Ann. Inst. Statist. Math.
Vol. 49, No. 4, 711–726 (1997)

RELATIONSHIPS BETWEEN POST-DATA ACCURACY MEASURES

CONSTANTINOS GOUTIS[1]* AND GEORGE CASELLA[2]**

[1] *Department of Statistical Science, University College London, Gower St., London, WC1E 6BT, U.K. and Departamento de Estadistica y Econometria, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe, Madrid, Spain*

[2] *Biometrics Unit, Cornell University, 434 Warren Hall, Ithaca, NY 14853, U.S.A.*

(Received March 27, 1995; revised April 3, 1996)

Abstract. In the usual frequentist formulation of testing and interval estimation there is a strong relationship between α-level tests and $1 - \alpha$ confidence intervals. Such strong relationships do not always persist for post-data, or Bayesian, measures of accuracy of these procedures. We explore the relationship between post-data measures of accuracy of both tests and interval estimates, measures that are derived under a decision-theoretic structure. We find that, in general, there are strong post-data relationships in the one-sided case, and some relationships in the two-sided case.

Key words and phrases: Posterior probability, coverage probability, null hypothesis, Bayes, p-value.

1. Introduction

A post-data measure of accuracy of a procedure is an estimate of the correctness of an inference. We differentiate between post data measures, constructed after the data have been seen, from pre-data measures constructed before seeing any data. For the accuracy of a test, for example, common post-data measures are the p-value or the posterior probabilities, as opposed to the α-level which is a pre-data measure.

The concern here is with inferences about the accuracy of tests and confidence intervals. There is a close relationship between the two procedures, as each one can be obtained by inverting the other. This duality results in a direct correspondence between some characteristics, such as the confidence coefficient of an interval and the level of a test. Hence, one may expect the associated accuracy measures to be also related. The major focus of the paper is to explore the existence of such relationships.

---

* This paper was completed before the untimely death of Costas Goutis in July 1996.

** Research Supported by NSF Grant DMS 9305547. This is paper BU-1347-M in the Biometrics Unit, Cornell University, Ithaca, NY 14853. File acc2.tex.

We will concentrate on two ways of constructing estimates of accuracy. The first is the straightforward Bayesian calculation of posterior probabilities of hypotheses or regions. The second is through classical decision theory. Since the accuracy of a procedure can be viewed as the negative of its loss, estimating accuracy is equivalent to estimating a loss function (see, e.g., Rukhin (1988*a*, 1988*b*), Johnstone (1988) or Lu and Berger (1989*a*, 1989*b*)).

In our general development we assume that $X = x$ is observed, where $X \sim f(\cdot \mid \theta)$ and θ is an unknown parameter of interest. Furthermore, there is a prior distribution $\pi(\theta)$ which can be combined with $f(\cdot \mid \theta)$ in the usual way to yield a posterior distribution $\pi(\theta \mid x)$. For a set estimator of θ, $C(x)$, the post-data accuracy is measured by $I(\theta \in C(x))$, where $I(\cdot)$ is the indicator function. This, in effect, measures the loss incurred by estimating θ with $C(x)$, since we can define

$$L_1(\theta, C(x)) = 1 - I(\theta \in C(x)) = \begin{cases} 1 & \text{if} \quad \theta \notin C(x) \\ 0 & \text{if} \quad \theta \in C(x). \end{cases}$$

To estimate the accuracy, we look for estimators $\gamma(x)$ that perform well against the loss

$$L_2(\theta, C, \gamma) = [I(\theta \in C(x)) - \gamma(x)]^2. \tag{1.1}$$

The pair $\langle C(x), \gamma(x)\rangle$ is referred to as a *confidence procedure*.

A similar approach was taken by Lu and Berger (1989*b*), Robert and Casella (1993), and George and Casella (1994) in the context of estimating a multivariate normal mean, while Goutis and Casella (1992) applied this methodology to Student's t interval. A discussion of the appropriateness of the loss (1.1) can be found in Hwang and Pemantle (1990) and Hwang *et al.* (1992).

For the testing problem a similar development is possible. In testing $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \notin \Theta_0$ we consider a rule of the form "accept H_0 if $x \in A$", where A is a subset of the sample space. The loss incurred by this procedure can be written

$$L_3(\theta, A) = \begin{cases} 1 & \text{if} \quad \theta \in \Theta_0, \ x \notin A \quad \text{or} \quad \theta \notin \Theta_0, \ x \in A \\ 0 & \text{if} \quad \theta \in \Theta_0, \ x \in A \quad \text{or} \quad \theta \notin \Theta_0, \ x \notin A. \end{cases}$$

Traditional decision theory implicitly uses $L_3(\theta, A)$, as the risk of a test under this loss is the Type I and Type II errors. To estimate the loss (or accuracy) of the procedure we use estimators $p(x)$ that perform well against

$$L_4(\theta, A, p) = [|I(\theta \in \Theta_0) - I(x \in A)| - p(x)]^2. \tag{1.2}$$

The procedures $\langle A, p(x)\rangle$ and $\langle C(x), \gamma(x)\rangle$ both provide a means of assessing post-data accuracy. To clarify the meaning of the "equivalence" of these post-data measures consider the typical pre-data situation. If A and C are an α-level acceptance region and its associated $1 - \alpha$ confidence set, then (writing $A(\theta)$ for clarity) we have

$$\theta \subset C(x) \Leftrightarrow x \subset A(\theta), \tag{1.3}$$

and hence

$$\sup_{\theta\in\Theta_0} P(X \notin A(\theta) \mid \theta) = 1 - \inf_{\theta} P(\theta \in C(X) \mid \theta), \tag{1.4}$$

which we refer to as pre-data equivalence of A and C.

We explore whether a similar equivalence exists between accuracy measures in a post-data setting. However, for the frequentist sets and tests related as in (1.3), there is no implicit definition of post-data accuracy measures in the construction, so it is not immediate how to derive the post-data quantities equivalent to the ones in (1.4). We are guided in this derivation by both decision-theoretical and Bayesian methodology, which lead us to accuracy measures derived from posterior distributions. Conditioning on $X = x$, instead of the Type I error for the test, we will consider p^π of the form

$$p^\pi(x) = P(\theta \in \Theta_0 \mid x)I(x \notin A) + P(\theta \notin \Theta_0 \mid x)I(x \in A) \tag{1.5}$$

as an estimate of the loss L_3 incurred, where $p^\pi(x)$ is the Bayes rule against the loss L_4 for a given acceptance region A and prior π. Similarly, instead of the coverage probability we will consider γ^π given by

$$\gamma^\pi(x) = \int_{C(x)} \pi(\theta \mid x)d\theta$$

as an estimate of the loss L_1 incurred.

The post-data version of (1.4) would involve the supremum of p^π and the infimum of γ^π, and in the remainder of the paper this relationship is explored. Section 2 considers the one-sided case, where a strong relationship exists. In the two-sided case, treated in Section 3, the relationship still exists but is not as strong. Finally, Section 4 contains a short discussion.

2. The one-sided case

There are two major points that we should take into account in establishing a post-data equivalence. The first is that for hypothesis testing π will usually depend on H_0, while for set estimation it does not. Hence, along with a family of tests we have a family of priors. Secondly there is a vast difference in the one-sided vs. two-sided problem (see Berger and Sellke (1987); Casella and Berger (1987); Hwang *et al.* (1992)). Any prior for testing $H_0 : \theta = \theta_0$ must put a point mass on θ_0, which would not be done if interest was in set estimation or in testing an interval null hypothesis. Therefore, we will consider the one-sided and two-sided cases separately.

We specialize to the case of a location parameter, and assume that we observe $X = x$, where $X \sim f(x - \theta)$ with $f(\cdot)$ continuous and having monotone likelihood ratio (mlr). Suppose also that the prior is continuous and of the form $\pi(\theta - \theta_0)$, where θ_0 is the parameter value specified in H_0, and that the supports of the density and prior do not depend on x or θ_0. Note that the results of this section will also apply to the scale parameter case.

The hypothesis of interest is

$$H_0 : \theta \leq \theta_0 \quad \text{vs.} \quad \theta > \theta_0, \tag{2.1}$$

which leads to the acceptance region and associated interval

$$A(\theta_0) = \{x : x \leq \theta_0 + c\}, \qquad C(x) = \{\theta : \theta \geq x - c\}, \tag{2.2}$$

where c is a constant, usually chosen to yield a specified α level. Using the loss function L_4, our measure of accuracy of $A(\theta_0)$ is

$$p_A(x, \theta_0) = P(\theta \leq \theta_0 \mid x) I(x \notin A(\theta_0)) + P(\theta > \theta_0 \mid x) I(x \in A(\theta_0))$$

where

$$P(\theta \leq \theta_0 \mid x) = \frac{\int_{-\infty}^{\theta_0} f(x-\theta)\pi(\theta - \theta_0) d\theta}{\int_{-\infty}^{\infty} f(x-\theta)\pi(\theta - \theta_0) d\theta}. \tag{2.3}$$

Similarly, using the loss L_2, our measure of accuracy of $C(x)$ is

$$\gamma_C(x, \theta_0) = \frac{\int_{x-c}^{\infty} f(x-\theta)\pi(\theta - \theta_0) d\theta}{\int_{-\infty}^{\infty} f(x-\theta)\pi(\theta - \theta_0) d\theta}.$$

The equivalence between p_A and γ_C is stated in the following theorem.

THEOREM 2.1. *Under the assumptions in this section*

$$\sup_{\theta_0 : \theta_0 \notin C(x)} p_A(x, \theta_0) = 1 - \gamma_C(c, 0). \tag{2.4}$$

Furthermore, if $\pi(\cdot)$ *has mlr, then*

$$\inf_{\theta_0 : \theta_0 \in C(x)} \gamma_C(x, \theta_0) = \gamma_C(c, 0), \tag{2.5}$$

and thus

$$\sup_{\theta_0 : \theta_0 \notin C(x)} p_A(x, \theta_0) = 1 - \inf_{\theta_0 : \theta_0 \in C(x)} \gamma_C(x, \theta_0). \tag{2.6}$$

Remark 1. Equation (2.6) can be thought of as the equivalent of (1.4). In words, equation (2.6) says that the maximum probability of not covering θ (the RHS) is the maximum error of the test when rejecting (the LHS), or equivalently

$$\sup_{\theta \in \Theta_0} P(X \notin A(\theta) \mid x) = 1 - \inf_{\theta} P(\theta \in C(x) \mid x).$$

Remark 2. The assumption of mlr of $\pi(\cdot)$ is somewhat restrictive, but it is a needed technical assumption, and is not a necessary condition for equation (2.6) to hold. Indeed, it can be shown that (2.6) holds for a Cauchy prior and Cauchy likelihood, and we have numerical evidence that it holds for a Cauchy prior and normal likelihood.

PROOF. We first calculate the LHS of (2.4). For $\theta_0 < x - c$, $p_A(x, \theta_0) = P(\theta \le \theta_0 \mid x)$. From (2.3), change variables to $t = \theta - \theta_0$, and define $y = x - \theta_0$. Then

$$P(\theta \le \theta_0 \mid x) = \frac{\int_{-\infty}^{0} f(y-t)\pi(t)dt}{\int_{-\infty}^{\infty} f(y-t)\pi(t)dt}, \tag{2.7}$$

and we want to calculate the supremum of (2.7) over the set $\{\theta_0 : \theta_0 < x - c\} = \{y : y > c\}$. Since f has mlr, the quantity $f(y-t)\pi(t)/f(c-t)\pi(t)$ is increasing in t for $y > c$. Applying Lemma A.1 of the Appendix we find that

$$\frac{\int_{x}^{\infty} f(y-t)\pi(t)dt}{\int_{x}^{\infty} f(c-t)\pi(t)dt},$$

is increasing in x for $y > c$. Setting $x = 0$ and $x = -\infty$ shows that the supremum of (2.7) over $\{y > c\}$ is achieved for $y = c$ and is equal to $1 - \gamma_C(c, 0)$, establishing (2.4).

To establish (2.5) we make the transformation $t = \theta - \theta_0$ and $y = x - \theta_0$ and show

$$\inf_{y : y \le c} \frac{\int_{y-c}^{\infty} f(t-y)\pi(t)dt}{\int_{-\infty}^{\infty} f(t-y)\pi(t)dt} = \frac{\int_{0}^{\infty} f(t-c)\pi(t)dt}{\int_{-\infty}^{\infty} f(t-c)\pi(t)dt}. \tag{2.8}$$

Apply Lemma A.1 with $g(t) = f(t-y)\pi(t)$, $h(t) = f(t-c)\pi(t)$, and $a(x) = x+y-c$. Since $\pi(x-\theta)$ has mlr and $c > y$, it follows that

$$\frac{g(a(x))a'(x)}{h(x)} = \frac{f(x-c)\pi(x+y-c)}{f(x-c)\pi(x)}$$

is increasing in x, and thus

$$\frac{\int_{x+y-c}^{\infty} f(t-y)\pi(t)dt}{\int_{x}^{\infty} f(t-c)\pi(t)dt}$$

is increasing in x. Evaluating the function at $x = 0$ and $x = -\infty$ establishes (2.8) and (2.5), proving the theorem. □

Allowing $\pi(\cdot)$ to depend on θ_0 is really a "testing-type" prior. The equivalence between the testing and set estimation errors is thus somewhat more striking. We can turn the problem around, and ask if the equivalence holds using an "interval-type" prior that does not depend on θ_0. In that case, the posterior probability that $\theta \in C(x)$ is equal to $\gamma_C(x, 0)$ and the following corollary is immediate.

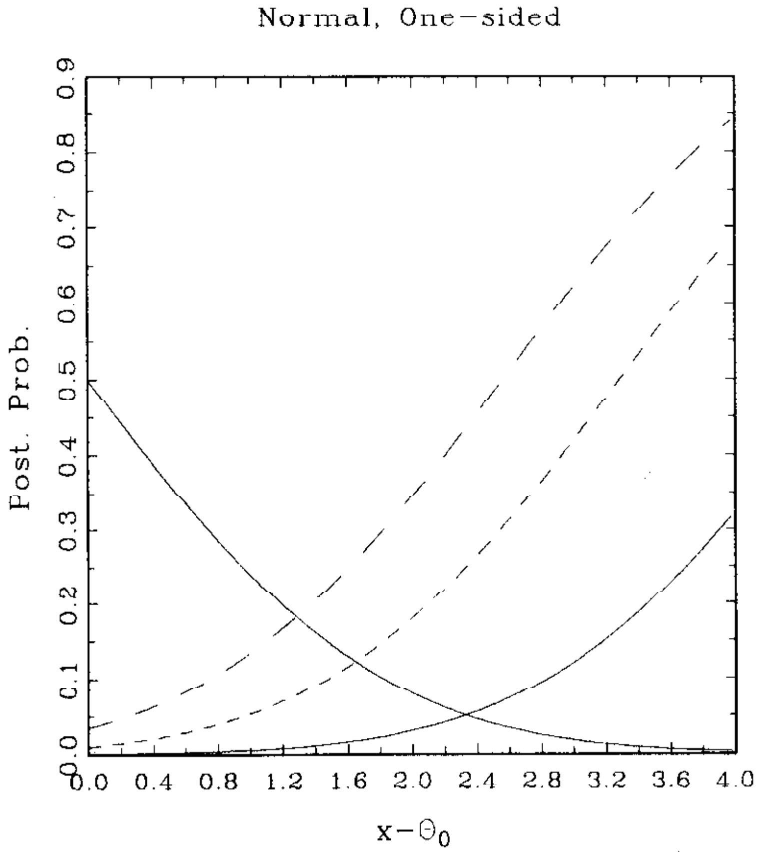

Fig. 1. The one-sided testing problem $H_0 : \theta \le \theta_0$ vs. $H_1 : \theta > \theta_0$, where $X \sim n(\theta, 1)$ and $\theta \sim n(\theta_0, 1)$. The solid line is $P(\theta \le \theta_0 \mid x)$. The posterior coverage probabilities of the interval $C(x) = \{\theta : \theta > x - c\}$ are given for $c = 1.282$ (long dashes), $c = 1.645$ (closely spaced dots), $c = 1.96$ (short dashes) and $c = 2.326$ (dots). The intersections are at the value of c.

COROLLARY 2.1. *If $\pi(\cdot)$ is independent of θ_0, then*

$$\sup_{\theta_0 \notin C(x)} p_A(x, \theta_0) = 1 - \gamma_C(x, 0).$$

To illustrate Theorem 2.1 we look at the following example.

Example 2.1. Suppose $X \sim n(\theta, \sigma^2)$ and $\theta \sim n(\theta_0, \tau^2)$ where σ^2 and τ^2 are known and (2.1) are the hypotheses of interest. The acceptance region and interval of (2.2) become $A(\theta_0) = \{x : x \le \theta_0 + c\}$ and $C(x) = \{\theta : \theta \ge x - c\}$. It is straightforward to calculate

$$P(\theta \le \theta_0 \mid x) = P\left(Z \le -\sqrt{B}\left[\frac{x - \theta_0}{\sigma}\right]\right) \tag{2.9}$$

and

$$\gamma_C(x, \theta_0) = P(Z \geq [(1 - B)(x - \theta_0) - c]/\sqrt{\sigma^2 B}), \tag{2.10}$$

where $Z \sim n(0, 1)$ and $B = \tau^2/(\sigma^2 + \tau^2)$. Since $P(\theta \leq \theta_0 \mid x)$ is increasing in θ_0 and $\gamma_C(x, \theta_0)$ is decreasing in θ_0, it follows that

$$\begin{aligned} \sup_{\theta_0 : \theta_0 < x - c} P(\theta \leq \theta_0 \mid x) &= P(\theta \leq x - c \mid x) \\ &= 1 - \inf_{\{\theta_0 : \theta_0 \geq x - c\}} \gamma_C(x, \theta_0) \\ &= P(Z \leq -c\sqrt{B}). \end{aligned} \tag{2.11}$$

The quantities in (2.9) and (2.10) are illustrated in Fig. 1. It can be seen that $P(\theta \leq \theta_0 \mid x)$ is decreasing in $x - \theta_0$ and always intersects the increasing (in $x - \theta_0$) $1 - \gamma_C(x, \theta_0)$ at c.

3. The two-sided case

In contrast with the frequentist situation, the Bayesian paradigm treats the two-sided hypothesis test in an entirely different way from the one-sided case. This reflects mainly on the form of the priors used in the point null case $H_0 : \theta = \theta_0$, where a point mass is put on θ_0. It turns out that the correspondence between two-sided testing and set estimation is weaker than in the one-sided case. However, if a set estimation-type prior distribution is used, then the correspondence is stronger.

Before proceeding we restate and strengthen our assumptions. As before, we assume that $f(\cdot)$ is continuous, symmetric and unimodal, but now we also assume that $f(\cdot)$ has a bounded derivative. Also, we assume that $f(\cdot)$ has the TP_3 property (Karlin (1968), Brown *et al.*, (1981)), which guarantees that the distribution of $|X|$ has mlr and implies TP_2, which is exactly the property of monotone likelihood ratio. We continue to make the same assumptions about $\pi(\cdot)$, but will sometimes also need $\pi(\cdot)$ to have TP_3. This will be explicitly stated.

3.1 *Priors putting mass on a point*

For testing

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta \neq \theta_0 \tag{3.1}$$

a typical prior distribution is of the form

$$\pi^*(\theta \mid \theta_0) = \begin{cases} \pi_0 & \text{if} \quad \theta = \theta_0 \\ (1 - \pi_0)\pi(\theta - \theta_0) & \text{if} \quad \theta \neq \theta_0, \end{cases} \tag{3.2}$$

where π_0 is a specified constant, and $\pi(\theta - \theta_0)$ is a density function. We assume that $\pi(\cdot)$ of (3.2) is continuous and symmetric, and that the value of π_0 is a constant independent of θ_0. An acceptance region and confidence interval corresponding to (3.1) are given by

$$A(\theta_0) = \{x : |x - \theta_0| \leq c\} \quad \text{and} \quad C(x) = \{\theta : |x - \theta| \leq c\}. \tag{3.3}$$

Under the prior of (3.2), the probability that H_0 is true is

$$P(\theta = \theta_0 \mid x) = \frac{\pi_0 f(x-\theta_0)}{\pi_0 f(x-\theta_0) + (1-\pi_0)\int_{-\infty}^{\infty} f(x-\theta)\pi(\theta-\theta_0)d\theta} \tag{3.4}$$

and substituting in (1.5) we obtain $p_A(x,\theta_0)$, the post-data accuracy of $A(\theta_0)$. Similarly, the post-data accuracy of $C(x)$, using the loss L_2, is

$$\gamma_C(x,\theta_0) = \frac{\pi_0 f(x-\theta_0)I(\theta_0 \in C(x)) + (1-\pi_0)\int_{C(x)} f(x-\theta)\pi(\theta-\theta_0)d\theta}{\pi_0 f(x-\theta_0) + (1-\pi_0)\int_{-\infty}^{\infty} f(x-\theta)\pi(\theta-\theta_0)d\theta}. \tag{3.5}$$

Unlike the one-sided case, there is no equivalence between $p_A(x,\theta_0)$ and $\gamma_C(x,\theta_0)$ as expressed in Theorem 2.1. However, we are able to calculate suprema and infima, which are given in the following theorem. Its proof is given in the Appendix.

THEOREM 3.1. *Under the assumptions in this section*

$$\sup_{\theta_0:|x-\theta_0|>c} p_A(x,\theta_0) = \frac{\pi_0 f(c)}{\pi_0 f(c) + (1-\pi_0)\int_{-\infty}^{\infty} f(c-t)\pi(t)dt}. \tag{3.6}$$

Furthermore, if $\pi(\cdot)$ has TP_3, then

$$\inf_{\theta_0:|x-\theta_0|\le c} \gamma_C(x,\theta_0) = \frac{\pi_0 f(c) + (1-\pi_0)\int_{-c}^{c} f(u)\pi(c-u)du}{\pi_0 f(c) + (1-\pi_0)\int_{-\infty}^{\infty} f(u)\pi(c-u)du}. \tag{3.7}$$

Remark 3. Depending on whether $\theta_0 \in C(x)$, the accuracy estimate in (3.5) falls on either side of the estimate

$$\gamma'_C(x,\theta_0) = \frac{\int_{C(x)} f(x-\theta)\pi(\theta-\theta_0)d\theta}{\int_{-\infty}^{\infty} f(x-\theta)\pi(\theta-\theta_0)d\theta},$$

which is what might have been used if the problem were treated as one of interval estimation that is, using an interval estimation prior. In fact, we can write

$$\gamma_C(x,\theta_0) = \frac{M(x-\theta_0)I(\theta_0 \in C(x)) + \gamma'_C(x,\theta_0)}{M(x-\theta_0)+1}, \tag{3.8}$$

where

$$M(y) = \frac{\pi_0}{1-\pi_0}\frac{f(y)}{\int_{-\infty}^{\infty} f(y-t)\pi(t)dt} \tag{3.9}$$

is the posterior odds ratio for the hypotheses (3.1). From (3.8) it is immediate that

$$\gamma_C(x,\theta_0)I(\theta_0 \notin C(x)) < \gamma'_C(x,\theta_0) < \gamma_C(x,\theta_0)I(\theta \in C(x)).$$

Example 3.1. Suppose $X \sim n(\theta, \sigma^2)$ and the prior on θ is

$$\pi(\theta \mid \theta_0) = \begin{cases} \pi_0 & \text{if } \theta = \theta_0 \\ (1-\pi_0) \times n(\theta_0, \tau^2) & \text{if } \theta \neq \theta_0 \end{cases}$$

where τ^2 is known. For the hypothesis of (3.1), we have $P(\theta = \theta_0 \mid x) = M/(M+1)$, where M is the posterior odds ratio, given by

$$M = \frac{\pi_0}{1-\pi_0}\sqrt{1-B}\exp\left\{\frac{-B}{2\sigma^2}(x-\theta_0)^2\right\},$$

where, again, $B = \tau^2/(\sigma^2+\tau^2)$. It is easy to see that $P(\theta = \theta_0 \mid x)$ is decreasing in $|x-\theta_0|$, so

$$\sup_{\{\theta_0:|x-\theta_0|>c\}} p_A(x,\theta_0) = P(\theta = x-c \mid x).$$

The post-data accuracy estimate for $C(x)$, $\gamma_C(x,\theta_0)$, is given by (3.8) with M of (3.1) and

$$\gamma_C'(x,\theta_0) = P\left(\sqrt{\frac{1}{\sigma^2 B}}[(1-B)(x-\theta_0)-c] \leq Z \leq \sqrt{\frac{1}{\sigma^2 B}}[(1-B)(x-\theta_0)+c]\right). \tag{3.10}$$

Since $\gamma_C(x,\theta_0)$ is also decreasing in $|x-\theta_0|$, we have

$$\inf_{\{\theta_0:|x-\theta_0|\leq c\}} \gamma_C(x,\theta_0) = P(\theta = x-c \mid x) + P(\theta \neq x-c \mid x) \times P\left(\frac{-c}{\sigma}\sqrt{B} \leq Z \leq \frac{c}{\sigma}\frac{1+B(\sigma^2/\tau^2)}{\sqrt{B}}\right). \tag{3.11}$$

Values of $P(\theta = \theta_0 \mid x)$ and $\gamma_C(x,\theta_0)$ are shown in Fig. 2 for various values of c. It can be seen that, for the most part, $P(\theta = \theta_0 \mid x)$ is smaller than $\gamma_C(x,\theta_0)$. However, for large c (small alpha) the order is reversed.

For fixed c, (3.6) and (3.7) are both increasing functions of π_0 and there is a unique $\pi_0^* = \pi_0^*(c)$ for which equality holds between (3.6) and one minus (3.7), and we have

$$\sup_{\theta_0:|x-\theta_0|>c} p_A(x,\theta_0) \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} 1 - \inf_{\theta_0:|x-\theta_0|\leq c} \gamma_C(x,\theta_0) \quad \text{as} \quad \pi_0 \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} \pi_0^*(c).$$

In general this relationship is artificial, since it is not usual for c and π_0 to be chosen in a dependent way. But it should come as no surprise, given the dependence of the posterior probability $P(\theta_0 \mid x)$ on π_0, and the subjective choice of π_0. In the next section, however, we look into another methodology that yields a less artificial equivalence.

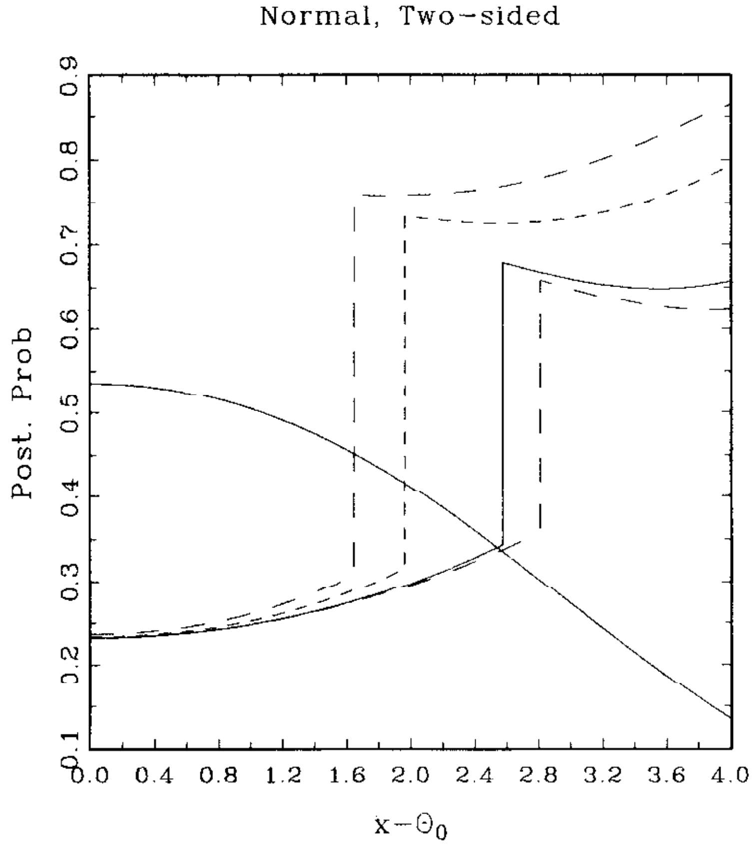

Fig. 2. The two-sided testing problem $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$, where $X \sim n(\theta, 1)$ and θ_0 has prior probability $\frac{1}{2}$, and $\theta \neq \theta_0$ has density $\frac{1}{2}n(\theta_0, 1)$ otherwise. The solid line is $P(\theta = \theta_0 \mid x)$. The posterior coverage probabilities of the interval $C(x) = \{\theta : |\theta - x| \leq c\}$ are given for $c = 1.645$ (long dashes), $c = 1.96$ (closely spaced dots), $c = 2.326$ (short dashes), $c = 2.574$ (dots) and $c = 2.807$ (dots and dashes).

3.2 *Continuous priors*

If interest were only in interval estimators, then a continuous prior would most likely be used. With such a prior, however, $P(\theta = \theta_0 \mid x)$ cannot be evaluated. Fortunately, in the decision-theoretic framework of Section 2, we can obtain accuracy estimates of a test based on a continuous prior. This is done by using an interval estimate to construct an estimate of testing accuracy.

For the hypothesis of (3.1), we can estimate the accuracy of $A(\theta_0)$ of (3.3) in the following way. Starting from $C(x)$ of (3.3), create the family of intervals

$$C_k(\theta_0) = \{\theta : |\theta_0 - \theta| \leq k\}$$

and the corresponding accuracy measure of $A(\theta_0)$:

$$p_k(x, \theta_0) = P(\theta \in C_k(\theta_0) \mid x) I(x \notin A(\theta_0)) + P(\theta \notin C_k(\theta_0) \mid x) I(x \in A(\theta_0)). \tag{3.12}$$

Of course, $p_k(x, \theta_0)$ is a Bayes rule for the interval hypothesis $H_0 : |\theta_0 - \theta| \leq k$. The advantage of the measure $p_k(x, \theta_0)$ is that it can be evaluated without using a prior that puts mass on θ_0. We consider a continuous prior $\pi(\theta - \theta_0)$, as might be used for interval estimation, calculate $P(\theta \in C_k(\theta_0) \mid x)$, and see that it is also the posterior probability of the hypothesis $H_0 : |\theta_0 - \theta| \leq k$. Thus, by constructing our probability from the confidence interval, we are, in effect, replacing the point null by an interval null. We have the following theorem, which shows the relationship between $p_k(x, \theta_0)$ and $\gamma_C(x, \theta_0)$.

THEOREM 3.2. *For the hypotheses in* (3.1), *where* $X \sim f(x - \theta)$, *continuous and symmetric with* TP_3, *and* $\pi(\theta - \theta_0)$ *continuous and symmetric, there exists a value* $k^* = k^*(c)$ *such that*

$$\sup_{\{\theta_0 : |x - \theta_0| > c\}} p_{k^*}(x, \theta_0) = 1 - \inf_{\{\theta_0 : |x - \theta_0| \leq c\}} \gamma_C(x, \theta_0). \tag{3.13}$$

PROOF. For $|x - \theta_0| > c$, $p_k(x, \theta_0) = 1 - P(\theta \in C_k(\theta_0) \mid x)$. Thus, to compute the left-hand side of (3.13) we can compute

$$\inf_{\theta_0 : |x - \theta_0| > c} P(\theta \in C_k(\theta_0) \mid x) = \inf_{y : |y| > c} \frac{\int_{-k}^{k} f(y - t)\pi(t)dt}{\int_{-\infty}^{\infty} f(y - t)\pi(t)dt}. \tag{3.14}$$

Write

$$\frac{\int_{-k}^{k} f(y - t)\pi(t)dt}{\int_{-k}^{k} f(c - t)\pi(t)dt} = \frac{\int_{0}^{k} [f(y - t) + f(y + t)]\pi(t)dt}{\int_{0}^{k} [f(c - t) + f(c + t)]\pi(t)dt}. \tag{3.15}$$

Since $f(\cdot)$ has TP_3, the function in square brackets (which is the density of $|X|$) has mlr. An application of Lemma A.1 shows that (3.15) is decreasing in k for $y > c > 0$. Thus the infimum in (3.14) is attained at $y = c$, and

$$\sup_{\theta_0 : |x - \theta_0| < c} p_k(x, \theta_0) = 1 - \frac{\int_{-k}^{k} f(c - t)\pi(t)dt}{\int_{-\infty}^{\infty} f(c - t)\pi(t)dt}. \tag{3.16}$$

The right-hand side of (3.16) is a monotone function of k that takes values from 0 to 1. Since the right-hand side of (3.13) is constant in k, equation (3.13) has a unique solution k^*. □

Note that the right-hand side of (3.13) is not directly computable, but it is not needed for the theorem. However, if the prior also has TP_3, we can establish that the infimum is attained for $|x - \theta_0| = c$ and hence compute it. This is illustrated in Example 3.2.

Example 3.2. Suppose $X \sim n(\theta, \sigma^2)$ and $\theta \sim n(\theta_0, \tau^2)$ where σ^2 and τ^2 are known. For the interval $C(x) = \{\theta : |x - \theta| \leq c\}$, an accuracy measure for $H_0 : \theta = \theta_0$ could be based on

$$P(\theta \in C_k(\theta_0) \mid x) = P\left(\frac{B(\theta_0 - x) - k}{\sqrt{\sigma^2 B}} < Z < \frac{B(\theta_0 - x) + k}{\sqrt{\sigma^2 B}}\right), \tag{3.17}$$

where $B = \tau^2/(\sigma^2 + \tau^2)$ and $Z \sim n(0,1)$. The post-data accuracy measure of $C(x)$ is $\gamma'_C(x, \theta_0)$ of (3.10) and, as both (3.10) and (3.17) are decreasing functions of $x - \theta_0$, we have

$$\sup_{\{\theta_0 : |x-\theta_0|>c\}} P(\theta \in C_k(\theta_0) \mid x) = P\left(\frac{-Bc-k}{\sigma\sqrt{B}} \leq Z \leq \frac{-Bc+k}{\sigma\sqrt{B}}\right)$$

$$\inf_{\{\theta_0 : |x-\theta_0|\leq c\}} \gamma'_C(x, \theta_0) = P\left(\frac{-Bc}{\sigma\sqrt{B}} \leq Z \leq \frac{c(2-B)}{\sigma\sqrt{B}}\right).$$

Using (3.2), it is now a simple matter to solve for the value of k^* to satisfy (3.13), and selected values are given in Table 3.1. There is a remarkable agreement between the values of Table 3.1 and Table 2 of Berger and Delampady (1987), which gives bounds on (standardized) ϵ such that the hypothesis $H_0 : |\theta - \theta_0| \leq \epsilon$ has approximately the same p-value as the exact hypothesis $H_0 : \theta = \theta_0$.

Table 3.1.

Values of k^* satisfying (3.13)

α	c	k^*
0.200	1.282	0.250
0.100	1.645	0.213
0.050	1.960	0.191
0.020	2.326	0.172
0.010	2.576	0.156
0.005	2.807	0.148
0.001	3.291	0.125

4. Discussion

In the frequentist paradigm, the testing/interval estimation duality has long been employed to both construct and evaluate these statistical procedures. This correspondence is a pre-data one, however, and did not translate to a post-data equivalence. On the other hand, Bayesians tend to treat intervals and testing (especially with point nulls) as two different entities, employing different priors in either situation. There has never been much effort directed toward establishing any Bayesian post-data testing/interval relationship.

We find that a strong post-data relationship exists in the one-sided testing case. By employing decision theory to develop the form of the post-data accuracy measures (as loss estimates), the relationship in the one-sided case is established. A similar relationship does not hold in the two-sided case, however, again demonstrating that Bayesian testing is vastly different from interval estimation in the two-sided case.

Most interestingly, it seems that the p-value occupies a middle ground. The testing/interval accuracy relationship is valid almost by definition for both one-sided and two-sided tests, as the p-value corresponds to a flat prior on the location

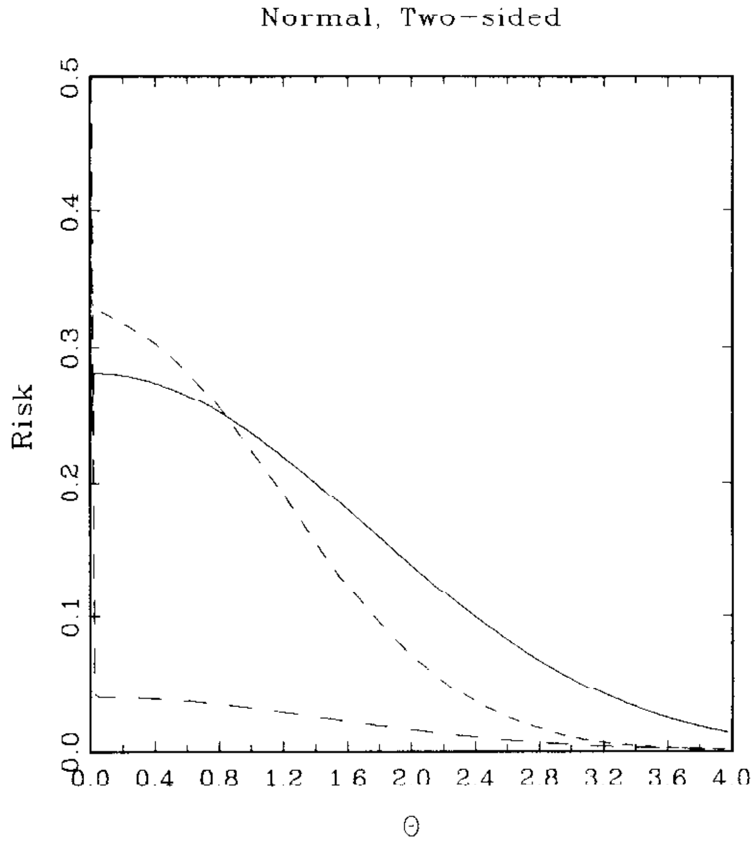

Fig. 3. The two-sided testing problem $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$, where $X \sim n(\theta, 1)$. Shown are risks, using the loss L_4 of (1.2) for estimates $p_A(x, \theta_0)$ based on a point-mass prior (solid line), $p_k(x, \theta_0)$ based on a continuous prior (dashes), and the p-value (dotted). Note that the risk at $\theta = 0$ is .23 for the point-mass prior estimate, .651 for the continuous prior estimate, and .33 for the p-value.

parameter. So we have a relationship between a post-data measure (the p-value) and a pre-data measure (the confidence coefficient) that yields a pre-data bound on the post-data accuracy measure. This equivalence is quite general, making essentially no assumptions about the form of the density or the forms of the tests or confidence sets. It follows directly from the frequentist equivalence of the construction of confidence sets as inverted tests.

This is a distinct difference from the fully Bayesian setup, and more closely mimics the frequentist correspondence. Thus, it seems that the p-value borrows properties from both the Bayesians and frequentists. It allows a post-data testing/interval correspondence in a manner that somewhat mimics the pre-data frequentist correspondence. This "middle ground" seems to carry over to risk considerations. Figure 3 shows the risk (using the loss L_4 of (1.2) of three estimators: the point-null estimator, the interval-null estimator, and the p-value. As can be seen, the p-value seems to occupy a middle risk between these other two estimators.

Acknowledgements

We would like to thank the editor and two referees for valuable suggestions that improved the paper.

Appendix

LEMMA A.1. *Let $h(\cdot)$ and $g(\cdot)$ be positive functions, and let $a(\cdot)$ be a differentiable function so that $g[a(x)]a'(x)/h(x)$ is a decreasing (increasing) function of x. Then, if the integrals exist,*

$$\frac{\int_{a(x)}^{\infty} g(t)dt}{\int_{x}^{\infty} h(t)dt} \tag{A.1}$$

is a decreasing (increasing) function of x.

PROOF. Differentiating (A.1) with respect to x and simplifying will show that the sign of the derivative is given by the sign of

$$\int_{x}^{\infty} \{h(x)g[a(t)]a'(t) - h(t)g[a(x)]a'(x)\}dt. \tag{A.2}$$

For $t > x$ the integrand in (A.2) is negative (positive) depending on whether $g[a(x)]a'(x)/h(x)$ is a decreasing (increasing) function. □

LEMMA A.2. *Under the assumptions of Section 3, the function $M(y)$ given by (3.9) is decreasing for $y > 0$ and increasing for $y < 0$.*

PROOF. By symmetry we only need to give details for $y > 0$. Differentiating, $M(y)$ is decreasing if

$$f'(y)\int_{\infty}^{\infty} f(y-t)\pi(t)dt - f(y)\int_{\infty}^{\infty} f'(y-t)\pi(t)dt < 0 \tag{A.3}$$

(interchange of derivative and integral is permissible since f' is assumed bounded). The left-hand side of (A.3) equals

$$\begin{aligned} &\int_{\infty}^{0} \{f'(y)f(y-t) - f(y)f'(y-t)\}\pi(t)dt \\ &\qquad + \int_{0}^{\infty} \{f'(y)f(y-t) - f(y)f'(y-t)\}\pi(t)dt \\ &= \int_{0}^{\infty} \{f'(y)f(y+u) - f(y)f'(y+u)\}\pi(u)du \\ &\qquad + \int_{0}^{\infty} \{f'(y)f(y-t) - f(y)f'(y-t)\}\pi(t)dt \\ &= \int_{0}^{\infty} \{f'(y)[f(y+t) + f(y-t)] - f(y)[f'(y+t) + f'(y-t)]\}\pi(t)dt, \end{aligned} \tag{A.4}$$

where in the above we have used the symmetry of $\pi(\cdot)$ and $f(\cdot)$. Now, if $f(y-t)$ is symmetric and has TP_3, then for $t > 0$ and $y > 0$, $[f(y+t) + f(y-t)]/[2f(y)]$ is increasing in y so its derivative with respect to y is positive. This implies

$$\frac{f'(y)}{f(y)} < \frac{f'(y+t) + f'(y-t)}{f(y+t) + f(y-t)}$$

so the integrand of (A.4) is negative for $t > 0$, and hence (A.3) holds. □

Proof of Theorem 3.1. Since $|x - \theta_0| > c$ is equivalent to $x \notin A(\theta_0)$, the post-data accuracy $p_A(x, \theta_0)$ equals $P(\theta = \theta_0 \mid x)$ of (3.4). Hence to compute (3.6) we need to compute

$$\sup_{\theta_0 : |x-\theta_0| > c} M(x - \theta_0), \tag{A.5}$$

where $M(y)$ is given by (3.9) for $t = \theta - \theta_0$. From Lemma A.2, $M(y)$ is increasing for $y < 0$ and decreasing for $y > 0$, so the supremum in (A.5) is attained at either boundary, establishing (3.6).

To establish (3.7), again let $y = x - \theta_0$, $u = x - \theta$. By symmetry it suffices to consider only $y \geq 0$. Substituting in the definition of $\gamma_C(x, \theta_0)$, it suffices to show that the function

$$\gamma(y) = \frac{\pi_0 f(y) + (1 - \pi_0) \int_{-c}^{c} f(u)\pi(y-u)du}{\pi_0 f(y) + (1 - \pi_0) \int_{-\infty}^{\infty} f(u)\pi(y-u)du}$$

is decreasing in y.

Under the assumption that $\pi(\cdot)$ is TP_3, the ratio

$$\frac{\int_{-c}^{c} f(u)\pi(y-u)du}{\int_{\infty}^{\infty} f(u)\pi(y-u)du} \tag{A.6}$$

is a decreasing function of y, so for $y_2 < y_1$ we obtain

$$\begin{aligned} &\int_{-c}^{c} f(u)\pi(y_1 - u)du \int_{-\infty}^{\infty} f(u)\pi(y_2 - u)du \\ &\quad \leq \int_{-c}^{c} f(u)\pi(y_2 - u)du \int_{-\infty}^{\infty} f(u)\pi(y_1 - u)du. \end{aligned} \tag{A.7}$$

Using this, a sufficient condition for the function $\gamma(y)$ to be decreasing is

$$\begin{aligned} &f(y_2)\int_{-c}^{c} f(u)\pi(y_1 - u)du + f(y_1)\int_{-\infty}^{\infty} f(u)\pi(y_2 - u)du \\ &\quad \leq f(y_1)\int_{-c}^{c} f(u)\pi(y_2 - u)du + f(y_2)\int_{-\infty}^{\infty} f(u)\pi(y_1 - u)du \end{aligned} \tag{A.8}$$

or, equivalently, that the function

$$N(y) = \frac{\dfrac{f(y)}{\int_{-\infty}^{\infty} f(u)\pi(y-u)du}}{1 - \dfrac{\int_{-c}^{c} f(u)\pi(y-u)du}{\int_{-\infty}^{\infty} f(u)\pi(y-u)du}}$$

is decreasing. The numerator of $N(y)$ is decreasing by an argument similar to the one used for the function $M(y)$ in Lemma A.2. Since (A.6) is decreasing, the denominator is increasing, hence $N(y)$ is decreasing, establishing (A.8) which implies (3.7) and completes the theorem. □

References

Berger, J. O. and Delampady, M. (1987). Testing precise hypotheses, *Statist. Sci.*, **2**, 317–352.

Berger, J. O. and Sellke, T. (1987). Testing a point null hypothesis: the irreconciliability of p-values and evidence (with discussion), *J. Amer. Statist. Assoc.*, **82**, 112–122.

Brown, L. D., Johnstone, I. M., and MacGibbon, K. B. (1981). Variation diminishing transformations: a direct approach to total positivity and its statistical applications, *J. Amer. Statist. Assoc.*, **76**, 824–832.

Casella, G. and Berger, R. L. (1987). Reconciling evidence in the one-sided testing problem (with discussion), *J. Amer. Statist. Assoc.*, **82**, 106–111.

George, E. I. and Casella, G. (1994). An empirical Bayes confidence report, *Statistica Sinica*, **4**, 617–638.

Goutis, C. and Casella, G. (1992). Increasing the confidence in Student's t interval, *Ann. Statist.*, **20**, 1501–1513.

Hwang, J. T. and Pemantle, R. (1990). Evaluation of estimators of statistical significance under a class of proper loss functions, *Statist. Decisions* (to appear).

Hwang, J. T., Casella, G., Robert, C., Wells, M. T., and Farrell, R. H. (1992). Estimation of accuracy in testing, *Ann. Statist.*, **20**, 490–509.

Johnstone, I. M. (1988). On inadmissibility of some unbiased estimates of loss, *Statistical Decision Theory and Related Topics IV* (eds. J. O. Berger and S. S. Gupta), 361–379, Springer, New York.

Karlin, S. (1968). *Total Positivity*, Stanford University Press, Stanford, CA.

Lu, K. L. and Berger, J. O. (1989*a*). Estimation of normal means: Frequentist estimation of loss, *Ann. Statist.*, **17**, 890–906.

Lu, K. L. and Berger, J. O. (1989*b*). Estimated confidence procedures for multivariate normal means, *J. Statist. Plann. Inference*, **23**, 1–19.

Robert, C. R. and Casella, G. (1993). Improved confidence statements for the usual multivariate normal confidence set, *Statistical Decision Theory and Related Topics V* (eds. J. O. Berger and S. S. Gupta), 351–368, Springer, New York.

Rukhin, A. L. (1988*a*). Estimated loss and admissible loss estimators, *Statistical Decision Theory and Related Topics IV* (eds. J. O. Berger and S. S. Gupta), 409–418, Springer, New York.

Rukhin, A. L. (1988*b*). Loss functions for loss estimation, *Ann. Statist.*, **16**, 1262–1269.