Sociedad de Estadística e Investigación Operativa
Test (2001) *Vol.* **10**, *No.* 2, pp. 000–000

On the Bayesianity of Pereira-Stern tests

M. Regina Madruga
Departamento de Estatística,
Universidade Federal do Pará, Brazil
Luís G. Esteves and Sergio Wechsler*
Instituto de Matemática e Estatística,
Universidade de São Paulo, Brazil

Abstract

C. Pereira and J. Stern have recently introduced a measure of evidence of a precise hypothesis consisting of the posterior probability of the set of points having smaller density than the supremum over the hypothesis. The related procedure is seen to be a Bayes test for specific loss functions. The nature of such loss functions and their relation to stylised inference problems are investigated. The dependence of the loss function on the sample is also discussed as well as the consequence of the introduction of Jeffreys's prior mass for the precise hypothesis on the separability of probability and utility.

Key Words: Bayesian Inference, Decision Theory, hypothesis test, loss functions.

AMS subject classification: 62C10, 62A15, 62F15, 62F03.

1 Introduction

Pereira and Stern (1999) have recently introduced a measure of evidence in favour of a precise hypothesis, i.e., a subset of the parametric space having null Lebesgue measure. The definition of their measure of evidence is now presented:

Definition 1.1. (Pereira and Stern) Consider a parametric statistical model, i.e., a quintet $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{A}}, \boldsymbol{F}, \boldsymbol{\Theta}, \pi)$, where $\boldsymbol{\mathcal{X}}$ is a sample space, $\boldsymbol{\mathcal{A}}$ is a suitable sigma-algebra of subsets of $\boldsymbol{\mathcal{X}}$, $\boldsymbol{F}$ is a class of probability distributions on $\boldsymbol{\mathcal{A}}$ indexed on a parametric space $\boldsymbol{\Theta}$ and π is a prior density over (a

*Correspondence to: Sergio Wechsler, Instituto de Matemática e Estatística, Universidade de São Paulo, Cx. Postal 66281, São Paulo-SP, Brazil, 05315-970. Email: sw@ime.usp.br

Received: September 2000; Accepted: May 2001

sigma-algebra of) $\mathbf{\Theta}$. Suppose a subset $\mathbf{\Theta}_0$ of $\mathbf{\Theta}$ having null Lebesgue measure (wrt $\mathbf{\Theta}$) is of interest. Let $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ be the posterior density of $\boldsymbol{\theta}$, given the sample observation $\boldsymbol{x}$, and $T(\boldsymbol{x}) = \{\boldsymbol{\theta} : \pi(\boldsymbol{\theta}|\boldsymbol{x}) > \sup_{\mathbf{\Theta}_0} \pi(\boldsymbol{\theta}|\boldsymbol{x})\}$. The *Pereira-Stern measure of evidence* is defined as $EV(\mathbf{\Theta}_0, \boldsymbol{x}) = 1 - Pr[\boldsymbol{\theta} \in T(\boldsymbol{x})|\boldsymbol{x}]$ and a Pereira-Stern test (or procedure) is to accept $\mathbf{\Theta}_0$ whenever $EV(\mathbf{\Theta}_0, \boldsymbol{x})$ is "large".

As we can see from Definition 1.1, the Pereira-Stern measure of evidence considers, in favour of a precise hypothesis, all points of the parametric space whose posterior density values are, at most, as large as the supremum over $\mathbf{\Theta}_0$; roughly speaking, it considers all points which are less "probable" than some point in $\mathbf{\Theta}_0$. Also, we should remember that, according to Pereira and Stern (1999), a large value of $EV(\mathbf{\Theta}_0, \boldsymbol{x})$ means that the subset $\mathbf{\Theta}_0$ lies in a high-probability region of $\mathbf{\Theta}$ and, therefore, the data support the null hypothesis; on the other hand, a small value of $EV(\mathbf{\Theta}_0, \boldsymbol{x})$ points out that $\mathbf{\Theta}_0$ is in a low-probability region of $\mathbf{\Theta}$ and the data would make us discredit the null hypothesis.

Pereira-Stern's procedures are in accordance with the "Principle of Least Surprise", as suggested by Good (1988), since it considers in the construction of the subset $T(\boldsymbol{x})$ those points in the parametric space less surprising ("more supported by the data", Good 1988) than the least surprising value in $\mathbf{\Theta}_0$ (for further details on this principle, see Good 1988). The posterior probability of $T(\boldsymbol{x})$, $Pr(\theta \in T(\boldsymbol{x})|\boldsymbol{x})$, may also be called "observed surprise", as indicated in Evans (1997).

Pereira and Stern (1999) claim that the use of $EV(\mathbf{\Theta}_0, \boldsymbol{x})$ to assess the evidence of $\mathbf{\Theta}_0$ is a "Bayesian" procedure, as only the posterior density is involved. Furthermore, the procedure has overcome the difficulty of dealing with a precise hypothesis (see Basu 1975, Berger and Delampady 1987, for a comprehensive analysis): unlike Jeffreys's tests (Jeffreys 1961), Pereira-Stern procedures do not introduce a prior positive probability for $\mathbf{\Theta}_0$.

The main purpose of this paper is to verify the existence of loss functions which render a Pereira-Stern procedure a Bayesian test of hypotheses of $\mathbf{\Theta}_0$ against $\mathbf{\Theta}_1$, the complementary set $\mathbf{\Theta} \backslash \mathbf{\Theta}_0$. For the reader's guidance, the content of this paper is presented as follows: In Section 2, we will exhibit such loss functions which confer "Bayesianity" to Pereira-Stern procedures and, in Section 3, shortcomings risen by the introduction of a prior positive probability for $\mathbf{\Theta}_0$ are pointed out. In Section 4, we will establish a relation between the Pereira-Stern solution to the problem of testing a precise hypothesis $\boldsymbol{\Theta}_0$ and the procedure for estimating $g(\theta) = \mathbf{1}(\theta \in \boldsymbol{\Theta}_0)$ as done in Hwang et al. (1992). Finally, we will discuss the aforementioned loss functions' unavoidable dependence on the sample data and examine loss functions depending on $\boldsymbol{x}$ in general. We should mention that, for simplicity, hereafter we let $\boldsymbol{\Theta}$ be the real line and $\boldsymbol{\Theta}_0$ have a single real number θ_0 (the general case $\boldsymbol{\Theta} \subseteq \mathbb{R}^n$, $n \in \mathbb{N}$, is similar and will be omitted in the present work; nevertheless, the necessary alterations will be commented as we go along). We will then consider the hypotheses: $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.

2 A Direct Bayesian Test

In this section, we will verify that Pereira-Stern procedures consist in direct Bayesian tests of hypotheses for specific loss functions. Here we call "direct" the tests of precise hypotheses that take into account only a probability density function over $\boldsymbol{\Theta}$, not introducing a positive probability for $\boldsymbol{\Theta}_0$ as in Jeffreys's tests. The latter will be considered in the next section. Let $\boldsymbol{D} = \{\textit{Accept } H_0\ (d_0), \textit{Reject } H_0\ (d_1)\}$ the decision space. We define the following loss function:

Definition 2.1. The loss function L on $\boldsymbol{D} \times \boldsymbol{\Theta}$ defined by $L(\textit{Reject } H_0,\ \theta) = a[1 - \mathbf{1}(\theta \in T(\boldsymbol{x}))]$ and $L(\textit{Accept } H_0,\ \theta) = b + c\mathbf{1}(\theta \in T(\boldsymbol{x})), a, b, c > 0$, is called a LP_1 *loss function.*

We should note that we will consider for the verification of the "Bayesianity" of Pereira-Stern procedures loss functions which depend on the sample observation $\boldsymbol{x}$. This fact hints the possibility that loss functions which turn Pereira-Stern procedures into Bayesian hypotheses tests unavoidably depend on the sample data. Such dependence may, at first sight, look odd. However, as we will examine in detail in the last section of the paper, loss functions depending on the sample data are able to incorporate some psychological aspects from an individual's preference ordering. In addition, this kind of loss function has already appeared in the literature, as in Bernardo and Smith (1994, pp. 395).

Another aspect of LP_1 loss functions is that they punish heavily the decision-maker who accepts H_0 when θ is, in fact, more "probable" than θ_0, that is, when θ belongs to $T(\boldsymbol{x})$. Now, let us prove the following:

Theorem 2.1. *Minimization of posterior expected LP_1 loss functions is a Pereira-Stern procedure.*

Proof. The posterior risk of acceptance is

$$\begin{aligned} E_\pi[L(d_0,\theta)|\boldsymbol{x}] &= E_\pi[L(\mathit{Accept}\ H_0,\theta)|\boldsymbol{x}] \\ &= \int_{\boldsymbol{\Theta}} [b + c\mathbf{1}(\theta \in T(\boldsymbol{x}))]\ \pi(\theta|\boldsymbol{x})d\theta \\ &= \int_{\boldsymbol{\Theta}} b\ \pi(\theta|\boldsymbol{x})d\theta + \int_{T(\boldsymbol{x})} c\ \pi(\theta|\boldsymbol{x})d\theta \\ &= b + c(1 - EV(\boldsymbol{\Theta}_0,\boldsymbol{x})). \end{aligned} \tag{2.1}$$

On the other hand, the posterior risk of rejection is

$$\begin{aligned} E_\pi[L(d_1,\theta)|\boldsymbol{x}] &= E_\pi[L(\mathit{Reject}\ H_0,\theta)|\boldsymbol{x}] \\ &= \int_{\boldsymbol{\Theta}} a[1 - \mathbf{1}(\theta \in T(\boldsymbol{x}))]\ \pi(\theta|\boldsymbol{x})d\theta \\ &= \int_{\boldsymbol{\Theta}} a\ \pi(\theta|\boldsymbol{x})d\theta - \int_{T(\boldsymbol{x})} a\ \pi(\theta|\boldsymbol{x})d\theta \\ &= aEV(\boldsymbol{\Theta}_0,\boldsymbol{x}). \end{aligned} \tag{2.2}$$

The test is, therefore, to accept $\boldsymbol{\Theta}_0$ if, and only if, $E_\pi[L(d_0,\theta)|\boldsymbol{x}] < E_\pi[L(d_1,\theta)|\boldsymbol{x}]$, that is, if

$$EV(\boldsymbol{\Theta}_0,\boldsymbol{x}) > \frac{b+c}{a+c} \tag{2.3}$$

□

From the above inequality, we note that if $a < b$, then the decision will be always to reject H_0, as $EV(\boldsymbol{\Theta}_0,\boldsymbol{x})$ takes values in the interval $[0,1]$. In particular, if for the decision-maker rejection of H_0 is preferable to acceptance of H_0 — whenever θ lies in $T^c(\boldsymbol{x})$ — then the decision will always be to reject H_0. On the other hand, if $a \gg b$ and c is "small", then the acceptance of H_0 does not require a large value for $EV(\boldsymbol{\Theta}_0,\boldsymbol{x})$. This happens whenever the decision-maker thinks that d_0 is preferable to d_1 and that θ belonging to $T(\boldsymbol{x})$ is not so embarrassing. As an example, if $a,b,c > 0$ satisfy $9a - 10b = c \quad (19a - 20b = c)$, then we will decide in favour of H_0 if $EV(\boldsymbol{\Theta}_0,\boldsymbol{x}) > 0.90$ (0.95), standard cutoff values in significance tests.

We should also emphasize that there are variations of LP_1 loss functions whose interpretations are different from the one presented here, but which still lead us to perform a Pereira-Stern test. For example, minimization of the posterior expectation of L' defined by $L'(\textit{Reject } H_0,\ \theta) = a - d\mathbf{1}(\theta \in T(\boldsymbol{x}))$ and $L'(\textit{Accept } H_0,\ \theta) = L(\textit{Accept } H_0,\ \theta)$, for $d > 0$, will again result in a Pereira-Stern test.

3 Pereira-Stern Procedures and Jeffreys's tests

In this section, we will verify that the introduction of a prior positive probability for the hypothesis H_0 will not render the Pereira-Stern measure of evidence a test statistic for the decision problem stated in Section 2 with LP_1 loss functions. More precisely, the Pereira-Stern measure of evidence will be just a term of this test statistic. For this purpose, let $f(\boldsymbol{x}|\theta)$ be the likelihood function, $g(\boldsymbol{x}) = \int_{\boldsymbol{\Theta}} f(\boldsymbol{x}|\theta)\pi(\theta)d\theta$ be the marginal density of the data and $\alpha \in [0,1]$ be the prior probability for H_0 (as done in Jeffreys's tests). Let us continue to solve the decision problem.

Suppose that the prior distribution on $\boldsymbol{\Theta}$ is given by

$$P(\theta) = \begin{cases} \alpha, & \theta = \theta_0 \\ (1-\alpha)\pi(\theta), & \theta \neq \theta_0 \end{cases}, \tag{3.1}$$

where $\pi(\theta)$ is the original density on the parametric space before specification of H_0 and H_1. Then, the posterior distribution on $\boldsymbol{\Theta}$ is

$$P(\theta|\boldsymbol{x}) = \begin{cases} \alpha f(\boldsymbol{x}|\theta_0)/f(\boldsymbol{x}), & \theta = \theta_0 \\ (1-\alpha) f(\boldsymbol{x}|\theta)\pi(\theta)/f(\boldsymbol{x}), & \theta \neq \theta_0 \end{cases}, \tag{3.2}$$

where $f(\boldsymbol{x}) = \alpha f(\boldsymbol{x}|\theta_0) + \int_{\theta\neq\theta_0}(1-\alpha)f(\boldsymbol{x}|\theta)\pi(\theta)d\theta$. The posterior risk of acceptance is

$$\begin{aligned} E_P[L(d_0,\theta)|\boldsymbol{x}] &= E_P[L(\textit{Accept } H_0,\theta)|\boldsymbol{x}] \\ &= \frac{b\ \alpha f(\boldsymbol{x}|\theta_0)}{f(\boldsymbol{x})} + \int_{\theta\neq\theta_0}[b + c\mathbf{1}(\theta \in T(\boldsymbol{x}))]\frac{(1-\alpha)f(\boldsymbol{x}|\theta)\pi(\theta)}{f(\boldsymbol{x})}d\theta \\ &= \frac{b\ \alpha f(\boldsymbol{x}|\theta_0)}{f(\boldsymbol{x})} + \frac{(1-\alpha)g(\boldsymbol{x})}{f(\boldsymbol{x})}[b + c - c\, EV(\Theta_0,\boldsymbol{x})]. \end{aligned} \tag{3.3}$$

On the other hand, the posterior risk of rejection is

$$\begin{aligned} E_P[L(d_1,\theta)|\boldsymbol{x}] &= E_P[L(\mathit{Reject}\ H_0,\theta)|\boldsymbol{x}] \\ &= \frac{a\ \alpha f(\boldsymbol{x}|\theta_0)}{f(\boldsymbol{x})} + \int_{\theta\neq\theta_0} [a - a\mathbf{1}(\theta \in T(\boldsymbol{x}))] \frac{(1-\alpha)f(\boldsymbol{x}|\theta)\pi(\theta)}{f(\boldsymbol{x})} d\theta \\ &= \frac{a\ \alpha f(\boldsymbol{x}|\theta_0)}{f(\boldsymbol{x})} + \frac{(1-\alpha)g(\boldsymbol{x})}{f(\boldsymbol{x})}\ a\ EV(\Theta_0,\boldsymbol{x}). \end{aligned} \tag{3.4}$$

The test is then to accept $\boldsymbol{\Theta}_0$ if, and only if, $E_P[L(d_0,\theta)|\boldsymbol{x}] < E_P[L(d_1, \theta)|\boldsymbol{x}]$, that is, if

$$EV(\boldsymbol{\Theta}_0,\boldsymbol{x}) + \frac{(a-b)\ \alpha\ f(\boldsymbol{x}|\theta_0)}{(a+c)(1-\alpha)g(\boldsymbol{x})} > \frac{b+c}{a+c}. \tag{3.5}$$

As we can see from the above inequality, $EV(\boldsymbol{\Theta}_0,\boldsymbol{x})$ is no more the single test statistic if we take into account Jeffreys's idea for testing precise hypothesis. In this case, the decision criterion will depend not only on the prior probability for H_0 (the larger the value of α is, the smaller $EV(\boldsymbol{\Theta}_0,\boldsymbol{x})$ needs to be in order to make us accept H_0), but also on the ratio of the likelihood of θ_0 to the mean likelihood $g(\boldsymbol{x})$, i.e., the statistic in (3.5) blends the Pereira-Stern evidence and the ratio of posterior probabilities of $\boldsymbol{\Theta}_0$ and $\boldsymbol{\Theta}_0^c$. We should mention that for the general case $\boldsymbol{\Theta} \subseteq \mathbb{R}^n$, $n \in \mathbb{N}$, a small modification should be done whenever $\boldsymbol{\Theta}_0$ is not a singleton: to substitute $g_0(\boldsymbol{x}) = \int_{\boldsymbol{\Theta}_0} f(\boldsymbol{x}|\theta)\pi_0(\theta)d\theta$ for $f(\boldsymbol{x}|\theta_0)$, with π_0 being a probability density function over $\boldsymbol{\Theta}_0$. If $\boldsymbol{\Theta}_0$ has only one element θ_0, π_0 is degenerate and no substitution is called for.

We should also note that when $\alpha = 0$ we return to the situation of Section 2, in which $EV(\boldsymbol{\Theta}_0,\boldsymbol{x})$ plays the role of the test statistic. For general $\alpha > 0$, in order to have the Pereira-Stern measure of evidence as the test statistic (that is, $EV(\boldsymbol{\Theta}_0,\boldsymbol{x})$ being the only term depending on $\boldsymbol{x}$, as in the case $\alpha = 0$), it seems that suitable loss functions depend not only on $\boldsymbol{x}$ but also on the original prior density π over $\boldsymbol{\Theta}$. An example of such loss function is given by $L(\mathit{Reject}\ H_0,\ \theta_0) = a$; $L(\mathit{Accept}\ H_0,\ \theta) = [b + \mathbf{1}(\theta \in T(\boldsymbol{x}))]f(\boldsymbol{x}|\theta_0)/g(\boldsymbol{x})$, for $\theta \neq \theta_0$; zero, otherwise, with $a, b > 0$. The dependence of the above loss function on π, when $\alpha > 0$, suggests that Pereira-Stern procedures — as Jeffreys's tests — do not separate probability and utility (Rubin 1987).

In the sequel, we will associate Pereira-Stern procedures with stylised inference problems.

4 Estimation of $\mathbf{1}(\boldsymbol{\theta} \in \boldsymbol{\Theta}_0)$

A different approach to a Pereira-Stern procedure is to consider it as a problem of estimation (Hwang et al. 1992, among others). More precisely, we consider $EV(\boldsymbol{\Theta}_0, \boldsymbol{x})$ as an estimator of $\mathbf{1}(\theta \in \boldsymbol{\Theta}_0)$. Thus, the new decision space, $\boldsymbol{D}'$, is formed by all $\boldsymbol{\mathcal{A}}$-measurable functions $\phi : \boldsymbol{\mathcal{X}} \to [0,1]$. We will show that the Pereira-Stern measure of evidence is a Bayesian solution for this estimation problem. In this context we define the following loss function:

Definition 4.1. Let $\phi(\boldsymbol{x})$ be an estimator of the function $\mathbf{1}(\theta \in \boldsymbol{\Theta}_0)$ and $T^c(\boldsymbol{x})$ the complementary set of $T(\boldsymbol{x})$. The loss function L on $\boldsymbol{D}' \times \boldsymbol{\Theta}$ defined by $L(\phi(\boldsymbol{x}), \theta) = [\mathbf{1}(\theta \in T^c(\boldsymbol{x})) - \phi(\boldsymbol{x})]^2$ is called a LP_2 *loss function.*

We should note that if we substitute $\mathbf{1}(\theta \in \boldsymbol{\Theta}_0)$ for the factor $\mathbf{1}(\theta \in T^c(\boldsymbol{x})) = 1 - \mathbf{1}(\theta \in T(\boldsymbol{x}))$ in the expression of LP_2 loss function, we will obtain the usual quadratic loss function (a proper scoring rule), whose optimal solution is the true Bayesian estimator $Pr(\theta \in \boldsymbol{\Theta}_0|\boldsymbol{x})$. The term $\mathbf{1}(\theta \in T^c(\boldsymbol{x}))$ incorporates Pereira-Stern's original idea that the points belonging to $T^c(\boldsymbol{x})$ should support the null hypothesis H_0, especially the point θ_0 itself, whereas values belonging to $T(\boldsymbol{x})$ should discredit H_0. The loss function defined above is then seen as the squared distance between the estimate $\phi(\boldsymbol{x})$ and the intuitive clairvoyant "estimate", $\mathbf{1}(\theta \in T^c(\boldsymbol{x}))$.

Theorem 4.1. *The Pereira-Stern measure of evidence minimizes the posterior expectation of* LP_2 *loss functions.*

Proof. The posterior risk is given by

$$\begin{aligned} E_\pi[L(\phi(\boldsymbol{x}), \theta)|\boldsymbol{x}] &= \int_{T^c(\boldsymbol{x})} (1 - \phi(\boldsymbol{x}))^2 \pi(\theta|\boldsymbol{x}) d\theta + \int_{T(\boldsymbol{x})} \phi^2(\boldsymbol{x}) \pi(\theta|\boldsymbol{x}) d\theta \\ &= (1 - \phi(\boldsymbol{x}))^2 EV(\boldsymbol{\Theta}_0, \boldsymbol{x}) + \phi^2(\boldsymbol{x})(1 - EV(\boldsymbol{\Theta}_0, \boldsymbol{x})) \\ &= \phi^2(\boldsymbol{x}) - 2\phi(\boldsymbol{x}) EV(\boldsymbol{\Theta}_0, \boldsymbol{x}) + EV(\boldsymbol{\Theta}_0, \boldsymbol{x}). \end{aligned} \tag{4.1}$$

Therefore $\phi^*(\boldsymbol{x}) = EV(\boldsymbol{\Theta}_0, \boldsymbol{x})$ is the optimal solution, minimizing the posterior risk. □

5 Discussion

It is easily seen that performance of a Pereira-Stern procedure as an inference about a precise hypothesis does not violate the Likelihood Principle. This being not sufficient for the "Bayesianity" of the procedure, we have proceeded to characterize it as a Bayesian test of hypotheses.

A loss function represents the preference of a decision-maker among consequences dependent on unknown values of the state of nature (Savage 1954). Assuming separability of probability and utility (see Rubin 1987, for a deeper approach), one would call "Bayesian" a procedure which minimizes expected loss functions — the coherent solution to the decision problem.

Only loss functions that depend on the factor $\mathbf{1}(\theta \in T(\boldsymbol{x}))$ lead to Pereira-Stern procedures. Pereira-Stern procedures, therefore, correspond to Bayes tests with loss functions which depend on $\boldsymbol{x}$. While not violating the Likelihood Principle — they are genuine "posterior" procedures — these procedures formally allow for consideration of the statistician's embarrassment (or pride!) on having accepted (or rejected) the null hypothesis when the value of θ is idealistically revealed to belong to $T(\boldsymbol{x})$, a "stylised form of statistical inference" (Bernardo and Smith 1994). The consideration of such psychological components in the construction of loss functions can only be welcomed. In a somewhat different scenario, Kadane (1992) has resolved Allais' paradox by using a utility function incorporating the statisticians suspicion that offers were too good to be true.

Another interesting feature of a Pereira-Stern procedure revealed by the examination of its "Bayesianity" is that the introduction of Jeffreys's prior probability for H_0 removes from $EV(\boldsymbol{\Theta}_0, \boldsymbol{x})$ the condition of full test statistic. A way out for this difficulty is to consider "loss" functions dependent on the original prior density $\pi(\theta)$. We arrive at the curious conclusion that performance of a Jeffreys's test in this setting does not separate utility (of rejection/acceptance) from probability (of θ). This phenomenon, which has connections with the classic problem of assigning positive probability to a precise hypothesis, calls for further investigation. In any case, it should be emphasized that the Pereira-Stern procedure - by avoiding Jeffreys's framework - separates utility from probability and keeps $EV(\boldsymbol{\Theta}_0, \boldsymbol{x})$ as the full test statistic. It is a most important alternative for Jeffreys's Bayesian tests of precise hypotheses, not requiring the probability mass for the precise hypothesis needed by the latter.

References

Basu, D. (1975). Statistical information and likelihood (with discussion). *Sankhyã A* **37**, 1-71.

Berger, J.O. and M. Delampady (1987). Testing precise hypotheses (with discussion). *Statistical Science*, **2**, 317-352.

Bernardo, J.M. and A.F.M. Smith (1994). *Bayesian Theory.* John Wiley and Sons, New York.

Evans, M. (1997). Bayesian inference procedures derived via the concept of relative surprise. *Communications in Statistics*, **26**, 1125-1143.

Good, I.J. (1988). Surprise index. *Encyclopedia of Statistical Sciences*, **9**, 104-109 (S. Kotz, N.L. Johnson and C.B. Reid, eds.) John Wiley and Sons, New York.

Hwang, J.T., G. Casella, C.P. Robert, M.T. Wells and R.G. Farrel (1992). Estimation of accuracy in testing. *Annals of Statistics*, **20**, 490-509.

Jeffreys, H. (1961). *Theory of Probability.* Oxford University Press.

Kadane, J.B. (1992). Healthy scepticism as an expected utility explanation of the phenomena of Allais and Ellsberg. *Theory and Decision*, **32**, 57-64.

Pereira, C.A. de B. and J. Stern (1999). Evidence and Credibility: a full Bayesian test of precise hypothesis. *Entropy*, **1**, 104-115.

Rubin, H. (1987). A weak system of axioms for 'rational' behaviour and the non-separability of utility from prior. *Statistics and Decisions*, **5**, 47-58.

Savage, L.J. (1954). *The Foundations of Statistics.* Wiley, New York.