The Annals of Statistics
1989, Vol. 17, No. 3, 1347–1359

INADMISSIBILITY OF THE EMPIRICAL DISTRIBUTION FUNCTION IN CONTINUOUS INVARIANT PROBLEMS

BY QIQING YU

University of California, Los Angeles and Zhongshan University

Consider the classical invariant decision problem of estimating an unknown continuous distribution function F, with the loss function $L(F, a) = \int (F(t) - a(t))^2 [F(t)]^\alpha [1 - F(t)]^\beta \, dF(t)$, and a random sample of size n from F. It is proved that the best invariant estimator is inadmissible when:

1. $n > 0$, $-1 < \alpha, \beta \le 0$ and $-1 \le \alpha + \beta$.
2. $n > 0$, $-1 < \alpha = \beta \le -\frac{1}{2}$.
3. $n > 1$, (i) $\alpha = -1$ and $\beta = 0$, or (ii) $\alpha = 0$ and $\beta = -1$.
4. $n > 2$, $\alpha = \beta = -1$.

Thus the empirical distribution function, which is the best invariant estimator when $\alpha = \beta = -1$, is inadmissible when $n \ge 3$. This extends some results of Brown.

1. Introduction. This paper presents results on the inadmissibility of the best invariant estimator of a continuous distribution function. The background of the problem is as follows.

Aggarwal (1955) introduced the problem of the invariant estimation of an unknown continuous distribution function $F(t)$, with the loss function

$$L(F, a) = \int \{F(t) - a(t)\}^2 h(F(t)) \, dF(t), \tag{1.1}$$

based on a sample of size n from $F(t)$. This decision problem is invariant under monotone transformations. It turns out that all the nonrandomized invariant estimators are of the form

$$d(t) = \sum_{i=0}^{n} u_i 1(Y_i \le t < Y_{i+1}), \tag{1.2}$$

where $1(E)$ represents the indicator function of the event E, $Y_0 = -\infty$, $Y_{n+1} = +\infty$ and $Y_1 < \cdots < Y_n$ are the order statistics of the sample $X_1, \ldots, X_n$, and $u_0, \ldots, u_n$ are constants. The best invariant estimator, denoted by $d_0(t)$, has constant risk and has the form (1.2) with

$$u_i = \int_0^1 t^{i+1}(1-t)^{n-i} h(t) \, dt \Big/ \int_0^1 t^i (1-t)^{n-i} h(t) \, dt, \qquad i = 0, \ldots, n. \tag{1.3}$$

Much study has been devoted to the theoretical properties of the best invariant estimator. A long outstanding open question [see, for example, Ferguson (1967)] has been "Is the best invariant estimator minimax?" Whether

Received April 1986; revised August 1988.
AMS 1980 *subject classifications*. Primary 62C15; secondary 62D05.
Key words and phrases. Admissibility, invariant estimator, empirical distribution function, nonparametric estimator, Cramér–von Mises loss.

or not the best invariant estimator is admissible is another interesting question. Read (1972) established asymptotic inadmissibility of the best invariant estimator for some special $h(t)$ with the loss (1.1). Brown (1988) gave an important result in this respect. He proved that, when $h(t) = 1$, the best invariant estimator is inadmissible for all sample sizes $n \geq 1$.

The most interesting case is when $h(t) = t^{-1}(1 - t)^{-1}$. In this case, the best invariant estimator, $\hat{F}(t) = 1/n\Sigma_{i=1}^{n} 1(X_i \leq t)$, is the empirical distribution function (empirical c.d.f.). Aggarwal (1955) pointed out that it is not admissible if $h(t) = t^{\alpha}(1 - t)^{\beta}$, $\alpha \geq -1$ and $\beta \geq -1$, with at least one inequality being strict, since it is not best invariant. Dvoretzky, Kiefer and Wolfowitz (1956) showed that it is asymptotically minimax for a wide variety of loss functions. Brown (1988) proved that it is admissible if the parameter space of continuous distribution functions is replaced by the family of all distribution functions.

If the loss function is taken to be

$$L(F, a) = \int (F(t) - a(t))^2 F(t)^{-1}(1 - F(t))^{-1}\, dW(t),$$

where $W(t)$ is a finite nonzero measure, and the parameter space is the family of all distribution functions, then the problem does not have an invariant structure. Phadia (1973) proved that the empirical distribution function is minimax for all such $W(t)$. Cohen and Kuo (1985) showed that the empirical distribution function is admissible for the loss

$$L(F, a) = \int (F(t) - a(t))^2 F(t)^{\alpha}(1 - F(t))^{\beta}\, dW(t),$$

where $-1 \leq \alpha$, $\beta < 1$, W is a finite nonzero measure and the parameter space is the family of all distribution functions.

In this paper, the classical setup, having the loss (1.1) with $h(t) = t^{\alpha}(1 - t)^{\beta}$, is considered. In Sections 2 and 3, the inadmissibility of the best invariant estimator is proved in the following cases, extending the result of Brown (1988):

1. $n > 0$, $-1 < \alpha$, $\beta \leq 0$ and $-1 \leq \alpha + \beta$.
2. $n > 0$, $-1 < \alpha = \beta \leq -\frac{1}{2}$.
3. $n > 1$, (a) $\alpha = -1$ and $\beta = 0$ or (b) $\alpha = 0$ and $\beta = -1$.
4. $n > 2$, $\alpha = \beta = -1$.

We conjecture that the best invariant estimator is inadmissible for:

(i) $n \geq 1$ and $-1 < \alpha$, $\beta \leq 0$.
(ii) (a) $n \geq 2$, $\alpha = -1$ and $-1 < \beta \leq 0$ or (b) $n \geq 2$, $-1 < \alpha \leq 0$ and $\beta = -1$.

In Section 2, we prove the inadmissibility of the empirical c.d.f. for $n > 2$. The estimator d_Q we used to improve on the empirical c.d.f. is displayed in (2.3.0) and explained in Remark 2.2. Also, we give the improved estimator d_1 [see (2.3.9)] for the best invariant estimator in case 3(a) above. In Section 3, we prove the inadmissibility results for cases 1 and 2 above. The improved estimator for the best invariant estimator is basically Brown's estimator d_B [see (2.1.4)]. In Section 4, we give a brief discussion of the estimators d_Q, d_1 and d_B.

2. Inadmissibility of the empirical distribution function.

2.1. *Notation and remarks.* Let $\Theta = \{F\colon F$ is a continuous distribution function on $R^1\}$ denote the parameter space and $X_1, \dots, X_n$ be a sample of size n from F in Θ. Let

$$A = \{a(t);\ a(t) \text{ is a nondecreasing function from } R^1 \text{ into } [0,1]\} \tag{2.1.1}$$

denote the action space. Let $L(F, a)$ be the loss function, where

$$L(F, a) = \int (F(t) - a(t))^2 h(F(t))\, dF(t), \qquad h(t) = t^{\alpha}(1-t)^{\beta},\ \alpha, \beta \ge -1. \tag{2.1.2}$$

Then the decision problem (Θ, A, L), with observations $X_1, \dots, X_n$, is invariant under monotone transformations. The best invariant estimator is

$$d_0(X, t) = \left[\alpha + 1 + \sum_{i=1}^{n} 1(X_i \le t)\right] \Big/ (n + 2 + \alpha + \beta), \tag{2.1.3}$$

with constant risk $R(F, d_0) = 1/(n + 2 + \alpha + \beta)$. When $\alpha = \beta = 0$, the best invariant estimator is $d_0(t) = [1 + \sum_{i=1}^{n} 1(X_i \le t)]/(n+2)$. Brown (1988) constructed an estimator

$$d_B(t) = d_0(t) + \sum_{i=1}^{n} \xi x_i(t)/[2(n+1)(n+2)], \tag{2.1.4}$$

where

$$\xi_x(t) = \begin{cases} 1, & \text{if } x \le 0 < t, \\ -1, & \text{if } t \le 0 < x, \\ 0, & \text{otherwise,} \end{cases} \tag{2.1.5}$$

and used it to improve on $d_0(t)$.

REMARK 2.1. Note that Brown's estimator has the form

$$d_B(t) = \left[\alpha + 1 + \sum_{i=1}^{n} 1(X_i \le t)\right] \Big/ (n + 2 + \alpha + \beta) + 2c \sum_{i=1}^{n} \xi x_i(t). \tag{2.1.6}$$

There is another equivalent expression for $d_B(t)$. Let $Y_0 = -\infty$, $Y_{n+1} = +\infty$ and $Y_1, \dots, Y_n$ be the order statistics of $X_1, \dots, X_n$. Given a fixed point s, say

$s = 0$ here, let

$$I = \max\{j \geq 0; Y_j \leq s\}, \tag{2.1.7}$$

i.e., $I + 1$ is the rank of s among $s, Y_1, \ldots, Y_n$.

Define $Y_0^I, \ldots, Y_{n+2}^I$ to be the order statistics of $(Y, -\infty, +\infty, s)$, i.e.,

$$Y_j^I = \begin{cases} Y_j, & \text{if } 0 \leq j \leq I, \\ s, & \text{if } j = I + 1, \\ Y_{j-1}, & \text{if } I + 1 < j \leq n + 2. \end{cases} \tag{2.1.8}$$

Then Brown's estimator can be expressed as

$$d_B(t) = \sum_{j=0}^{n+1} a_{Ij} 1\left(Y_j^I \leq t < Y_{j+1}^I\right), \tag{2.1.9}$$

where

$$a_{Ij} = \begin{cases} (j + \alpha + 1)/(n + 2 + \alpha + \beta) - c(n - I), \\ \qquad \text{if } 0 \leq j \leq I, \\ (j + \alpha)/(n + 2 + \alpha + \beta) + cI, \\ \qquad \text{if } I < j \leq n + 1. \end{cases} \tag{2.1.10}$$

Note that both Y_j^I and a_{Ij} are functions of I. In order to guarantee that $d_B(t)$ is a nondecreasing function of t, we need $a_{Ij} \leq a_{Ij+1}$, $j = 0, 1, \ldots, n$. In particular, when $I = j = i$, we have $i/n - c(n - i) \leq i/n + ci$, that is, $c \geq 0$.

REMARK 2.2. Brown's estimator does not improve on $d_0(t)$ when $\alpha = -1$ or $\beta = -1$. In particular, when $\alpha = \beta = -1$, it does not improve on the empirical distribution function. For example, when $\alpha = -1$, if there is an estimator d_B improving on d_0, then $c \neq 0$. Otherwise, they are identical [see (2.1.6)]. Let $I = 0$ and $t < Y_1^0$ $(= s)$. By (2.1.8) through (2.1.10), $d_B(t) = a_{00} 1(Y_0^0 \leq t < Y_1^0) = -cn$. By (2.1.1), $d_B(t) = -cn \geq 0$. Since $c \geq 0$ (see the end of Remark 2.1), it follows that $c = 0$, a contradiction.

As we can see, the best invariant estimator gives mass $1/(n + 2 + \alpha + \beta)$ to each of the observations and gives mass $(\alpha + 1)/(n + 2 + \alpha + \beta)$ and $(\beta + 1)/(n + 2 + \alpha + \beta)$ to $-\infty$ and $+\infty$, respectively. So, when $\alpha > -1$ and $\beta > -1$, Brown's estimator shrinks the best invariant estimator as follows. For each negative observation, it moves some positive mass from $+\infty$ or 0; for each positive observation, it moves the same amount of mass from $-\infty$ to 0. This is equivalent to assuming that there is a pseudoobservation at 0.

We can explain why Brown's estimator does not work when $\alpha = \beta = -1$ from this point of view. When $\alpha = \beta = -1$, the best invariant estimator is the empirical c.d.f., which gives no mass to $-\infty$ and $+\infty$. However, Brown's estimator still tries to move some mass from $-\infty$ or $+\infty$ to 0. Because of this, it is not a proper estimator of a distribution function ($\lim_{t \to -\infty} d(t) < 0$ or

$\lim_{t \to +\infty} d(t) > 1$), especially when all the observations have the same sign. If an estimator can improve on the empirical c.d.f. (by shrinking the latter in a manner similar to Brown's estimator), it is reasonable to believe that the shrinking should be done only when the pseudoobservation 0 is between the true observations, and we should consider shifting some mass from certain observations to 0 or to some other observations.

The improved estimator d_Q, proposed in (2.3.0), modifies the empirical c.d.f. as follows. The original observations $X_1, \dots, X_n$ are augmented with a pseudoobservation at 0 and the order statistics $Y_1^I, \dots, Y_{n+1}^I$ are formed. If I observations are negative, $0 < I < n$, then mass c_I is shifted down from Y_{n+1}^I to Y_n^I (toward 0) and mass c_{n-I} is shifted up from Y_1^I to Y_2^I (again toward 0). Values of $c_1, \dots, c_{n-1}$ are given in (2.3.2) or (2.3.3). Note that the estimator is always a distribution function.

This is certainly not the only improved estimator. For instance, one can easily construct different estimators that improve on d_0 when $n = 3$. The idea is to reassign weights to the order statistics $Y_1^I, \dots, Y_{n+1}^I$. Note that the weights might change if the rank $I + 1$ of 0 (among $Y_1^I, \dots, Y_{n+1}^I$) changes. However, if the sample size n is large, the complexity in adjusting the $(n+1) \times (n+2)$ [see (2.2.3)] variables makes the process extremely difficult. Finding a sample form that works and determining the constants $c_1, \dots, c_{n-1}$ are the points considered in the next section.

2.2. *A new class of estimators.* Consider a new class U of estimators which have the form

$$d(Y, t) = \sum_{j=0}^{n+1} a_{Ij} 1\{Y_j^I \le t < Y_{j+1}^I\}, \tag{2.2.1}$$

where

$$a_{Ij} = \begin{cases} u_j + c_{ij}, & \text{if } 0 \le j \le I = i, \\ u_{j-1} + c_{ij}, & \text{if } I = i < j \le n+1, \end{cases} \qquad u_j = \frac{\alpha + 1 + j}{n + 2 + \alpha + \beta}, \tag{2.2.2}$$

$j, i = 0, \dots, n$, and $a_{ij} \le a_{ij+1}$, for all possible i and j [so that $d(t)$ is really an estimator]. An alternative form of (2.2.1) is

$$d(Y, t) = \sum_{i=0}^{n} \sum_{j=0}^{n+1} a_{ij} 1\{Y_j^i \le t < Y_{j+1}^i, I = i\}. \tag{2.2.3}$$

Abusing notation, we will identify $d(Y, t)$ with the $(n+1) \times (n+2)$ matrix (a_{ij}), say, $d = d(Y, t) = (a_{ij})$.

We first need to compute the risk function of $d(Y, t)$.

LEMMA 2.1. *Assume $d = d(Y, t)$ has the form* (2.2.1) *and* $p = F(0)$. *Then*

$$R(F,d) = \sum_{i=0}^{n}\sum_{j=0}^{i}\int_0^p (t-a_{ij})^2 h(t)\binom{n}{j}\binom{n-j}{i-j} t^j (p-t)^{i-j}(1-p)^{n-j}\,dt$$

$$+\sum_{i=0}^{n}\sum_{j>i}^{n+1}\int_p^1 (t-a_{ij})^2 h(t)\binom{n}{j-1}\binom{j-1}{i}(1-t)^{n-j+1} \tag{2.2.4}$$

$$\times (t-p)^{j-1-i}p^i\,dt.$$

For proof, see Yu (1986).

2.3. *The main results.* For the cases $n = 1, 2$, it turns out that no estimators in U can improve on $\hat{F}(t)$. In fact, $\hat{F}$ is admissible when $n = 1, 2$ [Yu (1986)]. However, when $n \geq 3$, there are estimators in U that can improve on $\hat{F}(t)$.

THEOREM 2.2. *If $n \geq 3$ and $h(t) = t^{-1}(1-t)^{-1}$, then the empirical distribution function is not admissible. Indeed, the estimator d_Q is better than $\hat{F}(t)$, where*

$$d_Q = \sum_{j=0}^{n+1} a_{Ij} 1\{Y_j^I \leq t < Y_{j+1}^I\}. \tag{2.3.0}$$

Here

$$(a_{ij}) = \begin{pmatrix} 0 & 0 & \frac{1}{n} & \cdots & \frac{n-1}{n} & 1 \\ 0 & \frac{1}{n} - 2c_{n-1} & \frac{1}{n} & \cdots & \frac{n-1}{n} + 2c_1 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \frac{1}{n} - 2c_1 & \frac{2}{n} & \cdots & \frac{n-1}{n} + 2c_{n-1} & 1 \\ 0 & \frac{1}{n} & \frac{2}{n} & \cdots & 1 & 1 \end{pmatrix}, \tag{2.3.1}$$

$$c_{ij} = \begin{cases} 2c_i, & \textit{if } j = n,\ i = 1,\dots,n-1, \\ -2c_{n-i}, & \textit{if } j = 1,\ i = 1,\dots,n-1 \\ & [\textit{see } (2.2.2)],\ i = 0,\dots,n,\ j = 0,\dots,n+1, \\ 0, & \textit{otherwise}, \end{cases} \tag{2.3.2}$$

$$(c_{n-1},\dots,c_1) = \begin{cases} \varepsilon\left(1, \frac{1}{2}\right), & \textit{if } n = 3 \\ \varepsilon\left(\frac{1}{2}, \frac{1}{6}, \frac{1}{2}\right), & \textit{if } n = 4 \end{cases} \quad \textit{for a very small } \varepsilon > 0$$

and

$$c_k = \frac{\varepsilon k(n-k)[(n-k)(n-4)+n]}{(n-1)(n-2)}, \qquad k = 1, \ldots, n-1, \text{ if } n \geq 5. \tag{2.3.3}$$

Note that if we write $d_Q = (a_{ij})$, then $a_{ij} = 1 - a_{n-i, n+1-j}$ for all i, j. This results from the symmetry of the loss function.

PROOF. We first check that $d_Q(t)$ is nondecreasing in t, i.e., that the values in each row of (2.3.1) are nondecreasing. It suffices to check that if ε is small enough, we have $1/n - 2c_i < 2/n$, $i = 1, \ldots, n-2$, and $1/n - 2c_{n-1} < 1/n$ [see (2.3.1)]. The first inequality is true as long as c_i is small enough. The last inequality holds if $c_{n-1} > 0$ which is true by (2.3.2) and (2.3.3). So $d_Q \in U$ [see (2.2.1)].

Using (2.2.2), the difference in the risks $R(F, d_Q) - R(F, \hat{F})$ simplifies greatly because a lot of components in (2.3.1) are the same as those in $\hat{F}(t)$. In fact if $c_1 = \cdots = c_{n-1} = 0$, then $d_Q = \hat{F}$. Thus,

$$\begin{aligned}
R(F, d_Q) &- R(F, \hat{F}) \\
&= \sum_{i=1}^{n-1} \Bigg[\int_0^p -4c_{n-i}\left(t - \frac{1}{n} + c_{n-i}\right)\binom{n}{1}\binom{n-1}{i-1}(1-t)^{-1} \\
&\qquad\qquad \times (p-t)^{i-1}(1-p)^{n-i}\,dt \\
&\qquad + \int_p^1 -4c_i\left(t - \frac{n-1}{n} - c_i\right)\binom{n}{n-1}\binom{n-1}{i} t^{-1}(t-p)^{n-1-i}p^i\,dt \Bigg] \\
&= \sum_{i=1}^{n-1} \Bigg[\int_{1-p}^1 -4c_i\left(t - \frac{n-1}{n} - c_i\right) n \binom{n-1}{i} t^{-1} \\
&\qquad\qquad \times (t-(1-p))^{n-1-i}(1-p)^i\,dt \\
&\qquad + \int_p^1 -4c_i\left(t - \frac{n-1}{n} - c_i\right) n \binom{n-1}{i} t^{-1}(t-p)^{n-1-i}p^i\,dt \Bigg].
\end{aligned}$$

Define

$$\begin{aligned}
T(p) = \sum_{i=1}^{n-1} \Bigg[&\int_{1-p}^1 c_i^2 n \binom{n-1}{i} t^{-1}(t-(1-p))^{n-1-i}(1-p)^i\,dt \\
&+ \int_p^1 c_i^2 n \binom{n-1}{i} t^{-1}(t-p)^{n-1-i}p^i\,dt \Bigg],
\end{aligned} \tag{2.3.4}$$

$$B_i(p) = \int_p^1 n\left(t - \frac{n-1}{n}\right)\binom{n-1}{i} t^{-1}(t-p)^{n-1-i}p^i\,dt, \tag{2.3.5}$$

$$A(p) = \sum_{i=1}^{n-1} c_i[B_i(p) + B_i(1-p)]. \tag{2.3.6}$$

Then $R(F, d_Q) - R(F, \hat{F}) = -4A(p) + 4T(p)$. Note that $T(p) > 0$ for all $p \in (0,1)$, so we need to show that $A(p) > 0$ for all $p \in (0,1)$. For simplicity, we just verify the case $n \geq 5$. The proofs for the cases $n = 3$ and 4 are similar. It can be shown (see the Appendix) that

$$A(p) = \varepsilon p(1-p). \tag{2.3.7}$$

Now by (2.3.7) and (2.3.4),

$$\lim_{p \to 0^+} \frac{A(p)}{T(p)} = \varepsilon \Big/ \left[n \left(c_1^2 \frac{n-1}{n-2} + c_{n-1}^2 \right) \right] = \frac{1}{\varepsilon n[(n-2)(n-1)+4]} > 0.$$

Similarly, $\lim_{p \to 1^-} A(p)/T(p) > 0$. So we can extend $A(p)/T(p)$ to be continuous and positive on the closed interval $[0,1]$. Thus, given $\varepsilon > 0$, we can find a $\delta > 0$ such that $A(p)/T(p) > \delta$ for all $p \in [0,1]$. By (2.3.2), (2.3.4) and (2.3.7), $A(p) = O(\varepsilon)$ and $T(p) = O(\varepsilon^2)$, so we can find an ε small enough such that $A(p)/T(p) > 1$ for all $p \in [0,1]$, that is,

$$R(F, d_Q) - R(F, \hat{F}) = -4A(p) + 4T(p) \begin{cases} < 0, & \text{if } p \in (0,1), \\ \leq 0, & \text{otherwise.} \end{cases} \qquad \square$$

The proof of Theorem 2.2 only shows the existence of an $\varepsilon > 0$ such that the estimator d_Q improves on $\hat{F}(t)$. The following is an example of d_Q which improves on $\hat{F}(t)$ when $n = 3$.

EXAMPLE 2.3. When $n = 3$, let $\varepsilon = \frac{1}{4}$. Then

$$d_Q = \begin{pmatrix} 0 & 0 & \frac{1}{3} & \frac{2}{3} & 1 \\ 0 & \frac{1}{3} - \frac{1}{6} & \frac{1}{3} & \frac{2}{3} + \frac{1}{12} & 1 \\ 0 & \frac{1}{3} - \frac{1}{12} & \frac{2}{3} & \frac{2}{3} + \frac{1}{6} & 1 \\ 0 & \frac{1}{3} & \frac{2}{3} & 1 & 1 \end{pmatrix} \tag{2.3.8}$$

and

$$\hat{F} = d_0 = \begin{pmatrix} 0 & 0 & \frac{1}{3} & \frac{2}{3} & 1 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{2}{3} & 1 \\ 0 & \frac{1}{3} & \frac{2}{3} & \frac{2}{3} & 1 \\ 0 & \frac{1}{3} & \frac{2}{3} & 1 & 1 \end{pmatrix}.$$

Let $D(p) = R(F, d_Q) - R(F, \hat{F})$, where $p = F(0)$. Then, by Lemma 2.1 and tedious computation, we have

$$D(p) = \left(-\tfrac{1}{12}\right)(p - p^2) + \left(\tfrac{1}{24}\right)\left[-p^2 \log p - (1-p)^2 \log(1-p)\right].$$

Note that if $p \in (0,1)$:

(i) $(d^3/dp^3)D(p) = (2p-1)/[12p(1-p)] = 0$ only at $p = \frac{1}{2}$.
(ii) $D(p) = D(1-p)$.
(iii) $D(p) < 0$ near 0 or 1.
(iv) $D(\frac{1}{2}) = (-1 + \log 2)/48 < 0$.

So

$$R(F, d_Q) - R(F, \hat{F})\begin{cases} < 0, & \text{if } F(0) = \frac{1}{2}, \\ \leq 0, & \text{otherwise.} \end{cases}$$

It is clear that the largest improvement is at $p = \frac{1}{2}$ and the percentage improvement is about 1.3%.

Using similar methods, we can prove other inadmissibility results. The following theorem, whose proof is put in the Appendix, is one such result.

THEOREM 2.4. *If $n \geq 2$ and $h(t) = t^{-1}$, then the best invariant estimator d_0 is inadmissible. Furthermore d_0 can be improved by d_1, where*

$$(2.3.9) \qquad d_1 = \begin{pmatrix} 0 & 0 & \frac{1}{n+1} & \cdots & \frac{n-1}{n+1} & \frac{n}{n+1} \\ 0 & \frac{1}{n+1} + 2c_3 & \frac{1}{n+1} & \cdots & \frac{n-1}{n+1} & \frac{n}{n+1} + 2c_1 \\ 0 & \frac{1}{n+1} & \frac{2}{n+1} & \cdots & \frac{n-1}{n+1} & \frac{n}{n+1} + 2c_2 \\ 0 & \frac{1}{n+1} & \frac{2}{n+1} & \cdots & \frac{n-1}{n+1} & \frac{n}{n+1} \\ 0 & \frac{1}{n+1} & \frac{2}{n+1} & \cdots & \frac{n}{n+1} & \frac{n}{n+1} \end{pmatrix}.$$

Here $c_1 = c/n$, $c_2 = 2c/[n(n-1)]$, $c_3 = -2c/[n^2(n-1)]$ and c is a small positive number.

REMARK 2.3. The improved estimator d_1 [see (2.3.9)] can be interpreted as follows. If there is only one negative observation, then a small amount of mass $(2c/n)$ is moved from $+\infty$ to the largest observation and a small amount of mass $(4c/[n^2(n-1)])$ is moved from the smallest observation to 0. If there are only two negative observations, mass $4c/[n(n-1)]$ is moved from $+\infty$ to the largest observation. In the other cases the estimator remains the same as the best invariant estimator. Note that now the best invariant estimator gives mass $1/(n+1)$ to $+\infty$ and 0 to $-\infty$.

COROLLARY 2.5. *If $n \geq 2$ and $h(t) = (1-t)^{-1}$, then the best invariant estimator is not admissible.*

REMARK 2.4. When $n = 1$ and $\alpha = -1$ or $\beta = -1$, there is no estimator d in U such that d is better than d_0. In fact d_0 is admissible in these situations [Yu (1986)].

3. Some extensions of Brown's result. In Section 2 and Brown's paper, the inadmissibility of the best invariant estimator in the two most important cases, i.e., $\alpha = \beta = 0$ or -1, have been considered. Of interest also is the inadmissibility problem for general α and β. Note that Brown's estimator looks much simpler than the estimators we used in Section 2. Since, if $\alpha > -1$ and $\beta > -1$, the best invariant estimator assigns positive weight to $-\infty$ and $+\infty$, it is natural to raise the question: Is it possible to use Brown's estimator d_B [see (2.1.6)] to improve on d_0 for general α and β? We have some positive answers to this.

For convenience, in this section, we assume

$$h(t) = t^{\alpha}(1-t)^{\beta}, \qquad \alpha, \beta > -1. \tag{3.1}$$

The difference between the risks of $d_B(t)$ and $d_0(t)$ has a very nice form as the following lemma shows; its proof is tedious but not difficult. For proofs of the following lemmas and theorems, see Yu (1986).

LEMMA 3.1. *Under the above assumptions and notation,*

$$\begin{aligned} &R(F, d_0) - R(F, d_B) \\ &\quad = \frac{4nc}{n+2+\alpha+\beta} A(p, \alpha, \beta) - 4nc^2 T(p, \alpha, \beta), \end{aligned} \tag{3.2}$$

where $p = F(0)$,

$$\begin{aligned} A(p, \alpha, \beta) = {}& [p(\alpha+1)/(\alpha+\beta+2)] \int_0^1 h(t)\,dt - p\int_p^1 h(t)\,dt \\ & -(\alpha+\beta+3)\int_0^p th(t)\,dt + (\alpha+1)\int_0^p h(t)\,dt \end{aligned} \tag{3.3}$$

and

$$\begin{aligned} T(p, \alpha, \beta) = {}& \left[p + (n-1)p^2\right] \int_p^1 h(t)\,dt \\ & + \left[(1-p) + (n-1)(1-p)^2\right] \int_0^p h(t)\,dt. \end{aligned} \tag{3.4}$$

The next lemma provides a convenient tool to judge the inadmissibility of d_0.

LEMMA 3.2. *If* $-1 < \alpha$, $\beta \le 0$ *and* $A(p, \alpha, \beta) > 0$ [*see* (3.3)] *for all* $p \in (0,1)$, *then* d_0 *is inadmissible.*

By verifying the sufficient condition of Lemma 3.2, we have the following inadmissibility results.

THEOREM 3.3. *Suppose that* $n \ge 1$ *and* $-1 < \alpha$, $\beta \le 0$. *Then* d_0 *is inadmissible if either* (i) $-1 \le \alpha + \beta$ *or* (ii) $-1 < \alpha = \beta$.

REMARK 3.1. It may be that d_B improves on d_0 for all $\alpha, \beta \in (-1, 0]$. At least it is most likely that the best invariant estimator d_0 is inadmissible in the above cases.

REMARK 3.2. One might wonder whether Brown's estimator d_B works for $\alpha, \beta > 0$. But unfortunately the answer is "no." This can be seen as follows.

First note that c in (2.1.6) is nonnegative (see the end of Remark 2.1). Assume that $\alpha > 0$ and $\beta > -1$. For convenience, write $A(p) = A(p, \alpha, \beta)$ [see (3.3)]. It can be shown that

$$\lim_{p \to 0^+} A'(p) = \int_0^1 t^{\alpha+1}(1-t)^\beta \, dt - \int_0^1 t^\alpha (1-t)^\beta \, dt < 0.$$

Note that $A(0) = 0$, so $A(p) < 0$ for p near 0^+. For p near 0^+, Lemma 3.1 yields for $c > 0$,

$$R(F, d_0) - R(F, d_B) = \frac{4nc}{n + 2 + \alpha + \beta} A(p, \alpha, \beta) - 4nc^2 T(p, \alpha, \beta) < 0,$$

since $T(p, \alpha, \beta) \geq 0 \ \forall \ p \in (0, 1)$. Therefore, d_B cannot improve on d_0.

4. Summary. The estimators d_Q [see (2.3.0)], d_1 [see (2.3.9)] and d_B [see (2.1.4)], improving on d_0 in the cases $\alpha = \beta = -1$, $(\alpha, \beta) = (-1, 0)$ and $\alpha = \beta = 0$, respectively, coincide with our intuition. The unknown distribution function F is continuous, whereas d_0, being a step function, is discontinuous. If d_0 is inadmissible, then it is expected to be improved by an estimator which is somewhat smoother than d_0. Each of the three estimators above considers 0 as a pseudoobservation and readjusts the weights so that the new estimator becomes smoother. Each of them shrinks d_0 in a similar manner as follows. Given α and β (note that d_0 is essentially a function of α and β), let

$$S = \{e \in \{-\infty, +\infty, X_1, \dots, X_n\};\ d_0 \text{ assigns positive weight to } e\}.$$

If $\inf S < 0 < \sup S$, some weight is shifted from $\inf S$ or $\sup S$ toward 0, though not necessarily to 0 itself.

The difference between Brown's estimator d_B and the type of estimators d_Q and d_1 proposed in this paper is as follows.

1. Brown's estimator shifts weight to 0 only, whereas d_Q and d_1 shift weight to an observation among $(0, X_1, \dots, X_n)$. This observation may or may not be 0 [depending on I; see (2.1.7)].
2. Brown's estimator shifts weight only from $-\infty$ and $+\infty$, whereas d_Q and d_1 shift weight from Y_1 or Y_n also.

In general, there is little doubt that if d_0 does not assign positive mass to $+\infty$ and $-\infty$ simultaneously and if it is not admissible, then it can be improved only by estimators similar to d_Q or d_1. It is also likely that if d_0 does assign positive mass to $+\infty$ and $-\infty$ simultaneously, then d_0 is inadmissible iff an estimator of Brown's type improves on d_0.

APPENDIX

PROOF OF (2.3.7). It follows from (2.3.5) that

$$
\begin{aligned}
B_i(p)\Big/\left[n\binom{n-1}{i}\right]
&= \left[(-1)^{n-1-i}\frac{n-1}{n}\right]p^{n-1}\ln p + \left[\frac{(-1)^{n-i}}{n-i}\right]p^n \\
&\quad + \left[(-1)^{n-i-1} + \sum_{j=1}^{n-1-i}(-1)^{n-1-i-j}\binom{n-1-i}{j}\frac{n-1}{nj}\right]p^{n-1} \\
&\quad + \sum_{k=0}^{n-2-i}(-1)^k\binom{n-1-i}{k}\left(\frac{1}{n-i-k} - \frac{n-1}{n(n-1-i-k)}\right)p^{k+i},
\end{aligned}
$$

$i = 1, \ldots, n-1$. Thus [where c_i is as in (2.3.3)],

$$
\begin{aligned}
\sum_{i=1}^{n-1} c_i B_i(p) &= p^{n-1}\ln p \sum_{i=1}^{n-1} c_i(-1)^{n-1-i} n\binom{n-1}{i}\frac{n-1}{n} \\
&\quad + p^n \sum_{i=1}^{n-1} c_i n\binom{n-1}{i}\frac{(-1)^{n-i}}{n-i} \\
&\quad + p^{n-1}\sum_{i=1}^{n-1} c_i n\binom{n-1}{i}(-1)^{n-1-i} \\
&\qquad \times\left\{1 + \sum_{j=1}^{n-1-i}\binom{n-1-i}{j}(-1)^j\frac{n-1}{nj}\right\} \qquad \text{(A.1)} \\
&\quad + \sum_{m=1}^{n-2}\sum_{i=1}^{m} c_i n\binom{n-1}{i}(-1)^{m-i}\binom{n-1-i}{m-i} \\
&\qquad \times\left(\frac{1}{(n-m)} - \frac{n-1}{n(n-1-m)}\right)p^m \\
&= \varepsilon(-p + 4p^2 - 3p^3) \\
&= -\varepsilon p(1-p)(1-3p).
\end{aligned}
$$

By (2.3.6) and the above expression,

$$
\begin{aligned}
A(p) &= -\varepsilon p(1-p)(1-3p) - \varepsilon p(1-p)(1-3+3p) \qquad (2.3.7)\\
&= \varepsilon p(1-p).
\end{aligned}
$$

□

PROOF OF THEOREM 2.4. It is easy to check that if c [see (2.3.9)] is small enough, the values in each row of (2.3.9) are increasing. Thus $d_1 \in U$. Using Lemma 2.1, it can be shown that

$$R(F, d_1) - R(F, d_0) = -4A(p) + 4T(p),$$

where $A(p) = cp(1-p)^{n-1}/[n(n-1)(n+1)]$ and

$$T(p) = \int_0^p nc_3^2\, dt(1-p)^{n-1} + \int_p^1 nc_1^2 p(t-p)^{n-1} t^{-1}\, dt$$
$$+ \int_p^1 \tfrac{1}{2} n(n-1) c_2^2 p^2 (t-p)^{n-2} t^{-1}\, dt.$$

Thus

$$\lim_{p \to 0^+} \frac{A(p)}{T(p)} = c \Bigg/ \left[n(n^2-1) \left(\frac{4c^2}{n^3(n-1)^2} + \frac{c^2}{n(n-1)} \right) \right] > 0.$$

Also

$$\lim_{p \to 1^-} \frac{A(p)}{T(p)} = \frac{c/[n(n^2-1)]}{[c_3^2 \cdot n + c_2^2 \cdot (n(n-1))/2 + c_1^2 \cdot n]} > 0.$$

Thus we can extend $A(p)/T(p)$ to be continuous and positive on $[0,1]$. Hence we can find c small enough such that $A(p)/T(p) \geq 1$ on $[0,1]$. It follows that

$$R(F, d_1) - R(F, d_0) = -4A(p) + 4T(p) \begin{cases} < 0, & \text{if } F(0) \neq 0 \text{ or } 1, \\ \leq 0, & \text{otherwise.} \end{cases} \qquad \square$$

Acknowledgments. I am greatly indebted to my advisor T. S. Ferguson for all his help and suggestions in writing this paper, which is a part of my Ph.D. dissertation. Thanks also go to an Associate Editor and referees for their invaluable suggestions and comments. Furthermore, the help from Professor E. G. Phadia and the Associate Editor in polishing the final version of the paper is acknowledged.

REFERENCES

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

BROWN, L. D. (1988). Admissibility in discrete and continuous invariant nonparametric estimation problems and in their multinomial analogs. *Ann. Statist.* **16** 1567–1593.

COHEN, M. P. and KUO, L. (1985). The admissibility of the empirical distribution function. *Ann. Statist.* **13** 262–271.

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.

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

PHADIA, E. G. (1973). Minimax estimator 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.

YU, Q. (1986). Admissibility of the best invariant estimator of a distribution function. Ph.D. dissertation, Univ. California, Los Angeles.

BETSY ROSS COURT, APT. 14-D
BOUND BROOK, NEW JERSEY 08805