EVALUATING CONFIDENCE SETS USING LOSS FUNCTIONS

George Casella and Jiunn Tzon Hwang

Reprinted from
Statistica Sinica Vol. 1, No. 1, pp. 159–173
January 1991

ıtica Sinica 1(1991), 159-173

EVALUATING CONFIDENCE SETS USING LOSS FUNCTIONS

George Casella and Jiunn Tzon Hwang

Cornell University

Abstract: The value of a particular confidence set is traditionally measured by two quantities, its volume and probability of coverage. From a practical point of view, it is desirable to have a procedure that performs well against each measure. Such an approach, however, suffers in theory, for there is no one well-defined loss function which would place the problem in a true decision-theoretic setting. We derive necessary and sufficient conditions for minimax equivalence of the solution, using a linear combination loss function, to that of the component loss problem. Using this equivalence, it is possible to construct componentwise minimax estimators via the linear combination loss. We apply these results to estimation of a multivariate normal mean with unknown variance.

Key words and phrases: Confidence sets, multivariate normal density, minimax estimation.

1. Introduction

The theory of set estimation has progressed at a slower rate than that of point estimation. One possible reason for this is the lack of a reasonable, widely accepted loss function. Without such a loss function it is impossible to put the set estimation problem in a true decision-theoretic setting, and even defining such concepts as minimaxity can become somewhat contrived.

If C is a set estimator for a parameter γ, there are two natural measures of the worth of C. One is a measure of the size of the set C and the other is a measure of the containment of γ by C. The measure of size, which is taken to be a function of volume, is denoted by $\varphi(C)$, and the measure of containment, which is taken to be an indicator function, is denoted by I_γ, where $I_\gamma(C)= 1$ if $\gamma \in C$ and 0 otherwise. It is reasonable, when attempting to decide upon a particular estimator, C, to evaluate it using both $\varphi(C)$ and $I_\gamma(C)$, and choose a procedure that performs well against each measure. While this is appealing from a practical point of view, it is decidedly difficult to deal with in theory. From a theoretical point of view it is desirable to combine $\varphi(C)$ and $I_\gamma(C)$ into one loss function and choose a procedure which performs well against this loss function. This approach, however, can suffer in the eye of the practitioner, who may not see a reasonable way to combine $\varphi(C)$ and $I_\gamma(C)$. (For some suggestions see Casella, Hwang and Robert (1989).)

The purpose of this paper is to demonstrate when these approaches are equivalent. That is, for a particular linear combination of $\varphi(C)$ and $I_\gamma(C)$, to identify procedures which perform well against the linear combination loss function and also perform well against each measure. It will follow that the set estimation problem can be placed in a decision-theoretic framework and, whether or not one accepts the reasonableness of the loss function is irrelevant. The use of the loss function is justified through its connection to componentwise optimality. Thus, the loss function may provide a straightforward means of constructing set estimators that are componentwise optimal.

Most previous work on decision-theoretic set estimation has dealt with evaluation of procedures against either a linear combination loss or against $\varphi(C)$ and $I_\gamma(C)$ separately, with no investigation of a connection. For example, Winkler (1972) considers a variety of linear combination loss functions, and determined conditions that optimal intervals must satisfy. Cohen and Strawderman (1973), building on the work of Brown (1966), demonstrated that for a wide class of loss functions the best invariant confidence interval for a translation or scale parameter is an admissible minimax estimator.

Our main concern is with equivalence with respect to minimaxity. (It is straightforward to establish that if a confidence set is admissible against a loss function that is a linear combination of $\varphi(C)$ and $I_\gamma(C)$, then it is admissible componentwise. See Joshi (1969) for a complete discussion of this.) One of the only papers to explicitly deal with minimax equivalence is that of Blyth (1951). Although he was concerned mainly with sequential problems (and considered loss components of sampling cost and estimation loss), some of his results are quite general. He showed that, under certain conditions on the procedure C, if C is minimax against a linear combination loss, then C is minimax against a component loss. Our results both extend and strengthen those of Blyth.

The results reported in this paper have influenced other decision-theoretic investigations concerning the behavior of set estimators. (The results reported here were originally the basis for the technical report by Casella and Hwang (1982).) These include papers by Casella and Hwang (1983, 1987), who construct improved confidence sets using an empirical Bayes approach with loss functions; Cohen and Sackrowitz (1984), who investigate the relationship between componentwise loss and a random linear combination loss; Meeden and Vardeman (1985), who study the relationship between admissibility and Bayesianity; and Casella, Hwang and Robert (1989), who study nonlinear set estimation losses.

In Section 2 we derive the main equivalence results in a general setting, establishing a necessary and sufficient condition for minimax equivalence between the linear combination solution and the component solution. Section 3 considers the case of estimating the mean vector of a multivariate normal distribution with unknown variance. Using standard techniques, it is established that the usual confidence set is minimax against a linear combination loss, and the results of Section 2 are applied to establish minimaxity against the component loss. An appendix is included, which contains details of some of the calculations needed in Section 3.

2. Minimax Equivalence

Let x be an observation on the random variable X from the sample space $\mathcal{X} = R^p$, and let $\mathcal{B}$ be the σ-field of all subsets of $\mathcal{X}$. For each $\gamma \in \Omega$, let $P_\gamma(\cdot)$ be a probability measure on $\mathcal{B}$. Following Joshi (1969), we define a confidence procedure C to be a Borel measurable subset of the product space $\mathcal{X} \times \Omega$. Associated with each confidence procedure C are two cross sections, the γ section, a Borel subset of $\mathcal{X}$ given by

$$C_\gamma = \{x : (x, \gamma) \in C\},$$

and the X section (the confidence set for γ), a Borel subset of Ω given by

$$C_X = \{\gamma : (x, \gamma) \in C\}.$$

The coverage probability of a confidence procedure C, denoted by $P_\gamma(C)$, is the probability content of the set C_γ when γ is the true state of nature, i.e.,

$$P_\gamma(C) = P_\gamma(\gamma \in C_X) = P_\gamma(X \in C_\gamma) = \int_{C_\gamma} dP_\gamma(x).$$

With each confidence procedure C, we also associate a measure of size, $\varphi(C)$. This can be given a quite general definition, but it suffices to consider $\varphi(C)$ as a scaled measure of volume. Define $\varphi(C)$ by

$$\varphi(C) = k\mathrm{Vol}(C_X), \tag{2.1}$$

where $k > 0$ is a constant and $\mathrm{Vol}(C_X)$ is the Lebesgue measure of C_X. (The results of this section actually hold for a more general form of (2.1), where k is replaced by $k(x, \gamma)$, a nonnegative, measurable function on $\mathcal{X} \times \Omega$ and $\mathrm{Vol}(C_X)$ is the Lebesgue measure of C_X.) The expected size of C, $E_\gamma\varphi(C)$, is the expected value of $\varphi(C)$ when γ is the true state of nature, that is, $E_\gamma\varphi(C) = \int \varphi(C) dP_\gamma(x)$.

A confidence procedure C is unique only up to an equivalence relation. Two procedures C_1 and C_2 are equivalent if the set $(C_1 \backslash C_2) \cup (C_2 \backslash C_1)$ has

Lebesgue measure zero (where $C_1 \backslash C_2 = C_1 \cap C_2^c$, the superscript "c" denoting complement). Such a restriction is necessary for otherwise it would be possible for two seemingly identical procedures to have very different coverage probabilities. (One confidence procedure could have its coverage probability increased, without increasing its volume, by simply adding isolated points.) This would render many loss criteria meaningless in establishing a preference ordering among confidence sets.

We now give two definitions of a minimax confidence procedure.

Definition 2.1. A confidence procedure C^* is said to be *α-minimax* if, for a given α,

$$\begin{aligned} &\text{i)} \quad P_\gamma(C^*) \geq 1-\alpha \qquad \text{for all} \quad \gamma \\ &\text{ii)} \quad \sup_\gamma E_\gamma \varphi(C^*) = \inf_C \sup_\gamma E_\gamma \varphi(C) \ , \end{aligned}$$

where the infimum is taken over all confidence procedures C that satisfy i). Let $\mathcal{G}_\alpha$ denote the class of all α-minimax confidence procedures.

This is a more practical definition of minimaxity: an α-minimax confidence set minimizes the maximum expected volume among all $1-\alpha$ confidence sets.

For the other definition of minimaxity, we introduce the loss function

$$L(\gamma, C) = \varphi(C) - I_\gamma(C). \tag{2.2}$$

Such a loss function has many drawbacks, but has often been considered in the literature (see, for example, Joshi (1969), Winkler (1972), Cohen and Strawderman (1973), or Meeden and Vardeman (1985)). Recall, however, that our goal (as with Cohen and Sackrowitz (1984)) is to only use the loss function as a means to an end. That is, we are only interested in using it to establish connections with the component loss problem.

The risk of a confidence procedure C is then given by

$$R(\gamma, C) = E_\gamma L(\gamma, C) = E_\gamma \varphi(C) - P_\gamma(C). \tag{2.3}$$

Definition 2.2. A confidence procedure C^* is said to be *k-minimax* if

$$\sup_\gamma R(\gamma, C^*) = \inf_C \sup_\gamma R(\gamma, C),$$

where the infimum is taken over all confidence procedures C.

The "k" in the name k-minimax refers to the constant k of (2.1), the factor by which the volume is scaled. Let $\mathcal{G}_k$ denote the class of k-minimax confidence procedures.

Finally, we define a class of confidence procedures, $\mathcal{G}^*$, which provides an important link between α-minimaxity and k-minimaxity. This class is defined by

$$\begin{aligned}\mathcal{G}^* = \{C \in \mathcal{X} \times \Omega : \text{there exists} \quad \{\gamma_n\}_{n=1}^{\infty} \quad \text{such that} \\ \lim_{n\to\infty} E_{\gamma_n}\varphi(C) = \sup_\gamma E_\gamma\varphi(C) \quad \text{and} \\ \lim_{n\to\infty} P_{\gamma_n}(C) = \inf_\gamma P_\gamma(C) = 1-\alpha\}.\end{aligned}$$

$\mathcal{G}^*$ can be thought of as a class of confidence sets for which the maximum expected scaled volume and minimum coverage probability occur at the same parameter value. Note that $\mathcal{G}^*$ contains any procedure with constant volume or constant coverage probability. Under suitable conditions, invariant confidence procedures will be in $\mathcal{G}^*$. (Goutis and Casella (1989) provide an example of an invariant confidence procedure that does not have constant coverage probability.) This fact is explored in the next section.

The following theorem establishes an equivalence between k-minimax confidence sets and α-minimax confidence sets.

Theorem 2.1. *Suppose $\mathcal{G}_k \cap \mathcal{G}^* \neq \emptyset$. Let $C \in \mathcal{G}_k$ be a $1-\alpha$ confidence set. Then $C \in \mathcal{G}_\alpha$ if and only if $C \in \mathcal{G}^*$.*

Proof. To prove sufficiency, assume $C \in \mathcal{G}^*$. Then

$$\begin{aligned}\sup_\gamma R(\gamma, C) &= \sup_\gamma \{E_\gamma\varphi(C) - P_\gamma(C)\} \\ &= \sup_\gamma E_\gamma\varphi(C) - \inf_\gamma P_\gamma(C) \\ &= \sup_\gamma E_\gamma\varphi(C) - (1-\alpha),\end{aligned}$$

where the second equality follows from the fact that $C \in \mathcal{G}^*$. Now suppose C' has confidence coefficient at least $1-\alpha$. We have

$$\sup_\gamma E_\gamma\varphi(C) - (1-\alpha) \le \sup_\gamma R(\gamma, C') \le \sup_\gamma E_\gamma\varphi(C') - (1-\alpha),$$

and hence, $C \in \mathcal{G}_\alpha$. To prove the necessity let $C' \in \mathcal{G}_k \cap \mathcal{G}^*$, which implies that we also have $C' \in \mathcal{G}_\alpha$. Now

$$\sup_\gamma R(\gamma, C) = \sup_\gamma E_\gamma\varphi(C') - \inf_\gamma P_\gamma(C') = \sup_\gamma E_\gamma\varphi(C') - (1-\alpha). \tag{2.4}$$

Let γ_1 be a value that satisfies

$$\sup_\gamma R(\gamma, C) = R(\gamma_1, C) = E_{\gamma_1}\varphi(C) - P_{\gamma_1}(C). \tag{2.5}$$

Since $C \in \mathcal{G}_\alpha, E_{\gamma_1}\varphi(C) \le \sup_\gamma E_\gamma\varphi(C')$, it follows from (2.4) and (2.5) that $P_{\gamma_1}(C) \le 1-\alpha$. But $\inf_\gamma P_\gamma(C) = 1-\alpha$, so it must be the case that $P_{\gamma_1}(C) = 1-\alpha = \inf_\gamma P_\gamma(C)$. But then it also follows that $E_{\gamma_1}\varphi(C) = \sup_\gamma E_\gamma\varphi(C') = \sup_\gamma E_\gamma\varphi(C)$ and, hence, $C \in \mathcal{G}^*$.

The sufficiency part of this theorem is quite similar to Lemma 5 of Blyth (1951), although the conditions are stated somewhat differently. For reasonable choices of k it is usually possible to find α such that $\mathcal{G}_k \cap \mathcal{G}^* \neq \emptyset$. For instance, a minimax equalizer rule will be a member of $\mathcal{G}_k \cap \mathcal{G}^*$. In such situations, the procedure for verifying α-minimaxity of a confidence set is clear; find a k-minimax confidence set and verify that it is in $\mathcal{G}^*$. As will be seen in the next section, working with the linear combination loss function is a great advantage, yielding an easy characterization of Bayes sets, and straightforward methods of establishing minimaxity.

Although it is of less practical importance, it is of interest to inquire when an α-minimax confidence set is also k-minimax. The next theorem establishes this result.

Theorem 2.2. *Suppose $\mathcal{G}_k \cap \mathcal{G}^* \neq \emptyset$. If $C \in \mathcal{G}_\alpha$ then $C \in \mathcal{G}_k$.*

Proof. Note that $1-\alpha \le \inf_\gamma P_\gamma(C)$, and let $C' \in \mathcal{G}_k \cap \mathcal{G}^*$ (and hence $C' \in \mathcal{G}_\alpha$). Then

$$\begin{aligned} R(\gamma, C) = E_\gamma\varphi(C) - P_\gamma(C) &\le \sup_\gamma E_\gamma\varphi(C) - (1-\alpha) \\ &\le \sup_\gamma E_\gamma\varphi(C') - (1-\alpha) = \sup_\gamma R(\gamma, C'). \end{aligned}$$

Hence, $R(\gamma, C) \le \sup_\gamma R(\gamma, C')$ for all γ, so $C \in \mathcal{G}_k$.

3. The Multivariate Normal Distribution

In this section we specialize to the case of constructing a confidence set for the mean of a multivariate normal distribution with unknown variance. Using a standard argument, it is established that the usual confidence set is k-minimax for a particular choice of k. Theorem 2.1 is then applied to establish α-minimaxity.

Let X have a p-variate normal distribution with mean θ and covariance matrix $\sigma^2 I (X \sim \mathrm{N}(\theta, \sigma^2 I))$, where both θ and σ^2 are unknown. Let s^2 be an observation on S^2, an estimate of σ^2 (independent of X), with $S^2 \sim (\sigma^2/\nu)\chi^2_\nu$. Let C be a set estimator of θ, and consider the loss function

$$L_k(\theta, \sigma^2, C) = \frac{k}{\sigma^p}\mathrm{Vol}(C) - I_\theta(C), \tag{3.1}$$

a special case of (2.2) with k replaced by k/σ^p, k constant. As will be seen, there is a direct relationship between the value of k in (3.1) and the confidence coefficient $1-\alpha$ of C. Indeed, just as we require $0 \leq 1-\alpha \leq 1$, there is a range of reasonable values of k, namely $0 < k \leq (2\pi)^{-p/2}$. If $k > (2\pi)^{-p/2}$, the volume component of the loss overwhelms the indicator function, making the loss function useless.

The usual confidence set for θ is

$$C^0 = \{\theta : |\theta - x| \leq cs\}, \tag{3.2}$$

a p-sphere of radius cs centered at x. If σ^2 is known, then C^0 (with s replaced by σ) is both k-minimax and α-minimax for appropriate k and α. This result, however, has not been extended to the unknown variance case. The minimaxity of C^0 is an important benchmark for use in measuring the performance of other set estimators in the unknown variance case. Establishment of Stein-type domination results must start from a classically optimal estimator, and C^0 provides such a starting point.

We begin by establishing, for given k and an appropriate choice of c, that C^0 is the best invariant set estimator of θ against the loss (3.1). To establish that C^0 is best invariant, we must also consider randomized confidence procedures which, following Joshi (1969), we define in the following way.

Definition 3.1. A *randomized confidence procedure* $\lambda(x,s,\theta)$ is a Lebesgue measurable function on $R^p \times (0,\infty) \times R^p$, taking values in $[0,1]$. We interpret $\lambda(x,s,\theta)$ as the probability of including θ in the confidence set when $X = x$ and $S = s$ are observed.

It is straightforward to verify that the problem remains invariant under the transformation

$$(x,s) \to (ax+b, as), \quad (\theta,\sigma) \to (a\theta + b, a\sigma), \tag{3.3}$$

where $a > 0$ and $b \in R^p$. The invariant rules must satisfy $\lambda(x,s,\theta) = \lambda(ax + b, as, a\theta + b)$, which implies that the invariant rules are of the form

$$\lambda(x,s,\theta) = \lambda[(x-\theta)/s]. \tag{3.4}$$

It immediately follows that for any invariant rule λ, both its expected scaled volume and coverage probability are constant with respect to θ and σ. We are now ready to prove the following theorem.

Theorem 3.1. *If* $0 < k \leq (2\pi)^{-p/2}$, *then the best invariant estimator against the loss* $L(\theta,\sigma^2,C) = (k/\sigma^p)\mathrm{Vol}(C) - I_\theta(C)$ *is given by* (3.2) *where* c^2 *satisfies* $(2\pi)^{-p/2}[\nu/(\nu+c^2)]^{(p+\nu)/2} = k$.

Proof. Note that $E(S^p|\sigma) = \sigma^p(2/\nu)^{p/2}\Gamma[(p+\nu)/2][\Gamma(\nu/2)]^{-1} = \sigma^p M$ (say). Straightforward calculation shows that for any invariant rule λ,

$$R(\theta,\sigma^2,\lambda) = (k/\sigma^p)E(S^p|\sigma)\int \lambda(t)dt - \int \lambda(t)p_\nu(t)dt$$
$$= \int [kM - p_\nu(t)]\lambda(t)dt,$$

where $p_\nu(\cdot)$ is a multivariate Student's t density with ν degrees of freedom. From the Neyman-Pearson Lemma, this risk is minimized by choosing $\lambda(t) = \{t : p_\nu(t) \geq kM\}$. From the definition of M and the form of the multivariate t density, it is easily seen that this set is equivalent to C^0.

Remark. The bounds on k are quite important. Obviously, if $k = 0$ then the volume term is eliminated from the loss function, and the optimal set estimator is the entire parameter space. More important, if $k > (2\pi)^{-p/2}$, the loss function places too much weight on the volume term. The consequences of this will be seen later.

Since the invariant confidence sets have constant scaled volume and coverage probability, it immediately follows that the confidence set C^0 given in (3.2) has minimum volume (and minimum expected volume) among all invariant confidence procedures λ satisfying $P(\lambda|\theta,\sigma) \geq P(C^0|\theta,\sigma)$. This property is usually taken to be the definition of a best invariant confidence set (see, e.g., Stein (1962)), rather than the definition arising from the use of the loss function. As can be seen, the loss function approach implies best invariance according to this definition also. Furthermore, the relationship given between k and c in Theorem 3.1 uniquely determines the best invariant set.

Now that C^0 has been established to be the best invariant (equalizer) rule, establishing k-minimaxity can proceed in a straightforward manner. Indeed, there are two distinct ways to proceed. One way is to verify that the assumptions of the Hunt-Stein Theorem (Kiefer (1957)) are satisfied, and then apply the theorem to establish the k-minimaxity of C^0 (see also Hooper (1982) for a similar development). The other technique is to verify that $R(\theta,\sigma^2,C^0)$ is a limiting Bayes risk.

We choose to employ the more standard technique, that of establishing that $R(\theta,\sigma^2,C^0)$ is a limiting Bayes risk. This method seems to be the more instructive one, for it gives more information about the structure of the problem, and also some hint about the type of Bayes rules which might lead to k-minimax set estimators. It also avoids the tedious and difficult task of verifying the assumptions of the Hunt-Stein Theorem.

The Bayes rules against the loss (3.1) are particularly easy to characterize. Let $\pi(\cdot|\cdot)$ be the general notation for a conditional (posterior) distribution, with $\pi(\cdot)$ denoting an unconditional (prior) distribution. The proof of the following theorem is straightforward.

Theorem 3.2. *Let $L_k(\theta,\sigma^2,C) = (k/\sigma^p)\mathrm{Vol}(C) - I_\theta(C)$, and suppose $\pi(\theta,\sigma^2)$ is a prior distribution. The Bayes rule is given by*

$$C^\pi = \left\{\theta : \pi(\theta|x,s^2) \geq k\int_0^\infty \frac{1}{\sigma^p}\pi(\sigma^2|x,s^2)d\sigma^2\right\}.$$

To establish k-minimaxity of C^0 against the loss function L_k, it remains to find a sequence of Bayes rules with Bayes risks converging to the risk of C^0. Consider the prior density

$$\pi(\theta,\sigma^2)d\theta d\sigma^2 = (2\pi\tau^2)^{-p/2}e^{-|\theta|^2/2\tau^2}\frac{1}{\Gamma(a)b^a}\frac{1}{(\sigma^2)^{a-1}}e^{-1/b\sigma^2}d\theta d\sigma^2, \qquad (3.5)$$

i.e., θ is distributed as p-variate normal with zero mean and covariance matrix $\tau^2 I$, and $\sigma^2 \sim$ Inverse Gamma (a,b). The Bayes rule against L_k, along with evaluation of the limiting Bayes risk, is given in the appendix. A limiting Bayes risk is given by

$$\lim_{\substack{\tau^2\to\infty\\ a\to 0\\ b\to\infty}} r(\pi,C^\pi) = k\left(\frac{2\pi c^2}{\nu}\right)^{p/2}\frac{\Gamma(\frac{p+\nu}{2})}{\Gamma(\nu/2)\Gamma[(p/2)+1]} - P(F_{p,\nu} \leq c^2/p),$$

where $c^2 = \nu[k(2\pi)^{p/2}]^{\frac{-2}{p+\nu}} - \nu$. A limiting Bayes rule is

$$C^L = \left\{\theta : (2\pi)^{-p/2}\left(\frac{\nu}{\nu+|x-\theta|^2/s^2}\right)^{\frac{p+\nu}{2}} \geq k\right\}.$$

The k-minimaxity of C^0 can now be established.

Theorem 3.3. *If $0 < k \leq (2\pi)^{-p/2}$, then $C^0 = \{\theta : |\theta - x| \leq cs\}$ is k-minimax if and only if k and c satisfy the relationship in Theorem* 3.1.

Proof. The theorem is established by noting that C^0 is equivalent to C^L, and $R_k(\theta,\sigma^2,C^0)$ equals the limiting Bayes risk, if and only if k and c satisfy $(2\pi)^{-p/2}[\nu/(\nu+c^2)]^{(p+\nu)/2} = k$. We thus have

$$\sup_{\theta,\sigma^2} R_k(\theta,\sigma^2,C^0) = \lim r(\pi,C^\pi),$$

and it follows (see, for example, Berger (1985)) that C^0 is k-minimax. The necessity follows from Theorem 3.1.

It is interesting to note that the conjugate prior (which is the same as (3.5) with the exception that $\theta|\sigma^2 \sim N(0, \tau^2\sigma^2 I)$) cannot be used to establish the k-minimaxity of C^0. This is hinted at by the fact that C^0 is generalized Bayes against the improper prior $\pi(\theta, \sigma^2) = \sigma^{-2}$, and the conjugate prior cannot approach this as a limit. More precisely, if we denote the conjugate prior by $\pi^*(\theta, \sigma^2)$, there does not exist a sequence of functions $m(\tau^2, a, b)$ such that $m(\tau^2, a, b)\pi^*(\theta, \sigma^2) \to \sigma^{-2}$. Also, in the Bayes risk of π^*, the term $\Gamma[a - (p/2)]$ appears, so the parameter a cannot approach zero.

Thus, using standard techniques, the k-minimaxity of C^0 has been established. The major goal, however, is to establish α-minimaxity of C^0. This follows immediately, however, from Theorem 2.1, since $C^0 \in \mathcal{G}^*$.

Theorem 3.4. *If c^2 satisfies $P(F_{p,\nu} \leq c^2/p) = 1 - \alpha$, then the set $C^0 = \{\theta : |\theta - x| \leq cs\}$ is α-minimax.*

Throughout this section we have required the condition that $k \leq (2\pi)^{-p/2}$. This condition is needed so that the volume component of the loss does not overwhelm the indicator function in the loss. Also, the relation between c and k, that is,

$$k = (2\pi)^{-p/2}\left(\frac{\nu}{\nu + c^2}\right)^{\frac{p+\nu}{2}}, \tag{3.6}$$

shows that there is a $1 - 1$ correspondence between the confidence level of C^0 and k, and implicitly places this bound on k. It is, of course, possible to consider values of k greater than $(2\pi)^{-p/2}$, but as there is no relation with the component loss problem for such values of k, we would hope that this is reflected in the solution of the linear combination loss problem. We now investigate the case where $k > (2\pi)^{-p/2}$, and find a rather surprising result, which shows the importance of tying the loss function to the componentwise problem (or, more generally, the importance of not evaluating the loss function in a vacuum).

Theorem 3.5. *If $k > (2\pi)^{-p/2}$, the empty set, $\emptyset$ (or, equivalently, any set of Lebesgue measure zero), is the unique minimax set estimator of θ. Moreover, $\emptyset$ is the unique proper Bayes (hence admissible) set estimator.*

Remark. Another way of characterizing the difference between the situation $0 < k \leq (2\pi)^{-p/2}$ and $k > (2\pi)^{-p/2}$ is that, for $k > (2\pi)^{-p/2}$ the best invariant set estimator is the unique minimax admissible estimator, which is probably not the case if $k \leq (2\pi)^{-p/2}$.

Proof. The risk of $\emptyset$ is, of course, $R(\theta, \sigma^2, \emptyset) = 0$ for all θ and σ. The minimaxity of $\emptyset$ can be deduced from the proof in the Appendix, but it will also follow immediately when it is demonstrated that $\emptyset$ is proper Bayes (since its Bayes risk will also be equal to zero).

To show that $\emptyset$ is proper Bayes, we can consider the conjugate priors, which are similar to those given in (3.5) with the exception that the prior distribution on θ is $\theta|\sigma^2 \sim N(0, \tau^2\sigma^2 I)$. For this prior, the Bayes rule is given by (after some algebra)

$$C^B = \left\{\theta : |\theta - \delta^B(x)|^2 \le \left(\frac{\tau^2}{\tau^2+1}\right)(u^{-1}-1)T(|x|,s)\right\}, \qquad (3.7)$$

where $\delta^B(x) = [\tau^2/(\tau^2+1)]x, T(|x|,s) = \nu s^2 + (2/b) + |x|^2/(\tau^2+1)$, and

$$u = \left[k\left(\frac{2\pi\tau^2}{\tau^2+1}\right)^{p/2}\right]^{\frac{2}{p+\nu+2a}}.$$

Now suppose that $k = (2\pi)^{-p/2}(1+\epsilon)$ for some $\epsilon > 0$. Choose τ^2 to satisfy $\tau^2 > 1/[(1+\epsilon)^{2/p} - 1]$, which implies $u > 1$ and that the right-hand side of the inequality in (3.7) is negative. Hence $C^B = \emptyset$ for this choice of τ^2. If we let $\pi_\emptyset$ denote the prior for which $\emptyset$ is the Bayes rule, then it follows that any minimax set estimator must be Bayes against $\pi_\emptyset$.

Let C be any confidence procedure which is not equivalent to $\emptyset$, that is, there exists a set B with positive Lebesgue measure such that for all $(x,s) \in B$, the confidence set $C_{x,s}$ is a set of θ values with positive Lebesgue measure. We then have

$$r(\pi_\emptyset, C) - r(\pi_\emptyset, \emptyset) = r(\pi_\emptyset, C) - r(\pi_\emptyset, C \cap \emptyset) = r(\pi_\emptyset, C - \emptyset) > 0,$$

where the last inequality follows from Theorem 3.2. Therefore, C cannot be Bayes against $\pi_\emptyset$ and, hence, $\emptyset$ is unique Bayes and unique minimax, hence admissible.

Thus the correspondence between the linear combination loss and the component loss is complete. If $k > (2\pi)^{-p/2}$ there is no value of α which corresponds to k. This shows that this choice of weights in the loss is absurd, but the solution to the linear combination loss function problem reflects this: the best set estimator is the empty set. Thus, although it is possible to obtain unreasonable results when the loss function is used alone, such results will not happen if the correspondence with componentwise loss is maintained.

Lastly, we note that all the results in this section can be made to apply to the case of known σ^2, simply by replacing s^2 by σ^2 and letting $\nu \to \infty$. For this case the conjugate prior can be used to establish the minimaxity of C^0, making the calculations a bit simpler. Also, the restriction $0 \le k \le (2\pi)^{-p/2}$ remains the same, but the relation (3.6) becomes $k = (2\pi)^{-p/2} \exp(-c^2/2)$.

Acknowledgement

We thank the referees and Associate Editor for their very insightful comments. This research is supported by National Science Foundation under Grant No.s DMS89-0039(for Casella) and DMS88-09016(for Huang).

Appendix: Convergence of Bayes Risk

Let $E(\cdot|\cdot)$ be the general notation for conditional expectation, with $E(\cdot)$ denoting unconditional expectation. Let $\pi(\theta, \sigma^2)$ be the (proper) prior distribution given by (3.5), and let C^π be the Bayes rule against this prior using the loss (3.1), which is given by

$$C^\pi = \left\{\theta : \frac{e^{-|\theta|^2/2\tau^2}\Gamma\left(\frac{1}{2}(p+\nu)+a\right)}{\left(\frac{1}{2}|x-\theta|^2 + \frac{1}{2}\nu s^2 + \frac{1}{b}\right)^{\frac{1}{2}(p+\nu)+a}} \ge k \int_0^\infty \left(\frac{2\pi\tau^2}{\sigma^2+\tau^2}\right)^{p/2} \frac{e^{-\frac{1}{\sigma^2}\left(\frac{1}{2}\nu s^2+\frac{1}{b}\right)} e^{-\frac{|x|^2}{2(\sigma^2+\tau^2)}}}{(\sigma^2)^{\frac{1}{2}(p+\nu)+a+1}} d\sigma^2\right\}. \quad (A.1)$$

Theorem A1. *The Bayes risk, $r(\pi, C^\pi)$, satisfies*

$$\lim_{\substack{\tau^2\to\infty \\ b\to\infty \\ a\to 0}} r(\pi, C^\pi) = R_k(\theta, \sigma^2, C), \qquad \forall \ \theta, \quad (A.2)$$

where

$$C = \{\theta : |\theta - x| \le cs\}, \quad (A.3)$$

and

$$c^2 = \nu\left[\left(k(2\pi)^{p/2}\right)^{\frac{-2}{p+\nu}} - 1\right]^+, \quad (A.4)$$

with "+" denoting positive part.

Proof. Straightforward calculation shows

$$\begin{aligned} E[\mathrm{Vol}(C)/\sigma^p|\theta, \sigma^2] &= E[\mathrm{Vol}(C)/\sigma^p] \\ &= \left(\frac{2\pi c^2}{\nu}\right)^{p/2} \frac{\Gamma\left((p+\nu)/2\right)}{\Gamma(\nu/2)\Gamma[(p/2)+1]} \end{aligned} \quad (A.5)$$

and

$$E[I_\theta(C)|\theta, \sigma^2] = E[I_\theta(C)] = P(F_{p,\nu} \le c^2/p),$$

where $F_{p,\nu}$ is an F random variable with p and ν degrees of freedom. We will show that

$$\lim_{\substack{\tau^2\to\infty\\ b\to\infty\\ a\to 0}} E\big[\mathrm{Vol}(C^\pi)/\sigma^p\big] \geq E\big[\mathrm{Vol}(C)/\sigma^p\big] \tag{A.6}$$

and

$$\lim E\big[I_\theta(C^\pi)\big] \leq E\big[I_\theta(C)\big]. \tag{A.7}$$

Inequalities (A.6) and (A.7) establish that the limit in (A.2) is less than or equal to $R_k(\theta, \sigma^2, C)$. The opposite inequality is immediate, since C^π is Bayes against π for every τ^2, b, and a. Hence, the equality in (A.2) will be established.

To establish (A.6), we will bound $\mathrm{Vol}(C^\pi)$ from below with the set C_1 given by

$$C_1 = \{\theta : |\theta - x| \leq r\},$$

where we define the random variable

$$r^2 = \min\left\{(k^* - 1)^+(\nu S^2 + 2/b), A - \frac{\sigma^2}{\sigma^2+\tau^2}|X|^2\right\},$$

where A is an arbitrary positive constant and $k^* = [k(2\pi)^{p/2}]^{\frac{-2}{p+\nu+2a}}$.

It can be established that C_1 is contained in C^π. Moreover, the expected scaled volume of $C_1, \psi_p E(r/\sigma)^p$, is independent of τ^2, where $\psi_p = \pi^{p/2}/\Gamma(\frac{p}{2}+1)$ is the volume of a unit p-sphere. This follows because r/σ depends only on the random vector $[S^2, \sigma^2, |X|^2/(\sigma^2+\tau^2)]$, which has a distribution independent of τ^2. We can use monotone convergence to bring the limit inside the integral, and conclude

$$\lim_{A\to\infty}\lim_{\tau^2\to\infty} E\big[\mathrm{Vol}(C_1)/\sigma^p\big] = \psi_p[k^*-1]^{p/2}E\big\{[\nu S^2 + (2/b)]/\sigma^2\big\}^{p/2}.$$

The last expectation can be evaluated explicitly and, after collecting terms and using the fact that $[\nu S^2 + (2/b)]/\sigma^2 \sim \chi^2_{\nu+a}$, we have

$$E\left[\frac{\nu S^2 + (2/b)}{\sigma^2}\right]^{p/2} = 2^{p/2}\Gamma\left(\frac{p+\nu+a}{2}\right)\left[\Gamma\left(\frac{\nu+a}{2}\right)\right]^{-1}$$

independent of b. Now, as $a \to 0, k^* - 1 \to c^2/\nu$ (for c^2 given in (A.4)). Hence, we have

$$\begin{aligned}\lim_{\substack{\tau^2\to\infty\\ b\to\infty\\ a\to 0}} E[\mathrm{Vol}(C^\pi)/\sigma^p] &\geq (2\pi c^2/\nu)^{p/2}\Gamma[(p+\nu)/2]\{\Gamma(\nu/2)\Gamma[(p/2)+1]\}^{-1}\\ &= E\big[\mathrm{Vol}(C_1)/\sigma^p\big],\end{aligned}$$

establishing (A.6).

To establish (A.7), we proceed in a similar fashion, using the set $C_2 = \{\theta : g^*(x, s^2, \theta) \geq k\}$, where $g^*(x, s^2, \theta)$ is given by

$$g^*(x,s^2,\theta) = \frac{\Gamma\left(\frac{p+\nu}{2}+a\right)}{\left[\frac{|x-\theta|^2}{2}+\frac{\nu s^2}{2}+\frac{1}{b}\right]^{\frac{p+\nu}{2}+a}}\left[\int_0^\infty (2\pi)^{p/2}\frac{e^{-\frac{1}{\sigma^2}\left(\frac{\nu s^2}{2}+\frac{1}{b}\right)}}{(\sigma^2)^{\frac{p+\nu}{2}+a+1}}\,d\sigma^2\right]^{-1}$$
$$= \frac{1}{(2\pi)^{p/2}}\left[\frac{\frac{\nu s^2}{2}+\frac{1}{b}}{\frac{|x-\theta|^2}{2}+\frac{\nu s^2}{2}+\frac{1}{b}}\right]^{\frac{p+\nu}{2}+a}.$$

It can then be established that

$$\lim_{\tau^2\to\infty} E[I_\theta(C^\pi)] \leq E_\theta[I_\theta(C_2)].$$

Also, recall that $k^* \to 1 + c^2/\nu$ as $a \to 0$, for c^2 of (A.4). Hence, it can be shown that

$$\lim_{\substack{a\to 0\\ b\to\infty}} E_\theta[I_\theta(C_2)] = P[(|X-\theta|^2/\sigma^2) \leq (c^2S^2/\sigma^2)],$$

and it therefore follows that

$$\lim E[I_\theta(C^\pi)] \leq P[(|X-\theta|^2/\sigma^2) \leq (c^2S^2/\sigma^2)]$$
$$= P[F_{p,\nu} \leq (c^2/p)] = E[I_\theta(C)],$$

establishing (A.7) and completing the proof.

References

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, New York.

Blyth, C. R. (1951). On minimax statistical decision procedures and their admissibility. *Ann. Math. Statist.* **22**, 22-42.

Brown, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.* **37**, 1087-1136.

Casella, G. and Hwang, J. T. (1982). Decision-theoretic set estimation. Technical Report BU-773-M, Biometrics Unit, Cornell University, Ithaca, New York.

Casella, G. and Hwang, J. T. (1983). Empirical Bayes confidence sets for the mean of a multivariate normal distribution. *J. Amer. Statist. Assoc.* **78**, 688-698.

Casella, G. and Hwang, J. T. (1987). Employing vague prior information in the construction of confidence sets. *J. Multivariate Anal.* **21**, 79-104.

Casella, G., Hwang, J. T. and Robert, C. (1989). Loss functions for set estimation. Technical Report BU-1027-M, Biometrics Unit, Cornell University, Ithaca, New York.

Cohen, A. and Sackrowitz, H. B. (1984). Decision theory results for vector risks with applications. *Statistics and Decisions,* Supplement Issue No.1, 159-176.

Cohen, A. and Strawderman, W. E. (1973). Admissible confidence interval and point estimation for translation or scale parameters. *Ann. Statist.* **1**, 545-550

Faith, R. E. (1976). Minimax Bayes point and set estimators of a multivariate normal mean. Technical Report #66, University of Michigan.

Goutis, C. and Casella, G. (1989). Improved invariant confidence intervals for a normal variance. Technical Report BU-1038-M, Biometrics Unit, Cornell University, Ithaca, New York.

Hooper, P. (1982). Invariant confidence sets with smallest expected measure. *Ann. Statist.* **10**, 1283-1294. Correction **12**(1984), 784.

Joshi, V. M. (1969). Admissibility of the usual confidence sets for the mean of a univariate or bivariate normal population. *Ann. Math. Statist.* **40**, 1042-1067.

Kiefer, J. (1957). Invariance, minimax sequential estimation, and continuous time processes. *Ann. Math. Statist.* **28**, 573-601.

Meeden, G. and Vardeman, S. (1985). Bayes and admissible set estimation. *J. Amer. Statist. Assoc.* **80**, 465-471.

Stein, C. (1962). Confidence sets for the mean of a multivariate normal distribution. *J. Roy. Statist. Soc. Ser. B* **24**, 265-296.

Winkler, R. L. (1972). A decision-theoretic approach to interval estimation. *J. Amer. Statist. Assoc.* **67**, 187-191.

Biometrics Unit, Cornell University, Ithaca, NY 14853, U.S.A.
Department of Mathematics, Cornell University, Ithaca, NY 14853, U.S.A.

(Received November 1988; accepted March 1990)