



ELSEVIER

Available at
WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE @ DIRECT®Journal of Statistical Planning and
Inference 117 (2003) 185–198journal of
statistical planning
and inferencewww.elsevier.com/locate/jspi

Bayesian evidence test for precise hypotheses

M.R. Madruga, C.A.B. Pereira, J.M. Stern*

Mathematics and Statistics Institute, University of São Paulo, São Paulo, Brazil

Received 21 July 2000; received in revised form 15 May 2002; accepted 25 June 2002

Abstract

The full Bayesian significance test (FBST) for precise hypotheses is presented, with some illustrative applications. In the FBST we compute the evidence against the precise hypothesis. We discuss some of the theoretical properties of the FBST, and provide an invariant formulation for coordinate transformations, provided a reference density has been established. This evidence is the probability of the highest relative surprise set, “tangential” to the sub-manifold (of the parameter space) that defines the null hypothesis.

© 2002 Elsevier B.V. All rights reserved.

MSC: 62A15; 62F15; 62H15; 65C60

Keywords: Evidence; Full Bayesian significance test; Invariant formulation; Reference density; Relative surprise; Uninformative prior

1. Introduction

The full Bayesian significance test (FBST) is presented in [Pereira and Stern \(1999b\)](#) as a coherent Bayesian significance test. The FBST is intuitive and has a geometric characterization. It can be implemented using modern numerical optimization and integration techniques. The method is “fully” Bayesian and consists in the analysis of credible sets. By fully we mean that we need only the knowledge of the parameter space represented by its posterior distribution. The FBST needs no additional assumption, like a positive prior probability of the precise hypothesis, that leads to the Lindley’s paradox ([Lindley, 1957](#)). Like all Bayesian analysis, the FBST regards likelihoods as the proper means for representing statistical information, a principle stated by [Basu \(1988\)](#),

* Corresponding author.

E-mail address: jstern@ime.usp.br (J.M. Stern).

Birnbaum (1972), Finetti (1974, 1981/1991, 1993), Good (1983), Kempthorne (1976, 1980), Royall (1997), and others, to simplify and unify statistical analysis. Another important aspect of the FBST is its consistency with the “Onus Probandi” juridical principle (Gaskins, 1992; Kokott, 1998).

The FBST formulation presented in this paper provides explicit invariance under general coordinate transformations of the parameter space, provided a reference density has been established. The FBST is illustrated with some examples.

As shown in Madruga et al. (2001), the FBST is also in harmony with Bayesian decision theory of Rubin (1987), in the sense that there are specific loss functions which render the FBST decision theoretic aspects. Although, we can cast the FBST in a decision theoretic framework, it was originally defined in a pure operational form, based only on the Onus Probandi juridical principle, Pereira and Stern (1999b). Compliance with this juridical principle, also known as Benefit of the Doubt, Presumption of Innocence or (in accounting) Safe Harbor Liability Rule, was imperative in some of our consulting projects, Pereira and Stern (1999a). This kind of principle establishes that:

“There is no liability as long as there is a reasonable basis for belief, effectively placing the burden of proof (Onus Probandi) on the plaintiff, who, in a lawsuit, must prove false a defendant’s misstatement, without making any assumption not explicitly stated by the defendant, or tacitly implied by existing law or regulation”. Accordingly, a working hypothesis is not rejected if there is not sufficient evidence against it.

Interesting connections of some of the characteristics stated above, with ethics, epistemology, law, psychology and statistics can be found in Carnap (1962), Box and Tiao (1973), Cox (1977), DeGroot (1970), Finetti (1974, 1981/1991), Gaskins (1992), Kokott (1998), Lauretto et al. (2002), Lindley (1957, 1978), Madruga et al. (2001), Pereira and Wechsler (1993), Pereira and Stern (1999a, b), Popper (1989), Savage (1962), Sellke et al. (1999), and Stern and Zacks (2002).

2. Computing the evidence against H_0

Let X_1, \dots, X_n be random variables having a joint density $f(x; \theta)$, with respect to a σ -finite measure μ . θ is a parameter vector in a parameter space $\Theta \subseteq \mathbb{R}^p$ ($p \geq 1$). We are interested in a precise null hypothesis $H_0: \theta \in \Theta_0$, $\Theta_0 \subset \Theta$, and $\dim(\Theta_0) < \dim(\Theta)$.

Let $L(\theta; x)$ denote the likelihood function of θ on Θ . Let $p(\theta)$ be a prior density on Θ , and $r(\theta)$ a reference density on Θ . We denote by $p_n(\theta)$ the posterior density of θ on Θ , i.e.

$$p_n(\theta) \propto L(\theta; x)p(\theta), \quad x = [x_1, \dots, x_n], \quad \theta = [\theta_1, \dots, \theta_p]$$

and define

$$\theta^* = \arg \max_{\theta \in \Theta_0} \left\{ \frac{p_n(\theta)}{r(\theta)} \right\}, \quad s_n^* = \max_{\theta \in \Theta_0} \left\{ \frac{p_n(\theta)}{r(\theta)} \right\} = \left\{ \frac{p_n(\theta^*)}{r(\theta^*)} \right\}.$$

The function $s_n(\theta) = p_n(\theta)/r(\theta)$ is called the “relative surprise” (Good, 1983). We define now, in the space Θ , the highest relative surprise set (HRSS) Θ_n^* of points

$\theta \in \Theta$ with higher relative surprise $s(\theta)$ than any point in Θ_0 , i.e.,

$$\Theta_n^* = \left\{ \theta \in \Theta \mid \frac{p_n(\theta)}{r(\theta)} \geq s_n^* \right\}.$$

Notice that the set Θ_n^* is “tangential” to Θ_0 at Θ_n^* . The evidence against H_0 , given by the sample data x , is defined as the posterior probability of the tangential HRSS, i.e.

$$Ev_n = \int_{\Theta_n^*} p_n(\theta) d\theta.$$

This definition of the evidence against H_0 is invariant with respect to a proper reparameterization. For instance, let $\omega = \phi(\theta)$, where ϕ is a measurable and integrable function. For the purpose of illustration, assume that ϕ is bijective (one-to-one) and continuously differentiable. Let $J(\omega)$ denote the Jacobian of the transformation, i.e.,

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

The posterior density of ω , given x , is

$$\tilde{p}_n(\omega) = p_n(\phi^{-1}(\omega))|J(\omega)|.$$

Notice that the reference density under the reparameterization changes to

$$\tilde{r}(\omega) = r(\phi^{-1}(\omega))|J(\omega)|.$$

Thus, the new surprise function is

$$\tilde{s}_n(\omega) = \tilde{p}_n(\omega)/\tilde{r}(\omega) = p_n(\phi^{-1}(\omega))/r(\phi^{-1}(\omega)).$$

Let $\Omega_0 = \phi(\Theta_0)$. It follows that

$$\tilde{s}_n^* = \sup_{\omega \in \Omega_0} \tilde{s}_n(\omega) = \sup_{\theta \in \Theta_0} s_n(\theta) = s_n^*.$$

Accordingly, $\Theta_n^* \mapsto \phi(\Theta_n^*) = \Omega_n^*$.

The evidence under reparameterization is

$$\tilde{E}v_n = \int_{\Omega_n^*} \tilde{p}_n(\omega) d\omega = \int_{\Theta_n^*} p_n(\theta) d\theta = Ev_n.$$

This proves invariance by proper reparameterizations of Ev_n .

Remark. (1) The original definition of the evidence against H_0 , Pereira and Stern (1999a, b), did not employ the reference density $r(\theta)$. In the former definition, the “tangential” set Θ_n^* was the highest probability density set (HPDS) (whose points have posterior density $p_n(\theta)$ greater than that of any point in Θ_0), instead of the HRSS.

The evidence in that former definition is the credibility of the set Θ_n^* . If this evidence is sufficiently high, it is customary to reject H_0 . The former definition of evidence is not invariant under reparameterization, as can be shown by various examples.

Taking the reference density as the (possibly improper) uniform density, $r(\theta) = U(\theta)$, the former and present definitions of evidence define the same tangential set, i.e. the HRSS and the HPDS coincide. In a proper reparameterization $\omega = \phi(\theta)$, using the present definition, we are just automatically mapping to the new coordinates the tangential set computed in the original coordinates, $\Omega_n^* = \phi(\Theta_n^*)$.

(2) We can generalize the procedure using other reference densities. For example, we may use as reference density the uninformative prior (also known as neutral or reference prior), if one is available. This possibility is suggested by the paper of Evans (1997), in conjunction with Jeffreys' rules to obtain uninformative priors, Zellner (1971, appendix to Chapter 2).

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

Jeffreys suggests $\psi = \log(\sigma)$ as a suitable transformation for a parameter $\sigma \in]0, \infty[$. Using $d\psi = d\sigma/\sigma$ and assuming the uniform reference prior on ψ , we obtain the reference prior for σ , $r(\sigma) \propto 1/\sigma$. This transformation also has an interesting property of being invariant under transformations of the form $\rho = \sigma^n$, i.e., $r(\rho) \propto 1/\rho$. We will use this prior in following examples.

In order to be consistent with the Onus Probandi principle, we will generally choose as reference density on Θ , the uniform density or a non-informative prior which yields a proper posterior density $p_n(\theta)$. In the examples in the next sections, we often use the uniform reference density. It is possible to use other reference densities, although doing so may impair the adherence to the Onus Probandi principle, or change its interpretation.

(3) The determination of the "tangential" set Θ_n^* might have to be done numerically, since analytic solutions might not be available, see Stern and Zacks (2002). Efficient numerical methods for optimization (finding θ^*) and integration are readily available. The reader is referred to Gentle (2000), Liu (2001), Luenberger (1984), Ljung et al. (1992), McCormick (1983), Ökten (1999), Pflug (1999), Spall (2000). In addition, the evidence integral could be estimated by Monte Carlo techniques, as shown in Stern and Zacks (2002). The final computer implementation makes use of user friendly, interactive and extensible environments, like Matlab, or the open source softwares Scilab, and Python.

(4) We notice that the FBST is used in full dimensionality of the parameter space. In the way it is defined and to preserve all its properties, elimination of "nuisance" parameters is not recommended.

(5) As shown by Madruga et al. (2001), one can define a loss function with respect to which the optimal Bayesian decision is to reject H_0 if Ev_n is greater than a critical level $0 < \lambda < 1$.

In the next section we present several examples which illustrate the use of the methodology and its potential.

3. Testing the normal mean

3.1. Variance known

Let X_1, \dots, X_n be i.i.d. random variables, having a common normal distribution, i.e., $X_i \sim N(\mu, 1)$, $-\infty < \mu < \infty$. The null hypothesis is $H_0: \mu = \mu_0$. The minimal sufficient statistics is \bar{x} , $\bar{x} = (1/n) \sum_{i=1}^n x_i$. The likelihood function is

$$L(\mu; n, \bar{x}) = \exp(-(n/2)(\mu - \bar{x})^2).$$

We choose the uniform improper prior, i.e. $p(\mu)d\mu = d\mu$, and $r(\mu) = 1$. The posterior density is

$$p_n(\mu) = \sqrt{n/2\pi} \exp(-(n/2)(\mu - \bar{x})^2),$$

$\Theta_0 = \{\mu_0\}$ and $s_n^* = p_n(\mu_0)$. The “tangential” set is

$$\Theta_n^* = \begin{cases} [\mu_0, \mu_0 + 2\bar{x}] & \text{if } \bar{x} > \mu_0, \\ [\mu_0 - 2\bar{x}, \mu_0] & \text{if } \bar{x} < \mu_0. \end{cases}$$

The evidence against H_0 is

$$Ev_n = 2\Phi(\sqrt{n}|\bar{x} - \mu_0|) - 1,$$

where $\Phi(\cdot)$ is the standard normal integral. Notice that if $E(x) = \mu_0$ then, by the strong law of large numbers, $\lim_{n \rightarrow \infty} Ev_n = 0$ a.s. $[\mu_0]$. On the other hand, if $E(x) = \mu \neq \mu_0$, then $\lim_{n \rightarrow \infty} Ev_n = 1$ a.s. $[\mu]$. This is an example of a general result.

3.2. Variance unknown

Let $\rho = 1/\sigma^2$, and let X_1, \dots, X_n be i.i.d. $N(\mu, 1/\rho)$. The minimal sufficient statistic is (\bar{x}, Q) , where $Q = \sum_{i=1}^n (x_i - \bar{x})^2$. The likelihood function is, for $-\infty < \mu < \infty$, $0 < \rho < \infty$,

$$L(\mu, \rho; n, \bar{x}, Q) \propto \rho^{n/2} \exp\left(-\rho Q/2 - \frac{n}{2} \rho (\mu - \bar{x})^2\right).$$

Taking the non-informative prior $p(\mu, \rho) = (1/\rho) d\mu d\rho$, we obtain the posterior density

$$p_n(\mu, \rho; n, \bar{x}, Q) = c \rho^{n/2-1} \exp\left(-\rho \frac{Q}{2} \left(1 + \frac{n}{Q} (\mu - \bar{x})^2\right)\right),$$

$$c = \frac{2^{n/2} \sqrt{n}}{Q^{(n+1)/2} \Gamma(\frac{1}{2}) \Gamma((n-1)/2)}.$$

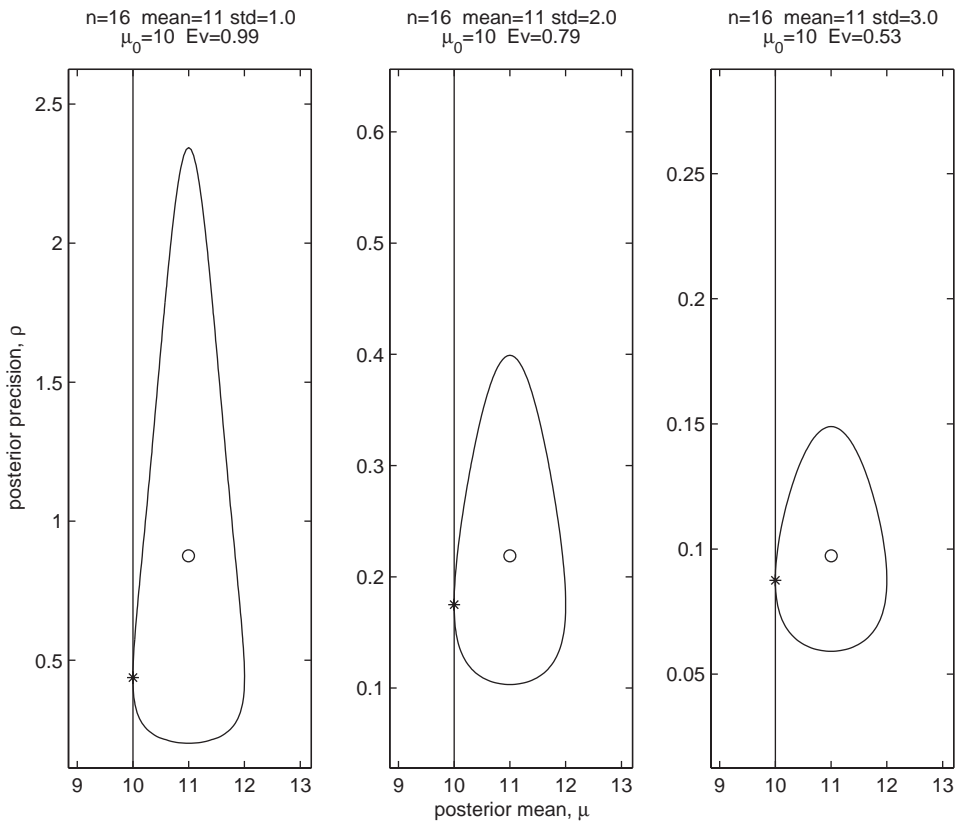


Fig. 1. FBST for $\mu_0 = 10$ on a normal pdf hypothesis and tangential set.

Hence the logposterior kernel is

$$f = \frac{n-2}{2} \ln \rho - \rho \frac{Q}{2} \left(1 + \frac{n}{Q} (\mu - \bar{x})^2 \right)$$

whose gradient is given by

$$\begin{bmatrix} \frac{\partial f}{\partial \mu} \\ \frac{\partial f}{\partial \rho} \end{bmatrix} = \begin{bmatrix} n\rho(\mu - \bar{x}) \\ \left(\frac{n}{2} - 1\right)\frac{1}{\rho} - \frac{Q}{2}\left(1 + \frac{n}{Q}(\mu - \bar{x})^2\right) \end{bmatrix}.$$

If we want to test $H : \mu = \mu_0$, using $r(\theta) = 1$, the constrained optimum is given by

$$\rho^* = \frac{n-2}{Q(1 + (n/Q)(\mu_0 - \bar{x})^2)}.$$

The integration step has to be performed numerically. Fig. 1 presents some examples showing the constrained and unconstrained optima, H and Θ^* .

4. Behrens–Fisher problem

At the Behrens–Fisher problem we want to test the hypothesis that two normal random variables, with unknown means, μ_1 and μ_2 , and unknown precisions, $\rho_1 = 1/\sigma_1^2$ and $\rho_2 = 1/\sigma_2^2$, have the same mean:

$$\Theta = \{[\mu_1, \rho_1, \mu_2, \rho_2] \in (R \times R_+)^2\},$$

$$\Theta_0 = \{[\mu_1, \rho_1, \mu_2, \rho_2] \in \Theta \mid \mu_1 = \mu_2\}.$$

We use the standard improper priors, uniform on $]-\infty, +\infty[$ for μ , and $1/\rho$ on $]0, +\infty[$ for the precision $\rho = 1/\sigma^2$, for the normal parameters, in order to get a fair comparison with p -values, DeGroot (1970), giving to each normal variate the posterior

$$p_n(\mu, \rho \mid x) \propto \sqrt{\rho} \exp(-n\rho(\mu - u)^2/2) \exp(-b\rho)\rho^{a-1},$$

$$x = [x_1 \dots x_n], \quad a = \frac{n-1}{2}, \quad m = \frac{1}{n} \sum_{i=1}^n x_i, \quad b = \frac{1}{2} \sum_{i=1}^n (x_i - m)^2.$$

At the numerical steps it is better, for numerical stability, to work with the loglikelihood. Given two samples, of size n_1 and n_2 ,

$$\begin{aligned} f(\mu_1, \rho_1, \mu_2, \rho_2 \mid n_1, m_1, b_1, n_2, m_2, b_2) \\ = (n_1/2 - 1) \log(\rho_1) - b_1 \rho_1 - (n_1 \rho_1/2)(\mu_1 - m_1)^2 \\ + (n_2/2 - 1) \log(\rho_2) - b_2 \rho_2 - (n_2 \rho_2/2)(\mu_2 - m_2)^2, \end{aligned}$$

the hypothesis being represented by the constraint

$$g(\mu_1, \rho_1, \mu_2, \rho_2) = \mu_1 - \mu_2 = 0.$$

The gradients of $f(\)$ and $g(\)$ have easy analytical expressions, that can be given to the optimizer:

$$\begin{aligned} df = [-n_1 \rho_1 (\mu_1 - m_1), (n_1/2 - 1)/\rho_1 - b_1 - (n_1/2)(\mu_1 - m_1)^2, \\ -n_2 \rho_2 (\mu_2 - m_2), (n_2/2 - 1)/\rho_2 - b_2 - (n_2/2)(\mu_2 - m_2)^2], \\ dg = [1, 0, -1, 0]. \end{aligned}$$

For the logsurprise relative to Jeffreys’ uninformative prior density and its gradient, we only have to replace the factors $(n/2 - 1)$ by $(n/2)$.

Table 1 presents the results for some examples, comparing the FBST evidence, with the uniform and Jeffreys’ uninformative prior reference densities, $Ev_U(H)$ and $Ev_J(H)$, and the p -values of the Welsh approximated t -test, tt , Lehmann (1986, pp. 208–209).

Table 1
Tests for Behrens–Fisher

m_2	$Ev_J(H)$	$Ev_U(H)$	$1 - tt$
100	0.00	0.00	0.00
101	0.08	0.07	0.66
102	0.53	0.50	0.94
103	0.89	0.87	0.99
104	0.98	0.98	1.00
105	1.00	1.00	1.00

$n_1 = 16, n_2 = 20.$
 $m_1 = 100, s_1 = s_2 = 3.$

While the FBST is a probability in the parameter space, the p -value is a probability in the sample space. Therefore, we can only check if they agree in tendency, but there would be no meaning in a direct comparison of the figures. We are comparing the mean of a first sample, of size, mean, and standard deviation $n_1 = 16, m_1 = 100$ and $s_1 = 3$, with a second sample, of size and standard deviation $n_2 = 20$ and $s_2 = 3$. The mean of the second sample, m_2 , is given in Table 1.

5. Coefficients of variation application

The coefficient of variation (CV) of a random variable X is defined as the ratio $CV(X) = Std(X)/E(X)$, i.e., the ratio of its standard deviation by its mean. We want to test the hypothesis that two normal random variables, with unknown mean and variance, have the same CV. Using the same notation of the last section,

$$\Theta_0 = \{[\mu_1, \rho_1, \mu_2, \rho_2] \in \Theta \mid \mu_1^2 \rho_1 = \mu_2^2 \rho_2\}.$$

The hypothesis is represented by the constraint

$$g(\mu_1, \rho_1, \mu_2, \rho_2) = \mu_1^2 \rho_1 - \mu_2^2 \rho_2 = 0$$

whose gradient can be given to the optimizer:

$$dg = [2\mu_1 \rho_1, \mu_1^2, -2\mu_2 \rho_2, -\mu_2^2].$$

Table 2 presents results for some examples. Some simpler hypothesis on the CV are analyzed in Lehmann (1986). However we are not aware of the exact p -values to compare with the FBST. We are comparing the coefficient of variation of a first sample, of size, mean, and standard deviation $n_1 = 16, m_1 = 100$ and $s_1 = 2$, with a second sample of size and mean $n_2 = 20$ and $m_2 = 200$. The standard deviation of the second sample, s_2 , is given in Table 2.

5.1. Hardy–Weinberg equilibrium law

In this biological application there is a sample of n individuals, where x_1 and x_3 are the two homozygote sample counts and $x_2 = n - x_1 - x_3$ is the heterozygote sample

Table 2
Tests for coefficients of variation

s_2	$Ev_U(H)$	$Ev_J(H)$	CV_1/CV_2
1	1.00	1.00	4.00
2	0.90	0.90	2.00
3	0.17	0.16	1.33
4	0.00	0.00	1.00
5	0.05	0.06	0.80
6	0.32	0.37	0.67
8	0.83	0.87	0.50
10	0.97	0.98	0.40
12	1.00	1.00	0.33

$n_1 = 16, n_2 = 20, s_1 = 2.$
 $m_1 = 100, m_2 = 200.$

count, and $\theta = [\theta_1, \theta_2, \theta_3]$ is the parameter vector. The posterior density for this trinomial model is

$$p_n(\theta | x) \propto \theta_1^{x_1} \theta_2^{x_2} \theta_3^{x_3}.$$

The parameter space and the null hypothesis set are:

$$\Theta = \{\theta \geq 0 \mid \theta_1 + \theta_2 + \theta_3 = 1\},$$

$$\Theta_0 = \{\theta \in \Theta \mid \theta_3 = (1 - \sqrt{\theta_1})^2\}.$$

The problem of testing the Hardy–Weinberg equilibrium law using the Bayes factor is discussed in detail by Montoya-Delgado et al. (2001) and Pereira and Rogatko (1984). The Bayes factor (BF), considering uniform priors over Θ_0 and $\Theta - \Theta_0$, and probability $\frac{1}{2}$ for H, is given by the following expression:

$$BF = \frac{(n + 2)!t!(2n - t)!2^{x_2}}{(2n + 1)!x_1!x_2!x_3!} \left[5/6 - \frac{2(t + 1)(2n - t + 1)}{(2n + 2)(2n + 3)} \right].$$

Here $t = 2x_1 + x_2$ is a sufficient statistic under H. This means that the likelihood under H depends on the data only through t .

Fig. 2 presents H and Θ^* for $x_1 = 5, y_3 = 10$ and $n = 20$. Fig. 3 compares Ev with the other standard measures: pv , the p -value for the standard χ^2 test, and the posterior probability of H using the BF, $pp = 1/(1 + 1/BF)$.

6. Holgate distribution: testing independence

This final example briefly describes a complex test where the FBST is considerably more powerful than competing tests, at least in the parts of the parameter space relevant to the application. The Holgate distribution appears in reliability theory, actuarial sciences and risk modeling, see for example Barlow and Prochan (1981), and Kocherlakota and Kocherlakota (1992).

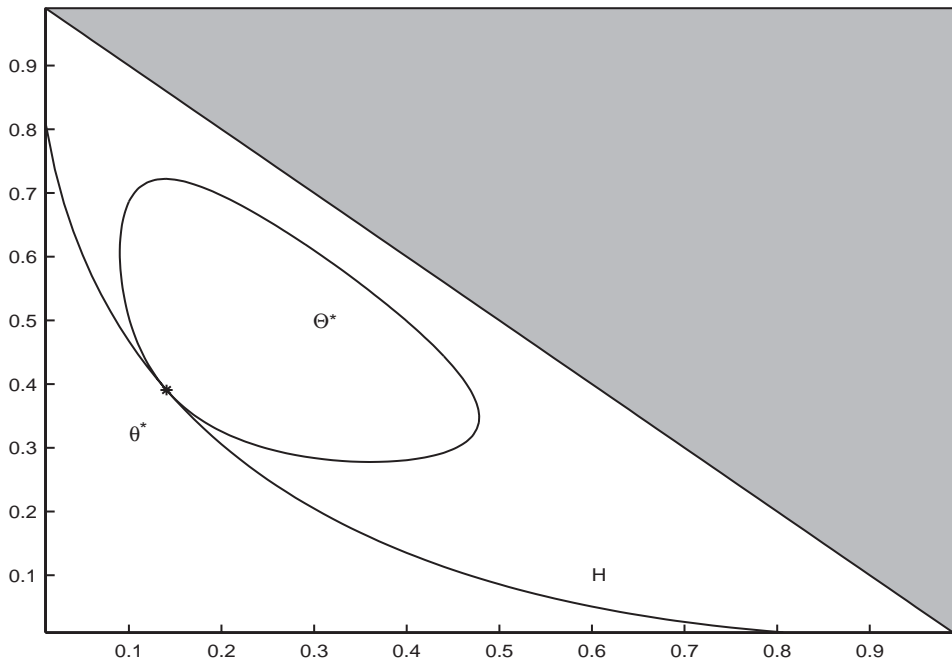


Fig. 2. H–W equil. hypothesis and tangential set for $x_1 = 5, x_2 = 5, x_3 = 10, Ev = 0.91$.

Let $p(j; \theta)$ denote the pdf of a Poisson distribution with mean θ . The pdf of the Holgate bivariate Poisson distribution can be derived by considering independent Poisson random variables J_1, J_2, J_3 with means, $\lambda_1, \lambda_2, \lambda_3$. Let $X = J_1 + J_3, Y = J_2 + J_3$ and $Z = J_3$. The marginal bivariate distribution of $W = (X, Y)$ is the Holgate distribution with parameters $\theta_1 = \lambda_1 + \lambda_3, \theta_2 = \lambda_2 + \lambda_3, \theta_3 = \lambda_3; 0 < \theta_1 < \infty, 0 < \theta_2 < \infty, 0 \leq \theta_3 < \min(\theta_1, \theta_2)$.

Thus, the pdf of the Holgate distribution is

$$f(x, y; \theta_1, \theta_2, \theta_3) = \sum_{l=0}^{\min(x,y)} p(x-l; \theta_1 - \theta_3) p(y-l; \theta_2 - \theta_3) p(l; \theta_3). \tag{1}$$

The variates X and Y are independent if and only if $\theta_3 = 0$, the null hypothesis.

Paul and Ho (1989) compared the power of several alternative tests, including Kocherlakota’s $C(x)$, Wald’s, and variations of t, F and the likelihood ratio. They established, by numerical simulations, that the modified-F test (MF) based on the sample correlation, ρ_n , is more powerful than the other tests. The MF test statistic is

$$MF = \frac{1}{2} \log \left(\frac{1 + \rho_n}{1 - \rho_n} \right) \left(1 - \frac{3}{4(n-1)} \right) + \frac{\rho_n}{4(n-1)}. \tag{2}$$

The efficient implementation of the FBST requires an efficient (in this case recursive) calculation of the Holgate distribution. The numerical integration of Ev is performed by Monte Carlo techniques, with careful estimation of the Monte Carlo precision. In

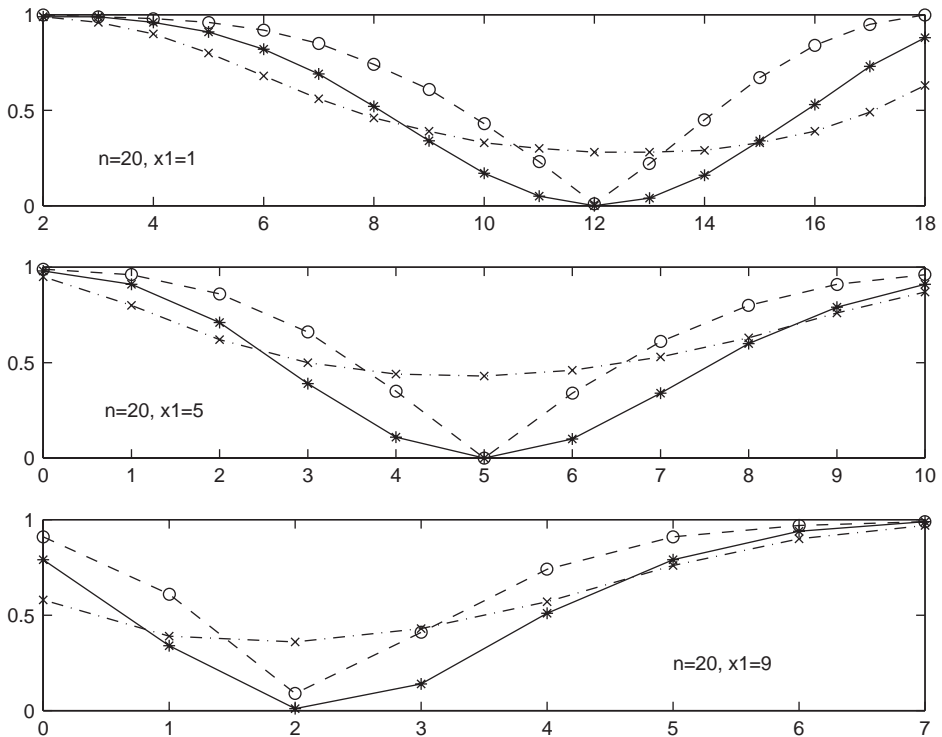


Fig. 3. FBST Ev (*), 1 - p.val. (o), and 1 - post.prob. (x), for H-W equil., vs. x_3 . $n = 20$ and $x_1 = 1, 5$ and 9.

order to compare the performance of the FBST and the MF tests we need estimates of quantiles under the null hypothesis, and power under the alternative. Careful iterative estimation-refinement procedures are necessary to obtain the desired accuracy in reasonable computation time. These techniques are detailed in Stern and Zacks (2002).

Since the correlation in the Holgate distribution is $\rho = \theta_3 / \sqrt{\theta_1 \theta_2}$, we can plot the power as a function of ρ . The tests' power changes slowly for $[\theta_1, \theta_2]$ in the area of interest, $\theta_1 = 1, \theta_2 \in [0.5, 2.0]$. Average interpolated power values, for significance level $\alpha = 0.05$, are presented in Fig. 4. As expected, Fig. 4 indicates that the power of the test is an increasing function of ρ . We also see that the FBST is considerably more efficient than the MF test when the sample size is about 20 and ρ is not too close to zero. Since both tests are consistent, the difference in power diminishes as the sample size n grows. This is shown in Fig. 4 for $n = 50$.

7. Final remarks

The FBST evidences presented were computed, to 1% precision, using Monte Carlo Simulation, as detailed in Stern and Zacks (2002). The power of FBST tests can be

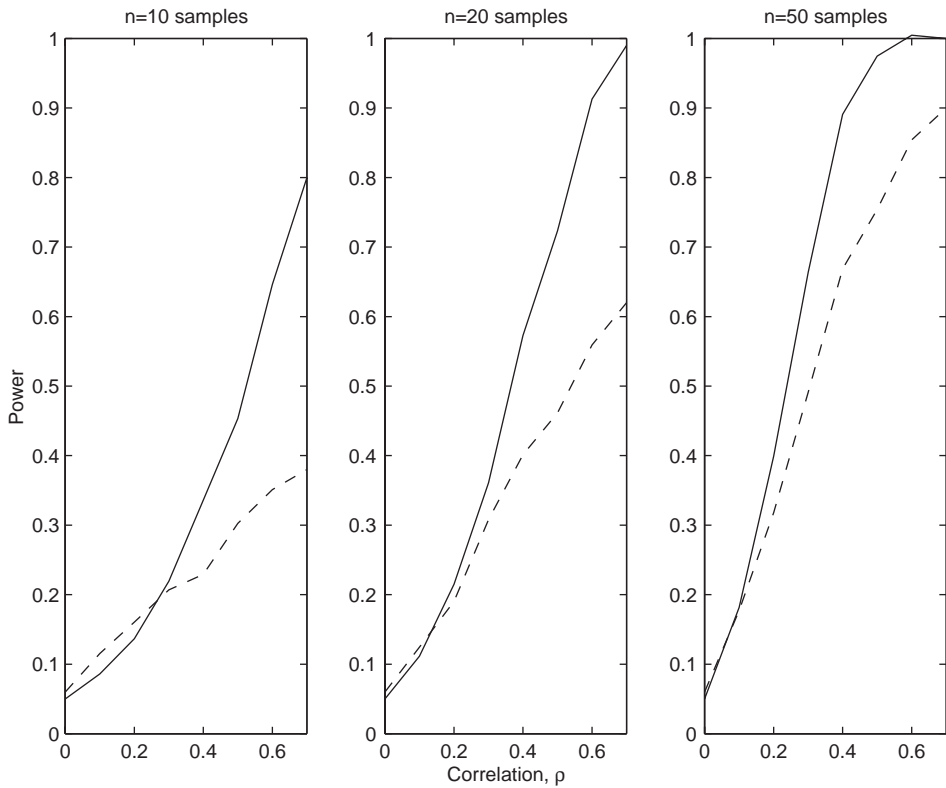


Fig. 4. Power as a function of ρ FBST: solid line, MF: dashed line. $\alpha = 0.05$.

computed using similar techniques. FBST tests can be far more powerful than existing procedures when testing some complex hypothesis, specially in higher dimensions, Irony et al. (2002), Lauretto et al. (2002), Pereira and Stern (2001), Stern and Zacks (2002).

Acknowledgements

The authors are grateful to the editor for several important suggestions. The authors are grateful for the support of the Departments of Computer Science and Statistics of the Institute of Mathematics and Statistics of the University of São Paulo. The last author is also grateful to Department of Mathematical Sciences at the State University of New York at Binghamton; this work was funded in part by grant FAPESP 2001-03484-1.

References

Barlow, R., Prochan, F., 1981. Statistical Theory of Reliability and Life Testing. To Begin With, Silver Spring, MD.

- Basu, D., 1988. Statistical Information and Likelihood. In: Ghosh, J.K. (Ed.), *Lecture Notes in Statistics*, Vol. 45. Springer, Berlin.
- Birnbaum, A., 1972. More on concepts of statistical evidence. *J. Amer. Statist. Assoc.* 67, 858–861.
- Box, G.E.P., Tiao, G.C., 1973. *Bayesian Inference in Statistical Analysis*. Wiley, New York.
- Carnap, R., 1962. *Logical Foundations of Probability*. University of Chicago Press, Chicago.
- Cox, D.R., 1977. The role of significance tests. *Scand. J. Statist.* 4, 49–70.
- DeGroot, M.H., 1970. *Optimal Statistical Decisions*. McGraw-Hill, New York.
- Evans, M., 1997. Bayesian inference procedures derived via the concept of relative surprise. *Comm. Statist.* 26, 1125–1143.
- Finetti, B. de, 1974. *Theory of Probability*, Vols. 1 and 2. Wiley, London.
- Finetti, B. de, 1981/1991. *Scritti*, Vol. 1, 1926–1930, Vol. 2, 1931–1936. CEDAM, Padova.
- Finetti, B. de, 1993. *Probabilità e Induzione*. CLUEB, Bologna.
- Gaskins, R.H., 1992. *Burdens of Proof in Modern Discourse*. Yale University Press, New Haven.
- Gentle, J.E., 2000. *Random Number Generation and Monte Carlo Methods*. Springer, New York.
- Good, I.J., 1983. *Good Thinking: The Foundations of Probability and its Applications*. University of Minnesota Press, Minnespotts, MN.
- Irony, T.Z., Lauretto, M., Pereira, C.A.B., Stern, J.M., 2002. A Weibull wearout test: full Bayesian approach. In: Hayakawa, Y., Irony, T., Xie, M. (Eds.), *Systems and Bayesian Reliability*. World Scientific, Singapore, pp. 287–300.
- Kempthorne, O., 1976. Of what use are tests of significance and tests of hypothesis. *Comm. Statist. Theory Methods* 8 (A5), 763–777.
- Kempthorne, O., 1980. Foundations of statistical thinking and reasoning. *Austral. CSIRO-DMS Newslett.* 68 (1–5) 69, 3–7.
- Klein, D.A., Rota, G.C., 1997. *Introduction to Geometric Probability*. Cambridge University Press, Cambridge.
- Kocherlakota, S., Kocherlakota, K., 1992. *Bivariate Discrete Distributions*. Marcel Decker, New York.
- Kokott, J., 1998. *The Burden of Proof in Comparative and International Human Rights Law*. Kluwer, The Hague.
- Lauretto, M., Pereira, C.A.B., Stern, J.M., 2002. Comparing parameters of two bivariate normal distributions using the invariant full Bayesian significance test. *Seventh Valencia International Meeting on Bayesian Statistics. Abstracts*, pp. 155. ISBA International Meeting, Tenerife, Canary Islands, Spain.
- Lehmann, E.L., 1986. *Testing Statistical Hypothesis*. Wiley, New York.
- Lindley, D.V., 1957. A statistical paradox. *Biometrika* 44, 187–192.
- Lindley, D.V., 1978. The Bayesian approach. *Scand. J. Statist.* 5, 1–26.
- Liu, J.S., 2001. *Monte Carlo Methods in Scientific Computing*. Springer, New York.
- Ljung, L., Pflug, G., Walk, H., 1992. *Stochastic Approximation and Optimization of Random Systems*. Birkhäuser, Basel.
- Luenberger, D.G., 1984. *Linear and Nonlinear Programming*. Addison-Wesley, Reading.
- Madruga, M.R., Esteves, L.G., Wechsler, S., 2001. On the Bayesianity of Pereira–Stern tests. *Test* 10, 291–299.
- McCormick, G.P., 1983. *Nonlinear Programming*. Wiley, New York.
- Montoya-Delgado, L.E., Irony, T.Z., Pereira, C.A.B., Whittle, M., 2001. Unconditional exact test for the Hardy–Weinberg law. *Genetics* 75–883.
- Ökten, G., 1999. *Contributions to the Theory of Monte Carlo and Quasi-Monte Carlo Methods*. www.dissertation.com.
- Paul, S.R., Ho, N.I., 1989. Estimation in the Biivariate Poisson Distribution and Hypothesis Testing Concerning Independence. *Communications in Statistics—Theory and Methods* 18, 1123–1133.
- Pereira, C.A.B., Rogatko, A., 1984. The Hardy–Weinberg equilibrium under a Bayesian perspective. *Braz. J. Genetics* 7, 689–707.
- Pereira, C.A.B., Stern, J.M., 1999a. A Dynamic Software Certification and Verification Procedure. *Proc. ISAS-99—International Conference on Information Systems Analysis and Synthesis II*, ISAS, Orlando, FL, USA 426–435.
- Pereira, C.A.B., Stern, J.M., 1999b. Evidence and credibility: full Bayesian significance test for precise hypotheses. *Entropy* 1, 69–80.

- Pereira, C.A.B., Stern, J.M., 2001. Model selection: full Bayesian approach. *Environmetrics* 12 (6), 559–568.
- Pereira, C.A.B., Wechsler, S., 1993. On the concept of p -value. *Braz. J. Probab. Statist.* 7, 159–177.
- Pflug, G.C., 1999. *Optimization of Stochastic Models: The interface Between Simulation and Optimization*. Kluwer, Dordrecht.
- Popper, K.R., 1989. *Conjectures and Refutations: The Growth of Scientific Knowledge*. Routledge, London.
- Royall, R., 1997. *Statistical Evidence: A Likelihood Paradigm*. Chapman & Hall, London.
- Rubin, H., 1987. A weak system of axioms for “Rational” behaviour and the non-separability of utility from prior. *Statist. Decisions* 5, 47–58.
- Santalo, L.A., 1976. *Integral Geometry and Geometric Probability*. In: *Encyclopedia of Mathematics and its Applications*, Vol. 1. Addison-Wesley, London.
- Savage, L.J., 1962. *The Foundations of Statistical Inference*. Methuen, London.
- Sellke, T., Bayarri, M.J., Berger, J., 1999. Calibration of p -values for testing precise null hypotheses. ISDS Discussion Paper 99-13.
- Spall, J.C., 2000. Adaptive stochastic approximation by the simultaneous perturbation method. *IEEE Trans. Automat. Control* 45, 1839–1853.
- Stern, J.S., Zacks, S., 2002. Testing the independence of poisson variates under the Holgate bivariate distribution The power of a new evidence test. RT-MAC-IME-USP 2002-03. Tech. Rep. Univ. of São Paulo. Submitted, *Statistics and Probability Letters*.
- Zellner, A., 1971. *An Introduction to Bayesian Inference in Econometrics*. Wiley, New York.