



---

Estimated Confidence Under the Validity Constraint

Author(s): Jiunn T. Hwang and Lawrence D. Brown

Source: *The Annals of Statistics*, Vol. 19, No. 4 (Dec., 1991), pp. 1964-1977

Published by: Institute of Mathematical Statistics

Stable URL: <http://www.jstor.org/stable/2241914>

Accessed: 25/03/2010 15:47

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=ims>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).



*Institute of Mathematical Statistics* is collaborating with JSTOR to digitize, preserve and extend access to *The Annals of Statistics*.

<http://www.jstor.org>

## ESTIMATED CONFIDENCE UNDER THE VALIDITY CONSTRAINT<sup>1,2</sup>

BY JIUNN T. HWANG AND LAWRENCE D. BROWN

*Cornell University*

We examine the decision theoretic estimated confidence approach proposed by Kiefer, Robinson and Berger, and focus on results under the frequentist validity constraint previously described by Brown and by Berger. Our main result is that the usual constant coverage probability estimator for the usual confidence set of a linear model is admissible under the frequentist validity constraint. Note that it is inadmissible without the frequentist validity constraint when the dimension is at least 5. The criterion of admissibility under the frequentist validity constraint is shown to be quite a reasonable one. Therefore the constant coverage probability estimator which has been widely used is justifiable from the post-data point of view.

**1. Introduction.** Let  $C_X$  denote a confidence set for a parameter  $\theta$  based on observation of the random variable  $X$ . In such a problem, interest centers on the coverage function  $I(\theta \in C_X)$ , where

$$(1.1) \quad I(\theta \in C_X) = \begin{cases} 1, & \text{if } \theta \in C_X, \\ 0, & \text{otherwise.} \end{cases}$$

To report a confidence for  $C_X$ , often a frequentist considers just the coverage probability, which is  $E_\theta I(\theta \in C_X) \equiv P(\theta \in C_X)$ . One says that  $\gamma = 1 - \alpha$  is a valid confidence for  $C_X$  if

$$(1.2) \quad E_\theta I(\theta \in C_X) \geq \gamma, \quad \forall \theta.$$

In place of the constant value  $\gamma$ , it is potentially more informative to provide a data dependent estimate  $\gamma(X)$  of the value of the coverage function, an approach formulated in Kiefer (1977b) and called the estimated confidence approach by Berger (1988). The estimator  $\gamma(X)$  is also considered to be a post-data report, since it depends on the outcome of the data. The frequentist promise (1.2) should then be translated into

$$(1.3) \quad E_\theta I(\theta \in C_X) \geq E_\theta \gamma(X), \quad \forall \theta.$$

This guarantees that the estimator  $\gamma(X)$  is conservative in the long run since on the average it will not overestimate the coverage probability. Condition

---

Received September 1989; revised October 1990.

<sup>1</sup>Research supported by NSF Grant DMS-88-09016.

<sup>2</sup>Dedicated to the memory of Jiunn Hwang's loving mother, Ruey May Guo, who passed away on June 18, 1990.

AMS 1980 subject classifications. Primary 62A20, 62F25, 62C05; secondary 62C15.

Key words and phrases. Coverage probability, usual confidence set, decision theory, relevant subsets, validity admissibility.

(1.3) will be referred to as the (frequentist) validity constraint. Imposition of this constraint in closely related settings has been proposed in Brown (1978) (who called it guaranteed confidence) and in Berger (1985).

Many statisticians have pointed out the danger of reporting only the unconditional coverage probability and ignoring totally the conditional issue. The best known example is perhaps that of Cox (1958). The confidence set version is precisely described in Bondar (1988). Similar behavior in a less artificial setting is discussed in Casella, Hwang and Robert (1989); see also Berger [(1985), Example 1, page 17].

Perhaps for some statisticians, the examples of Cox, Bondar and Berger point out a need for conditioning on the ancillary statistics or some other suitable partition of the sample space. There are situations, however, in which no ancillary statistics exist and yet a conditional inference is definitely needed. Fieller's confidence interval (1954) is one such example. However, in general, it is not clear how to choose the partition to condition on, although some theory has been developed in Kiefer (1976, 1977a, b), Brownie and Kiefer (1977) and Brown (1978). [Results in Brown (1990) describe the danger of conditioning on an ancillary in point estimation problems. However, such conditioning, if properly performed, is highly appropriate when making confidence statements, see Brown (1990), page 490, 535–536.]

To choose an estimator  $\gamma(X)$  with good conditional properties, it appears reasonable to require  $\gamma(X)$  to mimic  $I(\theta \in C_X)$  as much as possible. Therefore one sets up a loss function such as the squared error loss

$$(1.4) \quad L(I(\theta \in C_X), \gamma(X)) = (\gamma(X) - I(\theta \in C_X))^2,$$

and one requires  $\gamma(X)$  to minimize (in the sense of decision theory) its risk function

$$R(\theta, \gamma(X)) = E_{\theta} L(I(\theta \in C_X), \gamma(X)).$$

Terminologies such as “as good as,” “betterness” (equivalently “domination”), “admissibility” and “inadmissibility” can be defined similarly to those in decision theory. Additionally, we propose that reasonable estimators should satisfy the validity condition (1.3). The resulting criterion is called validity admissibility (*V*-admissibility): An estimator is *V*-admissible if it is (frequentist) valid [i.e., it satisfies (1.3)] and it is admissible among the class of (frequentist) valid estimators. Obviously, admissibility implies *V*-admissibility, but the converse need not hold true.

The aforementioned paradoxical examples can be resolved under this framework. In particular, in the examples of Bondar (1988) and Berger [(1985), page 17], it can be shown that each of the unconditional coverage probabilities is an unreasonable estimate of confidence in that it is *V*-inadmissible and is dominated by the coverage probability conditioning on the ancillary statistics.

Alternative frequentist criteria are based on nonexistence of relevant, or semirelevant subsets. For a survey of relevant subset literature, see Casella

(1988a, b). Technically, the admissibility (respectively,  $V$ -admissibility) criterion is stronger than the nonexistence of relevant (negatively biased relevant) subsets; see Robinson (1979a).

Ordinary admissibility as well as existence of relevant subsets were studied in Robinson (1979a, b). Among other results he proves, under (1.4), inadmissibility of  $\gamma$  as an estimator for  $I(|X - \theta| \leq c)$ , where  $X$  is a  $p$ -dimensional  $N(\theta, I)$ ,  $p = 5$  and  $c$  is the cutoff point for which  $P(|X - \theta| \leq c) = \gamma$ ; see Proposition 4.1 of Robinson [(1979b), page 766]. Lu and Berger (1989) extended this inadmissibility result to  $p \geq 5$ . Robinson's improved estimator violates the frequentist validity criterion and hence it remains open as to whether  $\gamma$  is  $V$ -admissible.

In this paper, under the normal setting, we prove in Theorem 1 that  $\gamma$  is  $V$ -admissible for any dimension  $p$  for the squared error loss function (1.4). When  $\gamma > \frac{1}{2}$ , we also prove the same result in Theorem 2 with respect to the absolute error loss function

$$(1.5) \quad L(I(\theta \in C_X), \gamma(X)) = |\gamma(X) - I(\theta \in C_X)|.$$

In either case, for  $p$  large enough, the estimator  $\gamma$  is inadmissible without the frequentist validity constraint.

The squared error loss function (1.4) appears to be an attractive criterion especially from the Bayesian's point of view. The Bayes rule against this loss is exactly the posterior coverage probability. Loss functions satisfying such a property are said to be proper; see, for example, Schervish (1989). It is anticipated that the admissibility results of this paper will remain unchanged for all proper loss functions. A result of this nature is verified in testing problems by Hwang and Pemantle (1990).

In addition, under the squared error loss, there is the following connection. It can be shown that an admissible estimator under (1.4), where  $C_X$  is an arbitrary confidence set, is also admissible for estimating  $E_\pi(I(\theta \in C_X)|X)$  under the loss

$$(1.6) \quad (\gamma(X) - E_\pi(I(\theta \in C_X)|X))^2.$$

(The converse is also true.) Here the "parameter" is the prior distribution  $\pi$ , which is assumed to be unknown. From a Bayesian's point of view,  $E_\pi(I(\theta \in C_X)|X)$  is the "answer," where  $E_\pi(\cdot|X)$  denotes the expectation with respect to the conditional distribution of  $\theta$  given  $X$ . However,  $\pi$  is unknown and hence the "answer" is unobtainable and has to be estimated. The connection between (1.4) and (1.6) can be established easily by noting that for any estimator  $\gamma(X)$ ,

$$\begin{aligned} E_\pi(\gamma(X) - I(\theta \in C_X))^2 &= E_\pi(\gamma(X) - E_\pi(I(\theta \in C_X)|X))^2 \\ &\quad + E_\pi((E_\pi I(\theta \in C_X)|X) - I(\theta \in C_X))^2. \end{aligned}$$

Here  $E_\pi$  denotes the expectation averaging out both  $X$  and  $\theta \sim \pi$ .

Our results are proven within a canonical normal framework. However, they all apply to the linear model

$$Y = X\beta + \varepsilon,$$

where  $X$  is  $n \times p$  and has full rank  $p$  and  $\varepsilon \sim N(0, \sigma^2 I)$ . Assuming  $\sigma^2$  is known, the usual  $\gamma$  confidence set is

$$\{\beta: (\hat{\beta} - \beta)(X'X)(\hat{\beta} - \beta) \leq \sigma^2 \chi_p^2(\gamma)\},$$

where  $\hat{\beta}$  is the least squared estimator and  $\chi_p^2(\gamma)$  is the upper  $\gamma$ -quantile of a chi-squared distribution with  $p$  degrees of freedom. After a direct transformation and an easy argument that allows us to discard the data other than  $\hat{\beta}$ , we can reduce the problem to the canonical form. Hence the conclusion is that  $\gamma$  is  $V$ -admissible.

One of Casella's (1988b) objections to the estimated confidence approach is that "the mathematical statistics problems can be much harder, perhaps insoluble." We hope that by solving the problem here, we can open up a possible outlet for the frequentists' fundamental problem.

**2. Admissibility under the frequentist validity constraint.** Assume that  $X$  is a  $p$ -dimensional  $N(\theta, I)$  random vector. A standard  $\gamma$ ,  $0 < \gamma < 1$ , confidence set is

$$C_X = \{\theta: |\theta - X| \leq c\},$$

where  $c$  is the cutoff point such that  $P(\theta \in C_X) = \gamma, \forall \theta$ .

**THEOREM 1.** *Under the squared error loss (1.4), the constant estimator  $\gamma$  is  $V$ -admissible for all  $p$ . Also,  $\gamma$  is admissible for  $p \leq 3$ .*

Previously the admissibility of  $\gamma$  has only been demonstrated for  $p = 1$  in Robinson (1979b). He also attempted to prove the inadmissibility for  $p = 3$  without success; see Robinson [(1979b), page 766]. Of course, there is a good reason that such an attempt has to fail since, as will be demonstrated,  $\gamma$  is admissible for  $p = 3$ . It is also true that  $\gamma$  is admissible even for  $p = 4$ , which, together with other admissibility results for general Bayes estimators, were proved in Brown and Hwang (1989). Previously it was established in Lu and Berger (1989) as well as Robert and Casella (1989) that  $\gamma$  is inadmissible (without the validity constraint) for  $p \geq 5$  and for  $p = 5$  in Robinson (1979b).

**THEOREM 2.** *Under the absolute error loss (1.5),  $\gamma > \frac{1}{2}$  is  $V$ -admissible.*

The proofs of both theorems are based on the following lemma. This lemma actually considers a more general context. The goal is to estimate the function  $G(X - \theta)$  under the loss function

$$L(G(X - \theta), \gamma(X)),$$

where  $L$  is a function convex with respect to the second coordinate. For

application to the case of Theorems 1 and 2, we will take  $G(X - \theta) = I_{|X-\theta|\leq c}$  and  $L(a, b) = (a - b)^2$  or  $L(a, b) = |a - b|$ , respectively.

The validity constraint in this more general context is

$$(2.1) \quad E_{\theta}\gamma(X) \leq B, \quad \forall \theta,$$

where  $B$  is a certain fixed constant bound which is assumed to be independent of  $\theta$ . We can similarly define the pointwise validity constraint as

$$(2.2) \quad \gamma(x) \leq B, \quad \forall x.$$

In the context of Theorems 1 and 2,  $B = \gamma$ .

Pointwise frequentist validity obviously implies frequentist validity. Consequently,  $V$ -admissibility implies  $PV$ -admissibility, that is, admissibility under the pointwise frequentist validity constraint. The converse is not always true. However, under the assumption of the following lemma, the converse does hold.

LEMMA 3. *Assume that  $Z = X - \theta$  has a distribution independent of  $\theta$ . Let  $\gamma = B$ , a constant. Assume that there exists a frequentist valid estimator  $\gamma(X)$  as good as  $\gamma$ . Let  $\gamma^*(\theta)$  denote  $E_{\theta}\gamma(X)$ . Then  $\gamma^*(X)$  is pointwise valid and is as good as  $\gamma$ .*

*Further, if the support of  $Z$  is the whole Euclidean space and the risk of any (bounded) estimator is continuous in  $\theta$ , then (i)  $\gamma^*(X)$  is better than  $\gamma$  and (ii)  $V$ -admissibility of  $\gamma$  is equivalent to its  $PV$ -admissibility.*

PROOF. Obviously,  $\gamma^*(X)$  is pointwise frequentist valid. To finish the proof for the first part, it suffices to prove that  $\gamma^*(X)$  is as good as  $\gamma$ .

Note that the risk function of  $\gamma$  is a constant and will be denoted by  $R^0$ . Hence

$$(2.3) \quad \begin{aligned} R^0 &\geq E_{\theta}L(g(X - \theta), \gamma(X)) \quad [\forall \theta] \\ &= EL(g(Z), \gamma(Z + \theta)). \end{aligned}$$

Let  $Z_1$  be another random vector independently identically distributed as  $Z$ . The last expression implies that

$$R^0 \geq E[L(g(Z), \gamma(Z + \theta + Z_1))|Z_1], \quad \forall \theta.$$

Taking expectation with respect to  $Z_1$ , we have

$$(2.4) \quad \begin{aligned} R^0 &\geq E[L(g(Z), \gamma(Z + \theta + Z_1))] \\ &\geq E[L(g(Z), E[\gamma(Z + \theta + Z_1)|Z])], \end{aligned}$$

by Jensen's inequality. The lower bound is

$$EL(g(Z), \gamma^*(Z + \theta)) = EL(g(X - \theta), \gamma^*(X)).$$

Hence  $\gamma^*(X)$  is as good as  $\gamma$ .

For (ii) in the second part, obviously  $V$ -admissibility implies  $PV$ -admissibility. Hence what is left to show is that  $V$ -inadmissibility of  $\gamma$  implies  $PV$ -inadmissibility of  $\gamma$ . This obviously follows from (i), which is established below.

By continuity assumption of the risk function, strict inequality in (2.3) holds for  $\theta$  in at least an open set. This and the assumption that  $Z$  has a whole support implies the first inequality in (2.4) is strict, completing the proof.  $\square$

Lemma 3 applies to Theorem 1 and 2 with  $B = \gamma$ . In particular, the continuity assumption of the risk function holds due to the following reason. Note that one can assume without loss of generality that  $\gamma(\cdot)$  is bounded above by one. This, together with the dominated convergence theorem and the fact that  $I_{|\theta-x|\leq c}$  is continuous in  $\theta$  for almost every  $x$  implies that the risk function is continuous in  $\theta$ .

Lemma 3 implies that in proving  $V$ -admissibility of  $\gamma$ , it suffices to prove  $PV$ -admissibility. In order to prove  $PV$ -admissibility, we adapt Blyth's method, taking into account the constraint (2.2), which reduces now to

$$(2.5) \quad \gamma(x) \leq \gamma, \quad \forall x.$$

To apply Blyth's method, we turn to a discussion of Bayes estimators. For a prior measure with density function  $\pi(\theta)$ , let

$$\pi(\theta|x) = \varphi(x - \theta)\pi(\theta) / \int \varphi(x - \theta)\pi(\theta) d\theta$$

be the formal posterior probability density of  $\theta$  given  $X = x$ . Here  $\varphi(x - \theta)$  denotes the p.d.f. of the  $p$ -variate  $N(\theta, I)$  distribution.

**SQUARED ERROR.** We will deal with the squared error loss function (1.4) first. The unrestricted Bayes rule  $\gamma^\pi$  chooses  $a$  to minimize

$$E\left[(a - I_{|\theta-X|\leq c})^2 | X\right]$$

and hence

$$\gamma^\pi(X) = E(I_{|\theta-X|\leq c} | X) = P(|\theta - X| \leq c | X).$$

However, under the constraint (2.5), the Bayes rule is  $\gamma^\pi(X) \wedge \gamma$ . This follows from the fact that  $E(a - I_{|\theta-X|\leq c})^2 | X$  is a convex function of  $a$ .

We will specifically consider a prior measure with the density

$$(2.6) \quad g(\theta) = \left( \frac{1}{b + |\theta|^2} \right)^d,$$

where  $b$  will be taken to be a very large number and  $d > 0$  is some constant to be specified in Lemma 6. In deriving the generalized Bayes estimator we need

the following two lemmas. Below, we let

$$\gamma(c) \equiv \int_{|\theta| < c} \varphi(\theta) d\theta \quad \text{and} \quad \rho(c) \equiv \int_{|\theta| \leq c} \theta_i^2 \varphi(\theta) d\theta / \gamma.$$

Note that here and later  $\gamma$  is also denoted as  $\gamma(c)$  whenever there is a need to emphasize the dependence of  $\gamma$  on  $c$ . It follows from this notation that  $\gamma(\infty) = 1$ .

LEMMA 4. *Let  $0 < c \leq \infty$  and  $Q(\theta) = s + v'\theta + \theta'M\theta$ , where  $s$  is a scalar,  $v$  a vector and  $M$  is a  $p \times p$  matrix; and these are all constants independent of  $\theta$ . Then*

$$(2.7) \quad \int_{|\theta| < c} Q(\theta)\varphi(\theta) d\theta = (s + \rho(c)\text{tr } M)\gamma(c),$$

where  $\text{tr } M$  denotes the trace of  $M$ . Furthermore  $\rho(c)$  strictly increases to one as  $c \rightarrow \infty$  and hence  $0 < \rho(c) < 1$  for  $c < \infty$ .

PROOF. The only statement which is not obvious is the integration of the third term of  $Q(\theta)$ . Now

$$\int (\theta'M\theta)\varphi(\theta) d\theta = \text{tr } M \int_{|\theta| \leq c} \theta\theta'\varphi(\theta) d\theta = \rho(c)\gamma(c)\text{tr } M. \quad \square$$

Using this lemma and a Taylor expansion, we can establish the following formula in Lemma 5. Below,  $o$  and  $O$  are little  $o$  and big  $O$  terms uniform in  $|x|$  as  $b \rightarrow \infty$ . As an example,  $o(1/(b + |x|^2))$  satisfies

$$\lim_{b \rightarrow \infty} \left[ \sup_{|x|} (b + |x|^2) o\left(\frac{1}{b + |x|^2}\right) \right] \rightarrow 0.$$

LEMMA 5. *For  $c \leq \infty$  and  $g$  as in (2.6),*

$$(2.8) \quad \begin{aligned} & \int_{|\theta-x| \leq c} g(\theta)\varphi(\theta-x) d\theta \\ &= \gamma(c)g(x) \left( 1 + d\rho(c) \frac{|x|^2(2d-2-p) - bp}{(b+|x|^2)^2} + o\left(\frac{1}{b+|x|^2}\right) \right). \end{aligned}$$

PROOF. Let  $z = \theta - x$  and hence the left-hand side of (2.8) is

$$\int_{|z| \leq c} g(z+x)\varphi(z) dz.$$

Note that

$$|z+x|^2 = |x|^2 + \Delta,$$



where  $\Delta = 2zx + |z|^2$ . By a one-dimensional Taylor expansion,

$$g(z+x) = g(x) \left( 1 - \frac{d\Delta}{b+|x|^2} + \frac{d(d+1)}{(b+|x|^2)^2} \frac{\Delta^2}{2} \right) - \frac{d(d+1)(d+2)}{(\tau)^{d+3}} \frac{\Delta^3}{3!},$$

where  $\tau$  is between  $b+|x|^2$  and  $b+|x|^2+\Delta = b+|z+x|^2$ . Lemma 4 then implies

$$\int_{|z| \leq c} \Delta \varphi(z) dz = p\rho(c)\gamma(c)$$

and

$$\int_{|z| \leq c} \Delta^2 \varphi(z) dz = 4|x|^2 \rho(c)\gamma(c) + O(1).$$

We will also establish in the Appendix that

$$(2.9) \quad \int_{|z| < \infty} \frac{|\Delta|^3}{\tau^{d+3}} \varphi(z) dz = g(x) o\left(\frac{1}{b+|x|^2}\right).$$

Putting this together and rearranging terms, we obtain (2.8).  $\square$

It follows directly from this lemma that the unrestricted generalized Bayes estimator with respect to  $g(\theta)$  has the following representation

$$(2.10) \quad \begin{aligned} \gamma^g(x) &= \frac{\int_{|\theta-x| \leq c} g(\theta) \varphi(\theta-x) d\theta}{\int_{|\theta-x| < \infty} g(\theta) \varphi(\theta-x) d\theta} \\ &= \gamma(c) \left( 1 + d(\rho(c) - 1) \left[ \frac{|x|^2(2d+2-p) - bp}{(b+|x|^2)^2} \right] \right) + o\left(\frac{1}{b+|x|^2}\right). \end{aligned}$$

Hence we come to the following technically important conclusion:

**LEMMA 6.** *Assume  $p \geq 3$  and hence there exists an  $\varepsilon > 0$  such that  $d = (p-2-\varepsilon)/2 > 0$ . Then there exists  $b$  large enough such that*

$$(2.11) \quad \gamma^g(x) > \gamma(c) \equiv \gamma$$

for all  $x$ .

**PROOF.** With such a choice of  $d$ , note that the leading term on the right-hand side of (2.10) is greater than  $\gamma$ . Furthermore,

$$\liminf_{b \rightarrow \infty} (b+|x|^2)(\gamma^g(x) - \gamma) \geq \gamma d(1 - \rho(c)) \min(\varepsilon, p) > 0,$$

establishing the lemma.  $\square$

REMARKS. Note that if  $g$  were a proper prior, then the unrestricted Bayes estimator could not satisfy (2.11) for all  $x$ . This is due to the identity

$$E\gamma^g(X) = EE(I_{|\theta-X| \leq c} | X) = EE(I_{|\theta-X| \leq c} | \theta) = \gamma.$$

Coming back to the situation where (2.11) holds, we note that the Bayes estimator under the restriction (2.5) is  $\gamma^g \wedge \gamma = \gamma$ . Therefore, unlike in many problems, there are many priors (including the Lebesgue prior) that lead to  $\gamma$  as a generalized Bayes estimator under the validity constraint. To prove  $V$ -admissibility of  $\gamma$ , we can work with any such prior  $g$  instead of the Lebesgue prior. It turns out that Blyth's method can be successfully applied to such a  $g$ , as shown below whereas it cannot conveniently be worked out with the Lebesgue prior under the restriction (2.5) except when  $p \leq 3$ .

As a side track remark, (2.11) actually holds for  $\varepsilon = 0$  and  $b = 0$ . To establish the assertion, one can use the correlation inequality, that is, the correlation of two increasing functions of a random variable is positive and Poisson's Law which asserts that

$$E|Y|^{p-2} = \left( \frac{1}{D} \wedge \frac{1}{r} \right)^{p-2},$$

where  $Y$  is a random vector uniformly distributed over the (surface of the) sphere in  $R^p$  of radius  $r$  centered at a point a distance  $D$  from the origin. However, we do not pursue this direction for two reasons: (i) Lemma 6 provides us flexibility in choosing  $b$  which leads to a simpler Blyth's type argument below, (ii) the present approach may more readily be generalized to other situations where Poisson's Law does not apply.

PROOF OF THEOREM 1. We will apply Blyth's method to the case  $p \geq 3$ . Consider a sequence of finite prior measures with the following densities

$$g_n(\theta) = h_n(\theta)g(\theta),$$

where  $g(\theta)$  is as in Lemma 6 and

$$h_n(\theta) = \left( \frac{n}{n + |\theta|^2} \right)^{3/2}.$$

Note that  $g_n(\theta)$  is integrable.

Suppose that  $\gamma$  is  $V$ -inadmissible. By Lemma 3, there exists  $\gamma^*(x) \leq \gamma$  such that

$$E(\gamma^*(X) - I_{|\theta-X| \leq c})^2 < E(\gamma - I_{|\theta-X| \leq c})^2, \quad \forall \theta.$$

[We can assume without loss of generality that strict inequality holds for every

$\theta$ , since otherwise replacing  $\gamma^*(X)$  by  $(\gamma^*(X) + \gamma)/2$  and Jensen's inequality will do.] Since  $h_n$  is increasing in  $n$ , there exists a positive number  $\delta$  independent of  $n$  such that

$$\begin{aligned} \delta &< \int \left[ E(\gamma - I_{|\theta-x|\leq c})^2 - E(\gamma^*(x) - I_{|\theta-x|\leq c})^2 \right] g_n(\theta) d\theta \\ (2.12) \quad &\leq \int \left[ E(\gamma - I_{|\theta-x|\leq c})^2 - E(\gamma^{g_n}(x) \wedge \gamma - I_{|\theta-x|\leq c})^2 \right] g_n(\theta) d\theta \\ &= \int [(\gamma^{g_n}(x) \wedge \gamma) - \gamma]^2 \varphi(\theta - x) g_n(\theta) d\theta dx. \end{aligned}$$

To lead to a contradiction, we will show that the last expression approaches zero, as  $n \rightarrow \infty$ . To accomplish this, all we have to do is to show:

(2.13) the limit can be passed inside the integral (2.12).

This approach is similar to that in Brown and Hwang (1982). It is based on the fact that the limit of the integrand is zero, since

$$\gamma^{g_n} \rightarrow \gamma^g \geq \gamma$$

by the monotone convergence theorem and consequently  $(\gamma^{g_n} \wedge \gamma) - \gamma \rightarrow 0$ .

To justify (2.13), it suffices to find an integrable upper bound of the integrand and use the dominated convergence theorem. We first follow a (lengthy) calculation similar to those leading to (2.8) and conclude that

$$\gamma^{g_n} - \gamma = O\left((b + |x|^2)^{-1}\right),$$

where the big  $O$  is uniform in both  $n$  and  $b$ ,  $n \geq b$ . Now the last integrand of (2.12) is bounded above by

(2.14) 
$$O\left((b + |x|^2)^{-2}\right) \varphi(x - \theta) g(\theta).$$

Let  $\text{int}$  denote the integral of the last expression against  $x$  and  $\theta$ . To show that  $\text{int}$  is finite, we first integrate (2.14) with respect to  $x$  and apply Lemma 5 to get

$$\text{int} = \int O\left((b + |\theta|^2)^{-2}\right) (b + |\theta|^2)^{-d} d\theta.$$

Using the usual spherical transformation and the fact  $d = (p - 2 - \epsilon)/2$ , it can be shown that  $\text{int}$  is finite, establishing the theorem for  $p \geq 3$ .

For  $p \leq 3$ , we show below that  $\gamma$  is admissible (without the validity constraint), which, of course, implies  $V$ -admissibility.

Proceed as before except that  $g(\theta) \equiv 1$  and there is no validity constraint. Use the same  $h_n(\theta)$ . Note that to apply Blyth's method, it suffices now to

prove that the following quantity approaches zero:

$$\begin{aligned} & \int (\gamma^{g_n} - \gamma)^2 \varphi(x - \theta) g_n(\theta) d\theta dx \\ &= \int O\left((n + |x|^2)^{-2}\right) \varphi(x - \theta) dx g_n(\theta) d\theta \\ &= \int O\left((n + |\theta|^2)^{-2}\right) g_n(\theta) d\theta \\ &\leq \int O\left((n + |\theta|^2)^{-2}\right) d\theta. \end{aligned}$$

The monotone convergence theorem can be applied to show that the upper bound converges to zero if

$$\int (1 + |\theta|^2)^{-2} d\theta < \infty.$$

Equivalently,

$$\int (1 + |\theta|^2)^{-2} |\theta|^{p-1} d|\theta| < \infty.$$

The last inequality holds if and only if  $p < 4$ , completing the proof.  $\square$

**ABSOLUTE ERROR.** Now we turn to the problem of the absolute error loss. We first note that the Bayes rule corresponding to a prior  $g$  is

$$\gamma^g(x) = \begin{cases} 1, & \text{if } P(|\theta - X| \leq c | X = x) > \frac{1}{2}, \\ 0, & \text{otherwise,} \end{cases}$$

where  $P(\cdot | X)$  is the posterior probability of  $\theta$  given  $X$ . Similarly the (generalized) Bayes rule under (2.5) is  $\gamma^g(X) \wedge \gamma$ .

We will consider a (proper) prior

$$g_n(\theta) = C_n \left( \frac{n}{n + |\theta|^2} \right)^d,$$

where  $n$  is an integer,  $d$  is such that  $g_n(\theta)$  has a finite integral and  $C_n$  is the normalizing constant. In Theorem 7 we will show that  $\gamma$  is Bayes under the constraint (2.5) with respect to  $g_n$  for  $n$  large enough. Due to the continuity of the risk function of an arbitrary estimator, this obviously implies  $PV$ -admissibility and hence  $V$ -admissibility.

**THEOREM 7.** *With respect to  $g_n$  where  $n$  is large enough,  $\gamma$  is the Bayes estimator under the constraint (2.5).*

**PROOF.** Since the Bayes estimate under the constraint (2.5) is  $\gamma^{g_n} \wedge \gamma$ , it suffices to prove that for  $n$  large enough,

$$P(|\theta - X| \leq c | X = x) > \frac{1}{2}, \quad \forall x.$$

By (2.10),

$$P(|\theta - X| \leq c | X = x) > \gamma \left( 1 - 2|X|^2(1 - \rho(c)) \frac{d(d+1)}{(n + |x|^2)^2} \right) + o\left(\frac{1}{n + |x|^2}\right),$$

where  $o$  is uniform in  $n \geq 1$ . The lower bound is greater than  $\frac{1}{2}$  if

$$\left(\gamma - \frac{1}{2}\right)(n + |x|^2) > 2\gamma d(d+1) + (n + |x|^2)o(1/(n + |x|^2)).$$

The last inequality holds for  $n$  large enough.  $\square$

The proof of the last theorem also implies that without the validity constraint, 1 is a Bayes estimator with respect to the proper prior  $g_n$  for  $n$  large enough. Hence it is admissible, even though it does not convey the correct picture as a confidence estimator. This striking example may be viewed either as providing justification for the validity constraint or as a criticism of the absolute error loss, or both.

**Acknowledgments.** The authors thank Professors James O. Berger, George Casella, K. L. Lu and Christian Robert for many valuable discussions. They also appreciate Professor Gideon Schwartz's help for pointing out Poisson's Law in relation to the paper.

## APPENDIX

PROOF OF (2.9). Assume that  $Z$  is a standard normal random vector. Write the left-hand side of (2.9) as

$$EI(Z \in R_1)|\Delta|^3/(\tau^{d+3}) + EI(Z \in R_1^c)|\Delta|^3/(\tau^{d+3}) \equiv A + B,$$

where  $R_1 = \{Z: |\Delta| > \frac{1}{2}(b + |x|^2)\}$  and  $R_1^c$  is its complement. We will show that  $A$  has an order smaller than the right-hand side of (2.9) by resorting to the normal exponential tail. For  $B$ ,  $\tau^{d+3}$  provides the same order.

We deal with  $A$  first. The inequality defining  $R_1$  implies

$$|Z|^2 + 2|x||Z| > \frac{1}{2}(b + |x|^2),$$

which implies

$$|Z| > \left[ (b + |x|^2)/2 + |x|^2 \right]^{1/2} - |x| \geq \left[ (b + |x|^2)/24 \right]^{1/2}.$$

Let  $R_2$  denote the region consisting of  $Z$  such that  $|Z| > [(b + |x|^2)/24]^{1/2}$ . Then  $R_1 \subset R_2$ . Noting that  $\tau \geq \min(b + |x|^2, b + |Z + x|^2) > b$ , we have

$$A \leq \frac{1}{b^{d+3}} EI(Z \in R_2)|\Delta|^3.$$

Consequently for some positive constants  $K_1, K_2$  and  $K_3$ ,

$$\frac{A}{g(x)} = \frac{K_1(|x|^3 + K_2)(b + |x|^2)^d}{b^{(d+3)}} O(e^{-(b+|x|^2)K_3}) = o((b + |x|^2)^{-1}),$$

uniformly in  $|x|^2$ .

For the  $B$  term, note that in  $R_1^c, |\Delta| < \frac{1}{2}(b + |x|^2)$  and hence

$$|\tau| \geq \min(b + |x|^2, b + |x|^2 + \Delta) \geq \frac{1}{2}(b + |x|^2).$$

This implies that for some constant  $K_4 > 0$ ,

$$\begin{aligned} B &\leq 2^{d+3} \frac{g(x)}{(b + |x|^2)^3} EI(Z \in R_1^c) |\Delta|^3 \\ &= g(x) O((b + |x|^2)^{-3}) (|x|^3 + K_4) \\ &= g(x) o((b + |x|^2)^{-1}), \end{aligned}$$

which completes the proof.  $\square$

*Note added in proof.* Eaton (1991) provides a different route to establish Theorem 1 which could replace the present argument subsequent to Lemma 6. A minor modification of Eaton’s result establishes that Theorem 1 is valid so long as the Markov chain with transition kernel

$$R(d\eta|\theta) = \left[ \int \left( \frac{\varphi(\eta - x)g(\eta)}{\int \varphi(\rho - x)g(\rho) d\rho} \right) \varphi(x - \theta) dx \right] d\eta$$

is recurrent. Further results in Eaton’s paper provide a mechanism with which it is possible, after some calculation, to establish recurrence of this Markov chain when  $g$  is as in Lemma 6.

### REFERENCES

BERGER, J. O. (1985). The frequentist viewpoint and conditioning. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. Le Cam and R. Olshen, eds.) 1 15–44. Wadsworth, Monterey, Calif.

BERGER, J. O. (1988). An alternative: The estimated confidence approach. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.) 1 85–90. Springer, New York.

BERGER, J. O. and WOLFERT, R. L. (1988). *The Likelihood Principle*, 2nd ed. IMS, Hayward, Calif.

BONDAR, J. V. (1988). Discussion of “Conditionally acceptable frequentist solutions” by George Casella. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.) 1 91–93. Springer, New York.

BROWN, L. D. (1967). The conditional level of student’s test. *Ann. Math. Statist.* 38 1068–1071.

BROWN, L. D. (1978). A contribution to Kiefer’s theory of conditional confidence procedures. *Ann. Statist.* 6 59–71.

BROWN, L. D. (1990). An ancillarity paradox which appears in multiple linear regression (with discussion). *Ann. Statist.* 18 471–538.

- BROWN, L. D. and HWANG, J. T. (1982). A unified admissibility proof. In *Statistical Decision Theory and Related Topics III* (S. S. Gupta and J. O. Berger, eds.) 1 205–230. Academic, New York.
- BROWN, L. D. and HWANG, J. T. (1989). Admissibility of confidence estimators. Statistics Center Technical Report, Cornell Univ.
- BROWNE, C. and KIEFER, J. (1977). The ideas of conditional confidence in the simplest setting. *Comm. Statist. A—Theory Methods* 6 697–751.
- CASELLA, G. (1988a). Conditional inference. Statistics Center Technical Report, Cornell Univ.
- CASELLA, G. (1988b). Conditionally acceptable frequentist solutions (with discussion). In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.) 1 73–117. Springer, New York.
- CASELLA, G., HWANG, J. T. and ROBERT, C. (1989). Loss function for set estimation. Statistics Center Technical Report, Cornell Univ.
- COX, D. R. (1958). Some problems connected with statistical inference. *Ann. Math. Statist.* 29 357–372.
- EATON, M. (1991). A statistical diptych: Admissible inferences—recurrence of symmetric Markov chains. Technical Report, Univ. Minnesota.
- FIELLER, E. C. (1954). Some problems in interval estimation. *J. Roy. Statist. Soc. Ser. B* 16 175–183.
- HWANG, J. T. and PEMANTLE, R. (1990). Evaluation of estimators of statistical significance under a class of proper loss functions. Statistics Center Technical Report, Cornell Univ.
- KIEFER, J. (1976). Admissibility of conditional confidence procedures. *Ann. Math. Statist.* 4 836–865.
- KIEFER, J. (1977a). Conditional confidence and estimated confidence in multidecision problems. In *Multivariate Analysis IV* (P. R. Krishnaiah, ed.) 143–158. North-Holland, Amsterdam.
- KIEFER, J. (1977b). Conditional confidence statements and confidence estimators (with discussion). *J. Amer. Statist. Assoc.* 72 789–827.
- LU, K. L. and BERGER, J. O. (1989). Estimated confidence procedures for multivariate normal means. *J. Statist. Plann. Inference* 23 1–19.
- ROBERT, C. and CASELLA, G. (1989). Improved confidence estimators for the usual multivariate normal confidence set. Statistics Center Technical Report, Cornell Univ.
- ROBINSON, G. K. (1979a). Conditional properties of statistical procedures. *Ann. Statist.* 7 742–755.
- ROBINSON, G. K. (1979b). Conditional properties of statistical procedures for location and scale parameters. *Ann. Statist.* 7 756–771.
- SCHERVISH, M. (1989). A general method for computing probability assessors. *Ann. Statist.* 17 1856–1879.

DEPARTMENT OF MATHEMATICS  
WHITE HALL  
CORNELL UNIVERSITY  
ITHACA, NEW YORK, 14853-7901