The Annals of Statistics
1981, Vol. 9, No. 6, 1289–1300

A COMPLETE CLASS THEOREM FOR STATISTICAL PROBLEMS WITH FINITE SAMPLE SPACES[1]

BY LAWRENCE D. BROWN

Cornell University

This paper contains a complete class theorem (Theorem 3.2) which applies to most statistical estimation problems having a finite sample space. This theorem also applies to many other statistical problems with finite sample spaces. The description of this complete class involves a stepwise algorithm. At each step of the process it is necessary to construct the Bayes procedures in a suitably modified version of the original problem. The complete class is a minimal complete class if the loss function is strictly convex.

Some examples are given to illustrate the application of this complete class theorem. Among these is a new result concerning the estimation of the parameters of a multinomial distribution under a normalized quadratic loss function. (See Example 4.5).

1. Notation, The Extended Problem, Examples.

1.1 GENERAL STATISTICAL DECISION PROBLEM. The general specification of a fixed sample statistical problem, $\mathscr{P}$, involves three measurable spaces, a family of distributions, and a loss function. More specifically, there is the *sample space*, $\mathscr{X}$, $\mathscr{B}_{\mathscr{X}}$; the *parameter space* Θ, $\mathscr{B}_\Theta$; and the *action space*, $\mathscr{A}$, $\mathscr{B}_{\mathscr{A}}$; the family of distributions, $\{F_\theta(\cdot): \theta \in \Theta\}$ on $\mathscr{X}$, $\mathscr{B}_{\mathscr{X}}$, which is a ($\mathscr{B}_\Theta$ measurable) Markov kernel of conditional probability measures given θ on $\mathscr{B}_{\mathscr{X}}$; and the loss function $L: \Theta \times \mathscr{A} \times \mathscr{X} \to [0, \infty]$, assumed to be $\mathscr{B}_\Theta \times \mathscr{B}_{\mathscr{A}} \times \mathscr{B}_{\mathscr{X}}$ measurable. (Customarily the loss function is chosen to be independent of the observation $x \in \mathscr{X}$; but the theory to be developed here extends easily to situations where the loss function depends also on $x \in \mathscr{X}$. Often when L is independent of $x \in \mathscr{X}$ we will simply write $L = L(\theta, a)$.)

A decision procedure is a Markov kernel, δ, of conditional probability measures given $x \in \mathscr{X}$ on $\mathscr{B}_{\mathscr{A}}$. Let D denote the set of all procedures. We use the conditional probability notation $\delta(\cdot \mid x)$. Every decision procedure defines a risk function $R(\cdot, \delta)$ by $R(\theta, \delta) = \int\int L(\theta, a, x)\delta(da \mid x) F_\theta(dx)$. This leads to the usual definitions; for example, δ is *admissible* if $(R(\theta, \delta') \le R(\theta, \delta)\ \forall \theta \in \Theta) \Rightarrow (R(\theta, \delta') \equiv R(\theta, \delta))$.

1.2 SPECIAL ASSUMPTIONS; THE SAMPLE SPACE. Assume throughout that $\mathscr{X}$ is a finite set with $\mathscr{B}_{\mathscr{X}}$ the usual discrete σ-field. Except where otherwise noted we let $\mathscr{X} \subseteq \{i, i = 0, \cdots, m\}$, without any further loss of generality. (The exceptions will occur in some of the examples where alternate specifications are more convenient.) In the remainder of this section, except for Example 1.14, we actually take $\mathscr{X} = \{i: i = 0, \cdots, m\}$.

1.3 SPECIAL ASSUMPTIONS; THE PARAMETER SPACE AND FAMILY OF DISTRIBUTIONS. Assume Θ, $\mathscr{B}_\Theta$ is a locally compact second countable space (Polish space) with Borel σ-field. Let Θ^k denote a suitable compactification, having properties set forth in Assumption 1.4, described later, with Borel field $\mathscr{B}_{\Theta^k}$.

Let ν denote counting measure on $\mathscr{X}$, so that $f_\theta(i) = (dF_\theta)/(d\nu)\ (i) = F_\theta(\{i\})$. Assume

Received December, 1979; revised February, 1981.

[1] Work supported in part by N.S.F. research grant MCS 7824175.

AMS 1980 *subject classifications.* Primary, 62C07, 62C15; secondary, 62F10, 62F11, 62C10.

Key words and phrases: Complete class theorem, finite sample space, admissible procedures, Bayes procedure, estimation, binomial distribution, multinomial distribution, strictly convex loss, squared error loss, maximum likelihood estimate.

the functions $f_\theta(i)$ are continuous on Θ for each $i = 0, \cdots, m$. To avoid trivialities, assume $\sup_{\theta\in\Theta} f_\theta(i) > 0$ for all $i = 0, \cdots, m$.

1.4 SPECIAL ASSUMPTIONS; THE ACTION SPACE AND LOSS FUNCTION. Assume $\mathscr{A}$ is compact and metrizable. Assume $L(\cdot, \cdot, \cdot)$ is bounded and $L(\cdot, a, x)$ is continuous on Θ for each $a \in \mathscr{A}$, $x \in \mathscr{X}$. Assume further that $L(\cdot, a, x)$ has a continuous extension to Θ^k for each $a \in \mathscr{A}$, $x \in \mathscr{X}$. This (unique) extension will also be denoted by the symbol $L(\cdot, \cdot, \cdot)$. Finally, assume that $L(\theta, \cdot, x)$ is also lower semicontinuous on $\mathscr{A}$ for each $\theta \in \Theta^k$, $x \in \mathscr{X}$.

1.5 THE EXTENDED PROBLEM. Let $\mathscr{P}$ be a statistical problem as above. Let

$$\mathbb{E} = \{(\theta, \pi): \theta \in \Theta, \quad \pi \in \mathbb{R}^{m+1}, \quad f_\theta(i) = \pi_i, \quad i = 0, \cdots, m\} \subset \Theta^k \times ([0, 1]^{m+1}).$$

Let $\mathscr{P}^*$, the extended problem, have $(\mathscr{X}, \mathscr{B}_{\mathscr{X}})$ and $(\mathscr{A}, \mathscr{B}_{\mathscr{A}})$ as in $\mathscr{P}$ but have parameter space $\Theta^* = \bar{\mathbb{E}}$, where the closure is taken in the product topology on $\Theta^k \times ([0, 1]^{m+1})$, and $\mathscr{B}_{\bar{\mathbb{E}}}$ is the Borel σ-field. Let $f^*_{(\theta,\pi)}(i) = F^*_{(\theta,\pi)}(\{x_i\}) = \pi_i$, $i = 0, \cdots, m$ for $(\theta, \pi) \in \Theta^*$ and $L^*((\theta, \pi), a, x) = L(\theta, a, x)$ for $(\theta, \pi) \in \Theta^*$, $a \in \mathscr{A}$, $x \in \mathscr{X}$. Let R^* denote the risk function in problem $\mathscr{P}^*$.

It is straightforward to check that $\mathscr{P}^*$ is a well defined statistical problem satisfying 1.2–1.4. In particular, if $\theta \in \Theta$ then there is a unique π, say $\pi(\theta)$, such that $(\theta, \pi) \in \Theta^*$ (namely $\pi_i = f_\theta(i)$, $i = 0, \cdots, m$), and $f^*_{(\theta,\pi)}(i) = f_\theta(i)$. However, if $\theta \in \Theta^k - \Theta$ then there may be many values of π such that $(\theta, \pi) \in \Theta^*$.

REMARK. In some settings Θ^* is isomorphic to Θ^k. This happens if and only if for each $\theta_0 \in \Theta^k$ the $\lim_{\theta\to\theta_0} f_\theta(i)$ exists, $i = 0, \cdots, m$. Another equivalent condition is that to each $\theta \in \Theta^k$ there exists a unique π such that $(\theta, \pi) \in \Theta^*$. See Examples 1.11 and 1.12 for some settings where this holds. If this is the case, then it suffices to take Θ^k as the parameter space for Θ^*, and $f_{\theta_0}(\cdot) = \lim_{\theta\to\theta_0} f_\theta(\cdot)$ for $\theta_0 \in \Theta^k$, etc.

The following propositions describe the three important properties of this extension process.

1.6 PROPOSITION. *Let δ by any procedure in problem $\mathscr{P}$. Then δ is also a procedure in problem $\mathscr{P}^*$, and conversely. Furthermore $R^*(\cdot, \delta)$ is a continuous function on Θ^*. Hence, $R^*(\cdot, \delta)$ on Θ^* is the unique continuous extension of $R(\cdot, \delta)$ on $\mathbb{E}$ to the* compactification Θ^* of $\mathbb{E}$.

PROOF. The continuity of $R^*(\cdot, \delta)$ follows from the continuity of the maps $(\theta, \pi) \to f_{(\theta,\pi)}(i)$ and the continuity of $L(\cdot, a, x)$ on Θ^*. The remainder of the proposition is also easy to check. □

REMARK. In view of Proposition 1.6, we will simply write $R(\cdot, \cdot)$ in place of $R^*(\cdot, \cdot)$.

1.7 PROPOSITION. *A procedure δ is admissible in problem $\mathscr{P}$ if and only if it is admissible in problem $\mathscr{P}^*$.*

PROOF. As a consequence of Proposition 1.6, $R(\theta, \delta') \leqq R(\theta, \delta)$ for all $\theta \in \Theta$ if and only if $R((\theta, \pi), \delta') \leqq R((\theta, \pi), \delta)$ for all $(\theta, \pi) \in \Theta^*$; and strict inequality holds for some θ in the first expresssion if and only if it holds for some (θ, π) in the second. The proposition then follows. □

1.8 BAYES PROCEDURES. Let P be a (prior) probability distribution on Θ, $\mathscr{B}_\Theta$. A Bayes procedure for P is a procedure δ such that

$$\int R\,(\theta, \delta)P\,(d\theta) = \inf\left\{\int R(\theta, \delta')P\,(d\theta) : \delta' \in D\right\}.$$

Given P, let

$$S(i) = S_P(i) = \left\{ a \in \mathscr{A}: \int L(\theta, a, i) f_\theta(i) P\,(d\theta) = \inf_{a' \in \mathscr{A}} \int L(\theta, a', i) f_\theta(i) P\,(d\theta) \right\}.$$

It is easily seen that a procedure δ is Bayes for P if and only if

$$\delta(S_P(i) \mid i) = 1, \qquad i = 0, \cdots, m.$$

Note that $S(i) \neq \phi$ because of the lower semicontinuity of $L(\theta, \cdot, x)$ (which implies the lower semicontinuity of $\int L(\theta, \cdot, i) f_\theta(i) P(d\theta)$ and the compactness of $\mathscr{A}$. Hence for every prior P there exists at least one Bayes procedure.

If $S_P(i)$ consists of exactly one point for $i = 0, \cdots, m$ then the Bayes procedure for P is uniquely determined. Furthermore, it must then be admissible since any uniquely determined Bayes procedure is admissible.

A set of conditions which will guarantee that $S_P(i)$ always consists of exactly one point—and hence that the corresponding Bayes procedure is admissible—is that $\mathscr{A}$ be a convex subset of Euclidean space, and that $L(\theta, \cdot, x)$ be strictly convex on $\mathscr{A}$ for every $\theta \in \Theta$, $x \in \mathscr{X}$ and that $\int f_\theta(i) P\,(d\theta) > 0$ for every $i = 0, \cdots, m$. (To show this, apply Jensen's inequality.)

1.9 Proposition. *A procedure δ is admissible in problem $\mathscr{P}^*$ only if there exists some prior distribution, P, on Θ^* such that δ is Bayes with respect to P. (In other words, the Bayes procedures form a complete class in problem $\mathscr{P}^*$.)*

Proof. This is a well-known decision theoretic result. Basically, the proof consists of two major steps. The first is to define $\hat{\Gamma} = \{r: \Theta^* \to [0, \infty]: \exists \delta \in D \text{ with } R(\theta, \delta) \leqq r(\theta) \forall \theta \in \Theta^*\}$ and to show that $\hat{\Gamma}$ is closed (hence compact) in the topology of pointwise convergence. The second major step involves defining $\tilde{\Gamma} = \{r: r \in \hat{\Gamma}, r \text{ continuous and bounded}\}$; checking that $\tilde{\Gamma}$ is closed and convex in the topology of uniform convergence of continuous functions of the compact set, Θ^*; and invoking the appropriate supporting hyperplane theorem. The complete result is proved in Wald (1950) for the case where L is independent of $x \in \mathscr{X}$. A more contemporary proof of the first major step is contained in LeCam (1955), again for the case where $L = L(\theta, a)$. These proofs can be adapted to the case where L depends on $x \in \mathscr{X}$; alternately one may invoke the more general closure theorem in Brown (1977) to complete the proof for the case where L depends on $x \in \mathscr{X}$. □

1.10 Example. Even in many simple and statistically natural settings, the converse of Proposition 1.9 is not valid. That is, there may exist Bayes procedures which are not admissible. Consider the well studied problem of estimating the mean of a binomial distribution with squared error loss. Here $\Theta = [0, m] = \mathscr{A}$ $f_\theta(i) = \binom{m}{i} (\theta/m)^i (1 - \theta/m)^{m-i}$, and $L(\theta, a) = (\theta - a)^2$. Consider, for example, the prior P which gives mass one to the point $\theta = 0$. Then $S_P(0) = 0$ but $S_P(i) = [0, m]$ for $i \geqq 1$ since $\int f_\theta(i) P\,(d\theta) \equiv 0$ for $i \geqq 1$. Hence any procedure which makes the estimate 0 when $i = 0$ is a Bayes procedure. Many such procedures are far from being admissible. For this reason the complete class characterization of Proposition 1.9 is not satisfactory even in this familiar problem. Johnson (1971) describes the minimal complete class for this problem. (Johnson's complete class is described in Example 4.1.) Our Theorem 3.2 considerably generalized Johnson's results.

The following examples may help clarify the concepts discussed so far. The first two are successive generalizations of the preceding binomial example.

1.11 Example; one parameter exponential family. Suppose

$$f_\theta(i) = h(i) \exp Q(\theta) T(i) - R(\theta)$$

where—mainly for convenience—Q and T are increasing real valued functions, Q is continuous and $h > 0$. The "natural parametrization" has $Q(\theta) \equiv \theta$, and $\Theta = (-\infty, \infty)$ is the "natural parameter space." However this is sometimes inconvenient in applications, for then $\Theta^k = [-\infty, \infty]$. A more convenient choice is usually $\Theta = (T(0), T(m))$ with

$$\theta = E_\theta(T(I)) = \frac{\partial R(\theta)}{\partial Q(\theta)} = \frac{R'(\theta)}{Q'(\theta)}. \tag{1.1}$$

In such a situation, with the parametrization (1.1), the remark in Definition 1.5 is operative. Thus to define problem $\mathscr{P}^*$, it suffices to let $\Theta^* = \Theta^k = [T(0), T(m)]$ and

$$f_{T(0)}(i) = 1 \quad \text{if and only if } i = 0,$$

and

$$f_{T(m)}(i) = 1 \quad \text{if and only if } i = m.$$

1.12 EXAMPLE MONOTONE LIKELIHOOD RATIO. Suppose $f_\theta(i+1)/f_\theta(i)$ is a nondecreasing extended real-valued function of θ, $i = 0, \cdots, m$. (Thus the problem has the MLR property.) Assumption 1.3 implies that the ratios $f_\theta(i+1)/f_\theta(i)$ are continuous in θ. A convenient parametrization is usually to have

$$\theta = E_\theta(i). \tag{1.2}$$

For convenience, assume (1.2) is the case and $\Theta = (a, b) \subset [0, m]$. Again the remark in Definition 1.5 is operative and it suffices to let $\Theta^* = \Theta^k = [a, b]$ with

$$f_a(i) = \lim_{\theta \downarrow a} f_\theta(i), \qquad f_b(i) = \lim_{\theta \nearrow b} f_\theta(i) \qquad i = 0, \cdots, m.$$

Note that here, as distinguished from the preceding example, it need not be the case that F_a and F_b are trivial distributions, concentrated on one point. However, it will be the case that $F_a(F_b)$ will have as support a set of the form $0 \leqq i \leqq m_a (m_b \leqq i \leqq m)$.

Note that in the preceding examples we have—for convenience of exposition—chosen Θ to be an interval. However, the theory to be developed applies equally well if Θ is actually some (measurable) subset of the intervals defined above. For instance, in Example 1.10, one might be given *a priori* that $\theta \in (0, 1/4) \cup (3/4, 1)$. This set could be chosen as Θ, then $\Theta^k = [0, 1/4] \cup [3/4, 1]$, etc.

1.13 REMARK; LOSS FUNCTIONS. The considerations of this paper are primarily appropriate to estimation problems. Thus, in particular, when $\Theta = \mathscr{A} \subset \mathbb{R}$, loss functions of the type $L(\theta, a) = \ell(\theta - a)$ with ℓ bowl-shaped and $\ell(0) = 0$ are suitable choices, though of course the theory is amenable to many other loss functions.

For some aspects of the theory it is required that $S_P(i)$ contain just one point whenever $\int f_\theta(i)P(d\theta) > 0$. This will be the case whenever $\mathscr{A}$ is convex and $L(\theta, \cdot, x)$ is strictly convex for all $\theta \in \Theta$, $x \in \mathscr{X}$. (Actually, in the later applications, this strict convexity is required for all $\theta \in \Theta^k$, $x \in \mathscr{X}$. See paragraphs preceding Theorem 2.4).

The theory to be developed is not suited to certain classical testing problems. In these $\mathscr{A} = \{0, 1\}$; $\Theta = \Theta_0 \cup \Theta_1$ and $L(\theta, a) = 0$ if $\theta \in \Theta_a$, and $= 1$ otherwise. However, if $\bar{\Theta}_0 \cap \bar{\Theta}_1 \neq \phi$ (for example if $\Theta_0 = (0, 1/2]$, $\Theta_1 = (1/2, 1)$) then this loss function does not satisfy the continuity conditions of Assumption 1.4; and the later considerations of this paper do not apply. Nor will it benefit to artificially separate Θ_0 and Θ_1, as is done for a sequential problem in Brown, Cohen, Strawderman (1980); the reason this will not benefit is discussed there. On the other hand, if $\bar{\Theta}_0 \cap \bar{\Theta}_1 = \phi$, then the results of this paper may apply. Similar comments apply to certain classical multiple decision formulations.

We conclude with a simple example further illustrating the construction of the extended problem.

1.14 EXAMPLE. Suppose $\mathscr{X} = \{(1,0) \cup (0, 1) \cup (1, 1)\} \subset \mathbb{R}^2$, $\Theta = (0, 1) \times (0, 1)$, and

$$F_\theta(\{x\}) = f_\theta(x) = (\theta_1 + \theta_2 - \theta_1\theta_2)^{-1} \Pi \theta_i^{x_i} (1 - \theta_i)^{1 - x_i}.$$

(Note that this is a simple two parameter exponential family. The significance of this family is further explained in Example 4.2.) Then

$$\liminf_{\theta\to 0} f_\theta((1, 0)) = 0 < 1 = \limsup_{\theta\to 0} f_\theta((1, 0)).$$

Hence the special remark in Definition 1.5 does not apply; and Θ^* cannot be defined as $\Theta^k = [0, 1] \times [0, 1]$. Instead, if we let π_1 correspond to $f_\theta((0, 1))$, π_2 to $f_\theta((1, 0))$ and π_3 to $f_\theta((1, 1))$, then $\Theta^* \subset \Theta^k \times ([0, 1]^3)$ is the set defined as (1.3)

$$\begin{aligned}\Theta^* = \{(\theta, \pi)&: \theta \in \Theta^k, \theta \neq (0, 0),\\ \pi_1 &= (\theta_1 + \theta_2 - \theta_1\theta_2)^{-1}(1 - \theta_1)\theta_2,\\ \pi_2 &= (\theta_1 + \theta_2 - \theta_1\theta_2)^{-1}\theta_1(1 - \theta_2),\\ \pi_3 &= (\theta_1 + \theta_2 - \theta_1\theta_2)^{-1}\theta_1\theta_2; \text{ or}\\ \theta &= (0, 0), \pi_2 = 1 - \pi_1,\\ 0 &\leqq \pi_1 \leqq 1, \pi_3 = 0\}\end{aligned}$$

and $f_{(\theta,\pi)}((0, 1)) = \pi_1$, $f_{(\theta,\pi)}((1, 0)) = \pi_2$, $f_{(\theta,\pi)}((1, 1)) = \pi_3$.

2. Sequences of Priors.

2.1 Restricted problems. Let $\mathscr{P}$ be a statistical problem as defined in Section 1. Let $\mathscr{X}' \subset \mathscr{X}$. Then the restricted problem, $\mathscr{P}' = \mathscr{P}(\mathscr{X}')$, of $\mathscr{P}$ to $\mathscr{X}'$ is defined to have sample space $\mathscr{X}'$, to have parameter space $\Theta' = \Theta(\mathscr{X}') = \{\theta \in \Theta: \sum_{i\in\mathscr{X}'} f_\theta(i) > 0\}$, to have

$$f'_\theta(i) = f_\theta(i)/\textstyle\sum_{i\in\mathscr{X}'} f_\theta(i) \quad \text{for } \theta \in \Theta', i \in \mathscr{X}', \tag{2.1}$$

and to have $\mathscr{A}$ as in problem $\mathscr{P}$, and loss $L'(\theta, a, x) = L(\theta, a, x)$, $\theta \in \Theta'$, $a \in \mathscr{A}$. Note that (2.1) merely defines the conditional probability function given $i \in \mathscr{X}' \subset \mathscr{X}$. It is easily checked that $\mathscr{P}'$ itself satisfies the assumptions of Section 1. Note that if $\mathscr{X}' \neq \phi$ then $\Theta' \neq \phi$ since we have assumed $\sup_{\theta\in\Theta} f_\theta(i) > 0$, $i = 0, \cdots, m$.

If $\mathscr{P}(\mathscr{X}')$, is the notation for the restriction of $\mathscr{P}$ to $\mathscr{X}'$, then we use the symbol $\mathscr{P}^*(\mathscr{X}')$ to denote the extension of the problem $\mathscr{P}(\mathscr{X}')$. Any procedure δ in problem $\mathscr{P}$ can also be considered as a procedure in problems $\mathscr{P}(\mathscr{X}')$ and $\mathscr{P}^*(\mathscr{X}')$ by taking the restriction of $\delta(\cdot \mid x)$ to the set $\{x \in \mathscr{X}'\}$.

2.2 Adapted and Full sequences of priors. This definition involves a related sequence of priors and sample spaces. Let $\mathscr{X} = \mathscr{X}_1 \supset \mathscr{X}_2 \supset \cdots \supset \mathscr{X}_\ell \neq \phi$, and let P_j be a prior distribution in problem $\mathscr{P}^*(\mathscr{X}_j)$ (i.e., a distribution on $\Theta_j^* = \Theta^*(\mathscr{X}_j)$ such that

$$\mathscr{X}_{j+1} = \left\{x \in \mathscr{X}_j: \int_{\Theta_j^*} f_\phi^{(j)}(x)P_j(d\phi) = 0\right\}, \quad j = 1, \cdots, \ell - 1. \tag{2.2}$$

Here, $f_\phi^{(j)}$ denotes the probability function in problem $\mathscr{P}^*(\mathscr{X}_j)$. Note that $\mathscr{X}_{j+1} \subsetneqq \mathscr{X}_j$. Such a sequence will be called *adapted*. An adapted sequence of length ℓ with

$$\int_{\Theta_\ell^*} f_\phi^{(\ell)}(x)P_\ell(d\phi) > 0 \ \forall x \in \mathscr{X}_\ell$$

will be called a *full* sequence.

2.3 Totally Bayes procedures. Let $\{P_j: j = 1, \cdots, \ell\}$ be a full sequence of priors (related to a sequence $\mathscr{X}_1 \supset \cdots \supset \mathscr{X}_\ell$). Then a procedure δ is called *totally Bayes* relative to $\{P_j\}$ if δ is Bayes relative to P_j in problem $\mathscr{P}^*(\mathscr{X}_j)$ for all $j = 1, \cdots, \ell$. This condition may be rephrased as

$$\delta(S_{P_j}(x)\,|\,x) = 1\ \forall x \in \mathscr{X}_j - \mathscr{X}_{j+1}, \quad j = 1, \cdots, \ell. \tag{2.2}$$

(By convention $\mathscr{X}_{\ell+1} = \phi$.) Let $\hat{S}(x) = S_{P_j}(x)$, $x \in \mathscr{X}_j - \mathscr{X}_{j+1}$, and (2.2) can be rewritten as $\delta(\hat{S}(x)\,|\,x) = 1$.

If $\hat{S}(x)$ consists of just one point for each $x \in \mathscr{X}$ then the totally Bayes procedure is said to be uniquely determined. Note that if $\mathscr{A}$ is a convex subset of Euclidean space and $L(\theta, \cdot, x)$ is strictly convex for each $\theta \in \Theta^k$, $x \in \mathscr{X}$, then any totally Bayes procedure is uniquely determined as noted in Remark 1.13. Such uniquely determined totally Bayes procedures are admissible, as the following theorem shows.

2.4 THEOREM. *If δ is a uniquely determined totally Bayes procedure, then δ is admissible.*

PROOF. Let δ be a uniquely determined totally Bayes procedure and suppose $R(\theta, \hat{\delta}) \leqq R(\theta, \delta)$ for all $\theta \in \Theta$. It will be shown (by induction on $j = 1, \cdots, \ell$) that $\hat{\delta}(\cdot\,|\,x) = \delta(\cdot\,|\,x)$ for $x \in \mathscr{X}_j - \mathscr{X}_{j+1}$.

First, suppose $x \in \mathscr{X}_1 - \mathscr{X}_2$. Then $\hat{\delta}$ is Bayes for P_1 since $R(\theta, \hat{\delta}) \leqq R(\theta, \delta)$ implies $\int R^{(1)}(\varphi, \hat{\delta})P_1(d\varphi) \leqq \int R^{(1)}(\varphi, \delta)P_1(d\varphi)$ by Proposition 1.6 and the remark following it, where $R^{(i)}$ denotes the risk function in Problem $\mathscr{P}^*(\mathscr{X}_i)$. It follows from the unicity assumption that $\hat{\delta}(\cdot\,|\,x) = \delta(\cdot\,|\,x)$ for $x \in \mathscr{X}_1 - \mathscr{X}_2$.

Now, suppose that $x \in \mathscr{X}_J - \mathscr{X}_{J+1}$ for some $J = 2, \cdots, \ell$; and that the induction hypothesis is true for all $j \leqq J - 1$. Because of the induction hypothesis, for $\theta \in \Theta$,

$$\begin{aligned} 0 \leqq R(\theta, \delta) - R(\theta, \hat{\delta}) &= \textstyle\sum_{x \in \mathscr{X}_J} \int_{\mathscr{A}} L(\theta, a, x)(\delta(da\,|\,x) - \hat{\delta}(da\,|\,x))f_\theta(x) \\ &= C^{(J)}(\theta) \textstyle\sum_{x \in \mathscr{X}_J} \int_{\mathscr{A}} L(\theta, a, x)(\delta(da\,|\,x) - \hat{\delta}(da\,|\,x))f_\theta^{(J)}(x) \\ &= C^{(J)}(\theta)(R^{(J)}((\theta, \pi_\theta^{(J)}), \delta) - R^{(J)}((\theta, \pi_\theta^{(J)}), \hat{\delta})) \end{aligned}$$

where $C^{(J)}(\theta) = (\sum_{x \in \mathscr{X}_J} f_\theta^{(J)}(x))^{-1}$ (and $\infty \cdot 0 = 0$), and $\pi_\theta^{(J)}$ is the point with coordinates $\{f_\theta^{(J)}(x): x \in \mathscr{X}_J\}$. It follows from Proposition 1.6 that $R^{(J)}((\theta, \pi), \hat{\delta}) \leqq R^{(J)}((\theta, \pi), \delta)$ for all $(\theta, \pi) \in \mathbb{E}$. Hence $\hat{\delta}$ is Bayes for prior P_j in problem $\mathscr{P}^*(\mathscr{X}_j)$. It follows from the unicity assumption that $\hat{\delta}(\cdot\,|\,x) = \delta(\cdot, x)$ for $x \in \mathscr{X}_j - \mathscr{X}_{j+1}$. The theorem now follows by induction. □

3. Complete Class Theorems. The first basic complete class theorem concerns the collection of all restrictions of $\mathscr{P}$.

3.1 THEOREM. *Let δ be an admissible procedure in problem $\mathscr{P}$. Let $\mathscr{X}' \subset \mathscr{X}$. Then δ is admissible in problem $\mathscr{P}(\mathscr{X}')$ and δ is Bayes in problem $\mathscr{P}^*(\mathscr{X}')$.*

Consequently, the collection of all procedures which are Bayes for every problem $\mathscr{P}^*(\mathscr{X}')$ $\mathscr{X}' \subset \mathscr{X}$, forms a complete class. (Let $\mathscr{C}_1$ denote this complete class.)

PROOF. Let δ be a procedure in problem $\mathscr{P}$ and let $\mathscr{X}' \subset \mathscr{X}$. Suppose δ is not admissible in problem $\mathscr{P}(\mathscr{X}')$. Then there is procedure δ' in $\mathscr{P}(\mathscr{X}')$ such that $R'(\theta, \delta') \leqq R'(\theta, \delta)$ for all $\theta \in \Theta(\mathscr{X}')$ with strict inequality for some $\theta' \in \Theta\,(\mathscr{X}')$. Let $\hat{\delta}$ be the procedure defined by

$$\hat{\delta}(\cdot\,|\,x) = \begin{cases} \delta'(\cdot\,|\,x) & x \in \mathscr{X}' \\ \delta(\cdot\,|\,x) & x \notin \mathscr{X}'. \end{cases}$$

Then
$$R(\theta, \hat{\delta}) = (\textstyle\sum_{x\in\mathscr{X}'} f_\theta(x))R'(\theta, \delta') + \sum_{x\notin\mathscr{X}'} \displaystyle\int L(\theta, a, x)\delta(x, da)f_\theta(x)$$
$$\leqq (\textstyle\sum_{x\in\mathscr{X}'} f_\theta(x))R'(\theta, \delta) + \sum_{x\in\mathscr{X}'} \displaystyle\int L(\theta, a, x)\delta(da \mid x)f_\theta(x)$$
$$= R(\theta, \delta)$$

with strict inequality for some $\theta' \in \Theta(\mathscr{X}) \subset \Theta$. It follows that δ is not admissible. This proves—by contraposition—the first assertion of the theorem. The remaining assertions follow from Propositions 1.7 and 1.9. □

The preceding theorem provides a reasonably economical complete class theorem since, as will be seen, $\mathscr{C}_1$ is often a minimal complete class. However, the description provided by the theorem is unnecessarily inefficient. In order to verify that $\delta \in \mathscr{C}_1$ it is necessary to check that δ is Bayes for every problem $\mathscr{P}^*(\mathscr{X}')$. In general there are a relatively large number $(2^{m+1} - 1)$ of such problems.

The following complete class theorem provides a description of $\mathscr{C}_1$ which can be verified much more efficiently.

3.2 THEOREM. *A procedure is in $\mathscr{C}_1$ if and only if it is totally Bayes. Consequently, the set of totally Bayes procedures is a complete class.*

If every totally Bayes procedure is uniquely determined, then this class is a minimal complete class.

PROOF. Suppose $\delta \in \mathscr{C}_1$. Then δ is Bayes for some prior, P_1, in problem $\mathscr{P}^*(\mathscr{X}_1)$. Let $\mathscr{X}_2 = \{x \in \mathscr{X}_1 : \int_{\Theta^*(\mathscr{X}_1)} f_\theta^{(1)}(x)P_1(d\theta) = 0\}$. Then δ is also Bayes for some prior P_2 in problem $\mathscr{P}^*(\mathscr{X}_2)$, etc. In this fashion one inductively constructs a full sequence $P_1, P_2, \cdots, P_\ell$ such that δ is totally Bayes with respect to this sequence. Thus the collection of totally Bayes procedures contains the complete class of Theorem 3.1.

Conversely, suppose δ is totally Bayes relative to a full sequence $\{P_j\}$. Let $\mathscr{X}' \subset \mathscr{X}$. Let

$$j' = \min\left\{j : \textstyle\sum_{x\in\mathscr{X}'} \displaystyle\int_{\Theta_j^*} f_\phi^{(j)}(x)P_j(d\phi) > 0\right\}. \tag{3.1}$$

Note that $P_{j'}$ is a prior distribution on the parameter space, $\Theta_{j'}^*$, of problem $\mathscr{P}^*(\mathscr{X}_{j'})$. Let Θ'^* denote the parameter space of problem $\mathscr{P}^*(\mathscr{X}')$. Observe that $\mathscr{X}_{j'} \supset \mathscr{X}'$ because of (3.1).

Let $U = \{(\theta, \pi) \in \Theta_{j'}^* : \sum_{i\in\mathscr{X}'} \pi_i > 0\}$. Define the map $m: U \to \Theta'^*$ by $m(\theta, \pi) = (\theta, \pi^*)$ where $\pi_i^* = \pi_i \chi_{\mathscr{X}'}(i)/(\sum_{j\in\mathscr{X}'} \pi_j)$. Define the prior P' in problem $\mathscr{P}^*(\mathscr{X}')$ as follows: Let Q' be the measure defined on $\Theta_{j'}^*$ by $Q'(d(\theta, \pi)) = (\sum_{j\in\mathscr{X}'} \pi_j)P_{j'}(d(\theta, \pi))$. (Note that Q' is concentrated on U and $Q' \neq 0$.). Let $P'(A) = kQ'(m^{-1}(A))$ where k is the appropriate normalizing constant.

Now, the posterior distribution of θ in problem $\mathscr{P}^*(\mathscr{X}_{j'})$ under $P_{j'}$ given $x \in \mathscr{X}'$ is the same as that in problem $\mathscr{P}^*(\mathscr{X}')$ under P' given $x \in \mathscr{X}'$ since

$$\int_{\theta\in A}\int_\pi \pi_i P_{j'}(d(\theta, \pi)) = \int_{\theta\in A}\int_\pi (\pi_i/(\textstyle\sum_{j\in\mathscr{X}'} \pi_j))(\sum_{j\in\mathscr{X}'} \pi_j)P_{j'}(d(\theta, \pi))$$
$$= \int_{\theta\in A}\int_\pi \pi_i^* Q'(d(\theta, \pi))$$
$$= k^{-1}\int_{\theta\in A}\int_{\pi^*} \pi_i^* P'(d(\theta, \pi^*))$$

for all $i \in \mathscr{X}'$. This completes the proof of the first assertion of the theorem.

The second assertion then follows from Theorem 2.4. □

3.3 REMARK; OTHER STEPWISE BAYES ALGORITHMS. The preceding complete class theorem involves the construction of admissible procedures via a stepwise algorithm. At each step a Bayes procedure is constructed for a suitably defined problem. As previously noted, such an algorithm was described in Johnson (1971) for some important special problems. This algorithm is generalized in Example 4.1 in the next section.

Stepwise Bayes procedures have also been utilized by Wald and Wolfowitz (1951) and by Hsuan (1979) in order to describe the (minimal) complete class when the parameter space is finite. Our results specialize to theirs when both parameter and sample space are finite. Otherwise, of course, our results do not overlap theirs.

Our stepwise algorithm (= "totally Bayes procedures") is somewhat more complex than that in the Wald-Wolfowitz-Hsuan construction. In their case the sample space and parameter space at each stage are always a suitable restriction of those in the original problem, and the family of distributions consists of the (well defined) conditional distributions given this restriction. In our algorithm the sample space at a given stage is also a restriction to the original sample space, but the parameter space may involve an extension as well as a restriction of the original parameter space, and the family of distributions may also have to be suitably extended.

4. Examples.

4.1 ONE PARAMETER EXPONENTIAL FAMILY. Consider the setting of Example 1.11 with $\Theta = (T(0), T(m))$ as in (1.1). Suppose that $L(\theta, \cdot, x)$ assumes its minimum on the set $M(\theta)$ for $\theta = T(0)$ and for $\theta = T(m)$. Then Theorem 3.2 and the remarks in Example 1.11 can be applied to yield that *the procedures of the following form comprise a complete class*:

There are constants $-1 \leqq \underline{m} < \bar{m} \leqq m + 1$ and a non-negative measure V (not necessarily finite) on $\Theta^k = [T(0), T(m)]$ with $\sum_{\underline{m}<i<\bar{m}} \int_{\Theta^k} f_\theta(i) V(d\theta) < \infty$ such that

$$(4.1) \qquad \begin{aligned} \delta(M(T(0)), i) &= 1 \qquad 0 \leqq i \leqq \underline{m} \\ \delta(M(T(m)), i) &= 1 \qquad \bar{m} \leqq i \leqq m \\ \delta(S(i), i) &= 1 \qquad \underline{m} < i < \bar{m} \end{aligned}$$

where

$$S(i) = \left\{ a: \int_{\Theta^k} L(\theta, a, i) f_\theta(i) V(d\theta) = \inf \int_{\Theta^k}^{\Theta} L(\theta, \alpha, i) f_\theta(i) V(d\theta) : \alpha \in \mathscr{A} \right\}.$$

(In the notation of preceding sections $S(i) = S_P(i)$ where $P(d\theta) = (\sum_{\underline{m}<i<\bar{m}} f_\theta(i))^{-1} V(d\theta)$.)

Suppose $\mathscr{A}$ is a convex subset of Euclidean space and $L(\theta, \cdot, x)$ is strictly convex. Then the preceding complete class is a minimal complete class. Johnson (1971) first described this minimal complete class in the case of the Binomial distribution with squared error loss, $L(\theta, a) = (\theta - a)^2$. In that case, of course, $M(T(0)) = \{0\}$ and $M(T(m)) = \{m\}$ in (4.1).

If F_θ is not an exponential family, but instead has monotone likelihood ratio, then Theorem 3.2 leads to a characterization very similar to (4.1), but not necessarily identical with it. (In particular, not all the procedures of (4.1) need be in the complete class.) The remarks in Example 1.12 should make it clear how to derive this class.

To see that $\mathscr{C}_1$ need not always be a minimal complete class, consider the binomial problem with $m = 2$ and $L(\theta, a) = |\theta - a|$. Consider the simple Bayes procedure for prior P with $P(\{1/4\}) = P(\{3/4\}) = 1/2$. Then $S(1) = [1/2, 3/2]$ and δ_1 is a Bayes procedure where $\delta_1(\{1/2\} \mid 0) = 1 = \delta_1(\{3/2\} \mid 2)$ and $\delta_1(\{1/2\} \mid 1) = 1/2 = \delta_1(\{3/2\} \mid 1)$. Hence $\delta_1 \in \mathscr{C}_1$. However δ_1 is not admissible since $R(\theta, \delta_2) \leqq R(\theta, \delta_1)$ with inequality for $\theta \in (\frac{1}{2}, \frac{3}{2})$ where $\delta_2(\{(i+1)/2\} \mid i) = 1$. The reader may have already noted that in this example certain randomized procedures may be Bayes or totally Bayes but can never be admissible. However, of course, there do exist estimation problems in which randomized estimators are admissible. Consider the preceding binomial problem, but with $L(\theta, a) = |\theta - a|^{1/2}$. Then the procedure δ_1 is Bayes (for prior P_1) and is also admissible—in fact no other procedure is even equally good; that is, $R(\theta, \delta) \leqq R(\theta, \delta_1)\ \forall\ \theta \in \Theta$ implies $\delta = \delta_1$.

4.2 Two independent Bernoulli variables—a two parameter exponential family. The description of $\mathscr{C}_1$ in Example 4.1 is greatly simplified by the fact that $\Theta^* = \Theta^k$ in every possible subproblem $\mathscr{P}(\mathscr{X}_j)$. Even in the simplest multiparameter exponential families this will not generally be true. Consider the case of two independent Bernoulli variables:

$$f_{(\theta_1,\theta_2)}(x_1, x_2) = \prod_{i=1}^{2} \theta_i^{x_i}(1-\theta_i)^{1-x_i};\ \theta_i \in (0, 1),\ x_i = 0, 1,\ i = 1, 2. \tag{4.2}$$

To be specific, take $\mathscr{A} = \Theta^k = [0, 1]^2$ and $L(\theta, a) = \|\theta - a\|^2$. If $P_1\{(0, 0)\} = 1$ then $\mathscr{X}_2 = \{(1, 0), (0, 1), (1, 1)\} \subset \mathscr{X}$. Problem $\mathscr{P}^*(\mathscr{X}_2)$ is described in Example 1.14. Note especially that $\Theta^*(\mathscr{X}_2)$ is described in (1.3) and $\Theta^*(\mathscr{X}_2) \neq \Theta^k$.

4.3 Example 4.2 continued; ordered parameter spaces. When $\Theta^k \neq \Theta^*$, as in the preceding example, there can be some rather unusual admissible procedures. Let f_θ be as in (4.2) but let

$$\Theta = \{(\theta_1, \theta_2): 0 \leqq \theta_1 \leqq \theta_2 \leqq 1\}.$$

This is an example of an ordered parameter space. Consider the non-randomized procedure δ_3 with

$$\delta_3((0, 0) \mid (0, 0)) = 1,$$

$$\delta_3((\tfrac{1}{4}, \tfrac{1}{4}) \mid (0, 1)) = 1,$$

$$\delta_3((\tfrac{3}{4}, \tfrac{3}{4}) \mid (1, 0)) = 1,$$

$$\delta_3((\tfrac{3}{4}, \tfrac{3}{4}) \mid (1, 1)) = 1.$$

It can be shown that δ_3 cannot be unique Bayes. Rather surprisingly, it is unique totally Bayes, and, hence, admissible. The requisite full sequence of priors is P_1, P_2 with $P_1(\{(0, 0)\}) = 1$ and

$$P_2(\{(\tfrac{3}{4}, \tfrac{3}{4}) \times (\tfrac{1}{5}, \tfrac{1}{5}, \tfrac{3}{5})\}) = \tfrac{5}{7},$$

$$P_2(\{(0, 0) \times (1, 0, 0)\}) = \tfrac{2}{7}.$$

(Note that δ_3 has another somewhat peculiar property—that $\delta_3(\cdot \mid (1, 0)) = \delta_3(\cdot \mid (1, 1))$ but $\delta_3(\cdot \mid (0, 0) \neq \delta_3(\cdot \mid (0, 1))$. The fact that $\delta_3(\cdot \mid (1, 0)) = \delta_3(\cdot \mid (1, 1))$ is, however, not necessary in order for a procedure similar to δ_3 to be totally Bayes.)

The problem of estimating ordered binomial parameters (under squared error loss) was considered in Sackrowitz and Strawderman (1974). They showed that the maximum likelihood estimator for such a problem is inadmissible, except for a few special cases where it is admissible. They used essentially Theorem 3.1 to prove inadmissibility; and Theorem 3.2 (with $\ell = 2$ and $\Theta^k = \Theta^*$) to prove admissibility in the exceptional cases.

4.4 Multinomial distribution, squared error loss. Another important multiparameter exponential family is the multinomial distribution, with

$$f_\theta(x) = \binom{N}{x} \prod_{i=1}^{r} \theta_i^{x_i},$$

$$\mathscr{X} = \{(x_1, \cdots, x_r): x_i \text{ non negative integer}, i = 1, \cdots, r; \textstyle\sum_{i=1}^{r} x_i = N\}$$

$$\Theta = \{(\theta_1, \cdots, \theta_r): 0 < \theta_i, i = 1, \cdots, r; \textstyle\sum\theta_i = 1\}.$$

Suppose it is desired to estimate $\theta \in \Theta$ by $a \in \mathscr{A} = \Theta^k$ with $L(\theta, a) = \| \theta - a \|^2 = \sum_{i=1}^r (\theta_i - a_i)^2$. The maximum likelihood estimator, δ_4, has $\delta_4(\{x/N\} \mid x) = 1$. *This estimator is unique totally Bayes, and hence admissible.* The requisite full sequence of priors is as follows: In each case $\Theta^* = \Theta^k$. Each P_i, $i = 1, \cdots, \ell = r$ is symmetric with respect to permutations of the coordinates. The spaces $\mathscr{X}_j$ are described as $\mathscr{X}_j = \{x \in \mathscr{X}$: at least j coordinates of x are nonzero$\}$. $P_j(\Theta_j) = 1$ where $\Theta_j = \{\theta \in \Theta^k$: exactly j of the coordinates $\theta_1, \cdots, \theta_r$ are nonzero$\}$. P_1 gives mass $1/r$ to each of the r unit vectors. The density of P_j with respect to Lebesgue measure over Θ_j is p_j where

$$p_j((\theta_1, \cdots, \theta_j, 0, \cdots, 0)) = K\, C(\theta)\, (\textstyle\prod_{i=1}^j \theta_i)^{-1} \text{ with } \quad C(\theta) = \sum_{x \in \mathscr{X}_j} f_\theta(x). \tag{4.3}$$

(The term $C(\theta)$ is required since $f_\theta^{(j)}(x) = C^{-1}(\theta) f_\theta(x)$ for $x \in \mathscr{X}_j$.) This admissibility result was stated in Johnson (1971). A complete, and short, proof is given in Alam (1979). The proof we have given above can be easily generalized to treat situations involving other suitable loss functions, or to produce different admissible estimators by altering the sequence described in (4.3). (A rather similar argument can be used to show that the maximum likelihood estimate of several independent binomial means is also admissible under squared error loss when the parameter space is the natural parameter space.) It is pertinent to the next example to note that the above results are equally valid if $\mathscr{A} = [0, 1]^r$.

4.5 MULTINOMIAL DISTRIBUTION, A NORMALIZED QUADRATIC LOSS. Let $\{f_\theta\}$ be as in (4.4); but consider the loss function

$$L_1(\theta, a) = \textstyle\sum_{i=1}^r (\theta_i - a_i)^2/\theta_i.$$

This loss function is discussed and motivated in Olkin and Sobel (1979). The maximum likelihood estimator, δ_4, of the previous section has a constant risk function.

Now, suppose $\mathscr{A} = [0, 1]^r = \{(a_1, \cdots, a_r): 0 \leqq a_i \leqq 1\}$. To avoid trivialities, assume $N \geqq r + 1$. Let $T = \{a = (a_1, \cdots, a_r) \in \mathscr{A}: \min\{a_i\} > 0, \sum_{i=1}^r a_i \geqq 1\}$. *Then, no estimator having* $\sum_{x \in \mathscr{X}} \delta(T - \{a\} \mid x) > 0$ *for every* $a \in \mathscr{A}$ *can be admissible. In particular,* δ_4 *is inadmissible.* This last assertion was proved by Alam (1978) who exhibited an estimator with smaller risk than δ_4.

PROOF. The loss function L_1 does not satisfy Assumption 1.4 since $L_1(\cdot, \cdot)$ is not bounded. Before the theory of Section 3 can be applied, it is necessary to reformulate the problem. To begin this reformulation, observe that admissibility under loss L_1 is equivalent to admissibility under loss L_2 where

$$L_2(\theta, a) = \{\textstyle\sum_{i=1}^r (\theta_i - a_i)^2/\theta_i\}/(\sum_{i=1}^r \theta_i^{-1}).$$

The loss $L_2(\cdot, \cdot)$ is bounded, but still does not clearly satisfy Assumption 1.4 since it cannot be continuously extended to the obvious compactification of Θ, which is

$$\Theta^k = \{\theta: 0 \leqq \theta_i, i = 1, \cdots, r; \Sigma\theta_i = 1\}.$$

This difficulty can be rectified by a reparametrization. Let

$$\lambda_i = \lambda_i(\theta) = (1/\theta_i)/(\textstyle\sum_{j=1}^r 1/\theta_j), \; i = 1, \cdots, r,$$

so that

$$\theta_i = \theta_i(\lambda) = (1/\lambda_i)/(\textstyle\sum_{j=1}^r 1/\lambda_j), \; i = 1, \cdots, r.$$

Consider the problem over the parameter space $\Lambda = \{\lambda: \theta(\lambda) \in \Theta\} = \{\lambda: 0 < \lambda_i; i = 1, \cdots, r; \Sigma\lambda_i = 1\}$. The loss function equivalent to L_2 is

$$\begin{aligned} L_3(\lambda, a) &= [\textstyle\sum_{i=1}^r \{\theta_i(\lambda) - a_i\}^2/\theta_i(\lambda)]/\{\sum_{i=1}^r \theta_i^{-1}(\lambda)\} \\ &= (1 - 2\textstyle\sum_{i=1}^r a_i)/\sum_{j=1}^r \lambda_j^{-1} + \sum_{i=1}^r \lambda_i a_i^2. \end{aligned}$$

Now, L_3 is bounded and has an obvious extension to $\Lambda^k = \{\lambda: 0 \leqq \lambda_1, i = 1, \cdots, r; \Sigma\lambda_i = 1\}$; namely

$$L_3(\lambda, a) = \sum_{i=1}^{r} \lambda_i a_i^2 \quad \text{if } \lambda \in \Lambda^k - \Lambda.$$

To complete the definition of the equivalent problem, we naturally take the probability function under λ to be $f_{\theta(\lambda)}$. Admissibility of a procedure in the original problem is equivalent to admissibility in the reformulated problem; and for the remainder of this proof we consider only the reformulated problem.

Suppose $\sum_{x\in\mathscr{X}} \delta(T - \{a\} \mid x) > 0$ for every $a \in \mathscr{A}$. Then there must exist points $\alpha^{(i)} \in T$ and $x^{(i)} \in \mathscr{X}$, $i = 1, 2$, with $\alpha^{(1)} \neq \alpha^{(2)}$ and $\alpha^{(i)} \in \operatorname{Supp} \delta(\cdot \mid x^{(i)})$, $i = 1, 2$. If $x^{(1)} = x^{(2)}$ then δ is not a nonrandomized estimator, and hence must be inadmissible since $L_3(\lambda, \cdot)$ is strictly convex for $\lambda \in \Lambda$ (and convex for $\lambda \in \Lambda^k$). Suppose $x^{(1)} \neq x^{(2)}$ and consider the extended subproblem $\mathscr{P}^* = \mathscr{P}^*(\{x^{(1)}, x^{(2)}\})$. $\mathscr{P}^*$ has parameter space $\Lambda^* \subset \Lambda^k \times \{(\pi_1, \pi_2): \pi_1 + \pi_2 = 1\}$ and $f^*_{(\lambda,\pi)}(x^{(i)}) = \pi_i$, $i = 1, 2$. If $\lambda \in \Lambda$ then $(\lambda, \pi) \in \Lambda^*$ if and only if

$$\pi_i = \pi_i(\lambda) = f_{\theta(\lambda)}(x^{(i)}) / (\textstyle\sum_{j=1}^{2} f_{\theta(\lambda)}(x^{(j)})).$$

By Theorem 3.1, if δ is admissible then δ must be Bayes for some prior in problem $\mathscr{P}^*$. So, let P^* be a prior distribution on Λ^*. The proof will be complete when we show that δ cannot be Bayes with respect to P^*.

There are three possibilities:

$$\text{(4.4)} \qquad \operatorname{Supp} P^* \subset \{(\lambda, \pi): \lambda \in \Lambda^* - \Lambda\}, \text{ or}$$

$$\text{(4.5)} \qquad \operatorname{Supp} P^* \cap \{(\lambda, \pi): \lambda \in \Lambda\} = (\lambda', \pi(\lambda')), \text{ or}$$

$$\text{(4.6)} \qquad \operatorname{Supp} P^* \cap \{(\lambda, \pi): \lambda \in \Lambda\} \text{ contains at least two points.}$$

Observe that $a^{(i)} \in S_{P^*}(x^{(i)})$ if and only if $a^{(i)}$ minimizes

$$\int L_3(\lambda, a)\pi_i P^*(d(\lambda, \pi))$$

and thus if and only if

$$\begin{aligned} a_j^{(i)} &= \int_{\Lambda^*} (\textstyle\sum_{k=1}^{r} \lambda_k^{-1})^{-1}\pi_i P^*(d(\lambda, \pi)) \Big/ \int_{\Lambda^*} \lambda_j \pi_i P^*(d(\lambda, \pi)) \\ &= \int_{\lambda\in\Lambda} (\textstyle\sum\lambda_k^{-1})^{-1}\pi_i P^*(d(\lambda, \pi)) \Big/ \int_{\Lambda^*} \lambda_j \pi_i P^*(d(\lambda, \pi)) \qquad \text{(4.7)} \\ &\leqq \int_{\lambda\in\Lambda} (\textstyle\sum\lambda_k^{-1})^{-1}\pi_i P^*(d(\lambda, \pi)) \Big/ \int_{\lambda\in\Lambda} \lambda_j \pi_i P^*(d(\lambda, \pi)). \end{aligned}$$

Here the second equality involves only the fact that $(\sum \lambda_k^{-1})^{-1} = 0$ for $\lambda \in \Lambda^*$. (If the denominator in (4.7) is zero, so will be the numerator, and $a_j^{(i)}$ can take any value in [0, 1].)

Suppose (4.4). There must be some index j such that $\int_{\Lambda^*} \lambda_j \pi_1 P^*(d(\lambda, \pi)) > 0$. For this index $a_j^{(1)} = 0$ because of (4.4) and (4.7). Hence $a^{(1)} \neq \alpha^{(1)}$ and δ cannot be Bayes in problem $\mathscr{P}^*$ for a prior satisfying (4.4).

Suppose (4.5). Then

$$\text{(4.8)} \qquad a_j^{(i)} \leqq (\textstyle\sum\lambda_k'^{-1})^{-1}/(\lambda_j') = \theta_j(\lambda')$$

since $\int_{\lambda\in\Lambda}(\Sigma\lambda_k^{-1})^{-1}\pi_i P^*(d(\lambda, \pi)) = (\Sigma\lambda_k'^{-1})^{-1}P^*(\{(\lambda', \pi(\lambda'))\})$ and $\int_{\lambda\in\Lambda}\lambda_j\pi_i P^*(d(\lambda, \pi)) = \lambda_j' P^*(\{(\lambda', \pi(\lambda'))\})$. But, since $\alpha^{(1)} \neq \alpha^{(2)}$ and $\Sigma\alpha_k^{(1)} = \Sigma\alpha_k^{(2)} = 1 = \Sigma\theta_k(\lambda')$ there must exist indices i, j such that $\alpha_j^{(i)} > \theta_j(\lambda')$. For such indices $a_j^{(i)} < \alpha_j^{(i)}$. Hence δ cannot be Bayes for a prior satisfying (4.5).

Suppose (4.6). Let Q be the distribution on $\{(\lambda, \pi) \in \Lambda^*: \lambda \in \Lambda\}$ defined by

$$Q(d(\lambda, \pi)) = (\sum \lambda_k^{-1})^{-1}\pi_1 P^*(d(\lambda, \pi)) \int_{\lambda \in \Lambda} (\sum \lambda_k^{-1})^{-1}\pi_1 P^*(d(\lambda, \pi)).$$

Then Supp Q contains at least two points, and

$$a_j^{(1)} \leqq \left(\int_{\lambda \in \Lambda} \lambda_j (\sum \lambda_k^{-1}) Q(d(\lambda, \pi)) \right)^{-1} \leqq \int_{\lambda \in \Lambda} \lambda_j^{-1} (\sum \lambda_k^{-1})^{-1} Q(d(\lambda, \pi)) \tag{4.9}$$

by Jensen's inequality. Furthermore, there is strict inequality in the second inequality of (4.9) for some indices $j = 1, \cdots, r$. (Actually, this is true for any index j such that Supp Q contains at least two points having different λ_j coordinates.) Hence

$$\sum_{j=1}^r a_j^{(1)} < \int \sum_{j=1}^r \lambda_j^{-1} (\sum_{k=1}^r \lambda_k^{-1})^{-1} Q(d(\lambda, \pi)) = 1.$$

It follows that $a^{(1)} \notin T$; hence, again, $a^{(1)} \neq \alpha^{(1)}$. Hence δ cannot be Bayes for a prior satisfying (4.6); or, indeed, for any prior P^* in problem $\mathscr{P}^*$. As previously noted, this proves that δ is not admissible. □

It should be remarked that if one considers the more natural action space $\mathscr{A}' = \{(a_1, \cdots, a_r): a_i \geqq 0, i = 1, \cdots, r; \sum a_i = 1\}$ then δ_4 is an admissible estimator. This result was proved in the paper of Olkin and Sobel (1979) by means of an argument involving the Cramer-Rao inequality. It can also be proven by verifying that δ_4 is unique Bayes under loss L_1 and action space $\mathscr{A}'$ for the uniform prior over Θ. See Ighodaro (1980).

The multinomial problem with an entropy loss function is considered in Ighodaro, Santner, and Brown (1980). This loss function is unbounded in a more fundamental way than the loss function in the preceding example; and the methods of Section 3 need some alterations in order to yield a description of the minimal complete class.

REFERENCES

ALAM, K. (1978). Estimation of multinomial probabilities, mimeographed preprint (this preprint is not a preliminary version of Alam (1979)).

ALAM, K. (1979). Estimation of multinomial probabilities. *Ann. Statist.* **7** 282–283.

BROWN, L. (1977). Closure theorems for sequential-design processes. *Statistical Decision Theory and Related Topics* II 57–91.

BROWN, L., COHEN, A., and STRAWDERMAN, W. (1980). Complete classes for sequential tests of hypotheses. *Ann. Statist.* **8** 377–398.

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

IGHODARO, A. (1980). Ridge and James-Stein methods for contingency tables. Ph.D. Thesis, Cornell University.

IGHODARO, A., SANTNER, T., and BROWN, L. D. (1980). Admissibility and complete class results for the multinomial estimation problem with entropy and squared error loss. To appear in *J. Multivariate Anal.*

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

LECAM, L. (1955). An extension of Wald's theory of statistical decision functions. *Ann. Math. Statist.* **26** 69–81.

OLKIN, J. and SOBEL (1979). Admissible and minimax estimation for the multinomial distribution and for k independent binomial distributions. *Ann. Statist.* **7** 284–290.

SACKROWITZ, H. and STRAWDERMAN, W. (1974). On the admissibility of the M.L.E. for ordered binomial parameters. *Ann. Statist.* **2** 822–828.

WALD, A. (1950). *Statistical Decision Functions.* Wiley, New York.

WALD, A. and WOLFOWITZ. J. (1951). Characterization of the minimal complete class of decision functions when the number of distributions and decisions is finite. *Proc. Second Berk. Symp. Math. Statist. and Prob.* 149–157.

GRADUATE FIELD OF STATISTICS
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853