Ann. Inst. Statist. Math.
Vol. 44, No. 4, 729–735 (1992)

MINIMAX INVARIANT ESTIMATOR OF A CONTINUOUS DISTRIBUTION FUNCTION*

QIQING YU

Department of Applied Mathematics, State University of New York, Stony Brook, NY 11794, U.S.A.

(Received February 1, 1991; revised August 14, 1991)

Abstract. Consider the problems of the continuous invariant estimation of a distribution function with a wide class of loss functions. It has been conjectured for long that the best invariant estimator is minimax for all sample sizes $n \geq 1$. This conjecture is proved in this short note.

Key words and phrases: Minimaxity, invariant estimator, nonparametric estimator, product measure, Lebesgue measure, uniform distribution on a set.

1. Introduction

This short note solves the conjecture that the best invariant estimator of a distribution function is minimax in the finite sample size invariant decision problem, involving a wide class of loss functions. This is a well-known conjecture. The formulation of the problem, introduced by Aggarwal (1955), is as follows.

Let $X_1, \dots, X_n$ be a sample of size n from an unknown continuous distribution function F. Let $Y_0, \dots, Y_{n+1}$ be the order statistics of $0, X_1, \dots, X_n, 1$. Write $\vec{Y} = (Y_1, \dots, Y_n)$. Let $A = \{a(t) : a(t)$ is a nondecreasing function from $(0,1)$ into $[0,1]\}$ be the action space; and $\Theta = \{F : F$ is a continuous distribution function with support in $(0,1)\}$ the parameters space. Let $L(F, a)$ be the loss function, where

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

with $k \in [1, \infty)$ and $h(t) \geq 0$.

The decision problem of estimating F is invariant under monotone transformations. The invariant estimators have the form

$$d(\vec{Y}, t) = \sum_{j=0}^{n} u_j 1(Y_j \leq t < Y_{j+1}), \tag{1.2}$$

* Partially supported by National Science Foundation Grant DMS 9001194.

where $1(E)$ is the indicator function of a set E. $d(\vec{Y}, t)$ is of constant risk. So the best invariant estimator, denoted by d_0, exists.

A special case of the loss (1.1) is the weighted Cramer-von Mises loss:

$$L(F, a) = \int (F(t) - a(t))^2 h(F(t)) dF(t) \tag{1.3}$$

where $h(t) = t^{-1}(1-t)^{-1}$. In this case, the best invariant estimator is the empirical distribution function $\hat{F}(t)$ (see Aggarwal (1955)). Also, it is asymptotically minimax (see Dvoretzky *et al.* (1956)), is admissible iff the sample size n is 1 or 2 (see Yu (1989*a*, 1989*b*, 1989*d*)), and is minimax (see Yu (1989*b*, 1989*d*) and Yu and Chow (1991)).

Much study has been devoted to the theoretical properties of the best invariant estimator under the above set-up with general $h(t)$ for the loss function (1.3). Whether or not the best invariant estimator is admissible was an interesting open question until 1984 (see, for example, Cohen and Kuo (1985)). When $h(t) = 1$, Brown (1988) proved that the best invariant estimator is inadmissible for all sample sizes $n \geq 1$. When $h(t) = t^{\alpha}(1-t)^{\beta}$, $\alpha, \beta \geq -1$ in the loss (1.3), Yu (1988, 1989*a*) extended Brown's result and proved the inadmissibility of the best invariant estimator in the case $\alpha, \beta \in (-1, 0]$ for $n \geq 1$. Also, Yu (1989*a*) proved the inadmissibility of the best invariant estimator in the case $n \geq 2$, $\alpha = -1$ and $\beta = 0$, or $\alpha = 0$ and $\beta = -1$.

Whether or not the best invariant estimator is minimax has been an outstanding open question (see, for example, Ferguson (1967)). It has been conjectured for a long time that the best invariant estimator d_0 is minimax. Yu (1989*c*) gave a proof of the minimaxity of the best invariant estimators under loss (1.1) for $n = 1$. Yu and Chow (1991) extended the idea used in Yu (1989*c*) and proved that under loss (1.3) with $h(t) = t^{-1}(1-t)^{-1}$, the empirical distribution function is minimax for $n \geq 3$. In Theorem 2.3 of Section 2, the conjecture is proved in the affirmative under loss (1.1) for $n \geq 2$. The main idea of the proof of the main result is the same as that in Yu (1989*c*). Note that the situation in study on the minimaxity of d_0 is different from that in the study on the admissibility. In the latter study, we can only get results for some special cases of the loss (1.3) and we have different conclusion if the $h(t)$ in (1.3) is different or the sample size n is different, whereas we give a unified proof for the minimaxity of d_0 under the whole class of loss functions (1.1) including (1.3).

2. Minimax results

In this section, the set-up is as in Section 1 with loss (1.1). It is proved in Theorem 2.3 that d_0 is minimax.

Given a distribution function $F(t)$, let dF denote the measure induced by F, i.e., $dF\{(a, b)\} = F(b) - F(a)$; let $(dF)^j$ denote the product measure $dF \times \cdots \times dF$ with j factors, $j = 2, 3, \ldots$. Given a one-dimension measurable set B, let B^k denote the product set $B \times \cdots \times B$ with k factors. We denote Lebesgue measure by m. By a.e.m, we mean almost everywhere w.r.t Lebesgue measure. Note that according to our notation, given a measurable set B in R^n, $m^n\{\vec{Y} \in B\} \neq$

$m^n\{(X_1, \ldots, X_n) \in B\}$. For example, when $n = 3$, $m^3\{(Y_1, Y_2, Y_3) : Y_1 < 1/2 < Y_2\} = 3! m^3\{(X_1, X_2, X_3) : X_1 < 1/2 < X_2 < X_3\}$. A uniform distribution function $F(t)$ on a positive-Lebesgue-measure subset I is defined by

$$F(t) = \int_{-\infty}^{t} 1(x \in I)/m(I)dx.$$

The proof of the main minimax result depends heavily on Theorems 2.1 and 2.2.

THEOREM 2.1. (Yu and Chow (1991)) *Suppose that the sample size is n (≥ 1) and $d = d(\vec{Y}, t)$ is a nonrandomized estimator with finite risk and is a (measurable) function of the order statistic $\vec{Y}$. For any $\delta, \eta > 0$, there exist a uniform distribution function $F(t)$ on a positive-Lebesgue-measure subset I, and an invariant estimator d_1 (of form (1.2)) such that*

$$(dF)^{n+1}(\{(\vec{Y}, t) : |d(\vec{Y}, t) - d_1(\vec{Y}, t)| \geq \delta\}) \leq \eta. \tag{2.1}$$

THEOREM 2.2. *Suppose that the sample size is n (≥ 1) and $d = d(\vec{Y}, t)$ is a nonrandomized estimator with finite risks and is a (measurable) function of the order statistic $\vec{Y}$. For any $\epsilon > 0$, there exist a continuous distribution function F and an invariant estimator d_1 of form (1.2) such that*

$$|R(F, d) - R(F, d_1)| < \epsilon. \tag{2.2}$$

To prove Theorem 2.2, we need the following lemmas.

LEMMA 2.1. *Suppose that $n \geq 2$ and $E \int_{Y_1}^{Y_n} h(F(t))dF(t) < \infty$ where $h(t) \geq 0$. For any $\epsilon > 0$, there exists an $\eta > 0$ such that for all $F \in \Theta$ and $B \subset R^{n+1}$, satisfying $(dF)^{n+1}(B) < \eta$, we have*

$$E \int_{Y_1}^{Y_n} 1(B)h(F(t))dF(t) < \epsilon. \tag{2.3}$$

PROOF. Given $F \in \Theta$, define a transformation by $x = F(t)$ and let S be the image of $T \subset R^{n+1} \cap \{y_1 < t < y_n\}$ under the transformation $(x_1, \ldots, x_n, s) = (F(y_1), \ldots, F(y_n), F(t))$. By continuity of the integral $\int \cdots \int ds d\vec{x}$, given $\epsilon > 0$, $\exists\, \eta$ (independent of F) such that

$$\int \cdots \int_T h(F(t))dF(t)dF(y_1) \cdots dF(y_n) = \int \cdots \int_S h(s) ds d\vec{x} < \epsilon$$

whenever $(dF)^{n+1}(T) = \int \cdots \int_T dF(t)dF(y_1) \cdots dF(y_n) = \int \cdots \int_S ds d\vec{x} < \eta$. □

The following two lemmas can be proved in a similar manner.

LEMMA 2.2. *Suppose $n \geq 1$ and $E\int_{-\infty}^{Y_1} h(F(t))dF(t)$ is finite, where $h(t) \geq 0$. For any $\epsilon > 0$, there exists an $\eta > 0$ such that for all $F \in \Theta$ and $B \subset R^{n+1}$, satisfying $(dF)^{n+1}(B) < \eta$, we have*

$$E\int_{-\infty}^{Y_1} 1(B)h(F(t))dF(t) < \epsilon. \tag{2.4}$$

LEMMA 2.3. *Suppose $n \geq 1$ and $E\int_{Y_n}^{\infty} h(F(t))dF(t)$ is finite. For any $\epsilon > 0$, there exists an $\eta > 0$ such that for all $F \in \Theta$ and $B \subset R^{n+1}$, satisfying $(dF)^{n+1}(B) < \eta$, we have*

$$E\int_{Y_n}^{\infty} 1(B)h(F(t))dF(t) < \epsilon. \tag{2.5}$$

Remark. The case $h(t) = t^{-1}(1-t)^{-1}$ is a good example that we need to consider the above three lemmas separately. In that case, $E\int_{Y_1}^{Y_n} h(F(t))dF(t) < \infty$ but $E\int_{Y_n}^{\infty} h(F(t))dF(t) = \int_0^1 h(t)t^n dt = \infty$ and $E\int_{-\infty}^{Y_1} h(F(t))dF(t) = \int_0^1 h(t)(1-t)^n dt = \infty$. On the other hand, to get finite risk, we need $\int_0^1 h(t)t(1-t)dt < \infty$.

PROOF OF THEOREM 2.2. Suppose that an estimator, d, satisfies the assumptions in Theorem 2.2 and d_1 is another estimator. Note first that

$$\begin{aligned} R(F,d) - R(F,d_1) = {} & E\int_{Y_1}^{Y_n} [|F-d|^k - |F-d_1|^k]h(F(t))dF(t) \\ & + E\int_{Y_n}^{\infty} [|F-d|^k - |F-d_1|^k]h(F(t))dF(t) \\ & + E\int_{-\infty}^{Y_1} [|F-d|^k - |F-d_1|^k]h(F(t))dF(t). \end{aligned} \tag{2.6}$$

We will show next that for any $\epsilon > 0$, there is an estimator, d_1, with form (1.2) such that the three integrals on the right-hand side of (2.6) are all bounded by 2ϵ and this basically completes the proof of Theorem 2.2.

Given $\epsilon > 0$ and $\delta = \epsilon/c_0$, where $c_0 = 2E\int_{Z_1}^{Z_n} h(t)dt$, by Lemma 2.1, there exists an $\eta > 0$ such that (2.3) holds. For this η, it follows from Theorem 2.1 that there exist a distribution function F and an estimator d_1 such that (2.1) holds, i.e.,

$$\begin{aligned} & (dF)^{n+1}(B) < \eta, \\ \text{where} \quad & B = \{(\vec{Y},t) : |d(\vec{Y},t) - d_1(\vec{Y},t)| \geq \delta, Y_1 < t < Y_n\}. \end{aligned} \tag{2.7}$$

Then,

$$\left| E\int_{Y_1}^{Y_n} [|F-d|^k - |F-d_1|^k]h(F(t))dF(t) \right| \tag{2.8}$$

$$
\begin{aligned}
&\leq E\int_{Y_1}^{Y_n} |[|F-d|^k - |F-d_1|^k]| \\
&\qquad \cdot [1(|d-d_1|\geq \delta) + 1(|d-d_1|<\delta)]h(F(t))dF(t) \\
&\leq E\int_{Y_1}^{Y_n} [1(|d-d_1|\geq \delta) \\
&\qquad + |[|F-d|^k - |F-d_1|^k]|1(|d-d_1|<\delta)]h(F(t))dF(t) \\
&= E\int_{Y_1}^{Y_n} 1(|d-d_1|\geq \delta)h(F(t))dF(t) \\
&\quad + E\int_{Y_1}^{Y_n} |[|F-d|^k - |F-d_1|^k]|1(|d-d_1|<\delta)h(F(t))dF(t) \\
&= E\int_{Y_1}^{Y_n} 1(|d-d_1|\geq \delta)h(F(t))dF(t) \\
&\quad + E\int_{Z_1}^{Z_n} |[|t-d|^k - |t-d_1|^k]|1(|d-d_1|<\delta)h(t)dt,
\end{aligned}
$$

where $Z_i = F(Y_i)$, $i = 1, \ldots, n$. Note that $F(X_1), \ldots, F(X_n)$ are random variables from a uniform distribution on $[0,1]$ and Z_i's are order statistics of them. It can be shown that, for $k \geq 1$ and $t \in [0,1]$, $|[|t-d|^k - |t-d_1|^k]| < 2|d-d_1|$. It follows that

$$
E\int_{Z_1}^{Z_n} |[|t-d|^k - |t-d_1|^k]|1(|d-d_1|<\delta)h(t)dt < \delta c_0, \tag{2.9}
$$

where $c_0 = 2E\int_{Z_1}^{Z_n} h(t)dt$. On the other hand, it follows from (2.7) and (2.3) that

$$
E\int_{Y_1}^{Y_n} 1(|d-d_1|\geq \delta)h(F(t))dF(t) = E\int_{Y_1}^{Y_n} 1(B)h(F(t))dF(t) < \epsilon. \tag{2.10}
$$

Note that $\delta = \epsilon/c_0$ by assumption. Thus, (2.9) and (2.10) imply that

$$
\left| E\int_{Y_1}^{Y_n} [|F-d|^k - |F-d_1|^k]h(F(t))dF(t) \right| < 2\epsilon. \tag{2.11}
$$

That is, the first integral on the right of (2.6) is bounded by 2ϵ. Similarly, we can show that the other two integrals on the right of (2.6) are bounded by 2ϵ, i.e.,

$$
\left| E\int_{Y_n}^{\infty} [|F-d|^k - |F-d_1|^k]h(F(t))dF(t) \right| < 2\epsilon; \tag{2.12}
$$

$$
\left| E\int_{-\infty}^{Y_1} [|F-d|^k - |F-d_1|^k]h(F(t))dF(t) \right| < 2\epsilon. \tag{2.13}
$$

To prove (2.12) and (2.13), without loss of generality, we can assume that $E\int h(F(t))dF(t)$ is finite. Otherwise, if $E\int_{-\infty}^{Y_1} 1(B)h(F(t))dF(t)$ is not finite for some $F \in \Theta$, since both d and d_1 have finite risks, due to the assumption on $h(t)$ (see previous Remark), it can be shown that $d = d_1 = 0$ for $t < Y_1$ a.e.m^{n+1}. (The proofs are similar to that of Theorem 3.1 in Yu (1989*b*).) Furthermore, if $E\int_{Y_n}^{\infty} 1(B)h(F(t))dF(t)$ is not finite for some $F \in \Theta$, since both d and d_1 have finite risks, it can be shown that $d = d_1 = 1$ for $t > Y_n$ a.e.m^{n+1}. In these cases, (2.12) and (2.13) are trivially true. Then, it follows from (2.6) that

$$\begin{aligned}|R(F,d) - R(F,d_1)| &\leq \left|E\int_{Y_1}^{Y_n} [|F-d|^k - |F-d_1|^k]h(F(t))dF(t)\right| \\ &\quad + \left|E\int_{Y_n}^{\infty} [|F-d|^k - |F-d_1|^k]h(F(t))dF(t)\right| \\ &\quad + \left|E\int_{-\infty}^{Y_1} [|F-d|^k - |F-d_1|^k]h(F(t))dF(t)\right| \\ &\leq 6\epsilon\end{aligned}$$

(the last inequality follows from (2.11)–(2.13)). This completes the proof. □

Now the minimaxity of d_0 follows.

THEOREM 2.3. *For sample size* $n \geq 1$, *under the assumptions* (1.1)–(1.2), *the best invariant estimator*, d_0, *is minimax.*

PROOF. Since d_0 has a constant risk, it suffices to show that $R(F,d) \geq R(F,d_0)$ for any $F \in \Theta$ and any estimator d which has finite risks for any $F \in \Theta$. Without loss of generality, we can assume that any estimator we consider is a function of order statistics $\vec{Y}$, since they form an essentially complete class.

Given an estimator d which is a function of order statistics $\vec{Y}$ and has finite risks for any $F \in \Theta$, by Theorem 2.2, there exists an $F \in \Theta$ and there exists an estimator d_1 of form (1.2) and thus of constant risk such that (2.2) holds. It follows that $2\epsilon + R(F,d) \geq R(F,d_1) \geq R(F,d_0)$, since d_0 and d_1 are both invariant and d_0 is the best invariant estimator. Note that ϵ and d are arbitrary, so $\inf_d \sup_{F\in\Theta} R(F,d) = R(F,d_0)$. □

Acknowledgements

The author wishes to thank two referees for valuable suggestions.

REFERENCES

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

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

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

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

Ferguson, T. S. (1967). *Mathematical Statistics, a Decision Theoretic Approach*, p. 197, Academic Press, New York.

Yu, Q. (1988). Inadmissibility of the best invariant estimator of a distribution function, *Sankhyā Ser. A* (to appear).

Yu, Q. (1989*a*). Inadmissibility of the empirical distribution function in continuous invariant problems, *Ann. Statist.*, **17**, 1347–1359.

Yu, Q. (1989*b*). Admissibility of the best invariant estimator of a distribution function, *Statist. Decisions*, **7**, 1–14.

Yu, Q. (1989*c*). Methodology for the invariant estimation of a continuous distribution function, *Ann. Inst. Statist. Math.*, **41**, 503–520.

Yu, Q. (1989*d*). Admissibility of the empirical distribution function in the invariant problem, *Statis. Decisions*, **7**, 383–398.

Yu, Q. and Chow, M. S. (1991). Minimaxity of the empirical distribution function in invariant problem, *Ann. Statist.*, **19**, 935–951.