The Annals of Statistics
1985, Vol. 13, No. 1, 262–271

THE ADMISSIBILITY OF THE EMPIRICAL DISTRIBUTION FUNCTION

BY MICHAEL P. COHEN AND LYNN KUO

Bureau of Labor Statistics and State University of New York at Stony Brook

Consider the problem of estimating an unknown distribution function F from the class of all distribution functions given a random sample of size n from F. It is proved that the empirical distribution function is admissible for the loss functions $L(F, \hat{F}) = \int (\hat{F}(t) - F(t))^2(F(t))^a(1 - F(t))^b \, dW(t)$ for any $a < 1$ and $b < 1$ and finite measure W. Related results for simultaneous estimation of distribution functions and for finite population sampling are also given.

1. Introduction. For the problem of estimating an unknown distribution function F based on a sample from F, the empirical distribution function $\hat{F}$ is an extensively used estimator. The decision theoretic properties of $\hat{F}$ have received considerable attention. Aggarwal (1955) shows that $\hat{F}$ is the best invariant estimator for continuous F when the loss function is

$$L(F, \phi) = \int (F - \phi)^2/[F(1 - F)] \, dF.$$

Dvoretzky, Kiefer, and Wolfowitz (1956) prove that $\hat{F}$ is asymptotically minimax for a wide variety of loss functions. Read (1972) indicates that $\hat{F}(t)$ is *not* asymptotically admissible as an estimator of $F(t)$ (in a pointwise sense) when the loss function is mean squared error and F is known to be absolutely continuous. Phadia (1973) shows that $\hat{F}$ is minimax when the loss function is

$$L(F, \phi) = \int (F - \phi)^2/[F(1 - F)] \, dW.$$

Here W is any finite, non-null measure and F may be any distribution function (not necessarily continuous).

In this paper we prove that the empirical distribution function is admissible for the class of loss functions

$$L_{a,b}(F, \phi) = \int (F - \phi)^2 F^a (1 - F)^b \, dW$$

where a and b are real numbers, $a < 1$ and $b < 1$, W is any finite, non-null measure, and F may be any distribution function. To show that the empirical distribution function is admissible, it suffices to show it is particularly good for a subfamily of the distribution functions. Let $\mathscr{F}_k$ denote a family of discrete

Received March 1983; revised September 1984.
AMS 1980 *subject classifications.* Primary 62C15; secondary 62G30, 62D05.
Key words and phrases. Admissibility, empirical distribution function, i.i.d. sample, weighted quadratic loss, simple random sampling without replacement, finite population, multinomial distribution, multivariate hypergeometric distribution.

distributions with at most k jumps. An inductive argument is presented which essentially shows that, for all k, $\hat{F}$ is a generalized Bayes estimator with respect to a class of priors concentrated on subsets of $\mathscr{F}_k$. From this, we can show no other estimators can be as good as $\hat{F}$.

Although our problem is nonparametric, it can be seen in Section 3 to be also closely related to the problem of finding admissible estimators for multinomial parameters. Johnson (1971), Alam (1979), Olkin and Sobel (1979), and Brown (1981) all treated this problem extensively with different loss functions and different techniques. A modification of the concise inductive argument given by Alam is used here for our purposes. The inductive argument is closely akin to the "stepwise Bayes" method developed by Hsuan (1979), Brown (1981), and Meeden and Ghosh (1981).

In Section 4, simultaneous estimation for many distributions is considered. The empirical distribution functions from each of the populations are shown to be admissible for a given loss function.

The problem of estimating the population distribution function of a finite population is also studied. Consider a simple random sample (without replacement) of size n from a finite population of size N. The unknown population distribution F is defined by letting $F(t) = N^{-1} \sum_1^N I[y_i \le t]$, where y_i are the unknown population values associated with unit i. The minimax estimators are obtained by Cohen and Kuo (1981) for four different loss functions. The admissibility of the empirical distribution function for these four loss functions is proved for any sampling plan of size n in Section 5. A recent paper of Meeden, Ghosh, and Vardeman (1983) provides a detailed development of the relationship between nonparametric estimation and finite population estimation.

2. Notation. The parameter space $\mathscr{F}$, the action space $\mathscr{A}$, and the loss function L are defined as follows:

$\mathscr{F} = \{F\colon F$ is a right-continuous distribution function on the real line $\mathbb{R}^1\}$

$\mathscr{A} = \{\phi\colon \phi$ is a nondecreasing right-continuous distribution function on $\mathbb{R}^1$ such that $0 \le \phi(-\infty) \le \phi(\infty) \le 1\}$

$$L_{a,b}(F, \phi) = \int (F(t) - \phi(t))^2 (F(t))^a (1 - F(t))^b \, dW(t) \tag{2.1}$$

where a and b are real numbers, and W is a given non-null, finite measure on $(\mathbb{R}^1, \mathscr{B})$ where $\mathscr{B}$ is the Borel field on $\mathbb{R}^1$. The integrand in (2.1) should be interpreted as zero if it has the form 0/0.

Given a sample of size n, $\mathbf{x} = (x_1, \cdots, x_n)$ from F, an estimator ϕ of $F(t)$ which depends on $\mathbf{x}$ is denoted by $\phi(t; \mathbf{x})$. In particular, the empirical distribution function is denoted by $\hat{F}$, that is $\hat{F}(t; \mathbf{x}) = n^{-1} \sum_1^n I[x_i \le t]$. The risk function $E_F L(F, \phi(t; \mathbf{x}))$ is denoted by $R(F, \phi)$.

The family of discrete distributions with at most k jumps is denoted by $\mathscr{F}_k$.

3. Admissibility. The key to the proof of admissibility is the correct choice of the subfamilies of distribution functions for which $\hat{F}$ is a particularly good estimator. The usual approach is to show $\hat{F}$ is Bayes with respect to a prior defined on this subfamily. One possibility might be the prior used by Phadia (1973). This prior concentrates on distribution functions with jumps at two points at most, say $-u$ and u, where the size of the jump at $-u$ is chosen from a beta distribution $\mathscr{B}_e(\alpha, \beta)$. Phadia shows that the empirical distribution function is Bayes with respect to this prior with $\alpha = \beta = 1$ for the loss function $L_{-1,-1}$. However, since this prior does not have full support, the Bayes property of $\hat{F}$ does not imply its admissibility. Essentially, any estimator which agrees with $\hat{F}$ when the data show at most two distinct values is Bayes with respect to this prior. It is clear that many of these estimators are not admissible.

A second possibility might be to show $\hat{F}$ is Bayes with respect to Ferguson's Dirichlet process prior (1973) with parameter α. It can be shown that the proper Bayes estimator $\phi_\alpha(t)$ against this prior for the loss in (2.1) is $(a + \alpha(-\infty, t] + \sum_1^n I(x_i \le t))/(a + b + \alpha(\mathbb{R}^1) + n)$. In order to have $\phi_\alpha(t) = \hat{F}(t)$ for $-u \le t < u$, it is necessary to choose an α which assigns no mass between $-u$ and u. Consequently, this approach encounters the same difficulties mentioned above in proving the admissibilty of $\hat{F}$. Instead, we consider priors which concentrate on distributions with more than two jumps. We prove that no other estimator is as good as $\hat{F}$ by an inductive argument which shows $\hat{F}$ is the essentially unique generalized Bayes rule with respect to a class of prior measures on $\mathscr{F}_k$. In fact, the following stronger result will be established:

THEOREM 3.1. (i) *If* $R(F, \phi) \le R(F, \hat{F})$ *for all* $F \in \mathscr{F}_{n+1}$, *then for all* $\mathbf{x}$, $\phi(t; \mathbf{x}) = \hat{F}(t; \mathbf{x})$ a.e. $dW(t)$ *whenever* $a < 1$ *and* $b < 1$.

(ii) *If* $a \le 0$ *and* $b \le 0$, *then* $R(F, \phi) \le R(F, \hat{F})$ *for all* $F \in \mathscr{F}_n$ *suffices.*

PROOF. Given an arbitrary vector $\mathbf{t} = (t_1, \cdots, t_k)$, such that $t_1 < t_2 < \cdots < t_k$, we define $\mathscr{F}_k(\mathbf{t})$ by $\mathscr{F}_k(\mathbf{t}) = \{F \mid F(t) = \sum_1^k p_i I[t_i \le t]$, where $p_i \ge 0$, $\sum_1^k p_i = 1\}$. (Note that $\mathscr{F}_k = \cup_{\mathbf{t} \in \mathbb{R}^k} \mathscr{F}_k(\mathbf{t})$). If $x_1, \cdots, x_n$ are a sample from $F \in \mathscr{F}_k(\mathbf{t})$, then all the observations are located at $t_1, \cdots, t_k$ with multiplicities denoted by $j_1, \cdots, j_k$ respectively. The empirical distribution function $\hat{F}$ reduces to $\hat{F}(t; \mathbf{x}) = \sum_{i=1}^k (j_i/n) I[t_i \le t]$, where $j_i \ge 0$, $\sum_{i=1}^k j_i = n$. Set $\mathbf{j}_k = (j_1, \cdots, j_k)$. The order statistics for the sample and hence $\mathbf{j}_k$ are sufficient statistics for this problem. It will be convenient to denote $\phi(t; \mathbf{x})$ by $\phi(t; \mathbf{j}_k)$. The lemma given below is key to the proof of the theorem.

LEMMA 3.1. *Let* $a < 1$ *and* $b < 1$. *Suppose* $R(F, \phi) \le R(F, \hat{F})$ *for all* $F \in \mathscr{F}_k(\mathbf{t})$. *Then for each* $\mathbf{j}_k$

(i) $\phi(t; \mathbf{j}_k) = \hat{F}(t; \mathbf{j}_k)$ a.e. $dW(t)$ *on the interval* $[t_1, t_k)$;

(ii) *if* $a \le 0$ *then* $\phi(t; \mathbf{j}_k) = 0$ *for* $t < t_1$; *and*

(iii) *if* $b \le 0$ *then* $\phi(t; \mathbf{j}_k) = 1$ *for* $t \ge t_k$.

Before proving the lemma, we indicate how the theorem follows from it.

Suppose $R(F, \phi) \le R(F, \hat{F})$ for all $F \in \mathscr{F}_{n+1}$ and $a < 1$, $b < 1$. Given $\mathbf{x}$, let $\mathbf{t}_{n+1}$ be a vector whose coordinates include all the distinct coordinates of $\mathbf{x}$ as well as an arbitrary point. From part (i) of the lemma, it follows easily that $\phi(t; \mathbf{x}) = \hat{F}(t; \mathbf{x})$, a.e. $dW(t)$. If $a \le 0$, $b \le 0$ and $R(F, \phi) \le R(F, \hat{F})$ for all $F \in \mathscr{F}_n$, let $\mathbf{t}_n$ be a vector whose coordinates include all the distinct coordinates of $\mathbf{x}$. From parts (i), (ii), and (iii) of the lemma, it follows that $\phi(t; \mathbf{x}) = \hat{F}(t; x)$ a.e. $dW(t)$. The theorem is thus a consequence of the lemma.

Proof of Lemma. Let us define $S_k = \{\mathbf{j}_k = (j_1, \cdots, j_k), j_i \ge 1$ for $i = 1, \cdots, k\}$. Then we can express the risk function for an $F \in \mathscr{F}_k(\mathbf{t})$ as the sum of two components

$$R(F, \phi) = C(F, \phi) + D(F, \phi) \tag{3.1}$$

where

$$C(F, \phi) = E \int (F(t) - \phi(t; \mathbf{j}_k))^2 F^a(t)(1 - F(t))^b I\{\mathbf{j}_k \in S_k\}\, dW(t)$$

and

$$D(F, \phi) = E \int (F(t) - \phi(t; \mathbf{j}_k))^2 F^a(t)(1 - F(t))^b I\{\mathbf{j}_k \in S_k^c\}\, dW(t).$$

Now define $\mathscr{B}_k(\mathbf{t}) = \{F \in \mathscr{F}_k(\mathbf{t}) \mid p_i = 0$ for some $i\}$. Note that if $F \in \mathscr{B}_k(\mathbf{t})$, then S_k is empty, so $C(F, \phi) = 0$.

We treat the case $a \le 0$ and $b \le 0$ first. For this case the lemma reduces to showing that if $R(F, \phi) \le R(F, \hat{F})$ for all $F \in \mathscr{F}_k$, then

$$\phi(t; \mathbf{j}_k) = \hat{F}(t; \mathbf{j}_k) \quad \text{a.e.} \quad dW(t). \tag{3.2}$$

We will prove (3.2) by induction on k. Notice that it holds for $k = 1$. Suppose that (3.2) holds for a particular k. We are going to show that if

$$R(F, \phi) \le R(F, \hat{F}) \quad \text{for all} \quad F \in \mathscr{F}_{k+1}, \tag{3.3}$$

then

$$\phi(t; \mathbf{j}_{k+1}) = \hat{F}(t; \mathbf{j}_{k+1}) \quad \text{a.e.} \quad dW(t). \tag{3.4}$$

We will prove (3.4) for any $\mathbf{t} \in \mathbb{R}^{k+1}$ by treating each of the two components of the risk separately. For $\mathbf{j}_{k+1} \in S_{k+1}^c$, we have

$$\phi(t; \mathbf{j}_{k+1}) = \hat{F}(t; \mathbf{j}_{k+1}) \quad \text{a.e.} \quad dW(t)$$

by considering $F \in \mathscr{B}_{k+1}(\mathbf{t})$ and applying (3.2). Therefore

$$D(F, \phi) = D(F, \hat{F}) \quad \text{for all} \quad F \in \mathscr{F}_{k+1}(\mathbf{t}).$$

Moreover, from (3.1) and (3.3), we have

$$C(F, \phi) \le C(F, \hat{F}) \quad \text{for all} \quad F \notin \mathscr{B}_{k+1}(\mathbf{t}).$$

That is, for $p_i > 0$, $i = 1, \cdots, k+1$,

$$\begin{aligned}
&\textstyle\sum_{i=0}^{k+1} \int_{[t_i,t_{i+1})} \sum_{\mathbf{j}_{k+1}} (\sum_{\ell=1}^{i} p_\ell - \phi(t; \mathbf{j}_{k+1}))^2 (\sum_{\ell=1}^{i} p_\ell)^a \\
&\qquad \cdot (1 - \textstyle\sum_{\ell=1}^{i} p_\ell)^b \binom{n}{j_1, \cdots, j_{k+1}} p_1^{j_1} \cdots p_{k+1}^{j_{k+1}} \, dW(t) \\
&\le \textstyle\sum_{i=1}^{k} \int_{[t_i,t_{i+1})} \sum_{\mathbf{j}_{k+1}} (\sum_{\ell=1}^{i} p_\ell - \sum_{\ell=1}^{i} j_\ell/n)^2 (\sum_{\ell=1}^{i} p_\ell)^a \\
&\qquad \cdot (1 - \textstyle\sum_{\ell=1}^{i} p_\ell)^b \binom{n}{j_1, \cdots, j_{k+1}} p_1^{j_1} \cdots p_{k+1}^{j_{k+1}} \, dW(t)
\end{aligned} \tag{3.5}$$

where $t_0 = -\infty$, $t_{k+2} = \infty$, $\sum_{\ell=1}^{0} p_\ell = 0$ and $\sum_{\mathbf{j}_{k+1}}$ means a k-fold summation over indices $j_1, \cdots, j_k$ such that $j_1 + j_2 + \cdots + j_{k+1} = n$.

It is clear that the left-hand side of (3.5) is made no larger by taking $\phi(t; \mathbf{j}_{k+1}) = 0$ for $t < t_1$ and $\phi(t; \mathbf{j}_{k+1}) = 1$ for $t \ge t_{k+1}$. We can now divide both sides of (3.5) by $p_1^{a+1} p_2 \cdots p_k p_{k+1}^{b+1}$ and integrate over $dp_1 dp_2 \cdots dp_k$, where $p_i > 0$, for all $i = 1, \cdots, k+1$, and $\sum_1^{k+1} p_i = 1$. Then interchanging $dW(t)$ and $dp_1 dp_2 \cdots dp_k$ (by Fubini), the left-hand side of the new inequality is uniquely minimized (up to sets of measure zero $dW(t)$) by

$$\begin{aligned}
&\phi(t; \mathbf{j}_{k+1}) \\
&= \frac{\int \cdots \int (\sum_{\ell=1}^{i} p_\ell)^{a+1} (1 - \sum_{\ell=1}^{i} p_\ell)^b p_1^{j_1-a-1} p_2^{j_2-1} \cdots p_k^{j_k-1} p_{k+1}^{j_{k+1}-b-1} \, dp_1 \cdots dp_k}{\int \cdots \int (\sum_{\ell=1}^{i} p_\ell)^{a} (1 - \sum_{\ell=1}^{i} p_\ell)^b p_1^{j_1-a-1} p_2^{j_2-1} \cdots p_k^{j_k-1} p_{k+1}^{j_{k+1}-b-1} \, dp_1 \cdots dp_k} \\
&= \frac{E(\sum_{\ell=1}^{i} p_\ell)^{a+1} (1 - \sum_{\ell=1}^{i} p_\ell)^b}{E(\sum_{\ell=1}^{i} p_\ell)^{a} (1 - \sum_{\ell=1}^{i} p_\ell)^b}, \\
&\qquad\qquad \text{where } \textstyle\sum_{\ell=1}^{i} p_\ell \sim \mathscr{B}_e(\sum_{\ell=1}^{i} j_\ell - a, \sum_{\ell=i+1}^{k+1} j_\ell - b) \\
&= \textstyle\sum_{\ell=1}^{i} j_\ell/n, \quad \text{when} \quad t_i \le t < t_{i+1}.
\end{aligned} \tag{3.6}$$

We also must have $\phi(t; \mathbf{j}_{k+1}) = 0$ for $t < t_1$ and $\phi(t; \mathbf{j}_{k+1}) = 1$ for $t \ge t_{k+1}$. The lemma has therefore now been proved for the case $a \le 0$ and $b \le 0$.

The cases $0 < a < 1$ or $0 < b < 1$ (or both) are somewhat more complicated. The proof is again by induction on k. Suppose that if $R(F, \phi) \le R(F, \hat{F})$ for all $F \in \mathscr{F}_k$, then for each $\mathbf{j}_k$, the conditions (i), (ii), and (iii) of Lemma 3.1 hold. We will show the corresponding result for $k+1$.

Suppose $\mathbf{j}_{k+1} \in S_{k+1}^c$. Then $j_i = 0$ for some i. Let $\mathbf{t}' = (t_1', \cdots, t_k')$ be the k-vector formed from $\mathbf{t} \in \mathbb{R}^{k+1}$ with t_i removed. Let $t_1' \le t_2' \le \cdots \le t_k'$. Then by (i)

$$\phi(t; \mathbf{j}_{k+1}) = \hat{F}(t; \mathbf{j}_{k+1}) \quad \text{a.e.} \quad dW(t) \quad \text{on} \quad [t_1', t_k'). \tag{3.7}$$

We wish to show that

$$\phi(t; \mathbf{j}_{k+1}) = \hat{F}(t; \mathbf{j}_{k+1}) \quad \text{a.e.} \quad dW(t) \quad \text{on} \quad [t_1, t_{k+1}). \tag{3.8}$$

Note that $t_1' = t_1$ or $t_k' = t_{k+1}$. If both equalities hold, then (3.8) is established.

Otherwise, suppose $t_1' \neq t_1$. Then $t_1' = t_2$. Let $\mathbf{j}_{k+1,1}$ denote the first coordinate of $\mathbf{j}_{k+1}$. It follows from (3.7), (ii), and (iii) that

$$\begin{aligned}
&E\int_{[t_1,t_2)} (F(t) - \phi(t;\mathbf{j}_{k+1}))^2F^a(t)(1-F(t))^bI\{\mathbf{j}_{k+1,1}=0\}\,dW(t)\\
(3.9)\quad &\leq E\int_{[t_1,t_2)} (F(t) - \hat{F}(t;\mathbf{j}_{k+1}))^2F^a(t)(1-F(t))^bI\{\mathbf{j}_{k+1,1}=0\}\,dW(t)\\
&\quad + E\int_{[t_1,t_{k+1})} (F(t) - \hat{F}(t;\mathbf{j}_{k+1}))^2F^a(t)(1-F(t))^bI\{\mathbf{j}_{k+1,1}\neq 0\}\,dW(t)
\end{aligned}$$

for any $F \in \mathscr{F}_{k+1}(\mathbf{t})$. Consider $F_\varepsilon \in \mathscr{F}_{k+1}(\mathbf{t})$ defined by

$$F_\varepsilon(t) = \varepsilon\, I[t_1 \leq t] + (p_2 - \varepsilon)I[t_2 \leq t] + \textstyle\sum_{\ell=3}^{k+1} p_\ell I[t_\ell \leq t]$$

where each $p_\ell > 0$, $\sum_{\ell=2}^{k+1} p_\ell = 1$ and $0 < \varepsilon < p_2$. Setting $F = F_\varepsilon$ in (3.9) and dividing both sides by ε^a, we have

$$\begin{aligned}
&E\int_{[t_1,t_2)} (\varepsilon - \phi(t;\mathbf{j}_{k+1}))^2(1-\varepsilon)^bI\{\mathbf{j}_{k+1,1}=0\}\,dW(t)\\
&\quad\leq E\int_{[t_1,t_2)} \varepsilon^2(1-\varepsilon)^bI\{\mathbf{j}_{k+1,1}=0\}\,dW(t)\\
&\qquad + E\int_{[t_1,t_{k+1})} (F_\varepsilon(t) - \hat{F}(t;\mathbf{j}_{k+1}))^2F_\varepsilon^a(t)\\
(3.10)\quad&\qquad\qquad\cdot(1-F_\varepsilon(t))^b\varepsilon^{-a}I\{\mathbf{j}_{k+1,1}\neq 0\}\,dW(t)\\
&\quad\leq \varepsilon^2(1-\varepsilon)^b\int_{[t_1,t_2)} dW(t)\\
&\qquad + \int_{[t_1,t_{k+1})} F_\varepsilon^a(t)(1-F_\varepsilon(t))^b\varepsilon^{-a}P\{\mathbf{j}_{k+1,1}\neq 0\}\,dW(t).
\end{aligned}$$

Note that $P\{\mathbf{j}_{k+1,1}\neq 0\} = 1-(1-\varepsilon)^n \sim n\varepsilon$ as $\varepsilon \downarrow 0$. Hence, as $\varepsilon \downarrow 0$, both terms of (3.10) vanish, so that

$$\phi(t;\mathbf{j}_{k+1}) = 0 \quad \text{a.e.} \quad dW(t) \quad \text{on} \quad [t_1, t_2) \quad \text{whenever} \quad \mathbf{j}_{k+1,1} = 0.$$

The situation $t_k' \neq t_{k+1}$ is handled similarly. Thus for $\mathbf{j}_{k+1} \in S_{k+1}^c$, we have

$$\phi(t;\mathbf{j}_{k+1}) = \hat{F}(t;\mathbf{j}_{k+1}) \quad \text{a.e.} \quad dW(t) \quad \text{on} \quad [t_1, t_{k+1});$$

if $a \leq 0$ then $\phi(t;\mathbf{j}_{k+1}) = 0$ for $t < t_1$; and if $b \leq 0$ then $\phi(t;\mathbf{j}_{k+1}) = 1$ for $t \geq t_k$. Therefore, $D(F,\phi) = D(F,\hat{F})$ for all $F \in \mathscr{F}_{k+1}(\mathbf{t})$. The rest of the proof is the same as that for the case $a \leq 0$ and $b \leq 0$.

Note that in order to have proper integrals in (3.6), we need $j_1 - a > 0$, $j_{k+1} - b > 0$, and $j_i > 0$ for $i = 2, \cdots, k$. When $\mathbf{j}_{k+1} \in S_{k+1}$, the above restrictions require the conditions $a < 1$ and $b < 1$.

For $a \geq 1$ or $b \geq 1$, the admissibility of $\hat{F}$ is still unknown. It can be shown that Lemma 3.1 holds for all a and b when $k \leq 2$. The cases $k \geq 3$ are still unresolved.

4. Estimating many distribution functions. In this section, simultaneous estimation for several distribution functions is considered. For $i = 1, \cdots, I$, suppose a sample $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{in_i})$ of size n_i is taken from an unknown distribution function F_i. Given $F_1, \cdots, F_I$, the observations $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_I$ are independent. Let $\mathbf{x}$, $\mathbf{F}$, ϕ denote $(\mathbf{x}_1, \cdots, \mathbf{x}_I)$, $(F_1, \cdots, F_I)$, and $(\phi_1, \cdots, \phi_I)$ respectively. Let $\hat{F}_i(t; \mathbf{x}) = \sum_{\ell=1}^{n_i} I(x_{i\ell} \leq t)/n_i$. The empirical distribution functions $\hat{F}_1, \cdots, \hat{F}_I$ are admissible for the following loss functions for all $a_i < 1$ and $b_i < 1$:

$$L(\mathbf{F}, \phi) = \sum_{i=1}^{I} \int (F_i(t) - \phi_i(t))^2 F_i^{a_i}(t)(1 - F_i(t))^{b_i}\, dW_i(t). \tag{4.1}$$

Hence, the Stein phenomenon does not occur in this problem.

THEOREM 4.1. (i) *If $R(\mathbf{F}, \phi) \leq R(\mathbf{F}, \hat{\mathbf{F}})$ for all $F_i \in \mathscr{F}_{n_i+1}$, $i = 1, \cdots, I$, then for all $\mathbf{x}$ and i, $\phi_i(t; \mathbf{x}) = \hat{F}_i(t; \mathbf{x})$ a.e. $dW_i(t)$ whenever $a_i < 1$ and $b_i < 1$.*

(ii) *If both $a_i \leq 0$ and $b_i \leq 0$ for any i, then $F_i \in \mathscr{F}_{n_i}$ suffices.*

The proof is omitted. The key idea is that the stepwise Bayes argument in Section 3 can be applied to each summand in (4.1) separately.

Gutmann (1982) shows that Stein's phenomenon is impossible in finite sample space problems. We need the somewhat stronger criterion of "essential uniqueness" (as in Lemma 3.1) as a step in the proof. Similar multivariate estimation problems are treated by Meeden, Ghosh, and Vardeman (1983, pages 21–24).

5. Finite populations. In this section, we consider the problem of sampling from a finite population. Suppose there is a population $U = \{1, 2, \cdots, N\}$ of N identifiable units with population value $y_i \in \mathbb{R}^1$ associated with the ith unit. Let s denote a subset of U containing n distinct elements and let S_n denote the set of all $\binom{N}{n}$ such s. A *sampling design* (with fixed sample size n) is a probability measure π on S_n. The survey sampler chooses a sample s with probability $\pi(s)$ and observes the data $\{(i, y_i);\ i \in s\}$. It is desired to estimate the *population distribution function* F where

$$F(t) = (1/N) \sum_{i=1}^{N} I[y_i \leq t].$$

The empirical distribution function is defined by

$$\hat{F}(t) = (1/n) \sum_{i \in s} I[y_i \leq t].$$

We will show that $\hat{F}$ is admissible for the four loss functions defined in (2.1) with $a = 0$ or -1 and $b = 0$ or -1. That is, there does not exist an estimator ϕ such that

$$\sum_{s \in S_n} \pi(s) L_{a,b}(F, \phi) \leq \sum_{s \in S_n} \pi(s) L_{a,b}(F, \hat{F}) \quad (a = 0 \text{ or } -1,\ b = 0 \text{ or } -1)$$

for all $\mathbf{y} = (y_1, \cdots, y_N)$(and hence F) with strict inequality for at least one $\mathbf{y}$.

In fact, we prove a stronger result. First define

$$\mathscr{F}_{N,k} = \{F\colon F(t) = \textstyle\sum_{i=1}^{k} (k_i/N)I[t_i \le t], \text{ where } k_i \ge 0, \sum_{i=1}^{k} k_i = N\}.$$

THEOREM 5.1. *If*

$$\textstyle\sum_{s\in S_n} \pi(s)L_{a,b}(F, \phi) \le \sum_{s\in S_n} \pi(s)L_{a,b}(F, \hat{F}) \quad (a = 0 \text{ or } -1,\ b = 0 \text{ or } -1)$$

for all $\mathbf{y}$ *such that the corresponding* $F \in \mathscr{F}_{N,n}$, *then for all samples* $s \in S_n$,

$$\phi(t; \{(i, y_i); i \in s\}) = \hat{F}(t; \{(i, y_i); i \in s\}) \quad \text{a.e.} \quad dW(t).$$

PROOF. We follow the proof of Theorem 3.1 with the following changes: Let θ_i denote the number of population units with value t_i. Let $\mathbf{y}^* = (y_1^*, \cdots, y_N^*)$, where $y_1^* \le y_2^* \le \cdots \le y_N^*$, be the order statistic of population values determined by the θ_i. Let $\gamma(1), \gamma(2), \cdots, \gamma(N)$ be a permutation of $1, 2, \cdots, N$. Then the θ_i and permutation γ together determine a population by $\mathbf{y} = (y^*_{\gamma(1)}, y^*_{\gamma(2)}, \cdots, y^*_{\gamma(N)})$. Let γ be chosen independently of the θ_i, each of the $N!$ permutations assigned equal probability $1/N!$ (The choice makes the observed y_i values ($i \in s$) sufficient. See Lehmann, 1983, pages 212–213). Now replace p_i in Theorem 3.1 by θ_i/N, and $\binom{n}{j_1, \cdots, j_{k+1}} p_1^{j_1} \cdots p_{k+1}^{j_{k+1}}$ in (3.5) by $\binom{\theta_1}{j_1}\binom{\theta_2}{j_2} \cdots \binom{\theta_{k+1}}{j_{k+1}}/\binom{N}{n}$. Divide both sides of (3.5) by $\theta_1^{a+1}\theta_2 \cdots \theta_k\theta_{k+1}^{b+1}$ and sum over the $\theta_1, \theta_2, \cdots, \theta_k$ and all permutations γ. Note that within these summations we can assume without loss of generality that $s = \{1, \cdots, n\}$. Also observe that $(\theta_1 - j_1, \cdots, \theta_{k+1} - j_{k+1})$ given $\mathbf{j}_{k+1}$ has a Dirichlet multinomial distribution with parameters $(N - n, j_1 - a, j_2, \cdots, j_k, j_{k+1} - b)$. A straightforward computation (see Blackwell and Girshick, 1954, page 168) shows, whenever $a = 0$ or -1, $b = 0$ or -1, that

$$\frac{E[(\theta_1 + \cdots + \theta_i)^{a+1}(N - (\theta_1 + \cdots + \theta_i))^b]}{N\,E[(\theta_1 + \cdots + \theta_i)^a(N - (\theta_1 + \cdots + \theta_i))^b]} = \frac{\sum_{\ell=1}^{i} j_\ell}{n},$$

as required.

REMARK. A unified theory of proving admissibility for both i.i.d. samples and samples from finite populations has been developed by Meeden, Ghosh, and Vardeman (1983). The connection between samples from finite populations and i.i.d. samples can be seen as follows. Suppose the finite population is chosen from a superpopulation with replacement with $P(x = t_i) = p_i$, $i = 1, \cdots, k$, where x denotes an outcome of a draw from this superpopulation. Let θ_i denote the number of elements in the finite population with value t_i. Then the probability mass function of $\theta_1, \cdots, \theta_k$ is $f(\boldsymbol{\theta} \mid \mathbf{p}) = \binom{N}{\theta_1 \cdots \theta_k} p_1^{\theta_1} \cdots p_k^{\theta_k}$.

If we choose a Dirichlet prior distribution on $p_1, \cdots, p_{k-1}$, then the posterior distribution of $j_1, \cdots, j_{k-1}$ given $\theta_1, \cdots, \theta_{k-1}$ can be shown as in Ericson (1969) to be the multivariate hypergeometric distribution

$$f(\mathbf{j} \mid \boldsymbol{\theta}) = \prod_{i=1}^{k} \binom{\theta_i}{j_i} \Big/ \binom{N}{n}, \quad \text{where} \quad \textstyle\sum_1^k j_i = n. \tag{5.1}$$

Note that (5.1) is the same sampling distribution used in the previous paragraph.

REMARK. For $a > 0$ or $b > 0$, $\hat{F}$ is inadmissible. It is dominated by $\hat{F}_U = \max((1/N), \hat{F})$ or $\hat{F}_L = \min(1 - (1/N), \hat{F})$, respectively.

REMARK. Results corresponding to those of Section 4 on simultaneous estimation of finite population distributions can also be obtained.

6. Final comments.

The loss functions

$$L(F, \phi) = \int (F(t) - \phi(t))^2 (F(t))^a (1 - F(t))^b \, dF(t), \tag{6.1}$$

with $dF(t)$ replacing $dW(t)$ in (2.1), are also of considerable interest. By an approach similar to that used here, Brown (1984) has shown that $\hat{F}$ is admissible, when $a = b = -1$, for the class of *all* distribution functions.

If $a \neq -1$ or $b \neq -1$ in (6.1) and the parameter space is the class of *continuous* distribution functions, the empirical distribution function is not the best invariant rule (Ferguson, 1967, pages 191–197). Hence, it is inadmissible. For $a = b = -1$, on the other hand, the empirical distribution function is a best invariant rule. A long-standing open question was whether or not the best invariant rule is also minimax (Ferguson, 1967, page 197). Brown (personal communication) has recently announced that $\hat{F}$ is minimax. He constructs an intricate sequence of priors on continuous distribution functions. The admissibility or inadmissibility of $\hat{F}$ in this situation is still unknown.

Acknowledgement. The authors wish to thank Professor Thomas S. Ferguson and the referees for many useful comments which helped to clarify and improve the presentation.

REFERENCES

AGGARWAL, O. P. (1955). Some minimax invariant procedures for estimating a cumulative distribution function. *Ann. Math. Statist.* **26** 450–462.

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

BLACKWELL, D. and GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions.* Wiley, New York. Dover (1979), New York.

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

BROWN, L. (1984). Admissibility in discrete and continuous invariant nonparametric problems, and in their multivariate analogs. Unpublished.

COHEN, M. P., and KUO, L. (1981). Minimax estimation of a distribution function for finite population sampling. University of Michigan Technical Report No. 110. *Statistics and Decisions*, to appear.

DVORETZKY, A., KIEFER, J., and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

ERICSON, W. A. (1969). Subjective Bayesian models in sampling finite populations. *J. Roy. Statist. Soc. Ser. B* **31** 195–224.

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

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230.

GUTMANN, S. (1982). Stein's paradox is impossible in problems with finite sample space. *Ann. Statist.* **10** 1017–1020.

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

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

LEHMANN, E. L. (1983). *Theory of Point Estimation.* Wiley, New York.

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

MEEDEN, G., GHOSH, M., and VARDEMAN, S. (1985). Some admissible nonparametric and related finite population sampling estimators. *Ann. Statist.* **13.** To appear.

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

PHADIA, E. G. (1973). Minimax estimation of a cumulative distribution function. *Ann. Statist.* **1** 1149–1157.

READ, R. R. (1972). The asymptotic inadmissibility of the sample distribution function. *Ann. Math. Statist.* **43** 89–95.

STATISTICAL METHODS DIVISION
OFFICE OF PRICES AND LIVING CONDITIONS
BUREAU OF LABOR STATISTICS
600 E STREET N.W.
WASHINGTON, D.C. 20212

DEPARTMENT OF APPLIED MATHEMATICS AND STATISTICS
STATE UNIVERSITY OF NEW YORK
STONY BROOK, NEW YORK 11794