The Annals of Statistics
1983, Vol. 11, No. 1, 296–305

CHOOSING BETWEEN EXPERIMENTS: APPLICATIONS TO FINITE POPULATION SAMPLING[1]

BY GLEN MEEDEN AND MALAY GHOSH

Iowa State University

Suppose that a statistician is faced with a decision problem involving an unknown parameter. Before making his decision he can carry out one of two possible experiments. Assume that he may choose at random which of the two experiments he will observe. For this problem a decision procedure for the statistician is a triple consisting of the randomizing probability measure he uses to choose between the experiments, the decision function he uses if he observes the first experiment and the decision function he uses if he observes the second experiment. The main theorem of this paper identifies the set of such admissible triples when the parameter space, and the sample spaces of the two experiments are finite. This result is then used to find some uniformly admissible procedures for some problems in finite population sampling.

1. Introduction. Suppose that θ, the true but unknown state of nature, is known to belong to some finite set Θ and the statistician is faced with the decision problem specified by the decision space D and the loss function $L(\theta, d)$, $d \in D$. Before making his decision, however, the statistician may choose to observe one of two possible experiments. For each experiment the family of possible distributions over a finite set of possible outcomes is indexed by the parameter θ. The statistician may even choose the experiment at random from these two. Suppose he observes the first experiment with probability γ and the second with probability $1 - \gamma$ where $\gamma \in [0, 1]$. Suppose that if the first experiment is chosen, he uses the decision function δ, while if the second experiment is chosen he uses the decision function ϕ. The problem for the statistician is how to choose the triple (γ, δ, ϕ). In this paper we study the admissibility of such triples.

Blackwell (1951, 1953) discussed the problem of comparing two experiments. He introduced the notion of one experiment being more informative than another and gave necessary and sufficient conditions for this notion to be true. Now if one experiment is more informative than another then it is clear that the statistician need only consider the more informative one. However, the more typical case is where neither experiment is more informative than the other. It is for this situation that the admissibility of the triples (γ, δ, ϕ) are of interest.

A naive first thought might be that if δ is admissible for the first experiment and ϕ is admissible for the second experiment then any choice of $\gamma \in [0, 1]$ yields an admissible triple. It can be easily shown that this is not true however. For example, in the trivial situation where the first experiment just consists in using δ alone and the second just consists in using ϕ alone with $r(\theta, \delta) = 1$ for all $\theta \in \Theta$ and $r(\theta, \phi) = 2$ for all $\theta \in \Theta$, any triple (γ, δ, ϕ) with $0 \le \gamma < 1$ is inadmissible. It is useful to think about how a Bayesian would approach the problem. Suppose λ, a probability distribution over Θ which assigns positive mass to every member of Θ, is the statistician's prior distribution. The statistician can then find his Bayes rule and overall Bayes risk for each experiment. He would then observe the experiment with the smaller overall Bayes risk. The only time he would

Received February 1981; revised June 1982.

[1] Research supported by the NSF Grant Number MCS-8005485.

AMS 1970 *subject classifications*. 62C15, 62D05, 62C10.

Key words and phrases. Choosing between experiments, admissibility, Bayes, orthogonal priors, finite population sampling, uniform admissibility, ratio estimator, Horvitz-Thompson estimator, Basu estimator, choice of designs.

randomly choose between the two experiments is when the Bayes risks were equal. Hence, the Bayesian would randomize only in the cases where he is indifferent between the two possible choices. It can be shown that for such a Bayesian, his resulting triple will always be admissible. This Bayesian idea of how to choose between the two experiments is well known and has often been used and discussed in the literature. It does not answer, in general, the admissibility question posed here. Because of the close relationship between admissibility and Bayesness, this idea is suggestive, however. In what follows, an extension of the Bayesian idea yields the answer to the question of admissibility for the triples (γ, δ, ϕ).

In Meeden and Ghosh (1981) the family of admissible decision rules was characterized for certain decision problems with a finite sample space and a finite parameter space. It was shown that a decision function is admissible if and only if it is "stepwise Bayes against a full family of mutually orthogonal prior distributions" (see Brown, 1981, and Hsuan, 1979, for a precise meaning of this statement). In this paper we show that a triple (γ, δ, ϕ) with $0 < \gamma < 1$ is admissible if and only if there exists a full family $\lambda^1, \cdots, \lambda^n$ of priors, such that for the first experiment, δ is stepwise Bayes against the family, and for the second experiment, ϕ is stepwise Bayes against the family and the Bayes risk of δ against λ^i is equal to the Bayes risk of ϕ against λ^i for all $i = 1, \cdots, n$. In this case any new triple that is formed by replacing γ with any other number from [0, 1] is admissible as well. Suppose now that δ and ϕ are stepwise Bayes against a full family of priors for their respective problems but the corresponding Bayes risks are not all equal. In this case there must exist a positive integer $j^* \leq n$, such that, the Bayes risk of δ against λ^i is equal to the Bayes risk of ϕ against λ^i for $i = 1, 2, \cdots, j^* - 1$ and their Bayes risks against λ^{j^*} are not equal. In this case, if the Bayes risk of δ against λ^{j^*} is smaller (larger) than the Bayes risk of ϕ against λ^{j^*}, then the triple $(1, \delta, -)((0, -, \phi))$ is admissible. These facts are summarized in Theorem 1. They can be easily generalized to the case where the statistician can choose, at random if he wishes, from a finite set of experiments. These and related points are discussed in Section 2.

In Section 3 these ideas are applied to finite population sampling. Suppose one is interested in estimating the population total of a population consisting of N units with squared error loss using a design p of fixed sample size n. Let W_n be the set of all designs of fixed sample size n. Given a design $p \in W_n$ and an estimator δ, one is often interested in knowing if δ is admissible when p is the design which will be used. Even of greater interest is deciding whether or not there exists a $p' \in W_n$ and a δ' such that risk function of the pair (p', δ') dominates the risk function of the pair (p, δ). If no such dominating pair exists then the pair (p, δ) is said to be uniformly admissible relative to W_n. This notion was first discussed in detail in Joshi (1966).

The problem of choosing a design and an estimator such that the pair is uniformly admissible relative to W_n is just a special case of the problem considered in this paper. In finite population sampling, however, the usual convention is to assume that the parameter space is N dimensional Euclidean space and the results of Section 2 are not directly applicable. In Section 3 it is shown how admissibility and uniform admissibility questions can sometimes be resolved by only considering finite subsets of the parameter space and a sufficient condition for uniform admissibility is given. From this condition, the uniform admissibility relative to W_n of the usual estimator of the population total along with any design from W_n follows easily. This gives an alternate proof of a fact first proved in Joshi (1966). Other uniform admissibility resuts of Godambe (1969), Ericson (1970) and Chaudhuri (1978) are simple consequences as well. Basu (1971) introduced an interesting estimator which can be thought of as a psuedo-Bayesian alternative to the classical ratio estimator and the Horvitz-Thompson estimator. First it is shown that Basu's estimator is admissible for any fixed design whatsoever. Then a subset of designs of W_n is identified with the property that when any one of them is used with Basu's estimator, the pair is uniformly admissible relative to W_n. It is seen that this set usually contains just one design which puts probability one on some sample.

2. Admissibility in the choice of an experiment. Let Θ, a finite set, denote the parameter space which contains the true but unknown state of nature θ. Let D be the decision space with generic element d. Let $L(\theta, d)$ be the nonnegative loss function. Assume that $L(\cdot, \cdot)$ is such that for any prior distribution λ on Θ, $\sum_\theta L(\theta, d)\lambda(\theta)$, as a function of d, is uniquely minimized by a member of D. Let X be a random variable with a family $\{f_\theta: \theta \in \Theta\}$ of possible probability functions. Let $\mathscr{X}$ be the sample space of X. Assume that $\mathscr{X}$ is finite and for each $x \in \mathscr{X}$ there exists a $\theta \in \Theta$ such that $f_\theta(x) > 0$. Let Y be a random variable with a family $\{p_\theta: \theta \in \Theta\}$ of possible probability functions. Let $\mathscr{Y}$ be the sample space of Y. Assume that $\mathscr{Y}$ is finite and for each $y \in \mathscr{Y}$ there exists a $\theta \in \Theta$ such that $p_\theta(y) > 0$. Finally let $\delta(\phi)$ denote a typical decision function (possible randomized) from $\mathscr{X}(\mathscr{Y})$ to D with risk function $r_x(\theta; \delta)(r_y(\theta; \phi))$.

Now the statistician can choose to observe X with probability γ and Y with probability $1 - \gamma$ where γ is any number contained in [0, 1]. If it turns out that $X(Y)$ is actually observed then he must use $\delta(\phi)$ to make his decision. Hence, for the statistician a decision procedure for this problem is a triple (γ, δ, ϕ). For such a triple its risk function is

$$r(\theta; \gamma, \delta, \phi) = \gamma r_x(\theta; \delta) + (1 - \gamma) r_y(\theta; \phi). \tag{2.1}$$

Theorem 1, the main result of this paper, essentially characterizes the class of admissible triples. Before stating the theorem, we need to introduce some more notations.

If λ is a prior distribution over Θ then

$$R(\gamma, \delta, \phi; \lambda) = \gamma R_x(\delta; \lambda) + (1 - \gamma) R_y(\phi; \lambda) \tag{2.2}$$

is the Bayes risk of the triple (γ, δ, ϕ) against λ where $R_x(\delta; \lambda)(R_y(\phi; \lambda))$ is the Bayes risk of $\delta(\phi)$ against λ. Let

$$g(x; \lambda) = \sum_\theta f_\theta(x)\lambda(\theta) \quad \text{and} \quad q(y; \lambda) = \sum_\theta p_\theta(y)\lambda(\theta)$$

be the marginal probability functions of X and Y respectively under the prior λ. For the prior λ let $\Theta(\lambda) = \{\theta; \lambda(\theta) > 0\}$. Two priors λ^i and $\lambda^j (i \neq j)$ are said to be orthogonal if $\Theta(\lambda^i) \cap \Theta(\lambda^j)$ is empty.

Let $\lambda^1 \cdots, \lambda^n$ be a set of priors on Θ. Let

$$\Lambda_x^1 = \{x: g(x; \lambda^1) > 0\}$$

and

$$\Lambda_x^j = \{x: x \notin \cup_{i=1}^{j-1} \Lambda_x^i \quad \text{and} \quad g(x; \lambda^j) > 0\}$$

for $j = 2, \cdots, n$. Note that some of the Λ_x^j's may be empty and that the set associated with a particular prior depends on the other priors in the sequence and its place in the sequence; $\Lambda_y^1, \cdots, \Lambda_y^n$ are defined in an analogous way.

A decision rule δ for the X problem is said to be stepwise Bayes (see Hsuan, 1979) against $\lambda^1, \cdots, \lambda^n$ if $\delta(x) = \delta_i^*(x)$ for all $x \in \Lambda_x^i$ for $i = 1, \cdots, n$ where δ_i^* is Bayes against λ_i. A decision rule ϕ for the Y problem is defined to be stepwise Bayes in a similar way.

THEOREM 1. (a) *Let* $\lambda^1, \cdots, \lambda^n$ *be a set of mutually orthogonal priors.*
(a.I) *If* $\lambda^1, \cdots, \lambda^n$ *are such that*

$$\begin{aligned} &\text{(i) } \Lambda_x^j \cup \Lambda_y^j \textit{ is nonempty for } j = 1, \cdots, n \\ &\textit{and} \\ &\text{(ii) } \cup_{j=1}^n (\Lambda_x^j \cup \Lambda_y^j) = \chi \cup \mathscr{Y} \end{aligned} \tag{2.3}$$

(2.4) *and if* δ *and* ϕ *are decision rules which are stepwise Bayes against* $\lambda^1, \cdots, \lambda^n$ *for the* X *and* Y *problems respectively, and if for some* $1 \le j^* \le n$ *it is the case that*

$$\begin{aligned} &R_x(\delta; \lambda^j) = R_y(\phi; \lambda^j) \quad \textit{for} \quad j = 1, \cdots, j^* - 1 \\ &\textit{and} \\ &R_x(\delta; \lambda^{k*}) < R_y(\phi; \lambda^{j*}) \end{aligned} \tag{2.5}$$

then $(1, \delta, -)$ *is admissible.*
(a.II) *If neither* $r_x(\theta; \delta) \le r_y(\theta; \phi)$ *for all* θ *with strict inequality for at least one* θ *nor vice versa, and if* (2.3) *and* (2.4) *are true and if*

$$R_x(\delta; \lambda^j) = R_y(\phi; \lambda^j) \quad \text{for} \quad j = 1, \cdots, n \tag{2.6}$$

then (γ, δ, ϕ) *is admissible for any* $\gamma \in [0, 1]$.
(b) *Conversely*
(b.I) *If* $(1, \delta, -)$ *is admissible then there exists a set of mutually orthogonal priors such that* (2.3) *and* (2.4) *are true and*

$$R_x(\delta; \lambda^j) \le R_y(\phi; \lambda^j) \quad \text{for} \quad j = 1, \cdots, n. \tag{2.7}$$

(b.II) *If* (γ, δ, ϕ) *is admissible where* $0 < \gamma < 1$ *then there exists a set of mutually orthogonal priors such that* (2.3), (2.4) *and* (2.6) *are true and* (γ', δ, ϕ) *is admissible as well where* $0 \le \gamma' \le 1$.

PROOF. Before proving part (a) we prove the following lemma.

LEMMA. *Let* $\lambda^1, \cdots, \lambda^n$ *be a set of priors and let* (γ, δ, ϕ) *be a triple for which* (2.4) *and* (2.6) *are true. If* $(\gamma_0, \delta_0, \phi_0)$ *is a triple which is at least as good as* (γ, δ, ϕ) *then* (2.4) *and hence* (2.6) *is true for* $(\gamma_0, \delta_0, \phi_0)$ *as well.*

PROOF. The proof of the lemma is by induction. The case $n = 1$ follows from (2.2), (2.4) and the uniqueness assumption on the loss function. We now prove the inductive step. Assume we have the priors $\lambda^1, \cdots, \lambda^n$. Now by induction

$$\delta_0(x) = \delta(x) \quad \text{for} \quad x \in \cup_{j=1}^n \Lambda_x^j \quad \text{and} \quad \phi_0(y) = \phi(y) \quad \text{for} \quad y \in \cup_{j=1}^n \Lambda_y^j.$$

If $\Lambda_x^{n+1} \cup \Lambda_y^{n+1}$ is empty then the inductive step is true trivially, so we assume this set is nonempty. Consider the restricted problem where $\theta \in \Theta(\lambda^{n+1})$ and the set of possible decision rules is all the rules which agree with (γ, δ, ϕ) on $(\cup_{j=1}^n \Lambda_x^j) \cup (\cup_{j=1}^n \Lambda_y^j)$. For this problem $(\gamma_0, \delta_0, \phi_0)$ is Bayes against λ^{n+1} and hence (2.4) holds for $\lambda^1, \cdots, \lambda^{n+1}$ and the lemma is proved.

We now prove (a.I) of the theorem. Suppose $(\gamma_0, \delta_0, \phi_0)$ is at least as good as $(1, \delta, -)$. Then by the lemma $\delta_0(x) = \delta(x)$ for $x \in \cup_{j=1}^{j^*-1}$ and $\phi_0(y) = \phi(y)$ for $y \in \cup_{j=1}^{j^*-1} \Lambda_y^j$ and

$$\begin{aligned} R_x(\delta, \lambda^{j*}) &\ge \gamma_0 R_x(\delta_0, \lambda^{j^*}) + (1 - \gamma_0) R_y(\phi_0, \lambda^{j^*}) \\ &\ge \gamma_0 R_x(\delta, \lambda^{j^*}) + (1 - \gamma_0) R_y(\phi, \lambda^{j^*}) \ge R_x(\delta, \lambda^{j^*}). \end{aligned}$$

By assumption (2.5) the last inequality is strict (a contradiction) unless $\gamma_0 = 1$. This implies that the triple $(1, \delta_0, -)$ dominates $(1, \delta, -)$ which is not possible by Theorem 1 of Meeden and Ghosh (1981). This proves (a.I).

To prove (a.II), note that if $(\gamma_0, \delta_0, \phi_0)$ is at least as good as (γ, δ, ϕ), then by the lemma $\delta_0 \equiv \delta$ and $\phi_0 \equiv \phi$. Since the risk function of δ does not dominate the risk function of ϕ and vice versa, $\gamma_0 = \gamma$ and (γ, δ, ϕ) is admissible.

To prove (b.II) we first prove by induction that there exists a set of mutually orthogonal priors satisfying (ii) of (2.3), (2.4) and (2.6). Since (γ, δ, ϕ) is admissible then there exists a prior λ^1 (see page 86 of Ferguson, 1967) such that (γ, δ, ϕ) is Bayes against λ^1, i.e., (2.4) holds for λ^1; (2.6) must be satisfied for λ^1 as well. If the Bayes risks are unequal, for example, if $R_x(\delta, \lambda^1) < R_y(\phi, \lambda^1)$, then $(1, \delta, -)$ has smaller Bayes risk against λ^1 than (γ, δ, ϕ) which is a contradiction.

Suppose now we have n mutually orthogonal priors $\lambda^1, \cdots, \lambda^n$, such that (2.4) and (2.6) are satisfied but (ii) of (2.3) is not. Consider now the restricted problem for $\theta \in \cup_{j=1}^n \Theta(\lambda^j)$ and $\tilde{\mathscr{X}} = \mathscr{X} - \cup_{j=1}^n \Lambda_x^j$ and $\tilde{\mathscr{Y}} = \mathscr{Y} - \cup_{j=1}^n \Lambda_y^j$ as the sample spaces with the corresponding probability functions rescaled if necessary. Note that for $\theta \in \cup_1^n \Theta(\lambda^i)$ neither $r_x(\theta, \delta)$ dominates $r_y(\theta, \phi)$ nor vice versa. Therefore, if (γ, δ, ϕ) is inadmissible for the restricted problem with $\theta \notin \cup_{j=1}^{n}\Theta(\lambda^j)$, it is inadmissible for the original problem as well, which is a contradiction. Hence, by page 86 of Ferguson (1967) the triple (γ, δ, ϕ) is Bayes against some prior λ^{n+1} for this restricted problem, i.e., (γ, δ, ϕ) and $\lambda^1, \cdots, \lambda^n$ satisfy (2.4) and (2.6). Since Θ, $\mathscr{X}$ and $\mathscr{Y}$ are all finite, there must exist a finite set of mutually orthogonal priors such that (ii) of (2.3), (2.4) and (2.6) are satisfied. After removing from this set all priors λ^i such that $\Lambda_x^i \cup \Lambda_y^i$ is empty, the remaining set satisfies (i) of (2.3) as well and (b.II) is proved.

The proof of (b.I) is similar to that of (b.II), and is omitted.

To see that in part (a.II) the assumption that neither $r_x(\theta; \delta)$ dominates $r_y(\theta; \phi)$ nor vice versa is needed, we consider a simple example.

Suppose that $\mathscr{X} = \mathscr{Y} = \{0, 1\}$ and $\Theta = \{0, \frac{1}{2}, 1\}$. Let $f_0(0) = p_0(0) = \frac{1}{3}$, $f_1(0) = p_1(0) = \frac{1}{2}$ and $f_{1/2}(0) = p_{1/2}(1) = \frac{3}{4}$. Let λ be the prior which puts mass $\frac{1}{2}$ on 0 and 1. Let $\delta(\phi)$ denote the unique Bayes estimator of θ with squared error loss for the $X(Y)$ problem. Then it is easy to check that $r_x(\theta; \delta) = r_y(\theta; \phi)$ for $\theta = 0$ and 1 while $r_x(\frac{1}{2}; \delta) > r_y(\frac{1}{2}; \phi)$. So even though $R_x(\delta; \lambda) = R_y(\phi; \lambda)$ the triple $(1, \delta, -)$ is not admissible.

In the theorem we have assumed that the parameter space is identical for the two problems. This is the most realistic assumption although it is possible to think of examples where the parameter spaces for the two problems would not be equal. The theorem remains true in the more general situation as well. To see this, let Θ_x and Θ_y be the parameter spaces for the X and Y problems respectively and assume that $\Theta_x \neq \Theta_y$. Let $\Theta = \Theta_x \cup \Theta_y$ and consider $\mathscr{X}' = \mathscr{X} \cup \{a\}$ as a new sample space for the X problem. Let the probability function for the $\mathscr{X}'$ sample space be given by

$$f'_\theta(x') = \begin{cases} f_\theta(x') & \theta \in \Theta_x,\ x' \neq a, \\ 0 & \theta \notin \Theta_x,\ x' = a, \\ 0 & \theta \in \Theta_x,\ x' \neq a, \\ 1 & \theta \notin \Theta_x,\ x' = a. \end{cases}$$

We let $\mathscr{Y}' = \mathscr{Y} \cup \{b\}$ and in a similar way define $p'_\theta(y')$ for $\theta \in \Theta$. Since the loss function is defined on $\Theta \times D$ we see that the assumptions of Theorem 1 are satisfied with the augmented sample spaces $\mathscr{X}'$ and $\mathscr{Y}'$.

Theorem 1 can be generalized to the situation where the statistician can choose his experiment from a set of k possible experiments where $k \geq 2$. Let X_i denote the random variable for the ith experiment. Let $\gamma = \{\gamma_1, \cdots, \gamma_k\}$ where $\gamma_i (\geq 0)$ denotes the probability that the statistician selects the ith experiment. Note that $\sum_1^k \gamma_i = 1$. We shall call γ a design. If δ_i denotes a decision function defined on the sample space of X_i then $(\gamma, \boldsymbol{\delta})$ denotes a typical decision strategy for the statistician where $\boldsymbol{\delta} = (\delta_1, \cdots, \delta_k)$. Theorem 1 can be generalized in the obvious way to characterize admissible decision strategies $(\gamma, \boldsymbol{\delta})$. For example, suppose $\lambda^1, \cdots, \lambda^n$ is a set of mutually orthogonal prior distributions with δ_i^* the unique stepwise Bayes decision rule for the ith problem. We assume that it is not the case that for some $i \neq j$ that $r_{x_i}(\theta, \delta_i^*)$ dominates $r_{x_j}(\theta, \delta_j^*)$. If this were so we would just remove the "inadmissible" experiments from consideration. This guarantees that the first assumption of (a.II) is satisfied in the general case. The set of designs γ^* such that $(\gamma^*, \boldsymbol{\delta}^*)$ is admissible can be found as follows. Let

$$\Phi_1(\lambda^1, \cdots, \lambda^n) = \{i: R(\delta_i^*; \lambda^1) = \min_{r=1,\cdots,n} R(\delta_r^*; \lambda^1)\}$$

and define $\Phi_2(\lambda^1, \cdots, \lambda^n), \cdots, \Phi_n(\lambda^1, \cdots, \lambda^n)$ inductively as follows. If $\Phi_{\ell-1}(\lambda^1, \cdots, \lambda^n)$ contains just one integer let $\Phi_\ell(\lambda^1, \cdots, \lambda^n), \cdots, \Phi_n(\lambda^1, \cdots, \lambda^n)$ all be empty and if $\Phi_{\ell-1}(\lambda^1, \cdots, \lambda^n)$ contains more than one integer let

$$\Phi_\ell(\lambda^1, \cdots, \lambda^n) = \{i: R(\delta_i^*; \lambda^\ell) = \inf_{r \in \Phi_{\ell-1}} R(\delta_r^*; \lambda^\ell)\}.$$

If there exists an ℓ' such that $\Phi_{\ell'}(\lambda^1, \cdots, \lambda^n)$ contains just one integer, say i_0, then the only design δ^* such that $(\gamma^*, \boldsymbol{\delta}^*)$ is admissible is the design that puts probability one on i_0. The only other possibility is that $\Phi_n(\lambda^1, \cdots, \lambda^n)$ contains several integers. In this case $(\gamma^*, \boldsymbol{\delta}^*)$ is admissible if and only if γ^* assigns probability one to the set $\Phi_n(\lambda^1, \cdots, \lambda^n)$. In either case let $\Phi(\lambda^1, \cdots, \lambda^n)$ be the class of designs such that (γ^*, δ^*) is admissible if and only if $\gamma^* \in \Phi(\lambda^1, \cdots, \lambda^n)$.

Note that in the case where $\cup_{i=1}^n \Theta(\lambda_i) = \Theta$, an "inadmissible" experiment will be automatically eliminated from consideration since it cannot belong to $\Phi_n(\lambda^1, \cdots, \lambda^n)$. In particular, in part (a.II) of Theorem 1, the assumption that neither $r_x(\theta, \delta)$ nor $r_y(\theta, \phi)$ dominates the other is superfluous when $\cup_{i=1}^n \Theta(\lambda_i) = \Theta$ by (2.6). This alternative version of (a.II) will be used in the proof of Theorem 2 and in the next section as well.

As a final generalization of Theorem 1, we note that part (a.II) of Theorem 1 can be extended to the case where Θ is no longer assumed to be finite. This is possible because some admissibility questions can be reduced to considering only finite subsets of the parameter space by using the following principle. Suppose Θ denotes the parameter space for a general decision problem with a, denoting a typical decision function with risk function $r(\theta; a)$. Suppose a^* is a decision function with the property that for every $\theta_0 \in \Theta$ there is a subset $\Psi(\theta_0)$ with $\theta_0 \in \Psi(\theta_0) \subset \Theta$ such that

$$\text{(2.8)} \qquad \begin{aligned} &\text{if } r(\theta, a') \le r(\theta, a^*) \quad \text{for all} \quad \theta \in \Psi(\theta_0) \\ &\text{then} \quad a' = a^* \text{ a.e. } (P_\theta) \quad \text{for all} \quad \theta \in \Psi(\theta_0). \end{aligned}$$

Not only does this mean that a^* is admissible for the problem having parameter space $\Psi(\theta_0)$, but it is admissible for the problem with the whole parameter space Θ. We now use this principle to prove Theorem 2.

THEOREM 2. *Let $\mathscr{X}$, $\mathscr{Y}$, and Θ be arbitrary. Assume for each $\theta \in \Theta$, $f_\theta(x) > 0$ for only finitely many $x \in \mathscr{X}$ and $p_\theta(y) > 0$ for only finitely many $y \in \mathscr{Y}$. Let (γ, δ, ϕ) be a triple with the property that for every $\theta_0 \in \Theta$ there exists a finite family of mutually orthogonal probability measures $\lambda_{\theta_0}^1, \cdots, \lambda_{\theta_0}^n$ (which may depend on θ_0) with finite support on Θ with the properties* (i) *$\theta_0 \in \Theta_{\theta_0} = \cup_{i=1}^n \Theta(\lambda_{\theta_0}^i)$,* (ii) *$\delta$ and ϕ are decision rules which are stepwise Bayes against $\lambda_{\theta_0}^1, \cdots, \lambda_{\theta_0}^n$ for the X and Y problems respectively when the parameter space is restricted to Θ_{θ_0}.*
If

$$\text{(2.9)} \qquad R_x(\delta; \lambda_{\theta_0}^j) = R_y(\phi; \lambda_{\theta_0}^j) \quad \text{for} \quad j = 1, \cdots, n,$$

then (γ, δ, ϕ) is admissible for any $\gamma \in [0, 1]$.

PROOF. If (γ, δ, ϕ) is not admissible then there exists a triple $(\gamma', \delta', \phi')$ such that

$$\text{(2.10)} \qquad r(\theta; \gamma', \delta', \phi') \le r(\theta; \gamma, \delta, \phi) \quad \text{for all} \quad \theta \in \Theta$$

with strict inequality at some θ, say θ_0. For θ_0 let $\lambda_{\theta_0}^1, \cdots, \lambda_{\theta_0}^n$ and Θ_{θ_0} be as given in the assumptions of the theorem. Just as in the proof of Theorem 1, we have that

$$P_\theta\{\delta(X) = \delta'(X)\} = P_\theta\{\phi(Y) = \phi'(Y)\} = 1 \text{ for } \theta \in \Theta_{\theta_0}.$$

Now, since $\Theta_{\theta_0} \cup_{i=1}^n \Theta(\lambda_{\theta_0}^i)$, from (2.9) and (2.10) it follows that $\gamma = \gamma'$ as well and hence, by the principle given in (2.8), Theorem 2 is proved.

Note that Theorem 2 can be easily extended to the case where we are choosing between n experiments instead of just two. In particular, this extended version will be used in the next section to find uniformly admissible strategies in finite population sampling.

3. Uniform admissibility in finite population sampling. Consider a finite population $\mathscr{U}$ with units labeled 1, 2, $\cdots$, N. Let y_i be the value of a single characteristic attached to the unit i. The vector $y = (y_1, \cdots, y_N)$ is the unknown state of nature and is assumed to belong to $\Theta = \mathscr{R}^N$. A subset of s of $\{1, 2, \cdots, N\}$ is called a sample. Let $n(s)$ denote the number of elements belonging to s. Let S denote the set of all possible samples. A design is a function p defined on S such that $p(s) \in [0, 1]$ for all $s \in S$ and $\sum_{s \in S} p(s) = 1$. Given $y \in \Theta$ and $s = \{i_1, \cdots, i_n\}$ where $1 \le i_1 < i_2 < \cdots < i_n \le N$, let $y(s) = (y_{i_1}, \cdots, y_{i_n})$. Suppose one wishes to estimate $\gamma(y)$, some real valued function of the parameter, with squared error loss. Let $e(s, y)$ denote an estimator $\gamma(y)$ where $e(s, y)$ depends on y only through $y(s)$. If the design p is used in conjunction with the estimator e then their risk function is

$$r(y; p, e) = \sum_s [e(s, y) - \gamma(y)]^2 p(s) = \sum_s r_s(y; e(s, \cdot))p(s). \tag{3.1}$$

An estimator e is said to be admissible for the design p if there does not exist any other estimator e' with $r(y; p, e') \le r(y; p, e)$ for all $y \in \Theta$ with strict inequality for some $y \in \Theta$. The pair (p, e) is said to be uniformly admissible relative to W a class of possible designs, if $p \in W$ and there does not exist any other pair (p', e'), with $p' \in W$, satisfying $r(y, p', e') \le r(y, p, e)$ for all $y \in \Theta$ with strict inequality for some $y \in \Theta$.

Now it is easy to see from (3.1) that the problem of choosing a uniformly admissible pair (p, e) is of the type considered in the previous section. Let W_n denote the set of all designs of fixed sample size n. Using a design $p \in W_n$ is equivalent to saying that there are $\binom{N}{n}$ possible experiments that the statistician can observe. That is, he must decide which sample of size n he wants to observe and he can make this choice by choosing at random, in any way that he desires, from among the $\binom{N}{n}$ samples.

Since the parameter space is not finite, Theorem 2 and its extension to choosing among any finite number of experiments will be used. In fact, the following theorem is a restatement of Theorem 2 in the framework of finite population sampling, and its proof will be omitted.

THEOREM 3. *Let $e(s, y)$ be an estimator of $\gamma(y)$ such that for every $y_0 \in \Theta$ there exists a finite family of mutually orthogonal probability measures $\lambda_{y_0}^1 \cdots, \lambda_{y_0}^n$ (which may depend on y_0) with finite support on Θ with the properties* (i) *$y_0 \in \Theta_{y_0} = \cup_{i=1}^n \Theta(\lambda_{y_0}^i)$ and* (ii) *for each sample s, $e(s, \cdot)$ is the unique stepwise Bayes estimator of $\gamma(y)$ against $\lambda_{y_0}^1, \cdots, \lambda_{y_0}^n$ when the parameter space is restricted to Θ_{y_0}. If p is a design belonging to W_n such that $p \in \Phi(\lambda_{y_0}^1, \cdots, \lambda_{y_0}^n)$ when the parameter space is restricted to Θ_{y_0} for every y_0, then $(p, e(s, y))$ is uniformly admissible for the original problem.*

COROLLARY 3.1. *Let $e(s, y)$ be the estimator of $\gamma(y)$ and suppose for every $y_0 \in \Theta$ there exists a prior λ_{y_0} (which may depend on y_0) with finite support $\Theta(\lambda_{y_0})$ such that $y_0 \in \Theta(\lambda_{y_0})$ and when the parameter space is restricted to $\Theta(\lambda_{y_0})$, $e(s, \cdot)$ is the unique Bayes estimator of $\gamma(y)$ for every sample s. If $\sum_{y\in\Theta(\lambda_{y_0})} r(y; p, e)\lambda_{y_0}(\theta)$ as a function of p, is constant over W_n then for any design $p \in W_n$, (p, e) is uniformly admissible relative to W_n.*

PROOF. For every $y_0 \in \Theta$ we have by the assumptions of the corollary and the extension of Theorem 1 that $\Phi(\lambda_{y_0}) = W_n$ and the corollary follows from Theorem 2.

In Godambe (1969) an estimator e^*, of the population total was proposed. Let V_n be the class of designs with average sample size n. It was shown that (p, e^*) is uniformly admissible relative to V_n for any $p \in V_n$. In the course of the proof of his result, it was shown that for every $y_0 \in \Theta$ there exists a λ_{y_0} which satisfies the conditions of the corollary. In addition, he demonstrated that $\sum_{y\in\Theta(\lambda_{y_0})} r(y; p, e^*)\lambda^*(y)$ as a function of p is constant over V_n. Since $W_n \subset V_n$ we have from the corollary that (p, e^*) is uniformly admissible relative to W_n for every $p \in W_n$. Although our result is not as general as Godambe's, it seems to throw further insight into the question of uniform admissibility.

In Ericson (1970) another estimator, say e', of the population total was proposed. Following Godambe it was shown that (p, e') is uniformly admissible relative to V_n for every $p \in V_n$. As before the weaker result with W_n replacing V_n follows from our corollary and some facts demonstrated in Ericson (1970).

Finally, in Chaudhuri (1978), in the spirit of Godambe, an estimator, e'', of the population variance was suggested. It was shown that (p, e'') is uniformly admissible relative to W_n for any $p \in W_n$. The proof can be simplified using our corollary.

For the rest of this section, consider the following estimator of the population total

(3.2) $$e_1(y, s) = \sum_{i \in s} y_i + \{n(s)\}^{-1}\{\sum_{i\in s}(y_i/m_i)\}(\sum_{i \notin s} m_i)$$

proposed by Basu (1971).

We shall now find designs such that if they are used with Basu's estimator, the resulting pairs are uniformly admissible. In this argument, the admissibility of Basu's estimator for any fixed design will be demonstrated.

For a set of r distinct real numbers $\alpha_1, \alpha_2, \cdots, \alpha_r$ with $1 \le r \le N$, let

(3.3) $$\bar{\mathscr{Y}}_m(\alpha_1, \cdots, \alpha_r) = \{y : y_i/m_i = \alpha_j \text{ for some } j = 1, \cdots, r, \text{ for all } i = 1, \cdots, N\},$$

where $m = (m_1, \cdots, m_N)$. Note that $\bar{\mathscr{Y}}_m(\alpha_1, \cdots, \alpha_r)$ is a subset of $\Theta = R^N$ containing finitely many points. Let

(3.4) $$\bar{\bar{\mathscr{Y}}}_m(\alpha_1, \cdots, \alpha_r) = \{y : y_i/m_i = \alpha_j \text{ for some } j = 1, \cdots, r, \text{ for all } i = 1, \cdots, N \text{ and each } \alpha_j \text{ appears at least once for } j = 1, \cdots, r\}.$$

If $y \in \bar{\bar{\mathscr{Y}}}_m(\alpha_1, \cdots, \alpha_r)$ we say that y is of order r for $\alpha_1, \cdots, \alpha_r$. Similarly if $y(s)$ is a sample point with $r \le n(s)$, we say that $y(s)$ is of order r for $\alpha_1, \cdots, \alpha_r$ if each y_i/m_i equals one of the r values $\alpha_1, \cdots, \alpha_r$, and if for each value α_j, there exists at least one i_ℓ for which $y_{i_\ell}/m_{i_\ell} = \alpha_j$. *If* $y \in \bar{\bar{\mathscr{Y}}}_m(\alpha_1, \cdots, \alpha_r)$, let $w_y(j)$ be the number of (y_i/m_i)'s which are equal to α_j. Note for each j, $w_y(j) \ge 1$ and $\sum_{j=1}^r w_y(j) = N$. If $y(s)$ is a sample point of order r for $\alpha_1, \cdots, \alpha_r$ let $w_y(j; s)$ be the number of observed (y_i/m_i)'s $(i \in s)$ which are equal to α_j.

We now exhibit a family of mutually orthogonal prior distributions on $\bar{\mathscr{Y}}(\alpha_1, \cdots, \alpha_r)$ against which Basu's estimator is the unique stepwise Bayes for any design p.

The first prior λ^1 puts mass $1/r$ on the r points $y = (m_1\alpha_j, m_2\alpha_j, \cdots, m_N\alpha_j)$ for $j = 1, \cdots, r$. For such a point all the observed ratios in a sample s are α_j and the Bayes estimator is just Basu's estimate in this case.

The second prior λ^2 is defined over the set $\cup_{\{i<i'\}} \bar{\bar{\mathscr{Y}}}_m(\alpha_i, \alpha_{i'})$. This set contains all parameter vectors of order two for some α_i and $\alpha_{i'}$. If y is of order two for α_i and $\alpha_{i'}$ with $i < i'$ then

$$\lambda^2(y) \propto \int_0^1 p^{w_y(i)-1}(1-p)^{w_y(i')-1}\,dp = \Gamma(w_y(i))\Gamma(w_y(i'))/\Gamma(N).$$

Note that for a sample $y(s)$ of order two for α_i and $\alpha_{i'}$, the marginal probability of $y(s)$ is given by

$$\lambda^2(y(s)) \propto \Gamma(w_y(i; s))\Gamma(w_y(i'; s))/\Gamma(n(s)).$$

Now the sample points which have positive marginal probability under λ^2 but not under λ^1 are just those of order two for some α_i and $\alpha_{i'}$ with $i < i'$. Let $y(s)$ be such a point and suppose $i^* \notin s$. Then

$$E\{(y_{i^*}/m_{i^*}) \mid y(s)\} = \{\alpha_i w_y(i; s) + \alpha_{i'} w_y(i'; s)\}/n(s).$$

It is easy to check that for such a sample s, the Bayes estimate of the population total for λ^2 at $y(s)$ is

$$\alpha_i \sum_{j\in s(i)} m_j + \alpha_{i'} \sum_{j \in s(i')} m_j + (\sum_{i^* \notin s} m_{i^*})\{\alpha_i w_y(i; s) + \alpha_{i'} w_y(i'; s)\}/n(s),$$

where $s(i) = \{j \in s : y_j/m_j = \alpha_i\}$. This is again Basu's estimate.

The third prior is defined over the set $\cup_{\{i<j<k\}} \bar{\bar{\mathscr{Y}}}_m(\alpha_i, \alpha_j, \alpha_k)$, and is given by

$$\lambda^3(y) \propto \int_0^1 \int_0^1 p_1^{w_y(i)-1} p_2^{w_y(j)-1}\,(1-p_1-p_2)^{w_y(k)-1}\,dp_1\,dp_2.$$

The sample points which have positive marginal probability under λ^3 but not under λ^1 and λ^2 are just those which are of order three for some α_i, α_j and α_k. It is easy to show that for such points, the Bayes estimate against λ^3 is Basu's estimate. Continuing in this way, it follows from Theorem 1 of Meeden and Ghosh (1981) that Basu's estimator is admissible when the parameter space is $\mathscr{Y}(\alpha_1, \cdots, \alpha_r)$, and hence is admissible when the parameter space is $\mathscr{R}^N$ as well.

When all the m_i's are equal, Basu's estimator is just the classical estimator and the above argument is an alternative proof of the admissibility of the classical estimator to that given in Joshi (1966). Recently Tsui (1982) has used Joshi's (1966) technique to prove the admissibility of estimators similar to e_1 when the parameter space is $\mathscr{R}^N$.

It is possible to use the above sequence of priors to construct an admissible estimator of the population variance (see Ghosh and Meeden, 1982). In addition, similar arguments can be used to study the admissibility of ratio and Horvitz-Thompson estimators (see Meeden and Ghosh, 1982).

We shall now use the above sequence of priors to study the uniform admissibility of Basu's estimator. Recall that W_n is the class of designs of fixed sample size n. Let S_n denote the set of all samples of size n.

Let s denote a sample of size n. For the design, p, with $p(s) = 1$ and the estimator e let $r_s(y; e(s))$ denote the risk function of e and p given in equation (3.1). In particular, for Basu's estimator, e_1, we have that

$$r_s(y, e_1(s)) = \{n^{-1}(\textstyle\sum_{i\in s} z_i)(\sum_{i\notin s} m_i) - \sum_{i\notin s} z_i m_i\}^2 = (\sum_1^N a_i z_i)^2, \tag{3.5}$$

where $z_i = y_i/m_i$ $(i = 1, \cdots, N)$, $a_i = n^{-1} \sum_{i\notin s} m_i$ for all $i \in s$ and $a_i = -m_i$ for all $i \notin s$. Let

$$\Gamma = \{s : s \in S_n \text{ and } r_s(y, e_1(s)) \text{ is not dominated by } r_{s'}(y, e_1(s')) \text{ for any other } s' \in S_n \text{ for } y \in \mathscr{Y}_m(\alpha_1, \cdots, \alpha_r)\}.$$

Now typically Γ is a proper subset of S_n and its actual composition depends on the vector $m = (m_1, \cdots, m_N)$. The next lemma identifies a subset of Γ which will be useful later.

LEMMA. *Let $m = (m_1, \cdots, m_N)$ be a vector of positive constants which are not all equal. If*

$$\Gamma(\max) = \{s : s \in S_n \text{ and } \textstyle\sum_{i\in s} m_i = \max_{s'\in S_n} \sum_{i\in s'} m_i\}$$

then $\Gamma(\max) \subset \Gamma$.

PROOF. Let $s_2 \in \Gamma(\max)$ and suppose s_1 $(\neq s_2) \in S_n$. To prove the lemma, it is enough to exhibit a parameter point $y^* \in \mathscr{Y}_m(\alpha_1, \cdots, \alpha_r)$ such that

$$r_{s_1}(y^*; e_1(s_1)) > r_{s_2}(y^*; e_1(s_2)). \tag{3.6}$$

It is enough to consider a parameter point which is of order two for some α_i and α_j. For notational convenience let $a = \alpha_i$ and $b = \alpha_j$ and $a_y(\tilde{s}) = \{\ell : \ell \notin s \text{ and } z_\ell = a\}$ where $s \in S_n$. If y is a parameter point of order two for a and b, it is easily seen that

$$r_s(y, e_1(s)) = (a - b)^2\{n^{-1} w_y(a, s) \textstyle\sum_{i\notin s} m_i - \sum_{i\in a_y(\tilde{s})} m_i\}^2. \tag{3.7}$$

Let y^* be the parameter point satisfying $y_i/m_i = a$ when $i \in s_1$ and $y_i/m_i = b$ when $i \notin s_1$. Then considering separately the two cases where $s_1 \cap s_2$ is empty and nonempty, it follows that (3.7) holds.

With this lemma we can identify some uniformly admissible designs for Basu's estimator.

THEOREM 4. *Let $\alpha_1, \cdots, \alpha_r$ be r distinct real numbers and let $m = (m_1, \cdots, m_N)$ be a vector of positive constants which are not all equal. Let p be a design which puts positive mass on samples belonging to $\Gamma(\max)$. Then Basu's estimator, e_1, as given in* (3.2) *and the design p is uniformly admissible relative to W_n when the parameter is* *assumed to be in $\bar{\mathscr{Y}}_m(\alpha_1, \cdots, \alpha_r)$. In addition, by Theorem 3, the pair is uniformly admissible relative to W_n when the parameter space is $\mathscr{R}^N$.*

PROOF. We will use the extension of Theorem 1 to find the designs in W_n which when used with Basu's estimator are uniformly admissible. These designs must concentrate their mass on the sets belonging to Γ. Since for any $s \in \Gamma$, Basu's estimator is stepwise Bayes against the sequence of priors introduced earlier, we need to compute the Bayes risks of Basu's estimator against these priors for the sample s. Under each of these priors, the ratios y_i/m_i are finitely exchangeable. Let Z_i denote the random variable y_i/m_i under a typical prior distribution in this sequence. If λ^ℓ denotes the ℓth prior in this sequence, then the Bayes risk of the estimator e_1 for the sample s against λ^ℓ is given by

$$R_s(e_1(s); \lambda^\ell) = (\textstyle\sum_1^N a_i^2)E(Z_1) + (\sum_{i\neq j} a_i a_j)E(Z_1 Z_2) \tag{3.8}$$

where the expectations are taken with respect to the marginal priors under λ^ℓ. Equation (3.8) follows from equation (3.5) and the fact that $\sum_1^N a_i = 0$. Since $\sum_{i\neq j} a_i a_j = -\sum_1^N a_j^2$ equation (3.8) becomes

$$R_s(e_1(s); \lambda^\ell) = (\textstyle\sum_1^N a_i^2)(EZ_1^2 - EZ_1 Z_2). \tag{3.9}$$

By the Schwarz inequality $E(Z_1Z_2) \leq (EZ_1^2)^{1/2}(EZ_2^2)^{1/2} = EZ_1^2$ and hence the minimization of the Bayes risk in (3.9) amounts to the minimization of $\sum_1^N a_i^2 = n^{-1}(\sum_{i\not\in s} m_i)^2 + \sum_{i\not\in s} m_i^2$. Since the m_i's are positive, by Theorem 1, the only designs which are admissible are those which put all their mass on the set $\Gamma(\max)$. This completes the proof of the theorem.

REFERENCES

BASU, D. (1971). An essay on the logical foundations of survey sampling, part one. *Foundations of Statistical Inference*, edited by V. P. Godambe and D. A. Sprott. Holt, Rinehart and Winston, Toronto.

BLACKWELL, DAVID. (1951). Comparison of experiments: *Proc. Second Berkeley Symposium on Math. Statist. and Probab.* Berkeley, Univ. Calif Press, pages 93–102.

BLACKWELL, DAVID. (1953). Equivalent comparisons of experiments. *Ann. Math. Statist.* **24** 265–272.

BROWN, LAWRENCE D. (1981). A complete class theorem for statistical problems with finite sample spaces. *Ann. Statist.* **9** 1289–1300.

CHAUDHURI, A. (1978). On estimating the variance of a finite population. *Metrika* **25** 65–76.

ERICSON, W. A. (1970). On a class of uniformly admissible estimators of a finite population total. *Ann. Math. Statist.* **41** 1369–1372.

FERGUSON, T. S. (1967). *Mathematical Statistics, A Decision Theoretic Approach.* Academic, New York.

GHOSH, M., and MEEDEN, G. (1982). Estimation of the variance in finite population sampling. *Sankhyā*, to appear.

GODAMBE, V. P. (1969). Admissibility and Bayes estimation in sampling finite populations V. *Ann. Math. Statist.* **40** 672–676.

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

JOSHI, V. M. (1966). Admissibility and Bayes estimation in sampling finite populations IV. *Ann. Math. Statist.* **37** 1658–1670.

MEEDEN, G., and GHOSH, M. (1981). Admissibility in finite problems. *Ann. Statist.* **9** 846–852.

MEEDEN, G., and GHOSH, M. (1982). On the admissibility and uniform admissibility of ratio type estimates. *Proceedings of the Golden Jubilee Conference* of the Indian Statistical Institute, to appear.

TSUI, K. W. (1982). A class of admissible estimators of a finite population total. *Ann. Inst. Statist. Math.*, to appear.

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

DEPARTMENT OF STATISTICS
NUCLEAR SCIENCES CENTER
UNIVERSITY OF FLORIDA
GAINESVILLE, FLORIDA 32611