,

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

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 weight) distribution,

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

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

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

In order to replace the 'proportional to' symbol, $\propto$, by an equality, it is necessary to divide the right hand side by the normalization constant, $c_1 = \int_\theta p(x \mid \theta)p_0(\theta)d\theta$. This is the *Bayes rule*, giving the (inverse) probability of the parameter given the data. That is the basic learning mechanism of Bayesian statistics. Computing

normalization constants is often difficult or cumbersome. Hence, especially in large models, it is customary to work with unormalized densities or *potentials* as long as possible in the intermediate calculations, computing only the final normalization constants. 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 i.i.d. (independent and identically distributed) the posterior distribution after n observations, $x^{(1)}, \dots x^{(n)}$, becomes,

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

If 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. For example, the conjugate priors for the Normal and Multivariate models are, respectively, Wishart and the Dirichlet distributions, see [55, 145].

The founding fathers of the Bayesian school, namely, Reverend Thomas Bayes, Richard Price and Pierre-Simon de Laplace, interpreted the Bayesian operation as a path taken for learning about probabilities related to unobservable causes, represented by the parameters of a statistical model, from probabilities related to their consequences, represented by observed data. Nevertheless, later interpretations of statistical inference, like those of Bruno de Finetti who endorsed the epistemological perspectives of empirical positivism, strongly discouraged such causal interpretations, see [128, 129] for further discussion of this controversy.

The 'beginnings and the endings' of the Bayesian learning process deserve 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. References [64, 69, 123, 145] explain how the concept of entropy can be used to unlock many of the mysteries related to the problems at hand. In particular, they discuss some fine details about criteria for prior selection and important properties of posterior convergence.

3 The Epistemic *e*-values

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

$$p_n(\theta) \propto p(x \mid \theta)\, p_0(\theta) .$$

A 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,

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

From now on, we use a relaxed notation, writing H instead of Θ_H. We are particularly interested in sharp (precise) hypotheses, i.e., those in which there is at least one equality constraint and, therefore, $\dim(H) < \dim(\Theta)$.

The FBST defines $\text{ev}(H)$, the *e*-value supporting (in favor of) the hypothesis H, and $\overline{\text{ev}}(H)$, the *e*-value against H, as

$$\begin{aligned}
s(\theta) &= p_n(\theta)/r(\theta) \, , \\
s^* = s(\theta^*) = \sup_{\theta \in H} s(\theta) \, , \ \ \widehat{s} = s(\widehat{\theta}) &= \sup_{\theta \in \Theta} s(\theta) \, , \\
T(v) = \{\theta \in \Theta \mid s(\theta) \leq v\} \, , \ \ W(v) &= \int_{T(v)} p_n(\theta) d\theta \, , \\
\overline{T}(v) = \Theta - T(v) \, , \ \ \overline{W}(v) &= 1 - W(v) \, , \\
\text{ev}\,(H) = W(s^*) \, , \ \ \overline{\text{ev}}(H) = \overline{W}(s^*) &= 1 - \text{ev}(H) \, .
\end{aligned}$$

The function $s(\theta)$ is known as the posterior *surprise function* relative to a given reference density, $r(\theta)$. $W(v)$ is the cumulative surprise distribution. Due to its interpretation in mathematical and philosophical logic, see [16], $W(v)$ is also known as (the statistical model's) *truth function* or *Wahrheitsfunktion*. The surprise function was used in the context of statistical inference by Good [56], Evans [48], Royall [103] and Schackle [109, 110], among others. Its role in the FBST is to make $\text{ev}(H)$ explicitly invariant under suitable transformations on the coordinate system of the parameter space, see next section.

The tangential (to the hypothesis) set $\overline{T} = \overline{T}(s^*)$, 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 H. When $r(\theta) \propto 1$, the possibly improper uniform density, $\overline{T}$ is the Posterior's Highest Density Probability Set (HDPS) tangential to the null set H. Small values of $\overline{\text{ev}}(H)$ indicate that the hypothesis traverses high density regions, favoring the hypothesis.

Notice that, in the FBST definition, there is an optimization step and an integration step. The optimization step follows a typical *maximum probability* argument, according to which, "a system is best represented by its highest probability realization". The integration step extracts information from the system as a probability weighted average. Many inference procedures of classical statistics rely basically on maximization operations, while many inference procedures of Bayesian statistics rely on integration (or marginalization) operations. In order to achieve all its desired properies, the FBST procedure has to use both operation types.

3.1 Nuisance parameters

Let us consider the situation where the hypothesis constraint, $H : \ h(\theta) = h(\delta) = 0 \, , \ \ \theta = [\delta, \lambda]$ is not a function of some of the parameters, λ. This situation is described in [11] by Debabrata Basu as follows:

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

Basu goes on listing at least 10 categories of procedures to achieve this goal, like using max_λ or $\int d\lambda$, the maximization or integration operators, in order to obtain a projected profile or marginal posterior function, $p(\delta \mid x)$. The FBST does not follow the nuisance parameters elimination paradigm, working in the original parameter space, in its full dimension.

3.2 Reference prior and invariance

In the FBST the role of the reference density, $r(\theta)$ is to make $\overline{\text{ev}}(H)$ 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 information in the parameter space, or the limit prior for no observations, or the neutral ground state for the Bayesian learning operation. Standard (possibly improper) uninformative priors include the uniform, maximum entropy densities, or Jeffreys' invariant prior. Finally, invariance, as used in statistics, is a metric concept, and the reference density can be interpreted as induced by the statistical model's information metric in the parameter space, $dl^2 = d\theta' G(\theta) d\theta$, see [2, 12, 17, 49, 55, 64, 69, 70, 145] for a detailed discussion. Jeffreys' invariant prior is proportional to the square root of the information matrix determinant, $p(\theta) \propto \sqrt{\det G(\theta)}$.

3.3 Proof of invariance

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

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

$$\widetilde{s}(\omega) = \frac{\widetilde{p}_n(\omega)}{\widetilde{r}(\omega)} = \frac{p_n(\phi^{-1}(\omega))|J(\omega)|}{r(\phi^{-1}(\omega))|J(\omega)|}$$

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

$$\widetilde{s}^* = \sup_{\omega\in\Omega_H} \widetilde{s}(\omega) = \sup_{\theta\in\Theta_H} s(\theta) = s^*$$

hence, the tangential set, $\overline{T} \mapsto \phi(\overline{T}) = \widetilde{\overline{T}}$, and

$$\widetilde{\overline{\text{ev}}}(H) = \int_{\widetilde{\overline{T}}} \widetilde{p}_n(\omega) d\omega = \int_{\overline{T}} p_n(\theta) d\theta = \overline{\text{ev}}(H).$$

3.4 Asymptotics and consistency

Let us consider the cumulative distribution of the evidence value against the hypothesis, $\overline{V}(c) = \Pr(\overline{\text{ev}} \le c)$, given θ^0, the true value of the parameter. Under appropriate regularity conditions, for increasing sample size, $n \to \infty$, we can say the following:

- If H is false, $\theta^0 \notin H$, then $\overline{\text{ev}}$ converges (in probability) to 1, that is, $\overline{V}(0 \le c < 1) \to 0$.
- If H is true, $\theta^0 \in H$, then $\overline{V}(c)$, the confidence level, is approximated by the function

$$QQ(t, h, c) = \text{Q}\left(t - h,\ \text{Q}^{-1}(t, c)\right), \quad \text{where}$$
$$\text{Q}(k, x) = \frac{\Gamma(k/2, x/2)}{\Gamma(k/2, \infty)}, \quad \Gamma(k, x) = \int_0^x y^{k-1} e^{-y} dy,$$

$t = \dim(\Theta)$, $h = \dim(H)$ and $\text{Q}(k, x)$ is the cumulative chi-square distribution with k degrees of freedom.

Under the same regularity conditions, an appropriate choice of threshold or critical level, $c(n)$, provides a consistent test, τ_c, that rejects the hypothesis if $\overline{\text{ev}}(H) > c$. The empirical power analysis developed in [76, 135] provides critical levels that are consistent and also effective for small samples.

3.5 Proof of consistency

Let $\overline{V}(c) = \Pr(\overline{\text{ev}} \le c)$ be the cumulative distribution of the evidence value against the hypothesis, given θ. We stated that, under appropriate regularity conditions, for increasing sample size, $n \to \infty$, if H is true, i.e. $\theta \in H$, then $\overline{V}(c)$, is approximated by the function

$$QQ(t, h, c) = \text{Q}\left(t - h,\ \text{Q}^{-1}(t, c)\right).$$

Let θ^0, $\widehat{\theta}$ and θ^* be the true value, the unconstrained MAP (Maximum A Posteriori), and constrained (to H) MAP estimators of the parameter θ.

Since the FBST is invariant, we can chose a coordinate system where, the (likelihood function) Fisher information matrix at the true parameter value is the identity, i.e., $J(\theta^0) = I$. From the posterior Normal approximation theorem, see [55], we know that the standarized total difference between $\widehat{\theta}$ and θ^0 converges in distribution to a standard Normal distribution, i.e.

$$\sqrt{n}(\widehat{\theta} - \theta^0) \to N\big(0, J(\theta^0)^{-1} J(\theta^0) J(\theta^0)^{-1}\big)$$
$$= N\big(0, J(\theta^0)^{-1}\big) = N(0, I)$$

This standarized total difference can be decomposed into tangent (to the hypothesis manifold) and transversal orthogonal components, i.e.

$$d_t = d_h + d_{t-h}\,, \quad dt = \sqrt{n}(\widehat{\theta} - \theta^0)\,,$$
$$d_h = \sqrt{n}(\theta^* - \theta^0)\,, \quad d_{t-h} = \sqrt{n}(\widehat{\theta} - \theta^*)\,.$$

Hence, the total, tangent and transversal distances (L^2 norms), $||d_t||$, $||d_h||$ and $||d_{t-h}||$, converge in distribution to chi-square variates with, respectively, t, h and $t - h$ degrees of freedom.

Also from, the MAP consistency, we know that the MAP estimate of the Fisher information matrix, $\widehat{J}$, converges in probability to true value, $J(\theta^0)$.

Now, if X_n converges in distribution to X, and Y_n converges in probability to Y, we know that the pair $[X_n, Y_n]$ converges in distribution to $[X, Y]$. Hence, the pair $[||d_{t-h}||, \widehat{J}]$ converges in distribution to $[x, J(\theta^0)]$, where x is a chi-square variate with $t - h$ degrees of freedom. So, from the continuous mapping theorem, the evidence value against H, $\overline{\text{ev}}(H)$, converges in distribution to $\overline{e} = \text{Q}\,(t, x)$, where x is a chi-square variate with $t - h$ degrees of freedom.

Since the cumulative chi-square distribution is an increasing function, we can invert the last formula, i.e., $\overline{e} = \text{Q}\,(t, x) \leq c \Leftrightarrow x \leq \text{Q}^{-1}(t, c)$. But, since x in a chi-square variate with $t - h$ degrees of freedom,

$$\Pr\,(\overline{e} \leq c) = QQ(t, h, c) = \text{Q.E.D.}$$

A similar argument, using a non-central chi-square distribution, proves the other asymptotic statement.

3.6 Decisions: reject *H*, remain neutral, or accept

In this subsection we briefly discuss the important question of deciding when to Accept, or Reject, or remain Neutral about a statistical hypothesis H, given observed data X. We start our discussion elaborating on the asymptotic results derived in the last sub-section.

If a random variable, x, has a continuous cumulative distribution function, $F(x)$, its probability integral transform generates a uniformly distributed random variable, $u = F(x)$, see [5]. Hence, the tranformation $\overline{\text{sev}} = QQ(t, h, \overline{\text{ev}})$, defines a "standarized e-value", $\text{sev} = 1 - \overline{\text{sev}}$, that can be used somewhat in the same way as a p-value of classical statistics. This standarized e-value may be a convenient value to report, since its asymptotically uniform distribution (under H) provides a large-sample limit interpretation, and many researchers will feel already familiar with consequent diagnostic procedures for scientific hypotheses based on adequately large empirical datasets. In particular, a researcher may use cut-off thresholds already familiar to the him when dealing with p-values. Efficient procedures for computing empirical cut-off thresholds that are effective for small size data sets are developed in [14, 76–79].

Traditionally, statisticians are used to establish a dichotomy: Reject/ Accept (technically, Not-Reject) H if the significance measure in use is below or above the established cut-off threshold. Nevertheless, a thorough analysis of consistent desiderata for logical properties of such a decision procedure take us to an unavoidable conclusion: The classical Reject/ Accept dichotomy must be replaced by a trichotomy, namely, Reject/ remain Neuter (a.k.a remain undecided or agnostic)/ Accept H if, respectively, $0 \leq \text{sev}(H \mid X) < c_1$, $c_1 \leq \text{sev}(H \mid X) < c_2$, or $c_2 \leq \text{sev}(H \mid X) \leq 1$, where $0 < c_1 < c_2 < 1$; For an extensive and detailed analysis of consistent desiderata for statistical test procedures, see [63, 112].

The study of such logical desiderata was in part motivated as a way to contrast the statistical properties of the FBST with other statistical tests of hypotheses. Surprisingly, it is possible to travel this path in the opposite direction, that is, it is possible to start from consistent desiderata for logical properties of statistical tests and, from those, derive a complete characterization of a class of statistical significance measures and hypothesis tests that coherently generalizes the FBST, see [46, 47, 131] for further details. Moreover, this Generalized FBST finds interesting applications in metrology and related fields, were reliable bounds for the precision of experimental measurements can be obtained from sources external to the statistical experiment designed to test the hypothesis under scrutiny. Finally, this kind of detailed error analysis for crucial scientific experiments finds valuable applications in the fields of metrology, epistemology, and philosophy of science, see [47] and future research.

4 A survey of FBST related literature

A systematic cataloging of all published articles related to this research program is beyond the scope of this article; in the next subsections we survey a selection of such articles. The selected articles provides a sample covering diverse areas like statistical theory and methods, applications to statistical modeling and operations research, and research in foundations of statistics, logic and epistemology. This selection is certainly biased, favoring the the authors' personal taste or involvement.

4.1 Statistical theory

Several authors have developed the statistical theory that provides the mathematical formalism and demonstrates the outstanding statistical properties FBST and its significance measure, the e-value. The following articles have explored and developed these themes of research:

- Reference [34] is the first article of this research program. It presents the basic definition of the e-value and the FBST, and gives several simple and intuitive applications. Reference [86] provides an explicitly invariant version of the inference procedures. After a long process in which the authors had to overcome objections raised by influential mainstream Bayesian thinkers, [37] was published in the flagship journal of ISBA - the *International Society for Bayesian*

Analysis. Reference [30] provides an entry on the FBST in the *International Encyclopedia of Statistical Science*.

- References [6, 7, 32, 77–79] give and extensive treatment for the case of non-nested and separate hypotheses, including a detailed analysis of some Bayesian classifiers.
- References [28, 44, 84, 94, 95, 113] establish several theoretical or empirical relations between the the *e*-value and alternative significance measures.
- References [19, 98, 99, 104, 105, 137–139] develop higher order asymptotic approximations of (log) likelihood and pseudo-likelihood functions that, in turn, are used do develop high-precision but fast computational algorithms for calculating *e*-values in parametric models. The availability of a good library of such fast and reliable computer programs will, in turn, we believe, facilitate the incorporation of the FBST in statistical softwares intended for end-users or routine applications.

4.2 Statistical modeling

Several authors have developed a wide range of applications of the FBST to statistical modeling and operations research. The following articles have explored and developed these themes of research:

- Reference [35] applies the FBST to software compliance testing and certification.
- Reference [76] provides a unified and coherent treatment to a large class of structural models based on the multivariate normal distribution. Previously, subclasses of these models had to be handled individually using tailor-made tests. Reference [141] gives some simple applications.
- References [24, 42, 43, 142] develop or use unit root and cointegration testing for time series. The FBST is shown to be more reliable and effective than several previously published tests, without the need of any ad hoc artifices, like specially designed artificial priors (an obvious oxymoron).
- References [61, 88, 102] apply the FBST to failure analysis and systems' reliability.
- Reference [22] applies the FBST to detect equilibrium conditions, or the lack thereof, in market prices of economic commodities or financial derivative contracts.
- Reference [25] applies the FBST in the context of empirical economic studies.
- Reference [54] use the FBST for selection and testing of statistical copulas.
- References [57, 135] consider applications using generalizations of the Poisson distribution.
- References [58–60] use the FBST for signal processing and detection of acoustic events.
- Reference [114] applies the FBST to model selection in statistical studies conducted under informative sampling conditions.

- References [18, 38, 62, 80, 83, 87, 91, 144] use the FBST to verify Hardy-Weinberg equilibrium conditions, and other applications of statistical modeling in the area of genetics.
- References [10, 3, 29, 75, 100, 101] use the FBST to test parametric hypotheses related to generalized Brownian motions, continuous or jump diffusions, extremal distributions, persistent memory and other stochastic processes.
- References [4, 14, 39, 93] develop theory or applications of the FBST for statistical hypotheses related to independence in contingency tables and other multinomial models.
- References [20, 1, 21, 31, 36, 41, 74, 82, 107, 108, 115] apply the FBST for checking hypotheses in statistical models applied to biological sciences, ecology, environmental sciences, medical diagnostics and efficacy evaluation, psychology and psychiatry.
- References [23, 65] apply the FBST to test hypotheses in astronomy and astrophysics.

4.3 Foundations of statistics, logic and epistemology

Traditional significance measures used in statistics are always designed to work in tandem with a specific epistemological framework that gives them an appropriate interpretive context and support. For example, *p*-values are usually presented in the context of the "judgment metaphor" and the deductive falibilism epistemological framework, as developed by the philosopher Karl Popper, among others. Meanwhile, Bayes factors are presented in the context of the "gambling metaphor" and utility based decision theory, as developed by Bruno de Finetti, see Ref. [40, 45, 66, 67]. Furthermore, the logic or algebraic properties of each significance measure, in its appropriate domain of statistical hypotheses, must be mutually supportive and compatible with intended interpretations. The following articles have explored and developed these themes of research:

- References [85, 112, 136] analyze the FBST from a decision-theoretic Bayesian perspective. The first paper proves the "Bayesianity" of the FBST, in the sense its inference procedures can be derived by minimization of an appropriate loss function.
- References [86, 133] compare the theoretical properties of the *e*-value with those of traditional significance measures, like the *p*-value and Bayes Factors. These articles analyze in great detail historical arguments given by celebrated statisticians against the use of procedures based on highest density probability sets. Among those that opposed such ideas is Dennis Lindley, an influential figure at IME-USP and a personal friend of the first author. Finally, Ref. [86, 133] analyze historical desiderata for an acceptable Bayesian significance test that were formulated by the frequentist statistician Oscar Kempthorne to the first author, and show how the FBST successfully achieves all these desired characteristics.

- Reference [16] analyzes the composition of hypotheses defined in independent statistical models and the corresponding composition rules for *e*-values and truth functions.
- Reference [140] studies significance measures for evidence amalgamation and meta-analysis.
- References [46, 47, 51, 63, 131] analyze conditions of logical consistency for significance measures and test procedures for several hypotheses defined in the same statistical model. Conversely, these articles fully characterize some (agnostic or trivalent) generalizations of the FBST as the only statistical tests satisfying such logical consistency conditions.
- References [119–121, 123–127] develop the Objective Cognitive Constructivism as an epistemological framework formally compatible and semantically amenable to the *e*-value significance measure and the FBST hypothesis test.
- Reference [27] analyzes solutions to the problem of (statistical) induction, including Bayesian perspectives in general and the FBST in particular.
- References [59, 97, 117, 118, 129] apply concepts related to the FBST or the Objective Cognitive Constructivism epistemological framework to the study of economic or legal systems.
- References [128, 130] analyze the philosophical premises used by Karl Pearson to define the *p*-value and to establish the epistemological foundations of frequentist statistics; why Pearson's work and the subsequent work of Bruno de Finetti reversed previous commitments of Bayesian statistics; and how the FBST can be seen a way to reenter the path envisioned by the founding fathers of the Bayesian school, namely, Reverend Thomas Bayes, Richard Price and Pierre-Simon de Laplace.
- References [15, 50, 81, 89, 106, 121] analyze the role of randomization procedures in the context of the Objective Cognitive Constructivism epistemological framework in particular, and in Bayesian statistics in general.

5 Future research and final remarks

The FBST research program has grown and spread far and wide, in some directions suggested by these authors, and also in other directions that were for us completely unforeseen and wonderfully surprising. We are confident that this research program will continue to flourish and expand, exploring new areas of theory and application. The authors would like to suggest a few topics (focusing on theoretical and applied statistics) worthy of further attention as possible entry points for those interested (be all welcome) in joining this research program:

(1) In the context of information based medicine, see [31], it is important to compare and test the sensibility and specificity of alternative diagnostic tools, access the bio-equivalence of drugs coming from different suppliers, identify and test the efficacy of possible genetic markers for clinical conditions, etc. How to combine fast and computationally inexpensive heuristic algorithms and reliable statistical test procedures to best handle these and similar problems?

(2a) Influence diagrams are a powerful tool for decision modeling, see [9, 33]. Nevertheless, it is often hard to select optimal diagrams to model complex applications, see for example [41, 111]. How can the FBST best be used for sequential or concomitant inclusion/ exclusion of links or edge selection in influence graphs?

(2b) The aforementioned questions also arise in the context of Bayesian networks. In this context, it is important not only to test the significance of individual edges, but also to test the integrity of higher level sparsity structures, like the network click structure or its block factors, see [116, 121, 132, 134].

(3) The e-value and the FBST were originally developed for parametric models. How can the e-value be used, interpreted, computed (and maybe generalized) in semi-parametric or non-parametric settings? For instance, in models using functional bases, how can we test speeds of convergence for series expansions?

(4a) The compositionality rules established in [16] are based on functional operations over the truth functions, $W(v)$. [8, 71] present similar rules (for serial-parallel composition) in the context of reliability theory. Can these theories be seen as particular cases of more general and abstract logical formalisms?

(4b) The same compositionality rules assume independence between distinct models in a given structure. Could statistical copulas, see [52, 54, 68, 92], be used to successfully capture weak dependencies between distinct truth functions?

(5) The conditions for pragmatic acceptance of sharp hypotheses stated in [47] depend on consensual bounds for background uncertainties. For universal physical constants, metrologists establish such bounds by aggregating results of diverse experiments; similar situations occur in meta-analysis studies. Several statistical methods have been proposed to aggregate such diverse data-sets, see [26, 53, 72, 73, 90]. What are the best ways to coherently establish and represent aggregate uncertainty bounds in the FBST framework?

Acknowledgements The authors are grateful to IME-USP - the Institute of Mathematics and Statistics of the University of São Paulo, and INMA-UFMS - the Institute of Mathematics of the Federal University of Mato Grosso do Sul. The authors are extremely grateful for the support received from their colleagues, collaborators, users and critics in the construction works of this research project.

Funding This research was funded by CNPq - the Brazilian National Counsel of Technological and Scientific Development (Grants PQ 302767/2017-7, PQ 301892/ 2015-6); and FAPESP - the State of São Paulo Research Foundation (Grants CEPID Shell-RCGI 2014/ 50279-4, CEPID CeMEAI 2013/07375-0).

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest. The funders had no role in this study's data analyses, in its methodological developments or conclusions, in writing this manuscript, or in deciding where to publish.

References

1. Ainsbury, E.A., Vinnikov, V.A., Puig, P., Higueras, M., Maznyk, N.A., Lloyd, D.C., Rothkamm, K.: Review of Bayesian statistical analysis methods for cytogenetic radiation biodosimetry, with a practical example. Radiat. Prot. Dosim. **162**(3), 185–196 (2013)
2. Amari, S.I.: Methods of Information Geometry. American Mathematical Society, Providence (2007)
3. Andrade, P., Rifo, L.L.R., Torres, S., Torres-Avilés, F.: Bayesian inference on the memory parameter for gamma-modulated regression models. Entropy **17**(10), 6576–6597 (2015)
4. Andrade, P.D.M., Stern, J.M., Pereira, C.A.D.B.: Bayesian test of significance for conditional independence: the multinomial model. Entropy **16**(3), 1376–1395 (2014)
5. Angus, J.E.: The probability integral transform and related results. SIAM Rev. **36**(4), 652–654 (1994)
6. Assane, C.C., Pereira, B.B., Pereira, C.A.B.: Bayesian significance test for discriminating between survival distributions. Commun. Stat. Theory Methods **47**(24), 6095–6107 (2018)
7. Assane, C.C., Pereira, B.B., Pereira, C.A.B.: Model choice in separate families: a comparison between the FBST and the Cox test. Commun. Stat. Simul. Comput. **48**(9), 2641–2654 (2019)
8. Barlow, R.E., Prochan, F.: Statistical Theory of Reliability and Life Testing Probability. Models. Silver Spring, To Begin With (1981)
9. Barlow, R.E., Pereira, C.: Influence diagrams and decision modelling. In: Barlow, R.E., Clarotti, C.A., Spizzichino, F. (eds.) Reliability and Decision Making, pp. 87–99. Springer, Dordrecht (1993)
10. Barahona, M., Rifo, L., Sepúlveda, M., Torres, S.: A simulation-based study on Bayesian estimators for the skew Brownian motion. Entropy **18**(7), 241 (2016)
11. Basu, D., Ghosh, J.K.: Statistical Information and Likelihood. Lecture Notes in Statistics, 45, (1988)
12. Bernardo, J.M.: Reference analysis. In: Dey, D.K., Rao, C.R. (eds.) Bayesian Thinking: Modeling and Computation. Handbook of Statistics, vol. 25, pp. 17–90. Elsevier, Amsterdam (2005)
13. Berger, J.O., Wolpert, R.L.: The Likelihood Principle, 2nd edn. Inst of Mathematical Statistic, Hayward, CA (1988)
14. Bernardo, G., Lauretto, M.S., Stern, J.M.: The full Bayesian significance test for symmetry in contingency tables. AIP Conf. Proc. **1443**, 198–205 (2012)
15. Bonassi, F.V., Nishimura, R., Stern, R.B.: In defense of randomization: a subjectivist Bayesian approach. AIP Conf. Proc. **1193**, 32–39 (2009)
16. Borges, W., Stern, J.M.: The rules of logic composition for the Bayesian epistemic E-values. Logic J. IGPL **15**(5/6), 401–420 (2007)
17. Box, G.E., Tiao, G.C.: Bayesian Inference in Statistical Analysis. Addison-Wesley, London (1973)
18. Brentani, H., Nakano, E.Y., Martins, C.B., Izbicki, R., Pereira, C.A.: Disequilibrium coefficient: a Bayesian perspective. Stat. Appl. Genet. Mol. Biol. **10**(1), 22 (2011)
19. Cabras, S., Racugno, W., Ventura, L.: Higher order asymptotic computation of Bayesian significance tests for precise null hypotheses in the presence of nuisance parameters. J. Stat. Comput. Simul. **85**(15), 2989–3001 (2015)
20. Cantinha, R.S., Borrely, S.I., Oguiura, N., Pereira, C.A.B., Rigolona, M.M., Nakano, E.: HSP70 expression in Biomphalaria glabrata snails exposed to cadmium. Ecotoxicol. Environ. Saf. **140**, 18–23 (2017)
21. Camargo, A.P., Stern, J.M., Lauretto, M.S.: Estimation and model selection in Dirichlet regression. AIP Conf. Proc. **1443**, 206–213 (2012)
22. Cerezetti, F.V., Stern, J.M.: Non-arbitrage in financial markets: a Bayesian approach for verification. AIP Conf. Proc. **1490**, 87–96 (2012)
23. Chakrabarty, D.: A new Bayesian test to test for the intractability-countering hypothesis. J. Am. Stat. Assoc. **112**(518), 561–577 (2017)
24. Chen, C.W.S., Lee, S.: A local unit root test in mean for financial time series. J. Stat. Comput. Simul. **86**(4), 788–806 (2015)
25. Chaiboonsri, C., Wannapan, S., Saosaovaphak, A.: Economic and business cycle of India: evidence from ICT Sector. In: Tsounis, N., Vlachvei, A. (eds.) Advances in Panel Data Analysis in Applied Economic Research, pp. 29–43. Springer Nature, Cham, Switzerland (2018)

26. Cohen, E.R., Crowe, K.M., Dumond, Jesse W. M.: The Fundamental Constants of Physics. NY, Interscience, (1957)
27. Cristofaro, R.: The analytical solution to the problem of statistical induction. Statistica **63**(2), 411–423 (2003)
28. D'Cunha, J.G., Rao, A.K.: Frequentist comparison of the Bayesian significance test for testing the median of the lognormal distribution. InterStat, **02**(001), 1–25 (2016)
29. de Bernardini, D.F., Rifo, L.L.R.: Full Bayesian significance test for extremal distributions. J. Appl. Stat. **38**(4), 851–863 (2011)
30. de Bragança Pereira, C.A.: Full Bayesian significant test (FBST). In: Lovric M. (ed.) International Encyclopedia of Statistical Science, pp. 551–554. Springer, Berlin (2011)
31. de Bragança Pereira, B., de Bragança Pereira, C.A.: A likelihood approach to diagnostic tests in clinical medicine. RevStat Stat. J. **3**(1), 77–98 (2005)
32. de Bragança Pereira, B., de Bragança Pereira, C.A.: Model Choice in Nonnested Families. Springer, Berlin (2016)
33. de Bragança Pereira, C.A., Barlow, R.E.: Medical diagnosis using influence diagrams. Networks **20**(5), 565–577 (1990)
34. de Bragança Pereira, C.A., Stern, J.M.: Evidence and credibility. Full Bayesian significance test for precise hypotheses. Entropy **1**, 99–110 (1999)
35. de Bragança Pereira, C.A., Stern, J.M.: A dynamic software certification and verification procedure. ISAS-SCI'99 Proc. **2**, 426–435 (1999)
36. de Bragança Pereira, C.A., Stern, J.M.: Model selection and regularization: full Bayesian approach. Environmetrics **12**(6), 559–568 (2001)
37. de Bragança Pereira, C.A., Stern, J.M., Wechsler, S.: Can a significance test be genuinely Bayesian? Bayesian Anal. **3**, 79–100 (2008)
38. del Rincón, S.V., Rogers, J., Widschwendter, M., Sun, D., Sieburg, H.B., Spruck, C.: Development and validation of a method for profiling post-translational modification activities using protein microarrays. PLoS ONE **5**(6), e11332 (2010)
39. de Bragança Pereira, C.A., Stern, J.M.: Special characterizations of standard discrete models. RevStat Stat. J. **6**, 199–230 (2008)
40. de Finetti, B.: Theory of Probability. Wiley, New York
41. de Mathis, M.A., do Rosario, M.C., Diniz, J.B., Torres, A.R., Shavitt, R.G., Ferrão, Y.A., Fossaluza, V., Pereira, C., Miguel, E.C.: Obsessive-compulsive disorder: influence of age at onset on comorbidity patterns. Eur. Psychiatry, 23(3): 187-194, (2008)
42. Diniz, M., Pereira, C.A.B., Stern, J.M.: Cointegration: Bayesian significance test. Commun. Stat. Theory Methods **41**(19), 3562–3574 (2012)
43. Diniz, M., Pereira, C., Stern, J.M.: Unit roots: Bayesian significance test. Commun. Stat. Theory Methods **40**(23), 4200–4213 (2012)
44. Diniz, M., Pereira, C.A.B., Polpo, A., Stern, J.M., Wechsler, S.: Relationship between Bayesian and Frequentist Significance Indices. Int. J. Uncertainty Quantif. **2**(2), 161–172 (2012)
45. Dubins, L., Savage, L.J.: How to Gamble If You Must. McGraw-Hill, Inequalities for Stochastic Processes. Dover Publications, New York (1965)
46. Esteves, L.G., Izbicki, R., Stern, J.M., Stern, R.B.: The logical consistency of simultaneous agnostic hypothesis tests. Entropy **18**, 256 (2016)
47. Esteves, L.G., Izbicki, R., Stern, J.M., Stern, R.B.: Pragmatic hypotheses in the evolution of science. Entropy **21**(9), 883 (2019)
48. Evans, M.: Bayesian inference procedures derived via the concept of relative surprise. Commun. Stat. **26**, 1125–1143 (1997)
49. Fang, S.C., Rajasekera, J.R., Jacob, H.S.: Entropy Optimization and Mathematical Programming. Kluwer, Dordrecht, Tsao (1997)
50. Fossaluza, V., Lauretto, M.S., Pereira, C.A.B., Stern, J.M.: Combining optimization and randomization approaches for the design of clinical trials. Springer Proc. Math. Stat. **118**, 173–184 (2015)
51. Fossaluza, V., Izbicki, R., Silva, G.M., Esteves, L.G.: Coherent hypothesis testing. Am. Stat. **71**(3), 242–248 (2017)
52. Fossaluza, V., Esteves, L.G., Pereira, C.: Estimating multivariate discrete distributions using Bernstein copulas. Entropy **20**(3), 194 (2018)
53. Garcia, M.V.P., Humes, C., Stern, J.M.: Generalized line criterion for Gauss–Seidel method. Comput. Appl. Math. **22**, 91–97 (2003)

54. García, J.E., González-López, V., Nelsen, R.B.: The structure of the class of maximum Tsallis-Havrda-Chavat entropy copulas. Entropy **18**(7), 264 (2016)
55. Gelman, A., Carlin, J.B., Stern, H.S., Donald, B.: Bayesian Data Analysis. Chapman-Hall/ CRC, Rubin (2004)
56. Good, I.J.: Good Thinking. Univ. of Minnesota (1983)
57. Hubert, P., Lauretto, M.S., Stern, J.M.: FBST for generalized Poisson distribution. AIP Conf. Proc. **1193**, 210–2019 (2009)
58. Hubert, P., Padovese, L., Stern, J.M.: A sequential algorithm for signal segmentation. Entropy **20**(1), 55 (2018)
59. Hubert, P., Stern, J.M.: Probabilistic equilibrium: a review on the application of MAXENT to macroeconomic models. Springer Proc. Math. Stat. **239**, 187–197 (2018)
60. Hubert, P., Killick, R., Chung, A., Padovese, L.R.: A Bayesian binary algorithm for root mean squared-based acoustic signal segmentation. J. Acoust. Soc. Am. **146**(3), 1799–1807 (2019)
61. Irony, T.Z., Lauretto, M., Pereira, C., Stern, J.M.: A Weibull wearout test: full Bayesian approach. In: Hayakawa, Y., Irony, T., Xie, M. (eds.) Systems and Bayesian Reliability, pp. 287–300. World Scientific, Singapore (2002)
62. Izbicki, R., Fossaluza, V., Hounie, A.G., Nakano, E.Y., Pereira, C.A.B.: Testing allele homogeneity: the problem of nested hypotheses. BMC Genet. **13**(103), 1–11 (2012)
63. Izbicki, R., Esteves, L.G.: Logical consistency in simultaneous statistical test procedures. Logic J. IGPL **23**, 732–758 (2015)
64. Jeffreys, H.: Theory of Probability, 1st edn. Clarendon Press, Oxford (1939)
65. Johnson, R., Chakrabarty, D., O'Sullivan, E., Raychaudhury, S.: Comparing x-ray and dynamical mass profiles in the early-type galaxy NGC 4636. Astrophys. J. **706**(2), 980–994 (2009)
66. Kadane, J.B.: Principles of Uncertainty. Chapman-Hall/CRC, New York (2011)
67. Kadane, J.B.: Pragmatics of Uncertainty. Chapman-Hall/CRC, New York (2016)
68. Kaplan, S., Lin, J.C.: An improved condensation procedure in discrete probability distribution calculations. Risk Anal. **7**, 15–19 (1987)
69. Kapur, J.N.: Maximum Entropy Models in Science and Engineering. John Wiley, New Delhi (1989)
70. Kapur, J.N., Kesavan, H.K.: Entropy Optimization Principles with Applications. Academic Press, Boston (1992)
71. Kaufmann, A., Grouchko, D., Cruon, R.: Mathematical Models for the Study of the Reliability of Systems. Academic Press, New York (1977)
72. Kelley, C.T.: Iterative Methods for Linear and Nonlinear Equations. SIAM, Philadelphia (1987)
73. Kelley, C.T.: Iterative Methods for Optimization. SIAM, Philadelphia (1987)
74. Kelter, R.: Analysis of Bayesian posterior significance and effect size indices for the two-sample t-test to support reproducible medical research. BMC Med. Res. Methodol. **20**(88), 1–18 (2020)
75. Kostrzewski, M.: On the existence of jumps in financial time series. Acta Phys. Polonica B **43**(10), 2001–2019 (2012)
76. Lauretto, M.S., Pereira, C.A.B., Stern, J.M., Zacks, S.: Full Bayesian signicance test applied to multivariate normal structure models. Braz. J. Probab. Stat. **17**, 147–168 (2003)
77. Lauretto, M.S., Faria, S., Pereira, B.B., Pereira, C.A.B., Stern, J.M.: The problem of separate hypotheses via mixtures models. AIP Conf. Proc. **954**, 268–275 (2007)
78. Lauretto, M.S., Julio Michael, S.: FBST for mixture model selection. AIP Conf. Proc. **803**, 121–128 (2005)
79. Lauretto, M.S., Stern, J.M.: Testing signicance in Bayesian classifiers. Front. Artif. Intell. Appl. **132**, 34–41 (2005)
80. Lauretto, M.S., Nakano, F., Faria, S., Pereira, C.A.B., Stern, J.M.: A straightforward multiallelic signicance test for the Hardy–Weinberg equilibrium law. Genet. Mol. Biol. **32**(3), 619–625 (2009)
81. Lauretto, M.S., Nakano, F., Pereira, C.A.B., Stern, J.M.: Intentional sampling by goal optimization with decoupling by stochastic perturbation. AIP Conf. Proc. **1490**, 189–201 (2014)
82. Lima, A.R., Mello, M.F., Andreoli, S.B., Fossaluza, V., Araújo, C.M. de., Jackowski, A.P., Bressan, R.A., Mari, J.J.: The impact of healthy parenting as a protective factor for posttraumatic stress disorder in adulthood: a case-control study. PLOS ONE. **9**(1), 1–9 (2014)
83. Loschi, R.H., Monteiro, J.V.D., Rocha, G.H.M.A., Mayrink, V.D.: Testing and estimating the nondisjunction fraction in Meiosis I using reference priors. Biom. J. **49**(6), 824–839 (2007)
84. Loschi, R.H., Santos, C.C., Arellano-Valle, R.B.: Test procedures based on combination of Bayesian evidences for H0. Braz. J. Probab. Stat. **26**(4), 450–473 (2012)

85. Madruga, M.R., Esteves, L.G., Wechsler, S.: On the Bayesianity of Pereira–Stern tests. Test **10**, 291–299 (2001)
86. Madruga, M.R., Pereira, C.A.B., Stern, J.M.: Bayesian evidence test for precise hypotheses. J. Stat. Plan. Inference **117**, 185–198 (2003)
87. Montoya-Delgado, L.E., Irony, T.Z., Pereira, C.A.D.B., Whittle, M.R.: An unconditional exact test for the Hardy-Weinberg equilibrium law: sample-space ordering using the Bayes factor. Genetics **158**(2), 875–883 (2001)
88. Maranhao, V.L., Lauretto, M.S., Stern, J.M.: FBST for covariance structures of generalized Gompertz models. AIP Conf. Proc. **1490**, 202–211 (2012)
89. Marcondes, D., Peixoto, P., Stern, J.M.: Assessing randomness in case assignment: the case study of the brazilian supreme court. Law, Probability and Risk **18**(2–3), 97–114 (2019)
90. Minka, T.: Divergence measures and message passing. Technical report MSR-TR-2005-173, Microsoft Research Ltd., Cambridge, UK (2005)
91. Nakano, F., Pereira, C.A.B., Stern, J.M., Whittle, M.R.: Genuine Bayesian multiallelic signicance test for the Hardy–Weinberg equilibrium law. Genet. Mol. Res. **4**, 619–631 (2006)
92. Nelsen, R.B.: An Introduction to Copulas, 2nd edn. Springer, New York (2006)
93. Oliveira, N.L., Pereira, C.A.B., Diniz, M.A., Polpo, A.: A discussion on significance indices for contingency tables under small sample sizes. PLoS ONE **13**(8), 1–19 (2018)
94. Patriota, A.G.: A classical measure of evidence for general null hypotheses. Fuzzy Sets Syst. **233**, 74–88 (2013)
95. Patriota, A.G.: On some assumptions of the null hypothesis statistical testing. Educ. Psychol. Meas. **77**(3), 507–528 (2017)
96. Pawitan, Y.: In All Likelihood: Statistical Modelling and Inference Using Likelihood. Oxford University Press, Oxford (2001)
97. Pigliucci, M., Boudry, M.: Prove it! The Burden of Proof Game in Science vs Pseudoscience Disputes. Philosophia **42**(2), 487–502 (2014)
98. Pinto, A., Ventura, L.: Approssimazioni Asintotiche di Ordine Elevato per Verifiche d'Ipotesi Bayesiani: Uno Studio per Dati di Sobrevvivenza. Università degli Studi di Padova, Dipartimento di Scienze Statistiche (2012)
99. Ranzato, G., Ventura, L.: Biostatistica Bayesiana con "Matching Priors". Università degli Studi di Padova, Dipartimento di Scienze Statistiche (2018)
100. Rifo, L.L.R., Torres, S.: Full Bayesian analysis for a class of jump-diffusion models. Commun. Stat. Theory Methods **38**(8), 1262–1271 (2009)
101. Rifo, L.L.R., González-López, V.: Full Bayesian analysis for a model of tail dependence. Commun. Stat. Theory Methods **41**(22), 4107–4123 (2012)
102. Rodrigues, J.: Full Bayesian significance test for zero-inflated distributions. J. Commun. Stat. Theory Methods **35**(2), 299–307 (2006)
103. Royall, R.: Statistical Evidence: A Likelihood Paradigm. Chapman and Hall, London (1997)
104. Ruli, E., Sartori, N., Ventura, L.: Robust approximate Bayesian inference. J. Stat. Plan. Inference **205**, 10–22 (2020)
105. Ruli, E., Ventura, L.: Can Bayesian, confidence distribution and frequentist inference agree? Stat. Methods Appl. https://doi.org/10.1007/s10260-020-00520-y
106. Saa, O., Stern, J.M.: Auditable blockchain randomization tool. Proceedings, **33**(1), 17.1–17.6 (2019)
107. Santos, N.C.L., Dias, R., Alvesc, D., Melo, B.G.M., Gomes, L., Severid, W., Agostinho, A.: Trophic and limnological changes in highly fragmented rivers predict the decreasing abundance of detritivorous fish. Ecol. Ind. **110**, 105933 (2020). MD, PhD, PhD, PhD, and Euripedes Constantino Miguel
108. Seixas, A.A.A., Hounie, A.G., Fossaluza, V., Curi, M., Alvarenga, P.G., de Mathis, M.A., Vallada, H., Pauls, D., Pereira, Carlos Alberto de Bragança., Miguel, E.C.: Anxiety disorders and rheumatic fever: is there an association? CNS Spectrums, **13**(12), 1039–1046 (2008)
109. Shackle, G.L.S.: Uncertainty in Economics and Other Reflections. Cambridge Univ. Press, London (1968)
110. Shackle, G.L.S.: Decision, Order and Time in Human Affairs. Cambridge Univ. Press, London (1969)
111. Shavitt, R.G., Requena, G., Alonso, P., Zai, G., Costa, D.L.C., Pereira, C., Rosário, M.C., Morais, I., Fontenelle, L., Cappi, C., Kennedy, J., Menchon, J.M., Miguel, E., Richter, P.M.A.: Quantifying

dimensional severity of obsessive-compulsive disorder for neurobiological research. Prog. Neuro-Psychopharmacology Biol. Psychiatry, **79**, 206–212 (2017)
112. Silva, G.M., Esteves, L.G., Fossaluza, V., Izbicki, R., Wechsler, S.: A Bayesian decision-theoretic approach to logically-consistent hypothesis testing. Entropy **17**(10), 6534–6559 (2015)
113. Silva, I.R.: On the correspondence between frequentist and Bayesian tests. Commun. Stat. Theory Methods **47**(14), 3477–3487 (2018)
114. Sikov, A., Stern, J.M.: Application of the full Bayesian significance test to model selection under informative sampling. Stat. Papers **60**, 89–104 (2019)
115. Spektor, M.S., Gluth, S., Fontanesi, L., Rieskamp, J.: How similarity between choice options affects decisions from experience: the accentuation-of-differences model. Psychol. Rev. **126**(1), 52–88 (2019)
116. Stern, J.M.: Simulated annealing with a temperature dependent penalty function. ORSA J. Comput. **4**, 311–319
117. Stern, J.M.: Signicance tests, belief calculi, and burden of proof in legal and scientic discourse. Front. Artif. Intell. Appl. **101**, 139–147 (2003)
118. Stern, J.M.: Paraconsistent sensitivity analysis for Bayesian significance tests. Lecture Notes in Artificial Intelligence **3171**, 134–143 (2004)
119. Stern, J.M.: Cognitive constructivism, Eigen-solutions, and sharp statistical hypotheses. Cybern. Hum. Knowing **14**(1), 9–36 (2007)
120. Stern, J.M.: Language and the self-reference paradox. Cybern. Hum. Knowing **14**(4), 71–92 (2007)
121. Stern, J.M.: Decoupling, sparsity, randomization, and objective Bayesian inference. Cybern. Hum. Know. **15**(2), 49–68 (2008)
122. Stern, J.M.: Cognitive constructivism and the epistemic significance of sharp statistical hypotheses in natural sciences. arXiv:1006.5471.Tutorial text for MaxEnt 2008 - The 28th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Boracéia, São Paulo, Brazil, July 6-11, (2008)
123. Stern, J.M.: Symmetry, invariance and ontology in physics and statistics. Symmetry **3**(3), 611–635 (2011)
124. Stern, J.M.: Constructive verification, empirical induction, and falibilist deduction: a threefold contrast. Information **2**, 635–650 (2011)
125. Stern, J.M.: Jacob's ladder and scientific ontologies. Cybern. Hum. Knowing **21**(3), 9–43 (2014)
126. Stern, J.M.: Cognitive-constructivism, quine, dogmas of empiricism, and Münchhausen's Trilemma. In: Adriano, P., Francisco, L., Laura L.R. Rifo, J.M.S., Marcelo L. (eds.). Interdisciplinary Bayesian Statistics: EBEB 2014, Springer Proceedings in Mathematics and Statistics, 118, 55-68, (2015)
127. Stern, J.M.: Puzzles: continuous versions of Haack's Equilibria, Eigen-states and ontologies. Logic J. IGPL **25**(4), 604–631 (2017)
128. Stern, J.M.: Jacob's Ladder: Logics of Magic, Metaphor and Metaphysics: Narratives of the Unconscious, the Self, and the Assembly. Sophia, published Online First June 7, (2017)
129. Stern, J.M.: Verstehen (causal/interpretative understanding), Erklären (law-governed description/prediction), and Empirical Legal Studies. J. Inst. Theor. Econ. **174**, 105–114 (2018)
130. Stern, J.M.: Karl Pearson on causes and inverse probabilities: renouncing the bride, inverted spinozism and goodness-of-fit. South Am. J. Logic **4**(1), 219–252 (2018)
131. Stern, J.M., Izbicki, R., Esteves, L.G., Stern, R.B.: Logically-consistent hypothesis testing and the hexagon of oppositions. Logic J. IGPL **25**, 741–757 (2018)
132. Stern, J.M., Colla, E.C.: Factorization of Bayesian networks. In: Nakamatsu, K., Phillips-Wren, G., Jain, L.C., Howlett, R.J. (eds). New Advances in Intelligent Decision Technologies, pp. 275–294. Heidelberg, Springer (2009)
133. Stern, J.M., Pereira, C.A.B.: Bayesian epistemic values. Focus on surprise: measure probability!. Logic J. IGPL **22**, 236–254 (2014)
134. Stern, J.M., Vavasis, S.A.: Active set methods for problems in column block angular form. Comput. Appl. Math. **12**, 199–226 (1994)
135. Stern, J.M., Zacks, S.: Testing the independence of Poisson variates under the Holgate bivariate distribution: the power of a new evidence test. Stat. Probab. Lett. **60**, 313–320 (2002)
136. Thulin, M.: Decision-theoretic justifications for Bayesian hypothesis testing using credible sets. J. Stat. Plan. Inference **146**, 133–138 (2014)
137. Ventura, L., Ruli, E., Racugno, W.: A note on approximate Bayesian credible sets based on modified loglikelihood ratios. Stat. Probab. Lett. **83**(11), 2467–2472 (2013)

138. Ventura, L., Reid, N.: METRON **72**, 231–245 (2014)
139. Ventura, L., Racugno, W.: Pseudo-likelihoods for Bayesian inference. In: DiBattista, T., Moreno, E., Racugno, W. (eds). Topics on Methodological and Applied Statistical Inference, pp. 205–220. Berlin, Springer (2016)
140. Vieland, V.J., Chang, H.: No evidence amalgamation without evidence measurement. Synthese **196**, 3139–3161 (2019)
141. Vikas, K., Rao, A.K.: Full Bayesian empirical likelihood significance test for equality of medians. InterStat, 2016, 01, 001,1-9 (2016)
142. Vosseler, A., Weber, E.: Bayesian analysis of periodic unit roots in the presence of a break. Appl. Econ. **49**(38), 3841–3862 (2016)
143. Wechsler, S., Pereira, C.A.B., Marques, P.C.: Birnbaum's theorem redux. AIP Conf. Proc. **1073**, 96–100 (2008)
144. Wittenburg, D.T.F., Klosa, J., Reinsch, N.: Covariance between genotypic effects and its use for genomic inference in half-sib families. G3 Genes Genomes Genet. **6**(9), 2761–2772 (2016)
145. Zellner, A.: Introduction to Bayesian Inference in Econometrics. Wiley, New York (1971)