Admissibility in Discrete and Continuous Invariant Nonparametric Estimation Problems and in their Multinomial Analogs
Author(s): Lawrence D. Brown
Source: *The Annals of Statistics*, Vol. 16, No. 4 (Dec., 1988), pp. 1567-1593
Published by: Institute of Mathematical Statistics
Stable URL: http://www.jstor.org/stable/2241780
Accessed: 25/03/2010 15:34

The Annals of Statistics
1988, Vol. 16, No. 4, 1567–1593

ADMISSIBILITY IN DISCRETE AND CONTINUOUS INVARIANT NONPARAMETRIC ESTIMATION PROBLEMS AND IN THEIR MULTINOMIAL ANALOGS[1]

BY LAWRENCE D. BROWN

Cornell University and Hebrew University

Discrete and multinomial analogs are defined for classical (continuous) invariant nonparametric problems of estimating the sample cumulative distribution function (sample c.d.f.) and the sample median. Admissibility of classical estimators and their analogs is investigated. In discrete (including multinomial) settings the sample c.d.f. is shown to be an admissible estimator of the population c.d.f. under the invariant weighted Cramér–von Mises loss function

$$L_1(F,\hat{F}) = \int\left[\left(F(t)-\hat{F}(t)\right)^2/\left(F(t)(1-F(t))\right)\right]\,dF(t).$$

Ordinary Cramér–von Mises loss—$L_2(F,\hat{F}) = \int[(F(t)-\hat{F}(t))^2]\,dF(t)$—is also studied. Admissibility of the best invariant estimator is investigated. (It is well known in the classical problem that the sample c.d.f. is not the best invariant estimator, and hence is not admissible.) In most discrete settings this estimator must be modified in an obvious fashion to take into account the end points of the known domain of definition for the sample c.d.f. When this is done the resulting estimator is shown to be admissible in some of the discrete settings. However, in the classical continuous setting and in other discrete settings, the best invariant estimator, or its modification, is shown to be inadmissible.

Kolmogorov–Smirnov loss for estimating the population c.d.f. is also investigated, but definitive admissibility results are obtained only for discrete problems with sample size 1. In discrete settings the sample median is an admissible estimator of the population median under invariant loss. In the continuous setting this is not true for even sample sizes.

1. Introduction. Very little is currently known about finite sample size decision–theoretic properties in invariant nonparametric estimation problems.

For reasons of aesthetics and convenience the usual formulation of these problems involves observation of independent identically distributed real observations having an unknown *continuous* cumulative distribution function (c.d.f.). However, it is also possible, and of interest, to study the formulation in which the unknown c.d.f. is assumed to correspond to an unknown *discrete* distribution. It is possible to simplify the problem further by assuming the unknown discrete distribution is *multinomial* on a *given* finite set of points E. Such a formulation is, of course, no longer invariant since the set E is given, and, hence, is not invariant under monotone transformations. However, in other

Received March 1984; revised March 1988.

[1]Research supported in part by NSF Grants MCS-82-00031 and DMS-85-06847.

AMS 1980 *subject classifications.* Primary 62C15; secondary 62D05.

Key words and phrases. Admissibility, nonparametric estimation, sample distribution function, sample median, multinomial distribution, Cramér–von Mises loss, Kolmogorov–Smirnov loss.

respects these multinomial problems share the flavor of their nonparametric progenitors.

There have been recent advances in decision-theoretic methodology for investigating admissibility in multinomial and other discrete problems. Brown (1981) generalizes techniques used earlier in specific settings by Johnson (1971) and Alam (1979), and these techniques have been used recently in Ighodaro, Santner and Brown (1982), Cohen and Kuo (1985) and Meeden, Ghosh and Vardeman (1985).

The latter two references are particularly relevant here since they treat multinomial versions of nonparametric problems. However, these versions, while of interest on their own merits, are not true analogs of the usual nonparametric problems since the loss functions used are discrete analogs of noninvariant loss functions, instead of being the discrete analogs of the invariant loss functions of classical nonparametric estimation problems. [A similar comment is relevant to Phadia (1972) which proves minimaxity of the sample c.d.f. relative to a noninvariant loss function.]

There are two main objectives of the current study. The first is to carefully formulate discrete and multinomial analogs of classical invariant nonparametric estimation problems. To the extent possible these formulations should preserve the flavor of the original classical formulations. (Hopefully without the use of artificial preservatives!) This is desirable for aesthetic reasons, and possibly also practical ones, as well as in the hope that decision-theoretic results in appropriately formulated discrete problems will transfer easily to the classical continuous problems.

The second objective is to investigate decision-theoretic results—primarily those concerning the fundamental property of admissibility—in these discrete formulations, and also in the classical, continuous formulation. As hoped, it has been possible to derive several admissibility results in the discrete formulations, and also a few in the classical formulation. In contrast to our original expectation it turns out that results in discrete and continuous settings may easily be different.

We can point to three major conclusions of this study:

I. Admissibility of the sample c.d.f. as an estimator of the population c.d.f. in discrete problems involving scaled Cramér–von Mises loss [L_1, as defined in (2.3.1)].

II. Inadmissibility of the best invariant estimator of the population c.d.f. in continuous problems involving ordinary Cramér–von Mises loss [L_2, defined in (2.3.2)]. This inadmissibility extends to some discrete reformulations of this problem but does not hold in others.

III. Several other admissibility and inadmissibility results and a number of open questions. One of the more interesting of these results is the admissibility of the (or, any) sample median as an estimator of the population median in discrete problems under a simple invariant loss L_4 [defined in (2.3.4)]. On the other hand, when the sample size is even any invariant sample median is inadmissible in continuous problems. Among the more interesting questions left open here are the admissibility of the sample c.d.f. in continuous problems using scaled

Cramér–von Mises loss referred to in I, a variety of questions involving admissibility under Kolmogorov–Smirnov loss, and the admissibility of the sample median in continuous problems having odd sample size. The first of these questions has recently been given a surprising answer in Yu (1986, 1987).

2. Formulation of the problem.

2.1. *Sample space and distributions.* The conventional formulation of a nonparametric estimation problem begins with a sample space corresponding to n independent identically distributed real random variables, $X_1, \ldots, X_n$, $n \geq 1$, on $I = (a, b) \subseteq (-\infty, \infty)$. Conventionally, it is assumed that each X_i has a continuous cumulative distribution function (c.d.f.) F about which nothing else is known. Thus, $I \subset (-\infty, \infty)$ is a specified interval and the space of possible distributions $\mathscr{F}$ for each X_i is

$$\mathscr{F}_C = \mathscr{F}_C(I) = \{F\colon F \text{ is a continuous c.d.f. on } I\}. \tag{2.1.1}$$

The problems to be considered are invariant under monotone, strictly increasing, transformations of the interval (a, b) onto their range. Hence, a problem with $\mathscr{F} = \mathscr{F}_C((a, b))$ is equivalent to one with $\mathscr{F} = \mathscr{F}_C((-\infty, \infty))$. Obviously, here the interval $[a, b]$, $-\infty < a < b < \infty$, may be substituted for (a, b) without changing the problem or the nature of the results.

A major focus of this study is on discrete reformulations of the preceding situation. One such reformulation involves specifying a set $E = \{\xi_i\colon i = 1, \ldots, m\} \subset (-\infty, \infty)$ and considering

$$\mathscr{F}_M = \mathscr{F}_M(E) = \{F\colon F \text{ is a discrete c.d.f. supported on } E\}. \tag{2.1.2}$$

Without loss of generality we assume $\xi_1 < \xi_2 < \cdots < \xi_m$.

In (2.1.2), the support set E of the multinomial distribution $F \in \mathscr{F}_M$ is specified in advance. A different discrete formulation involves the assumption that F be multinomial, with an unknown support contained in a specified interval $I \subset (-\infty, \infty)$. Then the set of possible distributions is

$$\mathscr{F}_D(I) = \{F\colon F \text{ is a discrete c.d.f. with finite support contained in } I\}. \tag{2.1.3}$$

As before, a formulation with distributions $\mathscr{F}_D((-\infty, \infty))$ is equivalent to one with distributions $\mathscr{F}_D((0,1))$. However, the formulation with distributions $\mathscr{F}_D([0,1])$ is not equivalent to that with distributions $\mathscr{F}_D((0,1))$. This (annoying!) technical fact must be kept in mind in handling nonparametric situations. A surprising instance of this nonequivalence is presented in Examples 4.1.4 and 4.1.5.

The remainder of the paper concentrates on formulations involving the sets of distributions described previously. It should be clear that there are other, nonequivalent formulations which may sometimes be of interest. For example, it may be that $\mathscr{F}$ is the subset of $\mathscr{F}_D$ consisting of distributions supported on at most k points, with k a known number specified in advance (see Remark 4.1.3); or it may be that $\mathscr{F} = \mathscr{F}_C \cup \mathscr{F}_D$, etc. Decision-theoretic results for such alternate formulations can often be easily deduced from corresponding results for the formulations (2.1.1)–(2.1.3), described previously. For this reason we make no further comments about these or other alternate formulations for $\mathscr{F}$ except for a few special remarks.

2.2. *Action space*. We wish to consider two varieties of estimation problems. The first variety involves estimation of the unknown c.d.f. The second involves estimation of an invariant functional of the c.d.f.—to be specific we consider estimation of the median of F.

The appropriate action space for estimation of $F \in \mathscr{F}$ is

$$\mathscr{A}_1 = \{a(\cdot) \ni a\colon R \to [0,1] \text{ and } a \text{ is nondecreasing}\}. \tag{2.2.1}$$

Two special features of this space are worth noting.

REMARK 2.2.1. Every $F \in \mathscr{F}$ is a c.d.f. and hence satisfies

$$\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to \infty} F(x) = 1. \tag{2.2.2}$$

However, $\mathscr{A}_1$ contains estimates $a(\cdot)$ which do not satisfy (2.2.2). Such estimates are often referred to as defective distribution functions. In order to construct a satisfactory theory it is necessary to include defective distribution functions in the action space. This necessity was recognized long ago. [See, for example, Aggarwal (1955).] For the conventional formulation with $\mathscr{F} = \mathscr{F}_C$ and invariant loss functions such as L_2, L_3 (defined later) the best invariant estimator $\mathscr{A}_1$ does not satisfy (2.2.2). It follows that no invariant estimator satisfying (2.2.2) can be admissible, even when one limits consideration to estimators taking only actions in $\mathscr{A}_1$ which do satisfy (2.2.2).

The desirability of allowing actions which do not satisfy (2.2.2) can be understood from another point of view. Since the loss functions to be adopted are bounded, the standard decision-theoretic formulation yields the existence of a minimal complete class if the action space is closed in a suitable topology. See, e.g., Brown (1977). For the problems to be considered, a suitable topology is either the topology of weak convergence of distribution functions or the topology of pointwise convergence. In either case the compactification of functions satisfying (2.2.2) includes also functions not satisfying (2.2.2). (Note that $\mathscr{A}_1$ is compact in these topologies.)

REMARK 2.2.2. The functions $\mathscr{F}$ are right continuous; however, the actions in $\mathscr{A}_1$ need not be right continuous. This freedom of action is not always required. If $a \in \mathscr{A}_1$ let a^ρ denote the right-continuous version of a. Suppose the loss function L satisfies

$$L(F, a) \geq L(F, a^\rho), \tag{2.2.3}$$

for every $F \in \mathscr{F}$, $a \in \mathscr{A}_1$. Then there is no loss of generality in restricting actions to be right continuous. (2.2.3) is satisfied in the conventional nonparametric formulations, which have $\mathscr{F} = \mathscr{F}_C$. However, in some discrete formulations (2.2.3) fails to hold. Some estimators which are not right continuous can then be admissible. See Example 7.1.3. Since we will consider such formulations we assume the action space is not limited to right-continuous functions. The desirability of allowing actions which are not right continuous is particularly clear in connection with the loss functions L_2' or L_2'' defined in Section 7.1. As was the case in Remark 2.2.1 one reason for allowing non-right-continuous actions when (2.2.3) is not satisfied is to guarantee the existence of a minimal complete class.

The second variety of estimation problem to be considered involves estimation of the median of F. For such problems the action space will be simply

$$A_2 = [-\infty, \infty]. \tag{2.2.4}$$

[There would be no difference in our result is we chose instead the more usual space $\mathscr{A}_2' = (-\infty, \infty)$; however, $\mathscr{A}_2$ seems technically preferable since it is compact in the natural topology.]

2.3. *Loss function.* Many loss functions have been proposed for the conventional problem of estimating an unknown continuous c.d.f. We have chosen to investigate the analogs in discrete formulations to three of the most popular of these loss functions.

The first two loss functions are of the Cramér–von Mises type; the first being a scaled version and the second being the standard version, as follows:

$$L_1(F, a) = \int \frac{(F(t) - a(t))^2}{F(t)(1 - F(t))}\, dF(t), \tag{2.3.1}$$

$$L_2(F, a) = \int (F(t) - a(t))^2\, dF(t). \tag{2.3.2}$$

The third loss function is the familiar Kolmogorov–Smirnov loss,

$$L_3(F, a) = \sup_t |F(t) - a(t)|. \tag{2.3.3}$$

We denote the risk function corresponding to L_i by the symbol R_i: thus $R_i(F, \delta) = E_F(L_i(F, \delta(\cdot)))$. When the value of i is clear from the context we write R instead of R_i.

All three of these loss functions are fully invariant under monotone transformations of the interval I when $\mathscr{F}$ is also invariant (i.e., when $\mathscr{F} = \mathscr{F}_C$ or $\mathscr{F}_D$).

All three of these loss functions are well defined in the continuous and discrete formulations to be considered. [In the integrand of (2.3.1) use the obvious convention $0/0 = 0$.] Admissibility results for these losses are given in Sections 3, 4, and 5, respectively. Section 7 contains some results for some natural variants L_2' and L_2'' of L_2. [See (7.1.2) and also (7.1.3).] In discrete formulations the admissibility results for L_2' and L_2'' differ from those for L_2.

For estimating the median we use the loss function

$$L_4(F, a) = \inf\{|b - \tfrac{1}{2}|: F(a^-) \le b \le F(a^+)\}. \tag{2.3.4}$$

The results described later would not be altered if we were instead to use

$$L_5(F, a) = l(L_4(F, a)), \tag{2.3.5}$$

where l is an increasing function. Section 6 contains admissibility results for estimation under the loss function L_4 or L_5.

2.4. *Estimators.*

Continuous problems. For problems involving continuous c.d.f.'s [i.e., $\mathscr{F} = \mathscr{F}_c(a, b)$] we will be investigating admissibility of the best invariant procedure. This procedure will be denoted by the generic symbol δ_0. The following paragraphs give a more precise description.

Let $x_{(1)} \le \cdots \le x_{(n)}$ denote the order statistics corresponding to the sample $x = (x_1, \dots, x_n)$ and let $x_{(0)} = -\infty$, $x_{(n+1)} = +\infty$. Then any nonrandomized invariant and right-continuous procedure has $\delta(x) = d_x(t)$, where

$$d_x(t) = \omega_i \quad \text{if } x_{(i)} \le t < x_{(i+1)}, i = 0, \dots, n. \tag{2.4.1}$$

For the best invariant procedure δ_0, the numbers ω_i are chosen to minimize

$$R(U, \delta) = \int \cdots \int L(U, \delta(x)) \sum_{i=1}^{n} dU(x_i), \tag{2.4.2}$$

where $U(t) = (0 \vee t) \wedge 1$ denotes the uniform c.d.f. on $(0, 1)$. Obviously, the choice $\{\omega_i\}$ yielding the best invariant procedure depends on L.

Aggarwal (1955) calculates the best invariant procedure under losses L_1 and L_2. For L_1 the procedure δ_0 is the sample c.d.f.,

$$F_n(t) = n^{-1} \sum_{i=1}^{n} \chi_{\{x_i \le t\}}(t) = \alpha_x(t) \quad \text{(say)}, \tag{2.4.3}$$

so that δ_0 is given by (2.4.1) with $\omega_i = i/n$.

For L_2 the procedure δ_0 is given by (2.4.1) with

$$\omega_i = (i + 1)/(n + 2). \tag{2.4.4}$$

Call the corresponding estimate $\beta_x(t)$. Note that, as already mentioned, this procedure corresponds to a defective distribution—i.e., $\lim_{t \to -\infty} \beta_x(t) = 1/(n + 2) > 0$ and also $\lim_{t \to \infty} \beta_x(t) = (n + 1)/(n + 2) < 1$.

The derivation of the best invariant estimator for the Kolmogorov–Smirnov loss is messier. Direct calculation (by hand) from (2.4.2) yields

$$\text{for } n = 1, \quad \omega_0 = \tfrac{3}{8}, \omega_1 = \tfrac{5}{8}. \tag{2.4.5}$$

Values of ω for $2 \le n \le 25$ have recently been numerically calculated and tabled by Friedman, Gelman and Phadia (1988). In this case we use the notation $\delta_0(x) = \gamma_x(t)$.

For estimating the median the best invariant procedure can easily be shown to be the sample median. When n is odd the sample median is uniquely defined, so that

$$\delta_0(x) = x_{((n+1)/2)}. \tag{2.4.6}$$

However, when n is even the best invariant estimator is not uniquely defined, and is either $x_{(n/2)}$ or $x_{((n+2)/2)}$, or any fixed randomization between them. Thus,

δ_0 is any (randomized) estimator of the form

$$\begin{aligned} \delta_0(x) &= x_{(n/2)}, && \text{with probability } \pi, \\ &= x_{((n+2)/2)}, && \text{with probability } (1-\pi). \end{aligned} \tag{2.4.7}$$

(Note that π is a constant. If π were to depend on x the resulting estimate would still be a sample median, but would not be invariant.) The problem of estimating the median also has a left–right symmetry. If this symmetry is taken into account, then for n even the only fully invariant procedure is the procedure of the form (2.4.7) with $\pi = \frac{1}{2}$. The admissibility results of Section 6 apply to any procedure of the form (2.4.7), not merely the procedure with $\pi = \frac{1}{2}$.

Discrete problems. For L_1 the natural analog of δ_0 in discrete problems is δ_0 itself. Thus, we will be studying (and, in fact, proving) the admissibility of the sample c.d.f. F_n under the loss L_1. [It should be noted that the definition (2.4.1) of δ_0 has been appropriately stated so that if, say, $x_{(i-1)} < x_{(i)} = \cdots = x_{(i+j)} < x_{(i+j+1)}$, then $\alpha_x(x_{(i)}^-) = \omega_{i-1} = (i-1)/n$ and $\alpha_x(x_{(i)}) = \omega_{i+j} = (i+j)/n$.]

For L_2 and $\mathscr{F} = \mathscr{F}_M(E)$ the immediate analog of δ_0 is again δ_0 itself. However, it is obvious that this estimator is inadmissible. Note that $F(\xi_m) = 1$. Hence, the estimator

$$\delta_0'(t) = \beta_x'(t) = \begin{cases} 0, & \text{if } t < \xi_1, \\ \beta_x(t), & \text{if } x_{(i)} \le t < x_{(i+1)},\ \xi_1 \le t < \xi_m, \\ 1, & \text{if } t \ge \xi_m, \end{cases} \tag{2.4.8}$$

is at least as good as δ_0, and is better whenever F gives positive probability to ξ_m. Consequently, when $\mathscr{F} = \mathscr{F}_M$ and $L = L_2$ we study (and prove) the admissibility of δ_0'.

Similarly, if $\mathscr{F} = \mathscr{F}_D([0,1])$ and $L = L_2$ the appropriate estimator for study is

$$\delta_0' = \beta_x'(t) = \begin{cases} 0, & t < 0, \\ \beta_x(t), & 0 \le t < 1, \\ 1, & t \ge 1. \end{cases} \tag{2.4.9}$$

When $\mathscr{F} = \mathscr{F}_D((0,1))$ one can use either δ_0 or δ_0' as defined in (2.4.9). They are equivalent since $R(F, \delta_0) = R(F, \delta_0')$ for all $F \in \mathscr{F}_D((0,1))$, because $\Pr_F(\{1\}) = 0$.

A similar pattern appears in connection with Kolmogorov–Smirnov loss L_3. If $\mathscr{F} = \mathscr{F}_M$ the appropriate estimator is

$$\delta_0' = \gamma_x'(t) = \begin{cases} 0, & t < \xi_1, \\ \gamma_x(t), & \xi_1 \le t < \xi_m, \\ 1, & t \ge \xi_m, \end{cases} \tag{2.4.10}$$

where, here, $\gamma_x(t)$ is the best invariant estimate when $\mathscr{F} = \mathscr{F}_C$ and $L = L_3$. Similarly, for $\mathscr{F} = \mathscr{F}_D([0,1])$ the appropriate estimator is also $\gamma_x'(t)$ [with $0 = \xi_1$ and $1 = \xi_m$ substituted in (2.4.9)]. When $\mathscr{F} = \mathscr{F}_D((0,1))$ the estimators δ_0 and δ_0', defined in (2.4.9), are equivalent since $L_3(F, \delta_0(x)) = L_3(F, \delta_0'(x))$ w.p.1.

For the problem of estimating the population median ($L = L_4$) we use the sample median δ_0, as defined in (2.4.6) and (2.4.7).

TABLE 1
Admissibility of the procedure δ_0 or δ_0' according to class of distributions and loss function

Loss function:	L_1	L_2	L_2' and L_2''	L_3	L_4
Defined in equation:	(2.3.1)	(2.3.2)	(7.1.2) and (7.1.3)	(2.3.3)	(2.3.4)
Admissibility results in:	Section 3	Section 4	Section 7	Section 5	Section 6
Distributions					
$\mathscr{F}_M$	A	A	I	A if $n = 1$? if $n \geq 2$	A
$\mathscr{F}_D([0,1])$	A	A	I	A if $n = 1$? if $n \geq 2$	A
$\mathscr{F}_D((0,1))$	A	? (See Remark 4.1.3)	I	?	A
$\mathscr{F}_C$	A if $n = 1, 2$[a] I if $n \geq 3$[a]	I	I	?	I if n even A if $n = 1$? otherwise

(A = admissible, I = inadmissible, ? = admissibility unknown.)
[a] These results were proved by Yu (1986, 1987) and appear in these papers.

2.5. *Summary of admissibility results.* Table 1 summarizes the admissibility results known to date according to class of distribution and loss function. With one very important exception the results mentioned in the table are proved in this paper. That exception is the result(s) for continuous distributions F_c and scaled Cramér–von Mises loss L_1. Those results were proved by Yu (1986, 1987), following the appearance of a preliminary version of the current manuscript.

The classes of distributions are as defined in Section 2.1. The procedures δ_0 and δ_0' are described in Section 2.3. As noted there the modified best invariant estimator δ_0' is to be used for the combinations $(\mathscr{F}_M, L_2)$, $(\mathscr{F}_M, L_3)$, $(\mathscr{F}_D([0,1]), L_2)$ and $\mathscr{F}_D([0,1], L_3)$; otherwise the best invariant estimator δ_0 is to be used. The loss function L_2' is a variant of L_2 and is defined in Section 7.

3. Admissibility results for loss L_1.

3.1. *Admissibility of the sample c.d.f. for $\mathscr{F}_M$ or $\mathscr{F}_D$.* It has already been noted in (2.4.3) that the sample c.d.f. is the best invariant estimator of the population c.d.f. under the scaled Cramér–von Mises loss L_1. The main result of this section is that this estimator is admissible under this loss in all of our discrete formulations.

THEOREM 3.1.1. *Let $L = L_1$, $\mathscr{F} = \mathscr{F}_M$. Then the sample c.d.f. δ_0 [defined in (2.4.3)] is admissible.*

PROOF. Recall the notation $\delta_0(x) = \alpha_x(t)$ and let $\delta(x) = d_x(t)$. It will be shown that

$$R(F,\delta) \leq R(F,\delta_0), \forall F \in \mathscr{F}_M \quad \text{implies } d_x(t) = \alpha_x(t) \text{ for } t = \xi_1, \dots, \xi_m \text{ and all possible } x. \tag{3.1.1}$$

Since $R(F, \delta)$ depends only on $d_x(t)$ for $t = \xi_1, \dots, \xi_m$ the assertion (3.1.1) implies $R(F, \delta) = R(F, \delta_0)$ and thus shows that δ_0 is admissible.

The proof of (3.1.1) proceeds by induction on m. The assertion is trivially true when $m = 1$, for then $\mathscr{F}_M$ contains only the one c.d.f. $F(t) = \chi_{\{t \geq \xi_1\}}(t)$. For this c.d.f. $R(F, \delta_0) = 0$ and $R(F, \delta) = 0$ if and only if $d_x(\xi_1) = 1 = \alpha_x(\xi_1)$ for all possible x.

Now assume (3.1.1) is valid for $(m-1)$ and $R(F, \delta) \leq R(F, \delta_0)$ for all $F \in \mathscr{F}_M$. It follows from the truth of (3.1.1) for $(m-1)$ that

$$d_x(t) = \alpha_x(t) \quad \text{whenever } \#\{x_1, \dots, x_n, t\} \leq m - 1, \tag{3.1.2}$$

where $\#\{\cdot\}$ denotes the cardinality of the set. To verify (3.1.2), apply (3.1.1) to $\mathscr{F}_M'$ having support set $\{\xi_1', \dots, \xi_{m-1}'\} \supset \{x_1, \dots, x_n, t\}$. Since $F \in \mathscr{F}_M'$ implies $F \in \mathscr{F}_M$, it follows that $R(F, \delta) \leq R(F, \delta_0)$ for all $F \in \mathscr{F}_M'$. Assertion (3.1.1) for $(m-1)$ then implies $d_x(t) = \alpha_x(t)$ for $\{x_1, \dots, x_n, t\} \subset \{\xi_1', \dots, \xi_{m-1}'\}$. This yields (3.1.2).

Note that if $n \leq m - 2$, then (3.1.2) implies (3.1.1). Hence, what follows concerns only the case where $n \geq m - 1$.

Suppose still that $R(F, \delta) \leq R(F, \delta_0)$, $F \in \mathscr{F}_M$. Write $\Pr_F\{\xi_i\} = p_i$; so that to each probability vector $p = (p_1, \dots, p_m)$ there corresponds a distribution $F_p \in \mathscr{F}_M$. Then

$$0 \leq \Delta = \int \cdots \int_P \left(R(F_p, \delta_0) - R(F_p, \delta)\right) \prod_{i=1}^{m-1} p_i^{-1}\, dp_i, \tag{3.1.3}$$

where $P = \{p_1, \dots, p_{m-1}\colon p_i \geq 0,\ i = 1, \dots, m-1, \text{ and } p_m = 1 - \sum_{i=1}^{m-1} p_i \geq 0\}$. Δ is thus the difference in risks integrated over the improper prior $\prod_{i=1}^{m-1} p_i^{-1}\, dp_i$. (In what follows we will thus be minimizing this integrated risk difference over that part of the sample space not previously determined by the induction hypothesis. This integrated risk difference will then be finite and so can be uniquely minimized by using the formal Bayes rule for each sample point currently under consideration.)

Let

$$l(p, a; \xi) = \frac{\left(F_p(\xi) - a\right)^2}{F_p(\xi)\left(1 - F_p(\xi)\right)}.$$

Then

$$\begin{aligned}\Delta &= \int \cdots \int_P E_{F_p}\left\{\sum_{j=1}^{m} p_j\left[l\left(p, \alpha_x(\xi_j); \xi_j\right)\right.\right. \\ &\qquad\qquad \left.\left. - l\left(p, d_x(\xi_j); \xi_j\right)\right]\right\} \prod_{i=1}^{m-1} p_i^{-1}\, dp_i \\ &\leq \int \cdots \int_P E_{F_p}\left\{\sum_{j=1}^{m-1} p_j\left[l\left(p, \alpha_x(\xi_j); \xi_j\right)\right.\right. \\ &\qquad\qquad \left.\left. - l\left(p, d_x(\xi_j); \xi_j\right)\right]\right\} \prod_{i=1}^{m-1} p_i^{-1}\, dp_i,\end{aligned} \tag{3.1.4}$$

since $l(p, \alpha_x(\xi_m); \xi_m) = 0 \le l(p, d_x(\xi_j); \xi_j)$. For each possible sample x, let $\eta_k = \eta_k(x) = \#\{x_i: x_i = \xi_k\}$. In the obvious way, let $\alpha_\eta(\xi) = \alpha_x(\xi)$ for $\eta(x) = \eta$, and similarly for $d_\eta(\cdot)$. Let $N_j = \{\eta: \ \eta_k \ge 1 \text{ for } 1 \le k \le m, \ k \ne j\}$, $j = 1, \dots, m-1$. Note now that if $\eta \notin N_j$, then $l(p, \alpha_\eta(\xi_j); \xi_j) = l(p, d_\eta(\xi_j); \xi_j)$ by (3.1.2). Consequently, the expectation appearing in (3.1.4) can be rewritten as

$$\sum_{j=1}^{m-1} \sum_{\eta \in N_j} \binom{n}{\eta}\left[l\left(p, \alpha_\eta(\xi_j); \xi_j\right) - l\left(p, d_\eta(\xi_j); \xi_j\right)\right] p_j \prod_{k=1}^{m} p_k^{\eta_k}, \tag{3.1.5}$$

where $\binom{n}{\eta}$ denotes the usual multinomial coefficient.

It is important that

$$\begin{aligned} &\int \cdots \int_P \binom{n}{\eta}\left[l\left(p, \alpha_\eta(\xi_j); \xi_j\right) - l\left(p, d_\eta(\xi_j); \xi_j\right)\right] \\ &\qquad \times p_j \prod_{k=1}^{m} p_k^{\eta_k} \prod_{i=1}^{m-1} p_i^{-1}\, dp_i < \infty, \end{aligned} \tag{3.1.6}$$

for every $\eta \in N_j$. To verify (3.1.6), observe that

$$\begin{aligned} p_j \prod_{k=1}^{m} p_k^{\eta_k} \prod_{i=1}^{m-1} p_i^{-1} &= p_j p_m^{\eta_m} \prod_{i=1}^{m-1} p_i^{\eta_i - 1} \\ &= p_j^{\eta_j} p_m^{\eta_m} \prod_{\substack{i=1 \\ i \ne j}}^{m-1} p_i^{\eta_i - 1}. \end{aligned} \tag{3.1.7}$$

For $\eta \in N_j$ the exponents of p_j, p_m and every p_i on the right of (3.1.7) are nonnegative. Calculating as in (3.1.9) then verifies (3.1.6). [For the case where $\eta = (0, \eta_2, \dots, \eta_m) \in N_1$ it is important to also note, as will be done, that $\alpha_\eta(\xi_1) = 0$, so that $l(p, \alpha_\eta(\xi_1); \xi_1) = p_1/(1 - p_1)$.]

It follows from (3.1.6) that

$$\begin{aligned} \Delta \le \sum_{j=1}^{m-1} \sum_{\eta \in N_j} \binom{n}{\eta} \int \cdots \int &\left[l\left(p, \alpha_\eta(\xi_j); \xi_j\right) - l\left(p, d_\eta(\xi_j); \xi_j\right)\right] \\ &\times p_j p_m^{\eta_m} \prod_{i=1}^{m-1} p_i^{\eta_i - 1}\, dp_i. \end{aligned} \tag{3.1.8}$$

Let $\sigma_j = \sum_{i=1}^{j} \eta_j$. The multiple integral in (3.1.8) can be evaluated when $\eta \in N_j$, $j = 1, \dots, m-1$, and $\eta_1 \ne 0$ as a standard exercise. [See, e.g., Ferguson (1973).] Making the substitution $u = \sum_{k=1}^{j} p_k$ yields

$$\begin{aligned} &\int \cdots \int_P \left[\frac{\left(\sum_{k=1}^{j} p_k - \alpha_\eta(\xi_j)\right)^2 - \left(\sum_{k=1}^{j} p_k - d_\eta(\xi_j)\right)^2}{\left(\sum_{k=1}^{j} p_k\right)\left(1 - \sum_{k=1}^{j} p_k\right)}\right] \\ &\qquad \times p_j p_m^{\eta_m} \prod_{i=1}^{m-1} p_i^{\eta_i - 1}\, dp_i \\ &\quad = C(\sigma_j, n) \int_0^1 \frac{\left(u - \alpha_\eta(\xi_j)\right)^2 - \left(u - d_\eta(\xi_j)\right)^2}{u(1-u)} u^{\sigma_j} (1-u)^{n - \sigma_j}\, du, \end{aligned} \tag{3.1.9}$$

where $C(\sigma_j, n)$ is an appropriate positive number. The expression on the right of (3.1.9) is uniquely maximized when

$$d_\eta(\xi_j) = \frac{\int_0^1 u^{\sigma_j}(1-u)^{n-\sigma_j-1}}{\int_0^1 u^{\sigma_j-1}(1-u)^{n-\sigma_j-1}\,du} = \frac{\sigma_j}{n} = \alpha_\eta(\xi_j). \tag{3.1.10}$$

When $\eta \in N_1$ and $\eta_1 = 0$ the multiple integral on the left of (3.1.9) has the value $-\infty$ unless $d_\eta(\xi_1) = 0 = \alpha_\eta(\xi_1)$. Hence, in this case also (3.1.9) is uniquely maximized by $d_\eta(\xi_1) = \alpha_\eta(\xi_1)$. Recall that, by assumption, $\Delta \geq 0$. Thus, the preceding results show that actually $\Delta = 0$ and $d_\eta(\xi_j) = \alpha_\eta(\xi_j)$, $j = 1,\dots, m-1$. It then follows trivially that also $d_\eta(\xi_m) = \alpha_\eta(\xi_m) = 1$. This verifies the induction hypothesis (3.1.1) and completes the proof. □

The preceding proof is a variant of the general stepwise Bayes argument described in Brown (1981). The primary variation in the argument occurs because the point t appears in the reinterpretation (3.1.2) of the basic induction hypothesis (3.1.1). [The appearance of t in (3.1.2) and the dependence there on $\#\{x_1,\dots, x_n, t\}$ rather than, say, on $\#\{x_1,\dots, x_n\}$, is consistent with the general results of Brown (1981). However, it was not explicitly observed there because no examples were considered in which the loss function has a structure like L_1, requiring integration over an additional variable (t).]

The remainder of the proof is actually fairly straightforward. The assumption of a formal multiple beta prior follows the pattern of previous proofs, such as Cohen and Kuo (1985), involving noninvariant loss functions. In fact, the calculation in (3.1.9) echoes a formally similar expression which appears in the derivation in Aggarwal (1955) of δ_0 as the best invariant procedure.

Admissibility when $\mathscr{F} = \mathscr{F}_D$ follows directly from Theorem 3.1.1 and a general observation about admissibility formally stated in Theorem 3.1.2.

THEOREM 3.1.2. *Suppose δ is an estimator such that for any $\{\xi_1,\dots,\xi_m\} \subset S$ the estimator δ is admissible for the problem with loss L and $\mathscr{F} = \mathscr{F}_M(\{\xi_1,\dots,\xi_m\})$. Then δ is admissible for the problem with loss L and $\mathscr{F} = \mathscr{F}_D(S)$.*

PROOF. The theorem follows immediately from the definition of admissibility and the fact that $\bigcup_{\{\xi_1,\dots,\xi_m\}\subset S} \mathscr{F}_M(\{\xi_1,\dots,\xi_m\}) = \mathscr{F}_D(S)$. □

COROLLARY 3.1.3. *Let $L = L_1$ and $\mathscr{F} = \mathscr{F}_D(S)$ for any $S \subset (-\infty, \infty)$. Then the sample c.d.f. δ_0 is admissible.*

PROOF. This follows directly from Theorems 3.1.1 and 3.1.2. □

3.2. *Concerning admissibility for $\mathscr{F}_C$.* Corollary 3.1.3 would seem to lend strong support to the conjecture that the sample c.d.f. δ_0 is admissible also when

$\mathscr{F} = \mathscr{F}_C$. However, it does not prove the conjecture since it is logically possible for an estimator to be admissible in all discrete problems and inadmissible in the continuous problem. The situations for L_2 and L_4 provide partial examples of this phenomenon. For L_2 the estimator δ_0' is admissible for $\mathscr{F}_D([0,1])$ but not admissible for $\mathscr{F}_C([0,1])$. For L_4 and n even the symmetric, invariant sample median δ_0, defined by (2.4.7) with $\pi = \frac{1}{2}$, is admissible in all our discrete formulations but is inadmissible when $\mathscr{F} = \mathscr{F}_C$. There are more trivial invariant problems in which this admissibility–inadmissibility phenonemon is obvious. Suppose, for example, that one wishes to test whether F is discrete or continuous under conventional 0–1 loss. Then the procedure which always decides that F is discrete is admissible (in fact, optimum) when $\mathscr{F} = \mathscr{F}_D$ but is inadmissible (in fact, worst possible) when $\mathscr{F} = \mathscr{F}_C$.

After the preceding was written Yu (1986, 1987) proved the surprising and significant results that δ_0 is inadmissible for L_1 and $\mathscr{F}_C$ when $n \geq 3$ but is admissible for $n = 1, 2$.

4. Admissibility results for loss L_2.

4.1. *Discrete settings.* For reasons already discussed in Section 2.4.2 we investigate in discrete problems admissibility under L_2 of the modified procedure δ_0' defined in (2.4.8) and (2.4.9). The first main result parallels Theorem 1.

THEOREM 4.1.1. *Let $L = L_2$, $\mathscr{F} = \mathscr{F}_M$. Then δ_0', defined in* (2.4.8), *is admissible.*

PROOF. The proof is extremely similar to that of Theorem 3.1.1, but with one subtle difference. The induction hypothesis (3.1.1) is replaced by the statement

$$\begin{aligned}&R(F,\delta) \leq R(F,\delta_0'),\ \forall\, F \in \mathscr{F}_M(\{\xi_1',\dots,\xi_{m'}'\}) \text{ with } \{\xi_1',\dots,\xi_{m'}'\}\\ &\subset \{\xi_1,\dots,\xi_m\} \text{ and } \xi_{m'}' = \xi_m \text{ implies } d_x(t) = \beta_x(t) \text{ for}\\ &\{x_1,\dots,x_n,t\} \subset \{\xi_1',\dots,\xi_{m'}'\}.\end{aligned} \tag{4.1.1}$$

The subtle difference here lies in the condition that $\xi_{m'}' = \xi_m$.

The proof now proceeds by induction on m. Each stage of the induction involves only values of x, t for which

$$\{x_1,\dots,x_n,t,\xi_m\} = \{\xi_1',\dots,\xi_{m'}'\}, \tag{4.1.2}$$

since values with $\{x_1,\dots,x_n,t,\xi_m\} \subsetneqq \{\xi_1',\dots,\xi_{m'}'\}$ will already have been considered at an earlier stage of the induction. One proceeds as in the proof of Theorem 3.1.1 The appropriate definition of l is now, of course, $l(p,d;\xi) = (F_p(\xi) - d)^2$. The expression (3.1.4) remains valid with ξ_j' and β_x' replacing ξ_j and α_x since $\xi_{m'}' = \xi_m$ so that $\beta_m'(\xi_{m'}') = 1$. Alter slightly the definition of N_j to become $N_j = \{\eta\colon \eta_k \geq 1 \text{ for } 1 \leq k \leq m-1,\ k \neq j\}$. In this manner one proceeds through the proof with only minor differences until (3.1.9). The right side of (3.1.9) now reads

$$(4.1.3)\quad C(\sigma_j, n)\int\left[\left(u - \beta_\eta'(\xi_j)\right)^2 - \left(u - d_\eta(\xi_j)\right)^2\right]u^{\sigma_j}(1-u)^{n-\sigma_j}\,du.$$

This expression is also valid when $\eta \in N_1$ and $\eta_1 = 0$. It is uniquely minimized when

$$(4.1.4)\quad d_\eta(\xi_j) = (\sigma_j + 1)/(n+2) = \beta'(\xi_j), \qquad j = 1, \dots, m' - 1.$$

The theorem then follows in the same manner as Theorem 3.1.1. □

When $\mathscr{F} = \mathscr{F}_D([0,1])$ admissibility follows by a variation of the argument used in Corollary 3.1.3., as follows.

COROLLARY 4.1.2. *Let $L = L_2$ and $\mathscr{F} = \mathscr{F}_D([0,1])$. Then δ_0', defined in (2.4.9), is admissible.*

PROOF. Suppose

$$(4.1.5)\qquad R(F,\delta) \le R(F,\delta_0') \quad \text{for all } F \in \mathscr{F}_D([0,1]).$$

Write $\delta(x) = d_x(t)$. Let $x = (x_1, \dots, x_n)$ be a possible sample point and let $t' \in [0,1]$. Let $E = \{\xi_1, \dots, \xi_m\} \supset \{x_1, \dots, x_n, t', 1\}$. Consider the problem with $\mathscr{F} = \mathscr{F}_M(E)$. δ_0' is admissible in this problem by the statement of Theorem 4.1.1, but the proof of the theorem shows even more—namely, that (4.1.5) for all $F \in \mathscr{F}_M(E) \subset \mathscr{F}_D([0,1])$ implies

$$(4.1.6)\qquad d_x(t) = \beta_x'(t) \quad \text{for all } t \in E.$$

Thus, $\delta = \delta_0'$ since (4.1.6) holds for all possible x and all $t \in [0,1]$. □

REMARK 4.1.3. The preceding proof does not verify that δ_0' is admissible if $\mathscr{F} = \mathscr{F}_D((0,1))$. It fails to apply because $1 \notin (0,1)$, so that $\mathscr{F}_M(E) \not\subset \mathscr{F}_D((0,1))$. [As previously noted, over $\mathscr{F}_D((0,1))$, δ_0' and δ_0 are equivalent so all assertions here concern both estimators.]

Intuition suggests that a procedure admissible in $\mathscr{F}_D([0,1])$ should also be admissible in $\mathscr{F}_D((0,1))$. Indeed, we have as yet found no natural examples where a procedure is admissible in $\mathscr{F}_D([0,1])$ and not in $\mathscr{F}_D((0,1))$. However, Example 4.1.5 suggests that this intuition may be faulty. Cognizant of Example 4.1.5 we nevertheless conjecture (somewhat uneasily!) that δ_0' is admissible for $\mathscr{F}_D((0,1))$ because we have failed to find an estimator dominating δ_0 for the cases $n = 1, 2, 3$.

Define $\mathscr{F}_D^{(m)}(S) \subset \mathscr{F}_D(S)$ to be the subset of $\mathscr{F}_D(S)$ consisting of distributions supported on at most m points. The same intuition which suggests that admissibility for $\mathscr{F}_D([0,1])$ implies admissibility for $\mathscr{F}_D((0,1))$ also suggests that admissibility for $\mathscr{F}_D^{(m)}([0,1])$ implies admissibility for $\mathscr{F}_D^{(m)}((0,1))$. However, Examples 4.1.4 and 4.1.5 taken together show this latter implication is false when $n \le m - 2$.

EXAMPLE 4.1.4. Admissibility of δ_0' for $\mathscr{F}_D^{(m)}([0,1])$ and $n \le m-2$.

Let $L = L_2$, $\mathscr{F} = \mathscr{F}_D^{(m)}([0,1])$, $n \le m-2$. Then δ_0' is admissible. To see this, consider $\mathscr{F} = \mathscr{F}_M(E)$ in Theorem 4.1.1, where $1 \in E \subset [0,1]$ and $\#E \le m$. Then the induction step (4.1.1) need be carried only through stage $m' = n+2 \le m$ since all possible values of x, t have $\#\{x_1, \ldots, x_n, t, 1\} \le n+2$. These priors are all concentrated on $\mathscr{F}_D^{(m)}([0,1])$. It follows that δ_0' is admissible in $\mathscr{F}_M(E) \subset \mathscr{F}_D^{(m)}([0,1])$. The reasoning of Corollary 4.1.2 can then be applied to prove admissibility of δ_0' in $\mathscr{F}_D^{(m)}([0,1])$.

[δ_0' is not admissible if $n \ge m-1$. The procedure $\delta(x) = d_x(t)$ with

$$d_x(t) = 1, \quad \text{if } \#\{x_1, \ldots, x_n\} \ge m-1 \text{ and } t > x_{(n)}, \qquad (4.1.7)$$
$$= \beta_x'(t), \quad \text{otherwise,}$$

is better.]

EXAMPLE 4.1.5. Inadmissibility of δ_0' (and δ_0) for $\mathscr{F}_D^{(m)}((0,1))$.

Let $L = L_2$ and $\mathscr{F} = \mathscr{F}_D^{(m)}((0,1))$, $m \ge 1$. Then δ_0' (and δ_0) is inadmissible even among invariant procedures. To see this, let $\delta''(x) = \beta_x''(\cdot)$, with

$$\beta_x''(t) = \beta_x'(t) + \frac{m^{-1}}{n+2} = \frac{1+m^{-1}}{n+2} + \frac{1}{n+2}\sum_{k=1}^{n} \chi_{\{t \ge x_k\}}(t). \qquad (4.1.8)$$

Then, with $\Pr_F(\{\xi_i\}) = p_i$ as before, elementary calculation yields

$$\begin{aligned} R(F, \delta_0') - R(F, \delta_0'') &= 2\left(\frac{m^{-1}}{n+2}\right)\left(\sum_{i=1}^{m} \frac{p_i^2}{n+2}\right) - \left(\frac{m^{-1}}{n+2}\right)^2 \\ &\ge \left(\frac{m^{-1}}{n+2}\right)^2 > 0, \end{aligned} \qquad (4.1.9)$$

since $\sum_{i=1}^m p_i^2 \ge m^{-1}$. Thus, δ'' is better than δ_0' and δ_0. I do not know whether δ'' is itself inadmissible or whether it is possible to improve on δ_0' by an amount significantly larger than $(m^{-1}/(n+2))^2$.

REMARK 4.1.6. The procedure δ_0' was motivated in Section 2.4 as the minimal modification of δ_0 necessary to compensate for an obvious inadequacy of δ_0. The preceding considerations suggest the possibility of instead using $\delta_1(x) = d_x(t)$, with

$$d_x(t) = \beta_x(t), \quad \text{if } t \le x_m, \qquad (4.1.10)$$
$$= 1, \quad \text{if } t > x_m.$$

(Note this estimator is not right continuous.)

Arguments like those in the proofs of Theorems 3.1.1 and 4.1.1 show that δ_1 is admissible for $\mathscr{F} = \mathscr{F}_M$. Hence, for any $S \subset \mathbb{R}$, it is also admissible for $\mathscr{F} = \mathscr{F}_D(S)$ by Theorem 3.1.2. (On the other hand, it is invariant in problems to which invariance applies; hence, when $\mathscr{F} = \mathscr{F}_C$ it is not admissible since then δ_0 is the best invariant estimator.)

4.2. *Inadmissibility of* δ_0 *for* $\mathscr{F}_C$. When $\mathscr{F} = \mathscr{F}_C$ the best invariant estimator is inadmissible. This is shown by the following theorem, which gives an explicit formula for an estimator that improves on δ_0.

Define, for $z, t \in \mathbb{R}$,

$$\begin{aligned} \zeta_z(t) &= 1, \qquad z \le 0 < t, \\ &= -1, \qquad t \le 0 < z, \\ &= 0, \qquad \text{otherwise}, \end{aligned} \tag{4.2.1}$$

and

$$\begin{aligned} \chi_z(t) &= 1, \qquad z \le t, \\ &= 0, \qquad z > t. \end{aligned} \tag{4.2.2}$$

Note that $\delta_0(x) = \beta_x(\cdot)$, where

$$\beta_x(t) = 1/(n+2) + \sum_{i=1}^{n} \chi_{x_i}(t)/(n+2). \tag{4.2.3}$$

THEOREM 4.2.1. *Let* $L = L_2$, $\mathscr{F} = \mathscr{F}_C$. *Define* $\delta(x) = d_x(\cdot)$ *by*

$$d_x(t) = \beta_x(t) + \sum_{i=1}^{n} \zeta_{x_i}(t)/2(n+1)(n+2). \tag{4.2.4}$$

Then

$$R(F, \delta_0) - R(F, \delta) = \frac{n \Pr_F(X \le 0)\Pr_F(X > 0)}{4(n+1)(n+2)^2} \ge 0. \tag{4.2.5}$$

Hence, δ_0 *is inadmissible.*

REMARK 4.2.2. Here is a way to visualize $\beta_x(t)$ and its relation to $d_x(t)$ defined in (4.2.4). Think of $\beta_x(t)$ as the c.d.f. corresponding to a distribution giving mass $1/(n+2)$ to each of the points $-\infty, x_1, x_2, \dots, x_n, \infty$. To produce $d_x(t)$, modify this distribution as follows: For each $x_i > 0$, $i = 1, \dots, n$, take mass $1/2(n+1)(n+2)$ from $-\infty$ and move it to 0. For each $x_i \le 0$ take this amount of mass from $+\infty$ and move it to 0. $d_x(t)$ is the c.d.f. of the resulting mass distribution.

The preceding description (due to a referee) shows that $d_x(t)$ can be interpreted as the result of a kind of "shrinkage" from $\pm\infty$ to 0 of the mass for $\beta_x(t)$.

REMARK 4.2.3. Since $R(F, \delta_0) = 1/6(n+2)$ the fractional saving in risk from using δ_0 is

$$0 \le \frac{R(F,\delta_0) - R(F,\delta)}{R(F,\delta_0)} = \frac{3n \Pr_F(X \le 0)\Pr_F(X > 0)}{2(n+1)(n+2)} \le \frac{3n}{8(n+1)(n+2)}. \tag{4.2.6}$$

This is 1/16 for $n = 1$ or 2 and decreases for larger n. Hence, the maximum fractional saving in risk is not large. We do not know whether δ is admissible or whether it is possible to find some other estimator dominating δ_0 which provides a significantly larger maximum fractional saving in risk.

PROOF OF THEOREM 4.2.1. Both $\beta_x(\cdot)$ and $d_x(\cdot)$ are equivariant under monotone transformations of the line which leave the origin fixed. And, of course, L_2 is invariant under such transformations. Hence, it suffices to verify (4.2.5) when

$$F(t) = U_p(t) = \min(1, \max(0, t+p)), \qquad p \ge 0,$$

the uniform distribution on $(-p, 1-p)$. The following formulas involve only routine, direct evaluations:

$$a_1 = \int\int\int \zeta_x(t)\chi_y(t)\, dU_p(x)\, dU_p(y)\, dU_y(t) = p(1-p)/2, \tag{4.2.7}$$

$$a_2 = \int\int \zeta_x(t)\chi_x(t)\, dU_p(x)\, dU_p(t) = p(1-p), \tag{4.2.8}$$

$$a_3 = \int\int \zeta_x(t)\, dU_p(x)\, dU_p(t) = 0, \tag{4.2.9}$$

$$a_4 = \int\int U_p(t)\zeta_x(t)\, dU_p(x)\, dU_p(t) = p(1-p)/2, \tag{4.2.10}$$

$$a_5 = \int\int\int \zeta_x(t)\zeta_y(t)\, dU_p(x)\, dU_p(y)\, dU_p(t) = p(1-p), \tag{4.2.11}$$

$$a_6 = \int\int \zeta_x^2(t)\, dU_p(x)\, dU_p(t) = 2p(1-p). \tag{4.2.12}$$

Let $\alpha = 1/2(n+1)(n+2)$. Then for $F = U_p$,

$$R(F,\delta_0) - R(F,\delta)$$

$$= E\Bigg(\int\Bigg[\Bigg(F(t) - 1/(n+2) - \sum_{i=1}^{n} \chi_{X_i}(t)/(n+2)\Bigg)^2$$

$$- \Big(F(t) - 1/(n+2) - \sum \chi_{X_i}(t)/(n+2) - \alpha \sum \zeta_{X_i}(t)\Big)^2\Bigg]\, dF(t)$$

$$= E\Bigg(\int\Big[2\alpha \sum \zeta_{X_i}(t)\Big(F(t) - 1/(n+2) - \sum \chi_{X_j}(t)/(n+2)\Big)$$

$$- \alpha^2\Big(\sum \zeta_{X_i}(t)\Big)^2\Big]\, dF(t)\Bigg) \tag{4.2.13}$$

$$= 2\alpha\big(na_4 - \big(na_3 + na_2 + n(n-1)a_1/(n+2)\big)\big)$$

$$- \alpha^2\big(na_6 + n(n-1)a_5\big)$$

$$= p(1-p)\big[2\alpha\big(n/2 - (n + n(n-1)/2)/(n+2)\big)$$

$$- \alpha^2(2n + n(n-1))\big]$$

$$= p(1-p)\big[\alpha n/(n+2) - \alpha^2 n(n+1)\big]$$

$$= p(1-p)n/4(n+1)(n+2)^2.$$

as claimed in (4.2.5). □

5. Results for Kolmogorov–Smirnov loss L_3. The only results we have for L_3 concern the case $n = 1$, of no interest in applications. Progress towards results for $n \geq 2$ was blocked in the first place by our ignorance of the precise numerical description of δ_0 when $n \geq 2$. After the first draft of this paper was written Friedman, Gelman and Phadia (1988) produced a numerical table describing δ_0 for $n \leq 25$. However, it is still not clear to me whether δ_0 is admissible for $n \geq 2$.

The proof when $n = 1$ of admissibility for $\mathscr{F} = \mathscr{F}_M$ is, as usual, a stepwise Bayes argument. The structure of this argument is slightly different from previous arguments in Theorems 3.1.1 and 4.1.1 because $n = 1$ and because of a qualitative difference between L_3 and the various Cramér–von Mises type losses considered earlier: When the support of F is given in the stepwise Bayes argument to be the two points $\{\xi_1, \xi_m\}$, then under L_3 the Bayes procedure is determined uniquely at all $\xi_i \in \{\xi_1, \ldots, \xi_m\}$, whereas under L_1 or L_2, etc., it is determined uniquely only at ξ_1 and ξ_m, and must be determined for ξ_i, $2 \leq i \leq m-1$, at future steps of the argument.

THEOREM 5.1.1. *Suppose $n = 1$, $L = L_3$, $\mathscr{F} = \mathscr{F}_M$. Then δ_0' is admissible.*

PROOF. Suppose $R(F, \delta) \leq R(F, \delta_0')$ for all $F \in \mathscr{F}_M$. Write $\delta(x) = d_x(t)$. As before, let $F_p(\xi_i)) = p_i$, $i = 1, \ldots, m$. Choose α such that

$$\int_0^{5/8} p^{\alpha+1}(1-p)^{\alpha}\, dp = \int_0^1 p^{\alpha+1}(1-p)^{\alpha}\, dp/2. \tag{5.1.1}$$

Consider $S_1 = \{F_p\colon p = (p_1, 0, \ldots, 0, 1-p_1)\}$. For $F \in S_1$,

$$L_3(F, d(\cdot)) \leq |p_1 - d(\xi_1)|, \tag{5.1.2}$$

with equality for all $F \in S_1$ if and only if

$$\begin{aligned} d(t) &= 0, & t &< \xi_1, \\ &= d(\xi_1), & \xi_1 &\leq t < \xi_m, \\ &= 1, & t &\geq \xi_m. \end{aligned} \tag{5.1.3}$$

Note also that $L_3(F, d(\cdot))$ is continuous in p_1 for $F \in S_1$. Thus, by a standard calculation, for $F_p \in S_1$ as above,

$$\begin{aligned} &\int R_3(F_p, \delta) p_1^{\alpha}(1-p_1)^{\alpha}\, dp_1 \\ &\geq \int_0^1 \left(|p_1 - d_{\{\xi_1\}}(\xi_1)| p_1 + |p_1 - d_{\{\xi_m\}}(\xi_1)|(1-p_1) \right) p_1^{\alpha}(1-p_1)^{\alpha}\, dp_1 \\ &\geq \int_0^1 \left(|p_1 - \tfrac{5}{8}| p_1 + |p_1 - \tfrac{3}{8}|(1-p_1) \right) p_1^{\alpha}(1-p_1)^{\alpha}\, dp_1 \\ &= \int_0^1 R_3(F_p, \delta_0) p_1 (1-p_1)^{\alpha}\, dp_1. \end{aligned} \tag{5.1.4}$$

[The second inequality in (5.1.4) follows from (5.1.1).] In view of (5.1.3) there is equality throughout (5.1.4) if and only if

$$d_{\{\xi_1\}}(\cdot) \equiv \gamma_{\xi_1}'(\cdot), \qquad d_{\{\xi_m\}}(\cdot) \equiv \gamma_{\xi_m}'(\cdot). \tag{5.1.5}$$

It follows that d satisfies (5.1.5) since $R_3(F, \delta) \leq R_3(F, \delta_0')$.

Now let

$$\begin{aligned} G_i(t) &= 0, & t &< \xi_1, \\ &= \tfrac{3}{8}, & \xi_1 &\leq t < \xi_i, \\ &= \tfrac{5}{8}, & \xi_i &\leq t < \xi_m, \\ &= 1, & t &\geq \xi_m,\ i = 2, \ldots, m-1. \end{aligned} \tag{5.1.6}$$

Then in view of (5.1.5)

$$\begin{aligned} 0 &\leq R_3(G_i, \delta_0') - R_3(G_i, \delta) \\ &= \tfrac{1}{4}\left(\sup|G_i(t) - \gamma_{\xi_i}'(t)| - \sup|G_i(t) - d_{\xi_i}(t)| \right) \\ &= \tfrac{1}{4}\left(-\sup|G_i(t) - d_{\xi_i}(t)| \right) \leq 0. \end{aligned} \tag{5.1.7}$$

It follows that $d_{\xi_i}(t) = \gamma'_{\xi_i}(t)$, $i = 2, \ldots, m-1$. This together with (5.1.5) shows $d = \gamma'$, so that δ' is admissible. □

COROLLARY 5.1.2. *Let* $n = 1$, $L = L_3$, $\mathscr{F} = \mathscr{F}_D([0,1])$. *Then* δ_0' *is admissible.*

PROOF. This corollary follows from Theorem 5.1.1 as did Corollary 4.1.2 from Theorem 4.1.1. □

REMARK 5.1.3. As was the case in Remark 4.1.3, admissibility of δ_0' when $\mathscr{F} = \mathscr{F}_D((0,1))$ does not follow from the proof used for Corollary 5.1.2, and we do not know whether δ_0' is admissible in this case. There seems to be some basis for thinking that the situation here parallels that in Section 4 and so for conjecturing that δ_0 (and δ_0') is inadmissible for $\mathscr{F}_C$.

6. Results for estimating the median with loss L_4.

6.1. *Admissibility of* δ_0 *in discrete settings.*

THEOREM 6.1.1. *Let* $L = L_4$, $\mathscr{F} = \mathscr{F}_M$. *Then the sample median* δ_0 *as defined in* (2.4.6) *and* (2.4.7) *is admissible.*

PROOF. A direct proof involves a stepwise Bayes argument of many steps. However, all these steps can be combined into a much simpler induction argument, proving a more general result.

It is convenient, as in the proof of Theorem 3.1.1, to consider the problem in multinomial form with $p_i = P_F(\{\xi_i\})$ and $\eta_j = \#\{x_i\colon x_i = \xi_j,\ i = 1, \ldots, n\}$. The vector $p = (p_1, \ldots, p_m)$ describes F, and the vector $\eta = (\eta_1, \ldots, \eta_m) \in N(n, m)$, which has a multinomial (n, p) distribution, is a sufficient statistic.

In the statement of the induction hypothesis we will consider loss functions which also depend on η. Specifically, we consider losses of the form

$$L(F, d, \eta) = l(\eta) L_4(F, d), \tag{6.1.1}$$

where $l(\eta) > 0$ for all possible η. We will also consider sample quantiles other than the sample median. It is necessary here to use a precise, and slightly restricted, definition of sample quantiles. For $0 \le \alpha \le 1$ the set of αth sample quantiles is $A_\alpha(\eta)$, as follows: If $\alpha = i/(n+1)$, $i = 1, \ldots, n$, then $A_\alpha(\eta)$ contains the unique point ξ_j for which $\sum_{i=1}^{j-1}\eta_i < i \le \sum_{i=1}^{j}\eta_i$. If $\alpha = 0$ or 1, respectively, then $A_\alpha(\eta) = \{\xi_1\}$ or $\{\xi_m\}$, respectively. If $i/(n+1) < \alpha < (i+1)/(n+1)$, $i = 0, \ldots, n$, then $A_\alpha(\eta) = A_{i/(n+1)}(\eta) \cup A_{(i+1)/(n+1)}(\eta)$.

An αth quantile estimator is any (randomized) procedure δ [to be also denoted as $\delta(\eta)$ and $\delta(\cdot|\eta)$], for which

$$\delta(A_\alpha(\eta)|\eta) = \Pr_\delta(A_\alpha(\eta)|\eta) = 1. \tag{6.1.2}$$

The sample median δ_0, defined in (2.4.6) and (2.4.7), is, of course, a $\frac{1}{2}$ quantile estimator. For later use in connection with randomized procedures, define $\bar{L}(F, \delta(\eta), \eta) = \int L(F, a, \eta)\delta(da|\eta)$.

Here is the induction hypothesis:

(6.1.3) $H(n, m)$: Fix n, m. Then for any loss function of the form (6.1.1) any αth quantile estimator δ^*, say, is admissible. Furthermore, if $R(F, \delta) \le R(F, \delta^*)$ under the loss (6.1.1) for all $F \in \mathscr{F}_M$, then $\delta = \delta^*$.

For any n, $H(n, 1)$ is trivially true since there is really only one quantile estimator, namely, $\delta(\{\xi_1\}|\eta) \equiv 1$. This estimator has risk 0 and is admissible, and no different estimator has risk 0.

Now consider $H(n, m)$, $m \ge 2$, and assume $H(n', m')$ is true for $n' = n$, $m' \le m - 1$ and for $n' \le n - 1$, all m'. Suppose δ^* is an αth quantile estimator, and

$$R(F, \delta) \le R(F, \delta^*), \quad \forall\, F \in \mathscr{F}_M. \tag{6.1.4}$$

By symmetry it suffices to consider the case $\alpha \le \frac{1}{2}$. In particular, (6.1.4) holds for all $F = F_p$ having $p_m = 0$. It then follows from the assumed validity of $H(n, m - 1)$ that

$$\delta(\eta) = \delta^*(\eta) \quad \text{whenever } \eta_m = 0. \tag{6.1.5}$$

[It is important here that $\delta^*(\{\xi_m\}, \eta) = 0$ when $\eta_m = 0$, because of the assumption that $\alpha \le \frac{1}{2}$.] Because of this,

$$\begin{aligned}
0 &\le R(F, \delta^*) - R(F, \delta) \\
&= \sum_{\eta \in N(n, m)} \binom{n}{\eta} \big(\bar{L}(F, \delta^*(\eta), \eta) - \bar{L}(F, \delta(\eta), \eta)\big) \prod_{i=1}^{m} p_i^{\eta_i} \\
&= p_m \sum_{\{\eta \in N(n, m):\ \eta_m \ge 1\}} \binom{n}{\eta} \big(\bar{L}(F, \delta^*(\eta), \eta) \\
&\qquad\qquad - \bar{L}(F, \delta(\eta), \eta)\big) \left(\prod_{i=1}^{m-1} p_i^{i} \right) p_m^{\eta_m - 1} \\
&= p_m \sum_{\eta \in N(n-1, m)} \binom{n-1}{\eta} \Bigg[\frac{n}{(\eta_m + 1)} \big(\bar{L}(F, \delta^*(\eta + e_m), \eta + e_m) \\
&\qquad\qquad - \bar{L}(F, \delta(\eta + e_m), \eta + e_m)\big) \Bigg] \prod_{i=1}^{m} p_i^{\eta_i},
\end{aligned} \tag{6.1.6}$$

where e_m denotes the mth unit vector. Define the new loss function $L'(F, a, \eta) = (n/(\eta_m + 1))L(F, a, \eta + e_m)$ and the procedures $\delta^{*\prime}$ and δ' on $N(n - 1, m)$ by $\delta^{*\prime}(\eta) = \delta^*(\eta + e_m)$ and $\delta'(\eta) = \delta'(\eta + e_m)$. Then (6.1.6) implies, by continuity, that

$$\begin{aligned}
0 \le \sum_{\eta \in N(n-1, m)} \binom{n-1}{\eta} \big(\bar{L}'(F, \delta^{*\prime}(\eta), \eta) \\
- \bar{L}'(F, \delta'(\eta), \eta)\big) \prod_{i=1}^{m} p_i^{\eta_i}.
\end{aligned} \tag{6.1.7}$$

The assumed validity of $H(n-1, m)$ then yields from (6.1.7) that $\delta^{*\prime}(\eta) = \delta'(\eta)$, $\eta \in N(n-1, m)$, since $\delta^{*\prime}(\eta)$ is a $\min((n+1)\alpha/n, 1)$ quantile estimator. Thus,

$$\delta^*(\eta) = \delta(\eta), \quad \forall\ \eta \in N(n, m) \ni \eta_m \geq 1. \tag{6.1.8}$$

(6.1.5) and (6.1.8) imply that $\delta = \delta^*$, which proves the validity of $H(n, m)$. The validity of $H(n, m)$ for all n, m obviously yields the assertion of the theorem as a special case. □

REMARK 6.1.2. The proof of Theorem 6.1.1 shows the validity of $H(n, m)$, a much more general fact than that actually claimed in the theorem. This added generality has an ironic backlash.

Note that the validity of $H(n, m)$ also implies, for example, that $x_{(1)} = \min\{x_i\colon\ i = 1, \ldots, n\}$ is an admissible estimator of the population median. Obviously, $x_{(1)}$, while admissible, is not a very worthwhile estimator. We have, consequently, proved δ_0 to be admissible in a manner which does not give any information as to whether δ_0 is also a worthwhile estimator. (It nevertheless probably is.)

The preceding observation emphasizes a commonplace fact. Merely to show that an estimator is admissible does not guarantee it is a worthwhile estimator. Other aspects of the performance of any admissible estimator must also be taken into account.

COROLLARY 6.1.3. *Let* $L = L_4$, $\mathscr{F} = \mathscr{F}_D$. *Then the sample median as defined in* (2.4.6) *and* (2.4.7) *is admissible.*

PROOF. This follows immediately from Theorem 6.1.1 and Theorem 3.1.2. □

6.2. *Concerning admissibility for* $\mathscr{F}_C$. Let $\mathscr{F} = \mathscr{F}_C$. For $n = 1$, δ_0 is admissible since it is the best location invariant estimator for the problem in which $\mathscr{F}$ is restricted to the set of uniform distributions on $(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, $-\infty \leq \theta \leq \infty$. For n odd, $n \geq 3$, it appears reasonable to conjecture that δ_0 is still admissible. For even n, δ_0 [defined by (2.4.7) with any π, $0 \leq \pi \leq 1$] is not admissible. This inadmissibility is closely related to the nonuniqueness of δ_0. δ_0 will be shown to be dominated by a noninvariant version of the sample median. It may be that some noninvariant version of the sample median is admissible.

THEOREM 6.2.1. *Let* $L = L_4$, $\mathscr{F} = \mathscr{F}_C$ *and let n be even. Let* δ_0 *be defined by* (2.4.7). *Then* δ_0 *is inadmissible. A better estimator is*

$$\begin{aligned} \delta(x) &= x_{((n+2)/2)}, && \text{if } x_{((n+2)/2)} < 0, \\ &= 0, && \text{if } x_{(n/2)} \leq 0 \leq x_{((n+2)/2)}, \\ &= x_{(n/2)}, && \text{if } x_{(n/2)} > 0. \end{aligned} \tag{6.2.1}$$

PROOF. The proof resembles that of a qualitatively analogous result in Farrell (1964). As in the proof of Theorem 4.2.1 it suffices to consider the case $F = U_p$, the uniform distribution on $(-p, 1-p)$, $0 \le p \le 1$. Note that $R(U_p, \delta_0)$ is independent of the choice of π in (2.4.7), as can easily be seen from the left–right symmetry of the integrals defining $R(U_p, \delta_0)$. By symmetry, it suffices to choose $p \le \frac{1}{2}$ and then to choose $\pi = 1$ in the definition of δ_0 and to show $R(U_p, \delta) \le R(U_p, \delta_0)$ with strict inequality for $p > 0$. Then $m = U_p^{-1}(\frac{1}{2}) \ge 0$ and for $0 < p \le \frac{1}{2}$,

$$(6.2.2)\qquad \begin{aligned} &R(U_p, \delta_0) - R(U_p, \delta) \\ &\quad = \int \cdots \int_{x_{(n/2)} \le 0 \le x_{((n+2)/2)}} \left(|m - x_{(n/2)}| - |m - 0|\right) \prod dU_p(x_i) \\ &\quad\quad + \int \cdots \int_{x_{((n+2)/2)} < 0} \left(|m - x_{(n/2)}| - |m - x_{((n+2)/2)}|\right) \prod dU_p(x_i) \\ &\quad > 0, \end{aligned}$$

since both integrands in (6.2.2) are positive. When $p = 0$ then $R(U_p, \delta_0) = R(U_p, \delta)$ since then $\delta = \delta_0$ with probability 1. □

7. Modified loss functions for discrete problems.

The loss functions L_1, L_2, L_3, considered previously, are conventionally defined only for $F \in \mathscr{F}_C$. In Section 2 we extended the conventional definition in the apparently obvious manner to also apply when $F \in \mathscr{F}_D$. Corresponding admissibility results were then presented in Sections 3–5. However, there are other ways to transfer the definitions of L_i from $\mathscr{F}_C$ to $\mathscr{F}_D$. We will consider in detail only the loss function L_2 since this is the most tractable of the three. First the modified loss functions will be defined and then admissibility results will be presented.

7.1. *Modified loss functions.* We have so far been discussing estimation of $F(\cdot)$, the right-continuous version of the c.d.f. It can be argued that it is more suitable (as well as more aesthetic) to estimate the symmetric version of the c.d.f., defined by

$$(7.1.1)\qquad \bar{F}(t) = (F(t^-) + F(t^+))/2.$$

Instead of the loss function $L_2(F, a) = \int (F(t) - a(t))^2 \, dF(t)$ one then considers

$$(7.1.2)\qquad L_2'(F, a) = \int \left(\bar{F}(t) - a(t)\right)^2 dF(t).$$

Of course, when $F \in \mathscr{F}_C$ it is true that $L_2(F, a) = L_2'(F, a)$. However, when $F \in \mathscr{F}_D$ the two losses are not equal [and they are not equivalent in the sense that $L_2(F, a) = \gamma(F) L_2(F, a) + \Delta(F)$ for some functions $\gamma(\cdot) > 0$, $\Delta(\cdot)$]. There is consequently no a fortiori reason to expect that admissibility under one loss should imply admissibility under the other; and we shall see that it does not.

The loss L_2 can be modified in a different fashion. Note that L_2 can be written for $F \in \mathscr{F}_C$ in the equivalent form

$$L_2''(F, a) = \int_0^1 \left(u - a\left(F^{-1}(u)\right)\right)^2 du, \tag{7.1.3}$$

by making the substitution $u = F(t)$ in the integrand. This expression also makes sense for discrete problems, with the obvious definition of F^{-1}, namely,

$$F^{-1}(t) = \sup\{x\colon F(x) \le t\} = \inf\{x\colon F(x) > t\}. \tag{7.1.4}$$

See Ferguson (1967), page 216, for a related formula.

If F is discrete it is no longer always true that $L_2(F, a) = L_2''(F, a)$. In fact, simple calculations show that for all c.d.f.'s,

$$L_2''(F, a) = L_2'(F, a) + \Delta(F), \tag{7.1.5}$$

where

$$\Delta(F) = \sum_{\{t\colon F(T^+) > F(t^-)\}} (F(t^+) - F(t^-))^3/12.$$

REMARK 7.1.1. It follows from (7.1.5) that admissibility under loss L_2' is equivalent to admissibility under loss L_2''. Consequently, in the next section we explicitly consider admissibility only for the loss L_2'.

REMARK 7.1.2. The same sort of arguments used to justify the modifications L_2' and L_2'' of L_2 could be used to motivate consideration of

$$L_3'(F, a) = \sup|\bar{F}(t) - a(t)|$$

or

$$L_3''(F, a) = \sup_{0 \le w \le 1} |w - a(F^{-1}(w))|$$

in preference to L_3. However, it is not the case here as it was in (7.1.5) that $L_3''(F, a) = L_3'(F, a) + \Delta_3(F)$ for some $\Delta_3(\cdot)$. Hence, admissibility under L_3' is not necessarily equivalent to admissibility under L_3''.

EXAMPLE 7.1.3. It was noted in Remark 2.2.2 that $\mathscr{A}_1$ contains estimates which are not right continuous. The possible desirability of including such estimates in the action space can easily be seen in connection with the losses L_2' or L_2'' since the problem can then be understood as one of estimating $\bar{F}$ which itself is not right continuous. However, even when the loss is L_2, so that one is estimating F, it is desirable to allow estimators which are not right continuous. One reason for this is illustrated by the following simple example.

Let $\mathscr{F} = \mathscr{F}_D(I)$ and $L = L_2$. Consider the no data decision problem ($n = 0!$). Let $x_0 \in I$ and

$$\begin{aligned} a_0(t) &= \tfrac{1}{2}, \qquad t \le x_0, \\ &= 1, \qquad t > x_0. \end{aligned} \tag{7.1.6}$$

Then the (nonrandomized) estimate $\delta^* \equiv a_0$ is admissible even though a_0 is not right continuous.

To prove this assertion, let

$$(7.1.7)\qquad \begin{aligned} F_{x,y}(t) &= 0, && t < x,\\ &= \tfrac{1}{2}, && x \le t < y,\\ &= 1, && t \ge y, \end{aligned}$$

and note that

$$R_2(F_{x,y}, \delta) = 0 \quad \text{for all } x \le x_0 < y$$

if and only if $\delta = \delta^*$.

With considerably more effort one can prove $\delta^{**} \equiv a_1(t)$ is admissible under loss L_2 with $\mathscr{F} = \mathscr{F}_D((-\infty, \infty))$, where

$$(7.1.8)\qquad \begin{aligned} a_1(t) &= \tfrac{1}{4}, && t < 0,\\ &= \tfrac{1}{2}, && t = 0,\\ &= \tfrac{3}{4}, && t > 0. \end{aligned}$$

[Note that this is the best invariant estimator for the special problem in which $\mathscr{F} = \mathscr{F}_D \cap \{F\colon F(0^-) < \frac{1}{2} < F(0^+)\}$.]

As previously noted we wish to consider carefully admissibility for discrete problems with $L = L_2'$. For L_2' and $\mathscr{F} = \mathscr{F}_C$ the best invariant procedure is, of course, still $\delta_0(x) = \beta_x(\cdot)$ since L_2 and L_2' are equal when F is continuous. However, when F is discrete L_2 and L_2' are no longer always equal. Since the problem with loss L_2' can be viewed as a problem of estimating $\overline{F}$, the symmetrized version of F, it thus seems natural to investigate admissibility of the symmetrized version of δ_0. It is also necessary to take into account the end points of the domain of $\mathscr{F}$. Thus, for $\mathscr{F} = \mathscr{F}_M$ we will investigate (and disprove) admissibility of

$$(7.1.9)\qquad \delta_0'' = \beta_x''(t) = \begin{cases} \beta_x(\xi_1^+)/2, & t = \xi_1,\\ (\beta_x(t^-) + \beta_x(t^+))/2, & \xi_1 < t < \xi_m,\\ (\beta_x(\xi_m^-) + 1)/2, & t = \xi_m. \end{cases}$$

When $\mathscr{F} = \mathscr{F}_D([a, b])$ [or $\mathscr{F}_D((a, b))$] the estimator is defined similarly with a in place of ξ_1 and b in place of ξ_m. [Of course, in the case of $\mathscr{F}_D((a, b))$ the special values at $t = a$ and $t = b$ are irrelevant.]

Remark 7.1.4. The choice to investigate δ_0'', as defined in (7.1.9), seems natural on the basis of symmetry, but is otherwise a somewhat arbitrary choice. We could instead have decided to investigate admissibility of δ_0 itself, or rather of δ_0' as defined in (2.4.9). [Incidentally, while we can prove the inadmissibility in this problem of the estimator in (7.1.9), we have not been able to prove the inadmissibility of the estimator (2.4.9) in this problem, although we suspect it is indeed inadmissible.]

One might observe that L_2' and $\mathscr{F}_D((a, b))$ are invariant under monotone transformations of (a, b), and ask, "Why not decide to investigate admissibility of the best invariant estimator with respect to L_2' and $\mathscr{F}_D(a, b)$?" The answer to this question is that for $\mathscr{F} = \mathscr{F}_D$ there is no best invariant estimator under L_2'. (This fact actually holds for all of our discrete problems involving estimation of the c.d.f., not merely for the L_2 problem. To understand this fact, note that the group of strictly increasing monotone transformations is far from transitive on $\mathscr{F}_D$ so one should not expect there to exist a best invariant estimator. It is then easy to produce examples showing that no best invariant estimator exists.)

7.2. *Inadmissibility of δ_0'' for loss L_2' and L_2''.* The results here in all cases are similar to those for L_2 and $\mathscr{F} = \mathscr{F}_C$. (Perhaps this supports the claim that L_2' is the best transfer to discrete settings of the loss L_2.) Recall that δ_0' is now replaced by $\delta_0''(x) = \beta_x''(\cdot)$, with β'' defined in (7.1.9). The function $\zeta(t)$ which appears later was defined in (4.2.1).

THEOREM 7.2.1. *Let $L = L_2'$ or L_2''. Define $\delta^*(x) = d_x^*(\cdot)$ by*

$$d_x^*(t) = \beta_x''(t) + \sum_{i=1}^{n} \zeta_{x_i}(t)/2(n+1)(n+2). \tag{7.2.1}$$

Then

$$R(F, \delta_0'') - R(F, \delta^*) \geq \frac{n \Pr_F(X \leq 0) \Pr_F(X > 0)}{4(n+1)(n+2)^2} \geq 0. \tag{7.2.2}$$

[Note that (7.2.2) is the same as (4.2.5).]

PROOF. Suppose $\mathscr{F} = \mathscr{F}_M$. Note that

$$\beta_x''(t) = h(t) + \sum_{i=1}^{n} \chi_{x_i}'(t)/(n+2), \tag{7.2.3}$$

where

$$\begin{aligned} h(t) &= \frac{1}{2(n+2)}, \quad \text{if } t = \xi_1, \\ &= \frac{1}{n+2}, \quad \text{if } \xi_1 < t < \xi_m, \\ &= \frac{3}{2(n+2)}, \quad \text{if } t = \xi_m, \end{aligned}$$

and

$$\chi_x'(t) = 0, \quad \text{if } t < x,$$
$$= \tfrac{1}{2}, \quad \text{if } t = x,$$
$$= 1, \quad \text{if } t > x.$$

Define $a_1', a_2', a_4', a_5', a_6'$ as a_1, a_2, a_4, a_5, a_6 in (4.2.7)–(4.2.12) but with an arbitrary dF replacing dU_p throughout, and with χ' replacing χ and $\bar{F}$ replacing $U_p(t)$ in the integrand of a_4; and define a_3' by

$$a_3' = \int \zeta_x(t) h(t)\, dF(x)\, dF(t). \tag{7.2.4}$$

Direct calculations yield that

$$\begin{aligned} a_1' &= p(1-p)/2 = a_4', \\ a_2' &= p(1-p) = a_5', \\ a_6' &= 2p(1-p) \end{aligned} \tag{7.2.5}$$

[as in (4.2.7)–(4.2.12), where $p = \Pr_F(X \le 0)$]. Also,

$$\begin{aligned} a_3' = &-[\Pr_F(\{\xi_1\})\Pr_F(X > 0) + \Pr_F(\{\xi_m\})\Pr_F(X \le 0)]/2(n+2) \\ &\le 0. \end{aligned} \tag{7.2.6}$$

Let R'' denote the risk function corresponding to L_2''. Let $\alpha = 1/2(n+1)(n+2)$, as in Theorem 4.2.1. Then, as there,

$$\begin{aligned} &R''(F, \delta_0') - R(F, \delta) \\ &\quad = 2\alpha\left(na_4' - \frac{na_3' + na_2' + n(n-1)a_1'}{n+2}\right) + \alpha^2(na_6 + n(n-1)a_5), \end{aligned} \tag{7.2.7}$$

so that

$$\begin{aligned} R''(F, \delta_0') - R(F, \delta) &= p(1-p)n/4(n+1)(n+2)^2 \\ &\quad -2\alpha n a_3'/(n+2) \\ &\ge p(1-p)n/4(n+1)(n+2)^2. \end{aligned} \tag{7.2.8}$$

This verifies (7.2.2) in this case. The result for L_2'' is identical because of the relation (7.1.5) between L_2'' and L_2'. The proof when $\mathscr{F} = \mathscr{F}_D([a, b])$ or $\mathscr{F}_D((a, b))$ is identical except that a, b replace ξ_1, ξ_m in the definition and evaluation of a_3'. [Of course, when $\mathscr{F} = \mathscr{F}_D((a, b))$, $a_3' = 0$ since $\Pr_F(\{a\}) = 0 = \Pr_F(\{b\})$.] □

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.

BROWN, L. D. (1977). Closure theorems for sequential-design processes. In *Statistical Decision Theory and Related Topics II* (S. Gupta and D. Moore, eds.) 57–91. Academic, New York.

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

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

FARRELL, R. (1964). Estimators of a location parameter in the absolutely continuous case. *Ann. Math. Statist.* **35** 949–998.

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

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

FRIEDMAN, Y., GELMAN, A. and PHADIA, E. (1988). Best invariant estimation of a distribution function under the Kolmogorov–Smirnov loss function. *Ann. Statist.* **16** 1254–1261.

IGHODARO, A., SANTNER, T. and BROWN, L. D. (1982). Admissibility and complete class results for the multinomial estimation problem with entropy and squared error loss. *J. Multivariate Anal.* **12** 469–479.

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

MEEDEN, G., GHOSH, M. and VARDEMAN, S. (1985). Some admissible non-parametric and related finite population sampling estimators. *Ann. Statist.* **13** 811–817.

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.

YU, Q. (1986). The inadmissibility of the best invariant estimator of a distribution function. Preprint.

YU, Q. (1987). Admissibility of the empirical distribution function in the invariant problem. I and II. Preprint.

DEPARTMENT OF MATHEMATICS
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853