Metrika (2002) 56: 31–42

Metrika
© Springer-Verlag 2002

Minimax invariant estimation of a continuous distribution function under entropy loss

Leila Mohammadi, Willem R. van Zwet

Mathematical Institute, University of Leiden, P.O. Box 9512, 2300 RA Leiden, The Netherlands. (e-mails: leila@math.leidenuniv.nl, vanzwet@math.leidenuniv.nl)

Abstract. For the invariant decision problem of estimating a continuous distribution function F with two entropy loss functions, it is proved that the best invariant estimators d_0 exist and are the same as the best invariant estimator of a continuous distribution function under the squared error loss function $L(F,d) = \int |F(t) - d(t)|^2 \, dF(t)$. They are minimax for any sample size $n \geq 1$.

1. Introduction

The best invariant estimators for a continuous distribution function under monotone transformations and the weighted Cramér-von Mises loss function or more general invariant loss functions were introduced by Aggarwal (1955). Yu (1992a) proved that the best invariant estimator is minimax for $n \geq 1$ under the loss

$$L(F,d) = \int |F(t) - d(t)|^r h(F(t)) \, dF(t), \tag{1.1}$$

where $r \geq 1$, h is a nonnegative weight function and d is a nondecreasing function from $\mathbf{R}$ into $[0,1]$.

Yu and Phadia (1992) proved that the best invariant estimator is minimax for $n \geq 1$ under the Kolmogorov-Smirnov loss function

$$L(F,d) = \sup_t \{|F(t) - d(t)|\}. \tag{1.2}$$

In order to evaluate the performance of an estimator of a distribution function, we consider another measure of the distance between a probability distribution and its estimator. For this we use the well-known Kullback-Leibler divergence or entropy loss functions.

First, let Y be a random variable with density $f_\theta(\cdot)$ and consider the problem of estimating θ under the entropy loss function. If the unknown value θ of the parameter is estimated by a number δ, then the loss is given by the formula

$$L(\theta,\delta) = E_\theta\left(\ln\left[\frac{f_\theta(Y)}{f_\delta(Y)}\right]\right). \tag{1.3}$$

James and Stein (1961) introduced this loss for estimation of the multinormal variance-covariance matrix. See Brown (1968), Dey and Srinivasan (1985), Ghosh and Yang (1988), Wieczorkowski and Zieliński (1992), Yang (1992) and Kearns et al. (1994).

In some problems, (1.3) reduces to

$$L(\theta,\delta) = \frac{\theta}{\delta} - \ln\frac{\theta}{\delta} - 1, \tag{1.4}$$

or a multiple of this loss (for example when Y has a gamma distribution with scale parameter θ).

When developing decision theoretic nonparametric results, one often converts the nonparametric problems to parametric problems with binomial or multinomial distributions. Yu (1992b) and Yu and Kuo (1995) investigated the connection between the nonparametric problem of estimating an unknown continuous (or discrete) distribution function and the parametric problem of estimating the probability of a success of a binomial distribution.

Let Z be a random variable such that

$$P_\theta(Z = z) = \binom{n}{i}\theta^z(1-\theta)^{n-z}, \quad z = 0, \ldots, n,$$

where $\theta \in [0,1]$ is the unknown parameter ($Z \sim B(n,\theta)$). In the problem of estimating θ by a number $\delta \in [0,1]$, the entropy loss function is given by the formula

$$\begin{aligned} L(\theta,\delta) &= \sum_{z=0}^{n} \ln\frac{P_\theta(Z=z)}{P_\delta(Z=z)} P_\theta(Z=z) \\ &= n\left\{\theta\ln\frac{\theta}{\delta} + (1-\theta)\ln\frac{1-\theta}{1-\delta}\right\}, \end{aligned} \tag{1.5}$$

where $c\ln 0$ is defined to be 0 and $+\infty$ for $c = 0$ and $c < 0$, respectively. See e.g., Kullback (1959), Renyi (1962), Sakamoto et al. (1986) and Zieliński (1990).

A general form of the entropy loss functions, in the nonparametric problem (which we call the binomial-entropy loss), is as follows

$$L(F,d) = n\int\left\{F(t)\ln\left(\frac{F(t)}{d(t)}\right) + (1-F(t))\ln\left(\frac{1-F(t)}{1-d(t)}\right)\right\} dW(t), \tag{1.6}$$

where W is a finite measure (cf. (1.5)).

In this paper, we consider the nonparametric problem of estimating an absolutely continuous distribution function F under two invariant entropy loss functions, which are modified versions of the losses (1.4) and (1.5), to wit

$$L(F,d) = \int \left(\frac{F(t)}{d(t)} - \ln\left(\frac{F(t)}{d(t)}\right) - 1 \right) dF(t) \tag{1.7}$$

and

$$L(F,d) = n \int \left\{ F(t) \ln\left(\frac{F(t)}{d(t)}\right) + (1 - F(t)) \ln\left(\frac{1 - F(t)}{1 - d(t)}\right) \right\} dF(t). \tag{1.8}$$

Note that the loss (1.8) is the binomial-entropy loss function (1.6) with $W = F$. It is symmetric (i.e., $L(F,d) = L(1-F, 1-d)$) whereas the loss (1.7) is asymmetric. Note also that these loss functions are of the form

$$L(F,d) = \int l(F(t), d(t))\, dF(t), \tag{1.9}$$

where $l(x,y)$ is a nonnegative, continuous function that is differentiable as a function of y and has a unique minimum at $y_0 \in (0,1)$.

In Section 2 we obtain the best invariant estimator of F under the losses (1.7) and (1.8). These estimators are the same as the best invariant estimator of F, under the loss (1.1) with $r = 2$ and $h \equiv 1$. In Section 3 we prove that the best invariant estimator is minimax under the losses (1.7) and (1.8) for $n \geq 1$.

2. The best invariant estimator

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random sample of size n from an unknown absolutely continuous distribution function F. Without loss of generality, we assume F to have support on $(0,1)$. Suppose $\mathbf{Y} = (Y_1, \ldots, Y_n)$ is the vector of order statistics of $X_1, \ldots, X_n$, $Y_0 = 0$ and $Y_{n+1} = 1$. The action space and parameter space are given by

$$\mathscr{A} = \{d(t) : d(t) \text{ is a nondecreasing function from } (0,1) \text{ into } [0,1]\}$$

and

$$\Theta_c = \{P : P \text{ is an absolutely continuous distribution on } (0,1)\}.$$

Alternatively, we shall indicate a distribution $P \in \Theta_c$ by its distribution function F and by an abuse of notation write e.g. $F \in \Theta_c$. We shall also restrict attention to nonrandomized estimators which are functions of the order statistics $\mathbf{Y}$, since they form an essentially complete class (see for example Yu & Chow (1991)). For convenience, we write $d = d(t) = d(\mathbf{X}, t) = d(\mathbf{Y}, t)$. Also, let $\mathscr{G}$ be the group of transformations

$$\mathscr{G} = \{g_\varphi : g_\varphi(y_1, \ldots, y_n) = (\varphi(y_1), \ldots, \varphi(y_n))\},$$

where φ is a continuous and strictly increasing function from $\mathbf{R}$ into $\mathbf{R}$. Then the losses (1.7) and (1.8) and decision problem are invariant under $\mathscr{G}$ (see Ferguson (1967)).

Any invariant estimator of F is of the form

$$d(\mathbf{Y}, t) = \sum_{i=0}^{n} u_i 1(Y_i \leq t < Y_{i+1}), \quad 0 \leq u_0 \leq u_1 \leq \cdots \leq u_n \leq 1, \tag{2.1}$$

where $1(A)$ is the indicator function of the set A and $u_1, \ldots, u_n$ are constants.

The best invariant estimator of F under the loss (1.7) is obtained as follows. For any invariant estimator d,

$$\begin{aligned} L(F,d) &= \int \left(\frac{F(t)}{d(t)} - \ln\left(\frac{F(t)}{d(t)}\right) - 1 \right) dF(t) \\ &= \sum_{i=0}^{n} \int_{Y_i}^{Y_{i+1}} \left(\frac{F(t)}{d(t)} - \ln\left(\frac{F(t)}{d(t)}\right) - 1 \right) dF(t) \\ &= \sum_{i=0}^{n} \int_{F(Y_i)}^{F(Y_{i+1})} \left(\frac{t}{u_i} - \ln\left(\frac{t}{u_i}\right) - 1 \right) dt \\ &= \sum_{i=0}^{n} L_i(F,d) \end{aligned}$$

with $Z_i = F(Y_i)$ and $L_i(F,d) = \int_{Z_i}^{Z_{i+1}} \left(\frac{t}{u_i} - \ln\left(\frac{t}{u_i}\right) - 1 \right) dt$. We have

$$R(F,d) = E(L(F,d)) = \sum_{i=0}^{n} E(L_i(F,d)) = \sum_{i=0}^{n} R_i(F,d),$$

where

$$\begin{aligned} R_i(F,d) &= E\left(\int_{Z_i}^{Z_{i+1}} \left(\frac{t}{u_i} - \ln\left(\frac{t}{u_i}\right) - 1 \right) dt \right) \\ &= \int_0^1 \int_0^{Z_{i+1}} \int_{Z_i}^{Z_{i+1}} \left(\frac{t}{u_i} - \ln\left(\frac{t}{u_i}\right) - 1 \right) dt \, dF_{Z_i, Z_{i+1}}(z_i, z_{i+1}) \\ &= \int_0^1 \int_t^1 \int_0^t \left(\frac{t}{u_i} - \ln\left(\frac{t}{u_i}\right) - 1 \right) dF_{Z_i, Z_{i+1}}(z_i, z_{i+1}) \, dt \\ &= \int_0^1 \left(\frac{t}{u_i} - \ln\left(\frac{t}{u_i}\right) - 1 \right) \int_t^1 \int_0^t dF_{Z_i, Z_{i+1}}(z_i, z_{i+1}) \, dt \\ &= \int_0^1 \left(\frac{t}{u_i} - \ln\left(\frac{t}{u_i}\right) - 1 \right) \binom{n}{i} t^i (1-t)^{n-i} \, dt \\ &= \frac{\binom{n}{i}}{u_i} \int_0^1 t^{i+1} (1-t)^{n-i} \, dt - \binom{n}{i} \int_0^1 (\ln t) t^i (1-t)^{n-i} \, dt \\ &\quad + [(\ln u_i) - 1] \binom{n}{i} \int_0^1 t^i (1-t)^{n-i} \, dt. \end{aligned} \tag{2.2}$$

Note also that

$$\frac{\partial}{\partial u_i} R_i(F,d) = -\frac{\binom{n}{i}}{u_i^2} \int_0^1 t^{i+1} (1-t)^{n-i} \, dt + \frac{\binom{n}{i}}{u_i} \int_0^1 t^i (1-t)^{n-i} \, dt.$$

If $\frac{\partial}{\partial u_i} R_i(F, d) = 0$, then

$$u_i = \frac{\int_0^1 t^{i+1}(1-t)^{n-i}\,dt}{\int_0^1 t^i(1-t)^{n-i}\,dt}$$

or

$$u_i = \frac{i+1}{n+2}, \quad i = 0, 1, \ldots, n,$$

and $\frac{\partial^2}{\partial u_i^2} R_i(F, d) > 0$. Thus the best invariant estimator of F under the loss (1.7) is

$$d_0 = d_0(\mathbf{Y}, t) = \sum_{i=0}^{n} \frac{i+1}{n+2} 1(Y_i \le t < Y_{i+1}), \quad i = 0, \ldots, n. \tag{2.3}$$

Similarly, we can obtain the best invariant estimator of F under the loss (1.8) which is the same as in (2.3). Let Θ be the class of all probability distributions on $\mathbf{R}$ with continuous distribution functions. Then d_0 defined by (2.3) is also the best invariant estimator of F under the losses (1.7) and (1.8) and the parameter space Θ. Notice that this estimator is the best invariant estimator of F under the quadratic loss (1.1) (with $r = 2$ and $h \equiv 1$) (see Yu (1992c) or Aggarwal (1955)).

3. Minimaxity of the best invariant estimator

In this section, it is proved that the best invariant estimator d_0 is minimax for all sample sizes $n \ge 1$ under the losses (1.7) and (1.8) and $P \in \Theta_c$. Here is the notation that we use hereafter.

Let U_0 be the set of all classical invariant estimators (the estimators of the form (2.1)) and V be the set of all estimators with finite risks for all $P \in \Theta_c$. Given a one-dimensional measurable set B, let B^k denote the k-fold product set $B \times \cdots \times B$. We denote the Lebesgue measure by m, and let m^k denote the k-fold product measure $m \times \cdots \times m$. By a.e. $[m]$, we mean almost everywhere with respect to Lebesgue measure.

The proof of our main result depends on the following results. A k-dimensional hypercube is a set $C = \{\mathbf{x} = (x_1, \ldots, x_k) : a_i \le x_i \le a_i + r,\ i = 1, \ldots, k\}$ with $r > 0$. The following lemma is an immediate consequence of Lebesgue's differentiation theorem (cf. Zaanen (1958), p. 156).

Lemma 3.1. *For every Lebesgue measurable set $B \subset \mathbf{R}^k$ with $m^k(B) > 0$ and every $\alpha \in (0, 1)$, there exists a hypercube $C \subset \mathbf{R}^k$ with $m^k(C \cap B) \ge (1 - \alpha)m^k(C) > 0$.*

For a Lebesgue measurable function $f : \mathbf{R}^k \to \mathbf{R}$, the L^∞-norm or the essential supremum of $|f|$ is defined by

$$\begin{aligned}\|f\|_\infty &= \inf\{M : |f| \le M \text{ a.e. } [m^k]\} \\ &= \inf\{M : m^k\{\mathbf{x} \in \mathbf{R}^k : |f(\mathbf{x})| > M\} = 0\},\end{aligned}$$

with the convention that $\|f\|_\infty = \infty$ if and only if for every $M > 0$, $m^k\{\mathbf{x} \in \mathbf{R}^k : |f(\mathbf{x})| > M\} > 0$. It is easy to see (see Dudley (1989), p. 118) that $|f| \le \|f\|_\infty$ a.e. $[m^k]$. We shall prove

Lemma 3.2. *Let $f : \mathbf{R}^k \to \mathbf{R}^+$ be a nonnegative Lebesgue measurable function with $\|f\|_\infty = \infty$. Then there exists an absolutely continuous probability measure P on $\mathbf{R}$ such that for the k-fold product measure $P^k = P \times \cdots \times P$ on $\mathbf{R}^k$ $\int f\, dP^k = \infty$.*

Remark 3.1. Note that Lemma 3.2 is trivial if P^k is allowed to be any absolutely continuous probability distribution on $\mathbf{R}^k$ rather than a product distribution with identical marginals.

Proof of Lemma 3.2: Choose a positive integer q and define the Lebesgue measurable set

$$B = \{\mathbf{x} \in \mathbf{R}^k : f(\mathbf{x}) \ge 2q(k2^q)^k\}.$$

As f is nonnegative and $\|f\|_\infty = \infty$, B has positive Lebesgue measure. According to Lemma 3.1, there exists a hypercube $C = \{\mathbf{x} = (x_1, \dots, x_k) : a_i \le x_i \le a_i + r, i = 1, \dots, k\}$ with $r > 0$, such that $m^k(C \cap B) \ge (1/2)m^k(C) > 0$. Let $A = \bigcup_{i=1}^k [a_i, a_i + r]$ denote the union of the 1-dimensional intervals $[a_i, a_i + r]$ and let μ be the absolutely continuous measure on $\mathbf{R}$ defined for any Lebesgue measurable set $D \subset \mathbf{R}$ by $\mu(D) = (2^q m(A))^{-1} m(D \cap A)$. Thus μ distributes a total measure 2^{-q} uniformly over the union of the intervals $[a_i, a_i + r]$ and because these intervals have equal lengths, each is assigned measure $\mu([a_i, a_i + r]) \ge (k2^q)^{-1}$, $i = 1, \dots, k$. It follows that the k-dimensional product measure μ^k assigns measure $\mu^k(C) = \prod_i \mu([a_i, a_i + r]) \ge (k2^q)^{-k}$ to the k-dimensional hypercube C and distributes this measure uniformly over C. In other words,

$$\mu^k(C') \ge (k2^q)^{-k}[m^k(C')/m^k(C)]$$

for any Lebesgue measurable set $C' \subset C$. Hence

$$\mu^k(B) \ge \mu^k(C \cap B) \ge (k2^q)^{-k}[m^k(C \cap B)/m^k(C)] \ge (1/2)(k2^q)^{-k}$$

and

$$\int f\, d\mu^k \ge \inf\{f(\mathbf{x}) : \mathbf{x} \in B\} \cdot \mu^k(B) \ge q.$$

If we now introduce the dependence on the positive integer q in our notation, we have constructed an absolutely continuous measure μ_q on $\mathbf{R}$ with $\mu_q(\mathbf{R}) = 2^{-q}$ and such that $\int f\, d\mu_q^k \ge q$. Carrying out this construction for $q = 1, 2, \dots,$ and defining $P = \sum_q \mu_q$, we see that P is an absolutely continuous probability measure with $\int f\, dP^k \ge \int f\, d\mu_q^k \ge q$ for any $q = 1, 2, \dots$. $\square$

Remark 3.2. Of course Lemma 3.2 continues to hold for $f : [0, 1]^k \to \mathbf{R}^+$ and in that case P is an absolutely continuous probability measure on $[0, 1]$.

Lemma 3.3. *Let $n \geq 1$ and d be a nonrandomized estimator of $F \in \Theta_c$. If $d \in V$ under the loss (1.7), then $\left\|\frac{1}{d}\right\|_{\infty} < \infty$. Hence there exists a number $l_d \in (0,1)$ such that $d(\mathbf{X},t) \geq l_d$ a.e. $[m^{n+1}]$.*

Proof: Let $\left\|\frac{1}{d}\right\|_{\infty} = +\infty$. According to Lemma 3.2, there exists an absolutely continuous distribution function F on $(0,1)$ such that $E_F\left(\int \frac{1}{d(\mathbf{X},t)}\, dF(t)\right) = +\infty$. Let $G(t) = \sqrt{F(t)}$. Then $G \in \Theta_c$ and $dF(t) = 2G(t)\, dG(t)$. Now,

$$
\begin{aligned}
E_G\left(\int \frac{G(t)}{d(\mathbf{X},t)} dG(t)\right) &\geq E_G\left(\int \frac{G(t)G(X_1)\ldots G(X_n)}{d(\mathbf{X},t)} dG(t)\right) \\
&= E_F\left(\int \frac{1}{2^{n+1}\, d(\mathbf{X},t)} dF(t)\right) \\
&= +\infty.
\end{aligned}
$$

Note that the function $1 - \frac{\ln x}{x}$ is minimized when $x = e$, hence

$$
\begin{aligned}
R(G,d) &= E_G\left\{\int_0^1 \left(\frac{G(t)}{d(t)} - \ln\frac{G(t)}{d(t)} - 1\right) dG(t)\right\} \\
&= E_G\left\{\int_0^1 \frac{G(t)}{d(t)}\left(1 - \frac{\ln\frac{G(t)}{d(t)}}{\frac{G(t)}{d(t)}}\right) dG(t)\right\} - 1 \\
&\geq \left(1 - \frac{1}{e}\right) E_G\left\{\int_0^1 \frac{G(t)}{d(t)} dG(t)\right\} - 1 \\
&= +\infty.
\end{aligned}
$$

This is a contradiction with $d \in V$. Consequently $\left\|\frac{1}{d}\right\|_{\infty} < \infty$. $\square$

Remark 3.3. Lemma 3.3 implies that under the loss (1.7),

$$
V = V_1 = \{d(t) : \exists l_d \in (0,1), l_d \leq d(\mathbf{X},t) \text{ a.e. } [m^{n+1}]\}.
$$

Lemma 3.4. *For any two estimators d and d^* in V_1 and $F \in \Theta_c$,*

$$
|l(F(t),d(t)) - l(F(t),d^*(t))| \leq \frac{2|d(t) - d^*(t)|}{(\min\{l_d, l_{d^*}\})^2}, \quad a.e.\ [m^{n+1}], \tag{3.1}
$$

with $l(x,y) = \frac{x}{y} - \ln\frac{x}{y} - 1$.

Proof: Let $x \in (0,1)$ and $y, z \in [r,1)$ with $r \in (0,1)$. Then

$$
\begin{aligned}
|l(x,y) - l(x,z)| &= \left|\frac{x}{y} - \ln\frac{x}{y} - \frac{x}{z} + \ln\frac{x}{z}\right| \\
&\le x\frac{|z-y|}{r^2} + |\ln y - \ln z| \\
&\le \frac{|z-y|}{r^2} + \frac{|z-y|}{r} \\
&\le \frac{2|z-y|}{r^2}.
\end{aligned}
$$

Taking $r = \min\{l_d, l_{d^*}\}$, (3.1) follows. $\square$

Lemma 3.5. Let $n \ge 1$ and d be a nonrandomized estimator of $F \in \Theta_c$. If $d \in V$ under the loss (1.8), then $\|-\ln d\|_\infty < \infty$ and $\|-\ln(1-d)\|_\infty < \infty$. Hence there exists a number $l_d \in (0, 1/2)$ such that $l_d \le d(\mathbf{X}, t) \le 1 - l_d$ a.e. $[m^{n+1}]$.

Proof: Let $\|-\ln d\|_\infty = \infty$. According to Lemma 3.2, there exists an absolutely continuous distribution function F on $(0,1)$ such that $E_F(\int \ln d(\mathbf{X}, t)\, dF(t)) = -\infty$. Let $G(t) = \sqrt{F(t)}$. Then $G \in \Theta_c$ and $dF(t) = 2G(t)\, dG(t)$. Now,

$$
\begin{aligned}
&-E_G\left(\int \ln d(\mathbf{X}, t) G(t)\, dG(t)\right) \\
&\qquad \ge -E_G\left(\int \ln d(\mathbf{X}, t) G(t) G(X_1) \dots G(X_n)\, dG(t)\right) \\
&\qquad = -\frac{1}{2^{n+1}} E_F\left(\int \ln d(\mathbf{X}, t)\, dF(t)\right) \\
&\qquad = +\infty.
\end{aligned}
$$

For this G, under the loss (1.8),

$$
\begin{aligned}
R(G, d) = E_G&\left\{\int_0^1 \left(G(t)\ln\left(\frac{G(t)}{d(t)}\right) + (1 - G(t))\ln\left(\frac{1-G(t)}{1-d(t)}\right)\right) dG(t)\right\} \\
&= -E_G\left\{\int_0^1 G(t)\ln d(t)\, dG(t)\right\} \\
&\quad - E_G\left\{\int_0^1 (1-G(t))\ln(1-d(t))\, dG(t)\right\} \\
&\quad + \int_0^1 (G(t)\ln G(t) + (1-G(t))\ln(1-G(t)))\, dG(t).
\end{aligned}
$$

Note that

$$-E_G\left\{\int_0^1 (1-G(t))\ln(1-d(t))\,dG(t)\right\} \ge 0$$

and

$$\left|\int_0^1 (G(t)\ln G(t) + (1-G(t))\ln(1-G(t)))\,dG(t)\right| < \infty.$$

Since $-E_G(\int \ln d(\mathbf{X},t)G(t)\,dG(t)) = +\infty$, it follows that $R(G,d) = +\infty$. This is a contradiction with $d \in V$. Consequently $\|-\ln d\|_\infty < \infty$. By the symmetry of the loss (1.8), $\|-\ln(1-d)\|_\infty < \infty$. $\square$

Remark 3.4. Lemma 3.5 implies that under the loss (1.8),

$$V = V_2 = \{d(t) : \exists l_d \in (0, 1/2), l_d \le d(\mathbf{X},t) \le 1 - l_d \text{ a.e. } [m^{n+1}]\}.$$

Lemma 3.6. *For any two estimators d and d^* in V_2, (3.1) holds with $l(x,y) = x\ln(x/y) + (1-x)\ln\{(1-x)/(1-y)\}$.*

Proof: Let $x \in (0,1)$ and $y, z \in [r, 1-r]$ with $0 < r < 1/2$. Then

$$\begin{aligned}|l(x,y) - l(x,z)| &= \left|x\ln\frac{x}{y} + (1-x)\ln\frac{1-x}{1-y} - x\ln\frac{x}{z} - (1-x)\ln\frac{1-x}{1-z}\right| \\ &\le x|\ln z - \ln y| + (1-x)|\ln(1-y) - \ln(1-z)| \\ &\le 2\frac{|z-y|}{r} \\ &\le 2\frac{|z-y|}{r^2}\end{aligned}$$

Taking $r = \min\{l_d, l_{d^*}\}$, (3.1) follows. $\square$

Theorem 3.1 (Yu and Chow 1991, Theorem 4). *Suppose that $d = d(\mathbf{Y},t) \in V$ is a nonrandomized estimator and that it is a (measurable) function of the order statistics $\mathbf{Y} = (Y_1, \ldots, Y_n)$. For any $\varepsilon, \delta > 0$, there exists a uniform distribution P on a Lebesgue-measurable subset $J \subset \mathbf{R}$ and an invariant estimator d_1 such that*

$$P^{n+1}\{(\mathbf{Y},t) : |d(\mathbf{Y},t) - d_1(\mathbf{Y},t)| \ge \varepsilon\} \le \delta. \tag{3.2}$$

Remark 3.5. Consider the loss (1.7). In Theorem 3.1, we may assume that $d_1 \in V_1$ and satisfies $d_1 \ge l_d$, where l_d is the a.e. lower bound for d in Lemma 3.3. To see this, suppose $d_1(t) = \sum_{i=0}^n v_i 1(Y_i \le t < Y_{i+1})$ and $d_1 \notin V$. Then obviously $v_0 = 0$ (see (2.2)). Take $w_i = \max\{v_i, l_d\}$ for $i = 0, \ldots, n$ and define

$$d_2(\mathbf{Y},t) = \sum_{i=0}^n w_i 1(Y_i \le t < Y_{i+1}).$$

Then $d_2 \in U_0 \cap V$ and

$$|d(\mathbf{Y},t) - d_1(\mathbf{Y},t)| \geq |d(\mathbf{Y},t) - d_2(\mathbf{Y},t)|, \quad \text{a.e. } [m^{n+1}].$$

Hence, for any $\varepsilon, \delta > 0$ and P as in Theorem 3.1,

$$\begin{aligned} P^{n+1}&\{(\mathbf{Y},t) : |d(\mathbf{Y},t) - d_2(\mathbf{Y},t)| \geq \varepsilon\} \\ &\leq P^{n+1}\{(\mathbf{Y},t) : |d(\mathbf{Y},t) - d_1(\mathbf{Y},t)| \geq \varepsilon\} \\ &\leq \delta. \end{aligned}$$

Similarly, we can assume that d_1 in Theorem 3.1 is in V_2 and satisfies $l_d \leq d_1 \leq 1 - l_d$, under the loss (1.8).

Lemma 3.7. *Suppose that the sample size $n \geq 1$ and $\varepsilon > 0$. Consider the loss (1.7) or (1.8). Then, given any $d \in V$ there exist $F^* \in \Theta_c$ and $d^* \in U_0 \cap V$ such that $|R(F^*,d) - R(F^*,d^*)| \leq \varepsilon$.*

Proof: We first consider the case when the loss is given by (1.7), so $L(F,d) = \int l(F,d)\,dF$ with $l(x,y) = \frac{x}{y} - \ln\frac{x}{y} - 1$. Let $d \in V$. By Theorem 3.1 and Remark 3.5, there exists $P^* \in \Theta_c$ and $d^* \in U_0 \cap V$ with $d^* \geq l_d$ such that

$$P^{*^{n+1}}\{(\mathbf{Y},t) : |d(\mathbf{Y},t) - d^*(\mathbf{Y},t)| \geq \varepsilon l_d^2/4\} \leq \varepsilon l_d^2/4. \tag{3.3}$$

Note that $d^* \geq l_d$ implies that we may choose $l_{d^*} = l_d$ in Lemma 3.4, so that (3.1) holds with $\min\{l_d, l_{d^*}\}$ replaced by l_d. Thus if $B = \{(\mathbf{Y},t) : |d(t) - d^*(t)| \geq \varepsilon l_d^2/4\}$ and F^* denote the distribution function corresponding to P^*, then

$$\begin{aligned} |R(F^*,d) - R(F^*,d^*)| &= |E\{L(F^*,d)\} - E\{L(F^*,d^*)\}| \\ &\leq E\left\{\int |l(F^*,d) - l(F^*,d^*)|\,dF^*(t)\right\} \\ &\leq \frac{2}{l_d^2} E\left\{\int |d - d^*|\,dF^*(t)\right\} \\ &\leq \frac{2}{l_d^2}\int_B |d - d^*|\,dF^*(t)\,dF^*(\mathbf{x}) + \frac{2}{l_d^2}\frac{l_d^2\varepsilon}{4} \\ &\leq \frac{2}{l_d^2} P^{*^{n+1}}(B) + \frac{\varepsilon}{2} \\ &\leq \varepsilon. \end{aligned}$$

The proof for the loss (1.8) uses Lemma 3.6 instead of Lemma 3.4. □

Now we can prove the minimaxity of d_0.

Theorem 3.2. *For sample size $n \geq 1$, under the losses (1.7) and (1.8) and for Θ_c and $\mathscr{A}$ as in Section 2, the best invariant estimator d_0 is minimax.*

Proof: Take $\varepsilon > 0$ and let $d(\mathbf{Y}, t)$ be an estimator with finite risk for all $F \in \Theta_c$. By Lemma 3.7, there exists $F \in \Theta_c$ and $d^* \in U_0 \cap V$ such that $|R(F,d) - R(F,d^*)| \leq \varepsilon$. Because $d_0, d^* \in U_0$, both have constant risk and as d_0 is the best invariant estimator, it follows that $R(F,d_0) \leq R(F,d^*) \leq R(F,d) + \varepsilon$. Now, since d and ε are arbitrary, it follows that $\inf_d \sup_{F \in \Theta} R(F,d) = R(F,d_0)$. $\square$

Remark 3.7. As d_0 also has constant risk for the larger class of probability distributions on $\mathbf{R}$ with continuous distribution functions, Theorem 3.2 implies that d_0 is minimax under the losses (1.7) and (1.8) for Θ and $\mathscr{A}$. Minimax results are usually formulated for Θ rather than Θ_c.

Remark 3.8. For the parametric problem of estimating a real valued parameter θ, several authors use the entropy loss function

$$L(\theta,\delta) = \frac{\delta}{\theta} - \ln\frac{\delta}{\theta} - 1$$

rather than (1.4). If, in analogy, we consider the loss (1.9) with $l(x,y) = \frac{y}{x} - \ln\frac{y}{x} - 1$, and d is an invariant estimator of the form (2.1), then (cf. (2.2))

$$\begin{aligned} R(F,d) &= \sum_{i=0}^{n} \binom{n}{i} \int_0^1 \left(\frac{u_i}{t} - \ln\frac{u_i}{t} - 1\right) t^i (1-t)^{n-i}\, dt \\ &\geq \int_0^1 \left(\frac{u_0}{t} - \ln\frac{u_0}{t} - 1\right)(1-t)^n\, dt \\ &= +\infty. \end{aligned}$$

Therefore the risk of the best invariant estimator is not finite.

References

Aggarwal OP (1955) Some minimax invariant procedures for estimating a cumulative distribution function. Ann. Math. Statist. 26:450–462

Brown LD (1968) Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters. Ann. Math. Statist. 39:29–48

Dey DK, Srinivasan C (1985) Estimation of a covariance matrix under Stein's loss. Ann. Statist. 13:1581–1591

Dudley RM (1989) Real analysis and probability. Wadsworth & Brooks/Cole. Pacific Grove, California

Ferguson TS (1967) Mathematical statistics. A Decision Theoretic Approach. Academic Press, New York

Ghosh M, Yang MC (1988) Simultaneous estimation of Poisson means under entropy loss. Ann. Statist. 16:278–291

James W, Stein C (1961) Estimation with quadratic loss. Proc. 4th Berkeley Symp. Math. Statist. Probab., 1. Univ. California Press, Berkeley, CA, 361–379

Kearns M, Mansour Y, Ron D, Rubinfeld R, Schapire RE, Sellie L (1994) On the learnability of discrete distributions. In proceedings of the twenty-sixth annual ACM symposium on the theory of computing, 273–282

Kullback S (1959) Information theory and statistics. Wiley, New York

Renyi A (1962) Wahrscheinlichkeitsrechnung mit einem Anbang uber Informationstheorie. VEB Deutscher Verlag der Wissenschaften, Berlin

Sakamoto Y, Ishiguro M, Kitagawa G (1986) Akaike information criterion statistics. KTK Sciientific Publisher Tokyo and D. Reidel Publishing Company

Wieczorkowski R, Zieliński R (1992) Minimax estimation of binomial probability with entropy loss function. Statist. Dec. 10:39–44

Yang MC (1992) Ridge estimation of independent Poisson means under entropy loss. Statist. Dec. 10:1–23

Yu Q, Chow MS (1991) Minimaxity of the empirical distribution function in invariant estimation. Ann. Statist. 19:935–951

Yu Q (1992a) Minimax invariant estimator of a continuous distribution function. Ann. Inst. Statist. Math. 44:729–735

Yu Q (1992b) A general method of finding a minimax estimator of a distribution function when no equalizer rule is available. Canad. J. Statist. 20:281–290

Yu Q (1992c) Inadmissibility of the best invariant estimator of a distribution function. Sankhyā, Series A 54:74–79

Yu Q, Phadia EG (1992) Minimaxity of the best invariant estimator of a distribution function under the Kolmogorov-Smirnov loss. Ann. Statist. 20:2192–2195

Yu Q, Kuo L (1995) An analogy between nonparametric problems of estimating a distribution function and their parametric versions. Sankhyā, Series A 57:472–485

Zaanen AC (1958) An introduction to the theory of integration. North-Holland Publishing Company, Amsterdam