The Annals of Statistics
1981, Vol. 9, No. 4, 846–852

ADMISSIBILITY IN FINITE PROBLEMS[1]

BY GLEN MEEDEN AND MALAY GHOSH

Iowa State University

Let X be a random variable which takes on only finitely many values $x \in \chi$ with a finite family of possible distributions indexed by some parameter $\theta \in \Theta$. Let $\Pi = \{\pi_x(\cdot): x \in \chi\}$ be a family of possible distributions (termed "inverse probability distributions") on Θ depending on $x \in \chi$. A theorem is given to characterize the admissibility of a decision rule δ which minimizes the expected loss with respect to the distribution $\pi_x(\cdot)$ for each $x \in \chi$. The theorem is partially extended to the case when the sample space and the parameter space are not necessarily finite. Finally a notion of "admissible consistency" is introduced and a necessary and sufficient condition for admissible consistency is provided when the parameter space is finite, while the sample space is countable.

1. Introduction. Let X be a random variable which takes on values in some finite sample space χ. Let $\{f_\theta, \theta \in \Theta\}$ be a family of possible probability functions for X where $\Theta = \{\theta_1, \cdots, \theta_k\}$. Assume that for each $x \in \chi$, $f_{\theta_i}(x) > 0$ for at least one $\theta_i \in \Theta$. Consider the decision problem specified by the decision space D and the nonnegative loss function L, i.e., for $d \in D$ and $\theta \in \Theta$, $L(\theta, d)$ denotes the nonnegative loss incurred in making the decision d when θ is the true value of the parameter.

For each $x \in \chi$, let $\pi_x(\cdot)$ denote a probability distribution on Θ, i.e., $\pi_x(\theta) \geq 0$ for all $\theta \in \Theta$ and $\sum_{\theta \in \Theta} \pi_x(\theta) = 1$. Following the terminology of Dawid and Stone (1972, 1973) a family

$$\Pi = \{\pi_x(\cdot): x \in \chi\} \tag{1.1}$$

will be called a family of inverse probability distributions on Θ. Further, such a family Π is said to be "Bayes" if there exists a probability distribution λ on Θ such that

$$\pi_x(\theta) \sum_{\theta \in \Theta} \lambda(\theta) f_\theta(x) = \lambda(\theta) f_\theta(x) \tag{1.2}$$

for all $\theta \in \Theta$, $x \in \chi$ i.e., $\pi_x(\cdot)$ is essentially a posterior distribution with respect to some prior λ for each $x \in \chi$. A decision function δ is said to be "optimal" with respect to the family Π if for each x, $\delta(x)$ minimizes the expected loss with respect to the distribution $\pi_x(\cdot)$. In the more common terminology such a decision rule δ is a Bayes decision rule with respect to the prior λ when $\pi_x(\cdot)$ is a bonafide posterior distribution for each $x \in \chi$.

In this note we study the problem of determining when an optimal decision function with respect to a family Π is admissible. It is well known that if Π is Bayes with respect to a prior λ which puts positive mass on each point of Θ, then the corresponding optimal decision function is admissible. In Section 2 we assume that D and L are such that $\sum_{i=1}^k L(\theta_i, d)\lambda_i$ is minimized uniquely with respect to d for any distribution $\lambda = (\lambda_1, \cdots, \lambda_k)$ on Θ where λ_i denotes the probability assigned to θ_i by the distribution λ. Theorem 1 of Section 2 shows that the family Π yields an admissible decision function if and only if Π is Bayes with respect to a set of mutually singular prior distributions. This notion will be made precise in Section 2. The result is essentially contained in the Theorem 2 and

Received July 1978; revised August 1980.

[1] Research partially supported by the NSF Grant #MCS 8005485

AMS 1970 *subject classifications*. 62C15, 62F10.

Key words and phrases. Inverse probability distributions, admissibility, Bayes rules, singular priors, admissible consistency, expectation consistency, discrete uniform, squared error loss.

Example 4 of Hsuan (1979) where a stepwise Bayesian procedure for obtaining admissible rules is introduced.

Suppose now that it is required that Π yields an admissible decision function for any decision problem based on χ, Θ and $\{f_\theta, \theta \in \Theta\}$ and not just the one satisfying the conditions of the previous paragraph. A family Π with this property is said to be "admissible consistent". This extends the notion of "expectation consistency" for a family Π which was introduced by Dawid and Stone (1972, 1973). Theorem 2 in Section 3 shows a family Π is admissible consistent if and only if it is Bayes against a prior which puts positive mass on each point of the parameter space. The relationship between Theorems 1 and 2 is also pointed out in this section. Finally, in Section 4, Theorem 3 extends the sufficiency part of Theorem 1 to cases where χ and Θ are arbitrary.

2. A characterization of admissible decision rules. Let $\lambda = (\lambda_1, \cdots, \lambda_k)$ denote a prior distribution on Θ with support $\Theta(\lambda) = \{\theta_i: \lambda_i > 0\}$. Two prior distributions λ and λ' are said to be singular if $\Theta(\lambda) \cap \Theta(\lambda') = \phi$. Theorem 1 characterizes the families of inverse probability distributions which lead to admissible optimal decision functions. Let $g(x; \lambda) = \sum_{i=1}^k f_{\theta_i}(x)\lambda_i$ denote the marginal probability function of X under λ.

THEOREM 1. *For the decision problem specified by the decision space D and loss function $L(\cdot, \cdot)$ assume that for each prior distribution λ, $\sum_{i=1}^k L(\theta_i, d)\lambda_i$ is minimized uniquely with respect to d. Let Π be a family of inverse probability distributions and let δ be optimal against* Π.

If δ is admissible, then there exists a nonempty set of mutually singular prior distributions $\lambda^1 = (\lambda_1^1, \cdots, \lambda_k^1), \cdots, \lambda^n = (\lambda_1^n, \cdots, \lambda_k^n)$ such that

(i) *if $*\Lambda^1 = \{x: g(x; \lambda^1) > 0\}$, and for $i = 2, \cdots, n$, $*\Lambda^i = \{x: g(x; \lambda^i) > 0$ and $x \notin \cup_{j=1}^{i-1} *\Lambda^j\}$, then each $*\Lambda^i$ is nonempty and $\cup_{i=1}^n *\Lambda^i = \chi$;*

(ii) *if $x \in *\Lambda^i$, then*

$$\pi_x(\theta)\, g(x; \lambda^i) = f_\theta(x)\lambda^i(\theta) \text{ for all } \theta \in \Theta, \qquad i = 1, \cdots, n.$$

Conversely, if for a given family Π of inverse probability distributions there exists a set of mutually singular prior distributions $\lambda^1, \cdots, \lambda^n$ satisfying (i) *and* (ii), *then δ the optimal decision function with respect to Π is admissible.*

Theorem 1 essentially says that a family of inverse probability distributions leads to an admissible rule if and only if it is Bayes with respect to several singular priors, since by (ii) of the theorem for $x \in *\Lambda^i$, $\delta(x)$ is the unique value of d minimizing

$$\textstyle\sum_{\theta \in \Theta(\lambda^i)} L(\theta, d) f_\theta(x)\lambda^i(\theta)/g(x; \lambda^i).$$

It should be noted that the order in which the λ^i's appear is important in the construction of $*\Lambda^1, \cdots, *\Lambda^n$. A different ordering of the λ^i's may result in a different set of $*\Lambda^i$'s.

As an example suppose $\chi = \{0, 1, 2\}$ and $\Theta = \{\theta_1, \theta_2, \theta_3\}$ with $\theta_1 < \theta_2 < \theta_3$. Let $f_{\theta_1}(x) =$ 1 or 0 as $x = 0$ or $x \neq 0$. Let $f_{\theta_2}(x) = 1/3$ for all x and let $f_{\theta_3}(0) = 0$ and $f_{\theta_3}(x) = 1/2$ for $x \neq 0$. First consider the problem of estimating θ when D is the closed interval $[\theta_1, \theta_3]$ and the loss is squared error. The family of priors $\lambda^1 = (1, 0, 0)$ and $\lambda^2 = (0, 1/2, 1/2)$ yields the family of inverse probability functions $\Pi_0(\theta_1) = 1$ and $\Pi_0(\theta_i) = 0$ for $i = 2$ and 3, $\Pi_1(\theta_2) = \Pi_2(\theta_2) = 2/5$, $\Pi_1(\theta_3) = \Pi_2(\theta_3) = 3/5$ and $\Pi_1(\theta_1) = \Pi_2(\theta_1) = 0$. The optimal estimator $\delta(0) = \theta_1$, $\delta(1) = \delta(2) = (2\theta_2 + 3\theta_3)/5$ against this family is admissible by Theorem 1. Now consider the decision problem with $D = \{d_1, d_2\}$ and the loss function $L(\cdot, \cdot)$ given by $L(\theta_1, d_i) = 1$ for $i = 1, 2$, $L(\theta_2, d_2) = L(\theta_3, d_1) = 0$ and $L(\theta_2, d_1) = 1/L(\theta_3, d_2) = 10$. It is easily checked that $\delta_1(x) = d_1$ or d_2 as $x = 0$ or $x \neq 0$ is an optimal decision rule against λ^1 and λ^2. This rule is not admissible, however, since it is dominated by the decision rule $\delta_2(x) = d_2$ for all x. Note that D and $L(\cdot, \cdot)$ do not satisfy the conditions of Theorem 1. Thus it is of interest to find when a family of inverse probability distributions yields optimal decision functions that are admissible for any decision problem whatsoever. This problem is discussed in Section 3.

This theorem was first suggested to the authors by some remarks of Johnson (1971).

The proof of the theorem is omitted because the theorem is essentially equivalent to Theorem 2 and the related discussion in Hsuan (1979) (see also Wald and Wolfowitz (1951), and Brown (1979)). The theorem can be reformulated to show that for this finite decision problem a decision function is admissible if and only if it is Bayes against a set of mutually orthogonal priors. The authors have used this technique in proving the admissibility and inadmissibility of various estimators in finite population sampling (see Meeden and Ghosh (1980), and Ghosh and Meeden (1980)). Also, the theorem can be extended to the case of a countable sample space χ(Θ being still finite) provided it is assumed that each admissible decision function has finite risk for each $\theta \in \Theta$. However, we have extended, in Section 4, the sufficiency part of Theorem 1 for arbitrary χ and arbitrary Θ under the assumption of finite Bayes risks of the "optimal" decision rules with respect to an ordered set of singular priors.

3. Admissible consistency and a characterization of inverse probability distributions. In this section we assume that χ is countable, while Θ is still finite. Suppose now that χ, Θ and $\{f_\theta, \theta \in \Theta\}$ are given and fixed. Let Π be a family of inverse probability distributions. We are interested in determining whether for each decision problem based on χ, Θ and $(f_\theta, \theta \in \Theta\}$ the corresponding optimal decision rule based on Π is admissible.

We need to introduce a few concepts before proving the main result. Let $L(\theta, \delta(x))$ denote the loss incurred by using a decision rule δ when θ is the true parameter value and x is the sample value. Suppose there exists a decision rule δ_0 such that for the family $\Pi = \{\pi_x(\cdot): x \in \chi\}$

$$\textstyle\sum_{\theta\in\Theta} L(\theta, \delta_0(x))\pi_x(\theta) = \inf_\delta \sum_{\theta\in\Theta} L(\theta, \delta(x))\pi_x(\theta),$$

for each $x \in \chi$. Then for any decision rule δ,

$$E_{\pi_x}[L(\tilde\theta, \delta_0(x)) - L(\tilde\theta, \delta(x))] \le 0, \tag{3.1}$$

for each x, where $\tilde\theta$ is a random variable assuming values θ in Θ. Does (3.1) imply that δ_0 is admissible, that is does there exist any δ such that

$$E_\theta[L(\theta, \delta_0(X)) - L(\theta, \delta(X))] \ge 0, \tag{3.2}$$

for all $\theta \in \Theta$ with strict inequality for some $\theta \in \Theta$?

With this in mind, we introduce the notion of *admissible consistency*. First let

$$t(\theta, x) = L(\theta, \delta_0(x)) - L(\theta, \delta(x)) - E_{\pi_x}[L(\tilde\theta, \delta_0(x)) - L(\tilde\theta, \delta(x))], \tag{3.3}$$

for each $\theta \in \Theta$ and $x \in \chi$. If (3.1) and (3.2) both hold, then for each x

$$E_{\pi_x} t(\tilde\theta, x) = 0, \tag{3.4}$$

while

$$E_\theta t(\theta, X) \ge 0, \tag{3.5}$$

for all $\theta \in \Theta$ with strict inequality for some $\theta \in \Theta$. Conversely, given any function $t(\theta, x)$ satisfying (3.4) one can construct loss functions $L(\theta, d)$ and decision rules δ and δ_0 such that $t(\theta, x) = L(\theta, \delta_0(x)) - L(\theta, \delta(x))$. We are now in a position to give a formal definition of "admissible consistency."

DEFINITION. Π is said to be admissible inconsistent with f if there exists a t such that (3.4) and (3.5) hold. Otherwise, Π is said to be admissible consistent with f.

Dawid and Stone (1972, 1973) have considered this situation when Θ is infinite and have defined the notion of *expectation inconsistency* for Π. When Θ is finite, Π is expectation inconsistent with f if there exists a t satisfying (3.4) and (3.5) with strict inequality in (3.5) for all $\theta \in \Theta$. (Actually Dawid and Stone have used utility rather than loss. But since utility can be interpreted as negative loss the two approaches are equivalent.)

They have shown that if Π is Bayes then it is expectation consistent with f. Conversely, if Π is expectation consistent with f, under very mild additional conditions it is Bayes. It is clear from the definition that admissible consistency implies expectation consistency.

Note, however, that the Bayesness of Π does not imply admissible consistency. The following simple example illustrates this.

Let $\chi = (x_1, x_2, x_3)$, $\Theta = \{\theta_1, \theta_2, \theta_3\}$; $f_{\theta_1}(x_1) = 1/3 = 1 - f_{\theta_1}(x_2)$, $f_{\theta_2}(x_1) = 3/4 = 1 - f_{\theta_2}(x_2)$, $f_{\theta_3}(x_3) = 1$, $\lambda(\theta_1) = 1/2 = \lambda(\theta_2)$. Then,

$$\pi_{x_1}(\theta_1) = 4/13 = 1 - \pi_{x_1}(\theta_2),\ \pi_{x_2}(\theta_1) = 8/11 = 1 - \pi_{x_2}(\theta_2). \tag{3.6}$$

Also, $\pi_{x_3}(\theta_j)$'s can be arbitrarily defined as long as they are nonnegative and $\sum_{j=1}^{3} \pi_{x_3}(\theta_j) = 1$. Take for example $\pi_{x_3}(\theta_1) = 1/4$, $\pi_{x_3}(\theta_2) = 1/3$, $\pi_{x_3}(\theta_3) = 5/12$. Now define $t(\theta_j, x_i) = 0$ $(i, j = 1, 2)$, $t(\theta_3, x_j)(j = 1, 2)$ arbitrary real numbers, $t(\theta_1, x_3) = -4/5$, $t(\theta_2, x_3) = -1/4$, and $t(\theta_3, x_3) = 1$. Then, one can easily verify that

$$E_{\pi_{x_i}} t(\tilde{\theta}, x_i) = 0\ (i +, 2, 3),\ E_{\theta_1} t(\theta_1, X) = E_{\theta_2} t(\theta_2, X) = 0,\ E_{\theta_3} t(\theta_3, X) = 1.$$

We now prove a theorem which shows that admissible consistency is equivalent to the Bayesness of the inverse probability distributions with respect to a "full" prior (i.e., a prior which puts positive mass on each point in the parameter space). It will be revealed in the proof that this theorem is in the spirit of Theorem 1 in Section 2. Recall the notations and definitions of Theorem 1.

THEOREM 2. *If Π is a family of inverse probability distributions which is admissible consistent with f, then there exists a prior distribution λ which puts positive mass on each parameter point and against which Π is Bayes. Conversely, if a family Π of inverse probability distributions is Bayes against a prior which puts positive mass on each parameter point, then Π is admissible consistent with f.*

PROOF. Suppose Π is admissible consistent with f. Then, proceed as in Lemma 1 of Dawid and Stone (1972) to get a prior λ^1 on Θ and a $\gamma_1 : \chi \to [0, \infty)$ such that

$$\pi_x(\theta)\gamma_1(x) = f_\theta(x)\lambda^1(\theta). \tag{3.7}$$

For $x \in \Lambda^1 = \{x : g(x; \lambda^1) > 0\}$, using $\sum_\theta \pi_x(\theta) = 1$, we find $\gamma_1(x) = g(x; \lambda^1) > 0$ and (3.7) can be rewritten as

$$\pi_x(\theta) g(x; \lambda^1) = f_\theta(x)\lambda^1(\theta) \tag{3.8}$$

Now if $\Theta(\lambda^1) = \Theta$ the proof is complete since $\lambda = \lambda^1$. So we suppose that $\Theta(\lambda^1) \neq \Theta$.

We first note that

$$\begin{aligned} &\text{if } x \in \Lambda^1 \quad \text{and} \quad \theta \notin \Theta(\lambda^1) \quad \text{then } \pi_x(\theta) = 0 \\ &\text{and} \\ &\text{if } x \notin \Lambda^1 \quad \text{and} \quad \theta \in \Theta(\lambda^1) \quad \text{then } \pi_x(\theta) = 0. \end{aligned} \tag{3.9}$$

The first equation of (3.9) follows from (3.8). To prove the second, it is enough to show that one cannot simultaneously have both $f_{\theta_0}(x_0) = 0$ and $\pi_{x_0}(\theta_0) > 0$. Now if this were true we could define $t(\theta, x) = 0$ for $x \neq x_0$, $= \pi_{x_0}(\theta_0)$ for $x = x_0$ and $\theta \neq \theta_0$, $= \pi_{x_0}(\theta_0) - 1$ for $x = x_0$ and $\theta = \theta_0$. Then t satisfies (3.4) and

$$\begin{aligned} E_\theta[t(\theta, X)] &= 0 && \text{for } \theta = \theta_0 \\ &= f_\theta(x_0)\pi_{x_0}(\theta_0) && \text{for } \theta \neq \theta_0. \end{aligned}$$

Since the family Π is admissible consistent, we must have $f_\theta(x_0) = 0$, for all θ, which has been excluded by assumption—a contradiction.

We now show that

$$\text{for } \theta \in \Theta(\lambda^1), \quad \text{if} \quad f_\theta(x) > 0 \quad \text{then} \quad x \in \Lambda^1,$$

(3.10) and

$$\text{for } \theta \notin \Theta(\lambda^1) \quad \text{if} \quad f_\theta(x) > 0 \quad \text{then} \quad x \notin \Lambda^1.$$

The first equation follows immediately from the definition of Λ^1. To prove the second, suppose that for some $x_0 \in \Lambda^1$ and $\theta_0 \notin \Theta(\lambda^1)$ that $f_{\theta_0}(x_0) > 0$. By taking $t(\theta_0, x_0) > 0$ and $t(\theta, x) = 0$ otherwise, it follows from (3.8) that the family Π is admissible inconsistent, which is a contradiction. Note that from (3.9) and (3.10), (3.8) must be true from all θ and all x. We see from the second equation of (3.10) that $x - \Lambda^1$ is nonempty since $\Theta - \Theta(\lambda^1)$ is nonempty.

We now consider the restricted problem where $x \in \chi - \Lambda^1$ and $\theta \in \Theta - \Theta(\lambda^1)$. From (3.10) we see that for $\theta \in \Theta - \Theta(\lambda^1)$, $f_\theta(\cdot)$ is a bonafide probability function for the restricted problem. Similarly from (3.9) we see that from $x \in \chi - \Lambda^1$, $\pi_x(\cdot)$ is an inverse probability distribution for this restricted problem. If $\{\pi_x,\ x \in \chi - \Lambda^1\}$ is admissible inconsistent with f for the restricted problem, then there exists a t^* defined on $(\chi - \Lambda^1) \times (\Theta - \Theta(\lambda^1))$ such that

$$E_{\pi_x} t^*(\tilde{\theta}, x) = 0 \qquad \text{for all } x \in \chi - \Lambda^1$$

and

$$E_\theta t^*(\theta, X) \geq 0 \qquad \text{for all } \theta \in \Theta - \Theta(\lambda^1),$$

with strict inequality for some θ. Defining

$$\begin{aligned} t(\theta, x) &= t^*(\theta, x) \qquad \text{for } \theta \in \Theta - \Theta(\lambda^1) \quad \text{and} \quad x \in \chi - \Lambda^1 \\ &= 0, \qquad \text{otherwise,} \end{aligned}$$

it follows that Π is admissible inconsistent with f for the original problem. Hence, similar to the first part of the proof there exist $\lambda^2(\theta)$ and $\gamma_2(x)$ such that

$$\pi_x(\theta)\gamma_2(x) = f_\theta(x)\lambda^2(\theta) \qquad \text{for } x \in \chi - \Lambda^1 \quad \text{and} \quad \theta \in \Theta - \Theta(\lambda^1). \tag{3.11}$$

For $x \in \Lambda^2 = \{x: g(x; \lambda^2) > 0\}$ we see that $\gamma_2(x) = g(x; \lambda^2) > 0$ since $\sum_\theta \pi_x(\theta) = 1$ and equation (3.11) becomes

$$\pi_x(\theta)\, g(x; \lambda^2) = f_\theta(x)\lambda^2(\theta) \tag{3.12}$$

for $x \in \Lambda^2$.

Now it is easily seen that the equations in (3.9) and (3.10) are true when λ^1 and Λ^1 are replaced by λ^2 and Λ^2. Hence, (3.12) is true for all x and all θ.

We now let $\lambda^* = (\frac{1}{2})\lambda^1 + (\frac{1}{2})\lambda^2$. We see that $\Theta(\lambda^*) = \Theta(\lambda^1) \cup \Theta(\lambda^2$ and $\Lambda = \{x: g(x; \lambda^*) > 0\} = \Lambda^1 \cup \Lambda^2$. Now if $\Theta(\lambda^*) = \Theta$ the proof is complete. Since (3.8) holds for all x when λ^2 is replaced by λ^*, we can take $\lambda = \lambda^*$. If $\Theta(\lambda^*) \neq \Theta$ then the equations in (3.9) and (3.10) hold for λ^* and Λ and we consider the restricted problem when $x \in \chi - \Lambda$ and $\theta \in \Theta - \Theta(\lambda^*)$. Since Θ is finite the proof is complete by induction. As noted in the introduction, the converse is well known, and so the theorem is proved.

Even though the statement of Theorem 2 involves just one prior, it is very similar in spirit to Theorem 1 as the proof demonstrates. Let $\lambda^1, \cdots, \lambda^n$ and $^*\Lambda^1, \cdots, {}^*\Lambda^n$ be as in Theorem 1. Let $\Lambda^i = \{x: g(x: \lambda^i) > 0\}$ for $i = 1, \cdots, n$. Now if it happens that $\Lambda^i = {}^*\Lambda^i$ for $i = 1, \cdots, n$ and $\cup_{i=1}^n \Theta(\lambda^i) = \Theta$ then each pair $(\Lambda^i, \Theta(\lambda^i))$ can be regarded as the sample space and parameter space of an isolated restricted problem, and the overall solution can be found by analyzing each restricted problem separately. In this case in Theorem 1 the order of the λ^i's is not important. In Theorem 2, we can take λ to be any convex combination of the λ^i's which puts positive weight on each λ^i for $i = 1, \cdots, n$.

4. Admissibility when the parameter space is not necessarily finite. Consider now the situation when the sample space χ and the parameter space Θ are arbitrary. For each θ let $f_\theta(\cdot)$ be a discrete probability function over χ. It is assumed that for each $x \in \chi$, $f_\theta(x) > 0$ for at least one $\theta \in \Theta$. Let λ be a discrete prior distribution on Θ putting mass on at most a countable number of points. For any such prior λ, let $\Theta(\lambda) = \{\theta \in \Theta:$

$\lambda(\theta) > 0\}$. The definition of singularity of two priors λ and λ' is the same as in Section 2. Also, the marginal distribution of X is denoted by $g(x; \lambda)$ with respect to the prior λ. The sufficiency part of Theorem 1 can now be extended as follows.

THEOREM 3. *Let $\Pi = \{\pi_x(\cdot): x \in \chi\}$ be a family of inverse probability distributions. Consider the decision problem specified by the decision space D and loss function $L(\cdot, \cdot)$. Suppose there exists a nonempty set $S = \{\lambda^\alpha: \alpha \in I\}$ of mutually singular prior* distributions where I is a well ordered set with smallest element $\alpha(1)$ such that

(i) *if* $^*\Lambda^{\alpha(1)} = \{x: g(x; \lambda^{\alpha(1)}) > 0\}$ *and* $^*\Lambda^\alpha = \{x: g(x; \lambda^\alpha) > 0$ *and* $x \notin \cup_{\alpha'<\alpha} {}^*\Lambda^{\alpha'}\}$ *for all* $\alpha > \alpha(1)$, *then each* $^*\Lambda^\alpha$ *is nonempty and* $\cup_{\alpha\in I} {}^*\Lambda^\alpha = \chi$;

(ii) if $x \in {}^*\Lambda^\alpha$, *then*

$$\pi_x(\theta)g(x; \lambda^\alpha) = f_\theta(x)\lambda^\alpha(\theta) \text{ for all } \theta \in \Theta$$

If, (a) *if for each* $x \in {}^*\Lambda^\alpha$, $\delta^\alpha(x)$ *uniquely minimizes* $\sum_{\theta\in\Theta} L(\theta, d)\pi_x(\theta)$ *with respect to* d, *and* (b) $\sum_{\theta\in\Theta(\lambda^\alpha)} \sum_{x\in{}^*\Lambda^\alpha} L(\theta, \delta^\alpha(x))f_\theta(x)\lambda^\alpha(\theta)/b^\alpha(\theta) < \infty$, *where* $b^\alpha(\theta) = \sum_{x\in{}^*\Lambda^\alpha} f_\theta(x)$, $\theta \in \Theta(\lambda^\alpha)$, *then the estimator* δ *given by* $\delta(x) = \delta^\alpha(x)$ *for* $x \in {}^*\Lambda^\alpha$ *is admissible.*

PROOF. Suppose not, then there exists some decision rule $\bar\delta^*$ such that

$$E_\theta L(\theta, \bar\delta^*(X)) \le E_\theta L(\theta, \delta(X)) \tag{4.1}$$

for all $\theta \in \Theta$ with strict inequality for some $\theta \in \Theta$. Now consider the restricted problem with x restricted to $^*\Lambda^{\alpha(1)}$ and θ restricted to $\Theta(\lambda^{\alpha(1)})$. In view of (ii), (a) and (b), δ restricted to $^*\Lambda^1$ is the unique "optimal" decision rule for the problem. So, $\bar\delta^*(x) = \delta(x)$ for $x \in {}^*\Lambda^{\alpha(1)}$ and equality holds in (4.1) for all $\theta \in \Theta(\lambda^{\alpha(1)})$.

Assume now that $\bar\delta^*(x) = \delta(x)$ for all $x \in \cup_{\alpha'<\alpha} {}^*\Lambda^{\alpha'}$ where α is a member of I. Consider now the restricted problem where x is restricted to the set $^*\Lambda^\alpha$ and θ is restricted to the set $\Theta(\lambda^\alpha)$ and for $\theta \in \Theta(\lambda^\alpha)$, $\bar f_\theta^*(x) = f_\theta(x)/b^\alpha(\theta)$ is the probability function. In view of (ii), (a) and (b), δ restricted to $^*\Lambda^\alpha$ is the unique optimal decision rule for the restricted problem and hence $\bar\delta^*(x) = \delta(x)$ for $x \in {}^*\Lambda^\alpha$ as well. Applying the principle of transfinite induction we have $\bar\delta^*(x) = \delta(x)$ for all $x \in \chi$; hence there is equality in (4.1) for all θ, which is a contradiction. This completes the proof.

Next we consider an application of this theorem in the discrete uniform model where the probability function of X is given by

$$f_\theta(x) = P_\theta(X = x) = \theta^{-1}, \qquad x = 1, 2, \cdots, \theta; \tag{4.2}$$

$\Theta = \{1, 2, \cdots\} = \chi$. In this case Blyth (1974) has proved the admissibility of X as an estimator of θ under squared error loss by using a Cramér-Rao type inequality. An alternate way to prove this result would be to use our Theorem 3 with the priors λ_θ putting unit mass at θ, $\theta = 1, 2, \cdots$.

Note that the unique minimum variance unbiased estimator (UMVUE) of θ in this case is $2X - 1$. It can be verified by direct computations that this estimator can be represented as the posterior mean with respect to the prior λ^1 with $\lambda^1(1) = 1$ when $x = 1$, and with respect to prior $\lambda^2(\theta) = 4/[3(\theta^2 - 1)]$, $\theta = 2, 3, \cdots$ when $x = 2, 3, \cdots$. Note that λ^1 and λ^2 are mutually singular. However, since the posterior risk of every estimator under $\lambda^2(\theta)$ and squared error loss is infinite, (a) is not satisfied, and Theorem 3 cannot be used to prove the admissibility of $2X - 1$. In fact, in this case one can directly verify that X dominates $2X - 1$ under the squared error loss.

Now, under the model (4.2), for every function $\gamma(\theta)$ of θ, the UMVUE of $\gamma(\theta)$ is given by $g(X)$ where

$$g(x) = x\gamma(x) - (x - 1)\gamma(x - 1), \qquad x = 1, 2, \cdots, \tag{4.3}$$

where $\gamma(0)$ can be arbitrarily defined. For $\theta \ge 2$ let

$$q(\theta; \gamma) = \{(\theta - 1)(\gamma(\theta) - \gamma(\theta - 1))\}^{-1} - \{(\theta + 1)(\gamma(\theta + 1) - \gamma(\theta))\}^{-1}$$

and suppose $q(\theta; \gamma) > 0$ for $\theta \geq 2$ and $\sum_{\theta=2}^{\infty} q(\theta; \gamma) < \infty$. If λ^1 and λ^2 are the mutually singular priors given by

$$\lambda^1(1) = 1 \quad \text{and} \quad \lambda^2(\theta) \propto q(\theta; \gamma) \quad \text{for} \quad \theta \geq 2, \tag{4.4}$$

then $g(x)$ is the posterior mean of $\gamma(\theta)$ with respect to the mutually singular priors λ^1 and λ^2 for $x = 1$ and $x \geq 1$ respectively. However, as already noted, in many of these cases Theorem 3 is not applicable when the loss function is $L(\theta, d) = (d - \gamma(\theta))^2$.

We consider below one example where Theorem 3 can be applied to prove the admissibility of $g(x) = x^{3/2} - (x-1)^{3/2}$ in estimating $\gamma(\theta) = \theta^{1/2}$ with squared error loss. In this case

$$E_\theta g^2(X) = E_\theta(X^3) + E_\theta(X-1)^3 - 2E_\theta(X^{3/2}(X-1)^{3/2}). \tag{4.5}$$

Using the inequality $x^{3/2}(x-1)^{3/2} \geq x^3 - (3/2)x^2$ for all $x = 2, 3, \cdots$ it follows from (4.5) that

$$E_\theta g^2(X) \leq 3E_\theta(X) = (3/2)(\theta + 1) < 3\theta \quad \text{for all} \quad \theta = 2, 3, \cdots. \tag{4.6}$$

Hence, the Bayes risk of $g(X)$ against the prior λ^2 is

$$\begin{aligned} r(\lambda^2, g) &\leq 3 \sum_{\theta=2}^{\infty} \theta \left\{ \frac{1}{(\theta-1)(\theta^{3/2} - (\theta-1)^{3/2})} - \frac{1}{(\theta+1)((\theta+1)^{3/2} - \theta^{3/2})} \right\} \\ &= 3 \sum_{\theta=2}^{\infty} \frac{\theta\{(\theta+1)^{5/2} + (\theta-1)^{5/2} - 2\theta^{5/2}\}}{(\theta^2-1)\theta^3(1 - (1-\theta^{-1})^{3/2})((1+\theta^{-1})^{3/2} - 1)} \\ &\leq 3 \sum_{\theta=2}^{\infty} \frac{\theta^{7/2} 15/4\theta^2(1 + 0(\theta^{-2}))}{(\theta^2-1)\theta^3\left(\dfrac{3}{2\theta} + 0(\theta^{-2})\right)\left(\dfrac{3}{2\theta}\right)} \leq C \sum_{\theta=2}^{\infty} \theta^{-3/2} < \infty, \end{aligned} \tag{4.7}$$

where C is some constant not depending on θ. Thus applying our Theorem 3, admissibility of g in estimating $\gamma(\theta) = \theta^{1/2}$ follows.

Acknowledgments. Thanks are due to the Associate Editor for his careful and constructive criticisms on earlier versions of the manuscript.

REFERENCES

BLYTH, C. R. (1974). Necessary and sufficient conditions for inequalities of Cramér-Rao types. *Ann. Statist.* **2** 464–473.

BROWN, LAWRENCE D. (1979). A complete class theorem for statistical problems with finite sample spaces. Unpublished manuscript.

DAWID, A. P. and STONE, M. (1972). Expectation consistency of inverse probability distributions. *Biometrika* **59** 486–489.

DAWID, A. P. and STONE, M. (1973). Expectation consistency and generalized Bayes inference. *Ann. Statist.* **1**. 478–485.

GHOSH, M. and MEEDEN, G. (1980). Biased estimation of the variance in finite population sampling. Unpublished manuscript.

HSUAN, F. (1979). A stepwise Bayesian procedure. *Ann. Statist.* **7** 860–868.

JOHNSON, BRUCE MCK. (1971). On the admissible estimators for certain fixed sample binomial problems. *Ann. Math Statist.* **42** 1579–1587.

MEEDEN, G. and GHOSH, M. (1980). Finite admissibility in finite population sampling. Unpublished manuscript.

WALD, A. and WOLFOWITZ, J. (1951). Characterization of the minimal complete class of decision functions. *Proc. 2nd. Berkeley Symp. Math. Statist. and Prob.* 149–157. (J. Neyman and E. Scott, eds.) Univ. California.

DEPARTMENT OF STATISTICS
IOWA STATE UNIVERSITY
SNEDECOR HALL
AMES, IOWA 50011