Statistica Sinica **8**(1998), 377-392

ADMISSIBILITY OF THE BEST INVARIANT ESTIMATOR OF A DISCRETE DISTRIBUTION FUNCTION

Yu, Qiqing

Binghamton University

Abstract: We consider the problem of invariant estimation of a discrete distribution function F under the Cramer-von Mises loss. It is proved that the best invariant estimator is admissible. This extends a result of Brown (1988) and settles an open question (Brown (1988)). The idea used in the proof of admissibility is a new refinement of the standard Bayes argument, which is different from the step-wise Bayes approach and Blyth's (1951) Lemma.

Key words and phrases: Admissibility, discrete distribution, invariant loss, non-parametric estimation.

1. Introduction

We study the admissibility of the best invariant estimator of a distribution function in the discrete set-up formulated by Brown (1988).

Let $\vec{X} = (X_1, \ldots, X_n)$ be a sample of size n from an unknown distribution function F. The action space is the set of all nondecreasing functions $a(t)$ from $(-\infty, \infty)$ into [0,1]. We consider two types of parameter spaces Θ. One is the family of all distribution functions, denoted by Θ_0. The other is the family of all discrete distribution functions, denoted by Θ_d. The family of all continuous distribution functions is denoted by Θ_c. The loss function is

$$L(F, a) = \int_{-\infty}^{+\infty} |F(t) - a(t)|^2 h(F(t)) dF(t), \text{ with } h(t) = t^{\alpha}(1-t)^{\beta} \text{ and } \alpha, \beta \geq -1. \tag{1.1}$$

The risk of an estimator, d (also written as $d = d(\vec{X}) = d(t) = d(\vec{X}, t)$), is $R(F, d) = EL(F, d(\vec{X}))$. Under the above formulation, decision problems of estimating F are invariant under monotone transformations (from R^1 onto R^1).

When $\Theta = \Theta_c$, Aggarwal (1955) showed that the best invariant estimator is

$$d_{\alpha,\beta}(t) = \frac{\alpha + 1 + \sum_{i=1}^{n} 1[t \geq X_i]}{n + 2 + \alpha + \beta}, \tag{1.2}$$

where $1[A]$ is the indicator function of a set A. A special case of the loss (1.1) is $\alpha = \beta = -1$. In this case, $d_{\alpha,\beta}(t)$ is the empirical distribution function (EDF) $\hat{F}(t)$ (Aggarwal (1955)). This estimator is minimax (Yu and Chow (1991)). It is admissible if the sample size n is 1 or 2 and is inadmissible if $n \geq 3$ (Yu (1989a,b,c)).

When the unknown distribution is discrete, Yu (1992) demonstrated that $\hat{F}$ is minimax under the loss (1.1) with $\alpha = \beta = -1$, and Brown (1988) and Yu (1993) showed that $\hat{F}$ is admissible under the loss (1.1) with $\alpha \in [-1, 0]$ and $\beta \in [-1, 1)$.

Another special case of the loss (1.1) is $\alpha = \beta = 0$, that is

$$L_1(F, a) = \int_{-\infty}^{+\infty} |F(t) - a(t)|^2 dF(t). \tag{1.3}$$

The best invariant estimator is

$$d_{0,0}(t) = \frac{1}{n+2} + \frac{1}{n+2} \sum_{i=1}^{n} 1[t \geq X_i]. \tag{1.4}$$

When $\Theta = \Theta_c$, Brown proved that $d_{0,0}(t)$ is inadmissible under the loss (1.3). When Θ is the family of discrete distributions with support on the interval $[0, 1]$, denoted by $\Theta_{d[0,1]}$, he showed that $d_{0,0}(t)$ is inadmissible under the loss (1.3), and $d_{0,0}(t)$ is improved by $d_b(t) = d_{0,0}(t)$ if $t < 1$; 1 otherwise. Under the loss (1.3) and the family of distributions $\Theta_{d[0,1]}$, $d_b(t)$ is admissible. If $\Theta = \Theta_d$ and the loss is as in (1.3), admissibility of $d_{0,0}(t)$ is an interesting open question. The major difficulty in this problem is that the standard step-wise Bayes procedure does not work. The step-wise procedure converts the problem to a multinomial distribution problem, where for each given observation $\vec{x}$, the estimate of interest is a posterior Bayes action (Berger (1985)). Unfortunately, in this problem, the estimate $d_{0,0}(\vec{x}, t)$ may not be a posterior Bayes action for each observation $\vec{x}(= (x_1, \ldots, x_n))$.

In this paper, we settle the open question in a more general setting. Denote

$$d_L(t) = \frac{s}{n+s+r} + \frac{1}{n+s+r} \sum_{i=1}^{n} 1[t \geq X_i], \quad \text{where } s, r \geq 0. \tag{1.5}$$

We show in Theorem 1 that if $\Theta = \Theta_d$ or Θ_0, then any d_L of the form (1.5) $(s, r \geq 0)$ is admissible under any loss function in the whole class of loss functions (1.1). Note that $d_{\alpha,\beta}(t)$ as in (1.2) is associated with the specific α and β in the loss (1.1) (but d_L is not). It follows that when $\Theta = \Theta_d$ the best invariant estimator $d_{\alpha,\beta}(t)$, as in (1.2), is admissible under the loss (1.1). Furthermore, when $\Theta = \Theta_d$ both the EDF and $d_{0,0}(t)$ are admissible under the loss (1.1) for any $\alpha, \beta \geq -1$. We use a new approach to attack the admissibility issue. The idea used in the proof of Theorem 1 can be summarized as the following proposition.

Proposition 1. *Suppose that $\mathcal{X}$ is the sample space, the m-dimensional parameter space Θ is open and that the risk $R(\theta, \delta)$ of any decision rule δ is continuous* *in* $\theta \in \Theta$. *Suppose that* $\{\tau_i\}$ *is a sequence of prior density functions on the parameter space* Θ *such that* $\lim_{i\to\infty} \tau_i = \tau$ *almost everywhere. Here* τ *is a positive prior density function. Suppose that* δ_0 *is a decision rule such that its Bayes risk* $r(\tau_i, \delta_0)$ *with respect to* τ_i *is finite for all* i *and for any other decision rule* δ *there exists a measurable partition* $\{B_k;\ k \in K\}$ *of the set* $\{\vec{x};\ \delta(\vec{x}) \neq \delta_0(\vec{x})\}$ *such that*

$$\lim_{j\to\infty} \frac{E[1[\vec{X} \in B_k] E_{\tau_j}(L(\theta, \delta_0(\vec{X}))|\vec{X})]}{E[1[\vec{X} \in B_k] E_{\tau_j}(L(\theta, \delta(\vec{X}))|\vec{X})]} < 1 \quad \textit{uniformly for all} \;\; k \in K. \tag{1.6}$$

Then δ_0 *is admissible.*

The proof of this statement is postponed to Section 5.

The differences between the new approach proposed in the proposition and the stepwise Bayes approach will be discussed in the end of the paper. A special case (when B_k's are singletons) of the approach mentioned in the proposition was used in Yu and Kuo (1995) to study admissibility issues for non-invariant loss functions.

If the family of distribution functions is replaced by $\Theta_{d[0,1]}$ considered in Brown (1988) and if $\beta > 0$, $d_L(t)$ is still admissible. The idea of the proof is similar. However, if $\beta \leq 0$, d_L is improved by $d_B(t) = d_L(t)$ if $t < 1$; 1 otherwise. Furthermore, it can be shown that $d_B(t)$ is admissible under the loss (1.1) with $\alpha, \beta \geq -1$ if $\Theta = \Theta_{d[0,1]}$.

Since we consider the problem of discrete invariant estimation, the integrand, $H(F, a) = (F - a)^2 h(F)$, in the loss function (1.1) needs to be properly modified when indeterminecy $\frac{0}{0}$ occurs (i.e., when $\beta < 0$ and $F = a = 1$). We define $\frac{0}{0}$ to be 0.

There are five sections in this paper. The main results are introduced in Section 2. The proofs of the main theorem are partitioned into two parts and are given in Sections 3 and 4, respectively. The significance of the approach proposed in Proposition 1 is discussed in Section 5.

2. Main Results

Our main result is stated as follows.

Theorem 1. *Suppose that the loss is as in* (1.1), $\Theta = \Theta_d$ *or* Θ_0. *Then the estimator* $d_L(t)$ *in* (1.5) *is admissible. Furthermore, if an estimator* d *is such that* $R(F, d) \leq R(F, d_L)$ *for all* $F \in \Theta_d$, *then* $d = d_L$.

We first outline the proof of Theorem 1. Given a discrete distribution function,

$$F(t) = \sum_j p_j 1[t \geq s_j], \text{ where } p_j\text{'s are weights and } s_j\text{'s are real numbers,} \tag{2.1}$$

and given arbitrary estimator $d(t)$, the loss function (1.1) can be written as

$$\int_{-\infty}^{+\infty}(F(t)-d(t))^2h(F(t))dF(t)=\sum_j p_j[F(s_j)-d(s_j)]^2(F(s_j))^\alpha(1-F(s_j))^\beta.$$

If $\beta>0$, then the risk of $d(t)$ does not depend on the value of d at $t_0=\sup_j\{s_j\}$ and is finite; on the other hand, if $\beta<0$ and if F gives a positive weight to t_0, then the risk of $d(t)$ does depend on the value of d at t_0 and equals $+\infty$, unless $d(t_0)=1$. Finally, if $\beta=0$, then the risk of $d(t)$ depends on the value of d at t_0, but the risk is always finite. Moreover, it turns out that the proof for the case $r=0$ is almost identical to that for the case $\beta>0$. Thus the proof of the theorem is divided into three cases: (1) $\beta>0$ or $r=0$; (2) $\beta=0$ and $r>0$; (3) $\beta<0$ and $r>0$.

To prove the theorem for case (1), we apply the idea described in Proposition 1 directly, in which $\{\tau_j\}_{j\geq 1}$ is a sequence of modified multivariate Beta priors on the family of distribution functions F of the form (2.1) which have at most $n+1$ different s_j's (the support of F), and B_k's are all singleton sets $\{(\vec{x},t)\}$ such that $d(\vec{x},t)\neq d_L(\vec{x},t)$.

The proof for case (2) is the most interesting one, since it settles an open question raised by Brown (1988). We first prove that it suffices to show that any estimator d which takes finitely many values cannot improve on d_L (see Lemma 2), and then prove the statement using the approach in Proposition 1. The priors $\{\tau_j\}_{j\geq 1}$ are similar to case (1), but the B_k's are not all singleton sets.

In case (3), in order to have a finite risk, we have to consider priors for the family of distribution functions F of form (2.1), for which t_0 does not belong to $\{s_1,s_2,\ldots\}$. Then we use the approximation and mimic the proof for case (2). Since this proof is very similar to that for $\beta=0$, for an easier presentation, we only prove Theorem 1 for cases (1) and (2). Readers who are interested in the proof for case (3) can find it in a technical report (Yu (1994)).

3. Proof of Theorem 1 for Case (1)

The proof of this section is similar to that in Yu and Kuo (1995). Hereafter, we denote the $(n+1)$-product set of $\{\xi_1,\ldots,\xi_m\}$ by $\{\xi_1,\ldots,\xi_m\}^{n+1}=\{(\vec{x},t): x_1,\ldots,x_n,t\in\{\xi_1,\ldots,\xi_m\}\}$.

We assume that d_1 is an estimator such that $d_1(\vec{x},t)\neq d_L(\vec{x},t)$ for $(\vec{x},t)=(\vec{x}_*,t_*)$, where $\vec{x}_*=(x_{*1},\ldots,x_{*n})$ is a fixed point in R^n, and $R(F,d_L)-R(F,d_1)\geq 0$ for all $F\in\Theta_d$; and reach a contradiction.

Denote $\xi_1<\cdots<\xi_{m-1}$ the distinct points in the set $\{x_{*1},\ldots,x_{*n},t_*\}$ and take $\xi_m>\xi_{m-1}$, say, $\xi_m=\xi_{m-1}+1$. Denote

$$V=\begin{cases}\{(\vec{x},t):(\vec{x},t)\in\{\xi_1,\ldots,\xi_m\}^{n+1},\ d_1(\vec{x},t)\neq d_L(\vec{x},t),t\neq\xi_m\} & \text{if } \beta>0;\\ \{(\vec{x},t):(\vec{x},t)\in\{\xi_1,\ldots,\xi_m\}^{n+1},\ d_1(\vec{x},t)\neq d_L(\vec{x},t)\} & \text{if } \beta\leq 0.\end{cases}\tag{3.1}$$

Then V is not empty (by the assumption about d_1). Hereafter in the proof, let

$$F(t) = \sum_{i=1}^{m} p_i 1[t \geq \xi_i] \tag{3.2}$$

be a discrete distribution function, where $p = (p_1, \ldots, p_m)$ is a probability vector; denote

$$d\tau^*(p) = p_1^{s-2-\alpha} p_2^{-1} \cdots p_{m-1}^{-1} p_m^{r-1-\beta} dp_1 \cdots dp_{m-1};$$
$$d\tau_\epsilon^* = d\tau^*(p) 1[p \in P_\epsilon], \text{ where } \epsilon > 0 \text{ and } P_\epsilon = \Big\{p : p_i \geq \epsilon/m, \sum_{i=1}^{m} p_i = 1\Big\}. \tag{3.3}$$

For each possible $\vec{x}$, let $\eta_k = \eta_k(\vec{x}) = \#\{i : x_i = \xi_k\}$ and $\eta = (\eta_1, \ldots, \eta_m)$, where $\#A$ is the cardinality of a set A. Let $d_1(\eta, t) = d_1(\vec{x}, t)$ for $\eta(\vec{x}) = \eta$ and similarly for $d_L(\eta, t)$. Write $\int \cdots \int_{P_\epsilon} = \int_{P_\epsilon}$ and $dp_1 \cdots dp_{m-1} = dp$. For any estimate d and for any $\epsilon > 0$, integrating $R(F, d)$ over the prior $d\tau_\epsilon^*$ yields a finite value. It can be shown that the integral is $\int_{P_\epsilon} R(F, d) d\tau_\epsilon^*(p) = \sum_{\vec{x}} \sum_{j=1}^{m} I(d(\vec{x}, \xi_j), \epsilon)$, where

$$I(d(\vec{x}, \xi_j), \epsilon) = \int_{P_\epsilon} p_j (\sum_{i=1}^{j} p_i - d(\vec{x}, \xi_j))^2 \prod_{k=1}^{m} p_k^{\eta_k - 1} h(\sum_{i=1}^{j} p_i) p_1^{s-1-\alpha} p_m^{r-\beta} dp. \tag{3.4}$$

In view of (3.4) and $R(F, d_L) - R(F, d_1) \geq 0$, integrating $R(F, d_L) - R(F, d_1)$ over the prior $d\tau_\epsilon^*$ yields

$$0 \leq \sum_{(\vec{x}, \xi_j) \in V} I(d_L(\vec{x}, \xi_j), \epsilon) - \sum_{(\vec{x}, \xi_j) \in V} I(d_1(\vec{x}, \xi_j), \epsilon) \text{ for any } \epsilon > 0, \tag{3.5}$$

(since $I(d_L(\vec{x}, \xi_j), \epsilon) = I(d_1(\vec{x}, \xi_j), \epsilon)$ if $(\vec{x}, \xi_j) \notin V$). That is

$$\frac{\sum_{(\vec{x}, \xi_j) \in V} I(d_L(\vec{x}, \xi_j), \epsilon)}{\sum_{(\vec{x}, \xi_j) \in V} I(d_1(\vec{x}, \xi_j), \epsilon)} \geq 1 \text{ for any } \epsilon > 0. \tag{3.6}$$

To get a contradiction, we show that $d_L(t)$ is a limit of a sequence of generalized Bayes estimators with respect to $\{\tau_{\epsilon_j}\}$, where $\epsilon_j \to 0$. In this regard, it would be more convenient to use the prior τ^* instead of the sequence of priors, as was done in Brown (1988) and Yu (1993). However, in general, replacing τ^* for τ_ϵ^* in (3.4) may not yield a finite value, since the exponent of p_1, $\eta_1 + s - 2 - \alpha$, in (3.4) may be < -1 (*e.g.*, $\eta_1 = n$, $s = 0$, $\alpha = 3n$).

It is worth noting that the multiple integral in (3.4) can be evaluated for each $(\vec{x}, \xi_j)$ (see, for example, Ferguson (1967)). Let $\sigma_j = \sigma_j(\eta) = \sum_{i=1}^{j} \eta_i$, $j = 1, \ldots, m-1$. Making the substitution $u_j = \sum_{i=1}^{j} p_i$, $j = 1, \ldots, m-1$,

$q_k = \frac{u_k}{u_{k+1}}$, for $k < j$ and $r_k = \frac{p_k}{1-u_{k-1}}$, for $k > j$, the multiple integral in (3.4) for each $(\vec{x}, \xi_j)$ becomes

$$\begin{aligned} I(d(\vec{x},\xi_j),\epsilon) = \Big[& \prod_{k=1}^{j-2} \int_{\frac{(k/m)\epsilon}{1-[(m-k-1)/m]\epsilon}}^{\frac{1-[(m-k)/m]\epsilon}{1-[(m-k-1)/m]\epsilon}} q_k^{\sigma_k+s-\alpha-2}(1-q_k)^{\eta_{k+1}-1} dq_k \Big] \\ & \times \int_{\frac{[(j-1)/m]\epsilon}{1-[(m-j)/m]\epsilon}}^{\frac{1-[(m-j+1)/m]\epsilon}{1-[(m-j)/m]\epsilon}} q_{j-1}^{\sigma_{j-1}+s-\alpha-2}(1-q_{j-1})^{\eta_j} dq_{j-1} \\ & \qquad\qquad\qquad \text{(note } p_j = (1-q_{j-1})u_j) \\ & \times \int_{\frac{j}{m}\epsilon}^{1-\frac{(m-j)}{m}\epsilon} (u_j - d(\vec{x},\xi_j))^2 u_j^{\sigma_j+s-1}(1-u_j)^{n-\sigma_j+r-1} du_j \\ & \times \Big[\prod_{k=j+1}^{m-1} \int_{\frac{\epsilon/m}{1-[(k-1)/m]\epsilon}}^{\frac{1-[(m-1)/m]\epsilon}{1-[(k-1)/m]\epsilon}} (1-r_k)^{n-\sigma_k+r-\beta-1} r_k^{\eta_k-1} dr_k \Big], \end{aligned} \tag{3.7}$$

where we define $\prod_{k=1}^{j-2} a_k = 1$ for $j = 1$ or 2 and for any real number a_k. Also define $\int f(q_j)dq_j = 1$ for $j < 1$ and for any function f. Note we can replace d by d_L or d_1 in (3.7). Since d_1 or d_L only occurs in one of the four factors in (3.7), this formula yields

$$\begin{aligned} \frac{I(d_L(\vec{x},\xi_j),\epsilon)}{I(d_1(\vec{x},\xi_j),\epsilon)} &= \frac{\int_{(j/m)\epsilon}^{1-[(m-j)/m]\epsilon} (u_j - d_L(\vec{x},\xi_j))^2 u_j^{s+\sigma_j-1}(1-u_j)^{n-\sigma_j+r-1} du_j}{\int_{(j/m)\epsilon}^{1-[(m-j)/m]\epsilon} (u_j - d_1(\vec{x},\xi_j))^2 u_j^{s+\sigma_j-1}(1-u_j)^{n-\sigma_j+r-1} du_j} \\ &< \frac{\int_0^1 (u_j - d_L(\vec{x},\xi_j))^2 u_j^{s+\sigma_j-1}(1-u_j)^{n-\sigma_j+r-1} du_j}{\int_\epsilon^{1-\epsilon} (u_j - d_1(\vec{x},\xi_j))^2 u_j^{s+\sigma_j-1}(1-u_j)^{n-\sigma_j+r-1} du_j} \\ &\to \frac{\int_0^1 (u_j - d_L(\vec{x},\xi_j))^2 u_j^{s+\sigma_j-1}(1-u_j)^{n-\sigma_j+r-1} du_j}{\int_0^1 (u_j - d_1(\vec{x},\xi_j))^2 u_j^{s+\sigma_j-1}(1-u_j)^{n-\sigma_j+r-1} du_j} \text{ as } \epsilon \downarrow 0. \end{aligned} \tag{3.8}$$

It is important to note that the last ratio is less than 1, since $d_1(\vec{x},\xi_j) \neq d_L(\vec{x},\xi_j)$ and

$$d = d_L(\vec{x},\xi_j) = \frac{s+\sigma_j}{n+s+r} \text{ minimizes } \int_0^1 (u_j - d)^2 u_j^{s+\sigma_j-1}(1-u_j)^{n-\sigma_j+r-1} du_j. \tag{3.9}$$

Furthermore, since $\#V$ is finite, it follows from (3.1) and (3.8) that for $\beta > 0$

$$\lim_{\epsilon\to 0} \frac{I(d_L(\vec{x},\xi_j),\epsilon)}{I(d_1(\vec{x},\xi_j),\epsilon)} < 1 \text{ uniformly in } m \text{ and for all } (\vec{x},\xi_j) \in V. \tag{3.10}$$

If $\beta \leq 0$ then the value of any estimator at the point $(\vec{x},\xi_m)$ would affect its risk. Furthermore, if $\beta \leq 0$ and $(\vec{x},\xi_m) \in V$, then (3.8) may not hold and thus

(3.10) may fail. However, if $r = 0$, (3.10) holds. To prove it, it is important to note that if $r = 0$, then $d_L(\vec{x}, \xi_m) = 1$ for all possible $\vec{x}$ under F (defined as in (3.2)), and

$$\frac{I(d_L(\vec{x}, \xi_m), \epsilon)}{I(d_1(\vec{x}, \xi_m), \epsilon)} = \frac{(\sum_{i=1}^m p_i - d_L(\vec{x}, \xi_m))^2}{(\sum_{i=1}^m p_i - d_1(\vec{x}, \xi_m))^2} = \frac{(1-1)^2}{(1 - d_1(\vec{x}, \xi_m))^2} \tag{3.11}$$
$$< 1 \text{ uniformly in } m \text{ and for all } (\vec{x}, \xi_m) \in V.$$

Thus (3.10), remains true, and it implies that there exists $\epsilon > 0$ such that

$$\frac{I(d_L(\vec{x}, \xi_i), \epsilon)}{I(d_1(\vec{x}, \xi_i), \epsilon)} < 1 \text{ uniformly in } m \text{ and for all } (\vec{x}, \xi_i) \in V. \tag{3.12}$$

There are only finitely many elements in V, so that

$$\frac{\sum_{(\vec{x},\xi_i)\in V} I(d_L(\vec{x}, \xi_i), \epsilon)}{\sum_{(\vec{x},\xi_i)\in V} I(d_1(\vec{x}, \xi_i), \epsilon)} < 1 \text{ for some } \epsilon > 0, \tag{3.13}$$

which contradicts (3.6). This concludes the proof of the theorem.

Remark 1. Note that the most interesting case $\alpha = \beta = 0$ is not included in case (1). The proof of Theorem 1 in the latter case is more difficult than the proof for the case $\beta > 0$ since (3.10) does not hold for $(\vec{x}, \xi_m) \in V$. In fact, using the notation as in (3.11), we have

$$\frac{I(d_L(\vec{x}, \xi_m), \epsilon)}{I(d_1(\vec{x}, \xi_m), \epsilon)} < 1 \text{ if and only if } |1 - d_L(\vec{x}, \xi_m)| < |1 - d_1(\vec{x}, \xi_m)|. \tag{3.14}$$

Thus, (3.10) fails if $d_1(\vec{x}, \xi_m) = 1$, $\beta \leq 0$ and $r > 0$, since $d_L(\vec{x}, \xi_m) = \frac{n+s}{n+s+r} < 1$. Note that $\frac{I(d_L(\vec{x},\xi_m),\epsilon)}{I(d_1(\vec{x},\xi_m),\epsilon)} < 1$ in (3.14) is a necessary condition that $d_L(\vec{x}, \xi_m)$ be a posterior Bayes action or a pointwise limit of posterior Bayes actions w.r.t the multivariate Beta prior τ^* or its modification τ^*_ϵ. In other words, if $\beta \leq 0$ and $r > 0$, then $d_L(\vec{x}, \xi_m)$ cannot be a posterior Bayes action or a pointwise limit of posterior Bayes actions w.r.t the multivariate Beta prior or its modification.

4. The Proof of Theorem 1 Under the Loss (1.1) with $\beta = 0$ (case 2)

Both the proofs for case (1) and for case (2) are applications of Proposition 1. In the proof for case (1), the B_k's (see (1.6)) are all singletons, whereas in the proof for case (2), this is not true. A major reason for the difference in B_k's is that d_L cannot be a posterior Bayes action or a pointwise limit of posterior Bayes actions w.r.t. τ^* in case (2) (see Remark 1), as it is in case (1). Other than that, these two proofs are very similar.

To simplify our proof, without loss of generality (WLOG), we can assume that all the estimators are functions of the order statistics, since these form an essentially complete class. Furthermore, given a d_L of form (1.5), since we study the admissibility of d_L, WLOG, we can assume that all estimators d considered belong to the class

$$U = \{d : R(F, d) \le R(F, d_L) \text{ for all discrete } F\}.$$

We outline the proof of the theorem for case (2).

Lemma 1. (Yu and Phadia, Theorem 4. (1993)) *Suppose that $n \ge 1$ and $d(\vec{x}, t) \in U$, but $d \ne d_L$ on a positive Lebesgue-measure set. Then one can find an estimator d_n which takes on finitely many values, $d_n \ne d_L$ such that $\forall \epsilon, \delta > 0$, there is a subset I, of positive measure, satisfying:*

$$\mu^{n+1}(\{(x_1, \ldots, x_n, t) : x_1, \ldots, x_n, t \in I, |d(\vec{x}, t) - d_n(\vec{x}, t)| > \epsilon\}) < \delta(\mu(I))^{n+1},$$

where μ^{n+1} is the $(n + 1)$-dimensional (product) Lebesgue measure.

It follows from Lemma 1 that any estimator d can be approximated by an estimator d_n which takes on finitely many values. Thus it suffices to show that the estimator d_n which takes on finitely many values cannot improve on d_L. Hereafter, given $\vec{x} = (x_1, \ldots, x_n)$, we assume that $x_1 \le \cdots \le x_n$ since we restrict our attention to functions of the order statistics only.

Lemma 2. *Assume the case (2). Suppose that d_1 is another estimator which takes on finitely many values on the product set $(\{\xi_i\}_{i\ge1})^{n+1}$. If $R(F, d_L) - R(F, d_1) \ge 0$ for any $F(t)$ with support on $\{\xi_i\}_{i\ge1}$, then $d_1 = d_L$ on $(\{\xi_i\}_{i\ge1})^{n+1}$.*

Proof. We assume that $d_1 \ne d_L$ at $(\vec{x}, t) \in \{\xi_1, \ldots, \xi_J\}^{n+1}$, $d_1 \in U$, and reach the contradiction.

Since $d_1(\vec{x}, \xi_j)$ takes on finitely many values, by choosing a subsequence of $\xi_J, \xi_{J+1}, \ldots$, WLOG, we can assume that $d_1(\vec{x}, \xi_j)$ is constant for $\xi_j > \xi_J$, say $d_1(\vec{x}, \xi_j) = l_n(x_1, \ldots, x_n)$ if $\xi_j > \xi_J$.

Treating $l_n(x_1, \ldots, x_n)$ as a function of x_n ($x_n \ge x_{n-1}$, since $x_1 \le \cdots \le x_n$) for each $x_1, \ldots, x_{n-1}$, by assumption, there are finitely many values of $l_n(x_1, \ldots, x_n)$. Thus, by taking subsequences of $\{\xi_{J+1}, \xi_{J+2}, \ldots\}$, WLOG, we can assume that $l_n(x_1, \ldots, x_{n-1}, x_n) = l_{n-1}(x_1, \ldots, x_{n-1})$ if $x_n \in \{\xi_{J+1}, \xi_{J+2}, \ldots\}$. By taking further subsequences inductively, WLOG, we can assume that

$$l_n(x_1, \ldots, x_{n-1}, x_n) = l_i(x_1, \ldots, x_i) \text{ if } x_{i+1}, \ldots, x_n \in \{\xi_{J+1}, \xi_{J+2}, \ldots\}, \quad i = 0, 1, \ldots, n.$$

In other words, we can assume that

$$d_1(\vec{x}, t) = l_i(x_1, \ldots, x_i) \text{ if } x_1, \ldots, x_i \le \xi_J \text{ and } x_{i+1}, \ldots, x_n, t \in \{\xi_{J+1}, \xi_{J+2}, \ldots\}, \quad i = 1, \ldots, n, \text{ and if } x_1 \le \cdots \le x_n < t. \tag{4.1}$$

$$d_1(\vec{x}, t) = l_0 \text{ if } x_1, \ldots, x_n, t \in \{\xi_{J+1}, \xi_{J+2}, \ldots\} \text{ and if } x_1 \le \cdots \le x_n < t. \tag{4.2}$$

Furthermore, treating $d_1(\vec{x}, x_n)$ as a function of $(x_1, \ldots, x_{n-1})$, in the same manner as in deriving (4.1) and (4.2), WLOG, we can assume that

$$d_1(\vec{x}, x_n) = l'_i(x_1, \ldots, x_i) \text{ if } x_1, \ldots, x_i \leq \xi_J \text{ and } x_{i+1}, \ldots, x_n \in \{\xi_{J+1}, \xi_{J+2}, \ldots\} \quad (4.3)$$
$$i = 1, \ldots, n-1;$$
$$d_1(\vec{x}, x_n) = l'_0 \text{ if } x_1, \ldots, x_n \in \{\xi_{J+1}, \xi_{J+2}, \ldots\}. \quad (4.4)$$

In view of case (1), WLOG, we can assume that $r > 0$. By assumption, $d_1 \neq d_L$ and d_L improves on d_1. Therefore the same arguments in the proof for case (1) yield (3.6) for any $m > J$ and for any F as in (3.2). Partition the set V into three subsets:

$$\begin{aligned} V_m &= \{(\vec{x}, t) \in V : t = \xi_m\} \\ V_1 &= \{(\vec{x}, t) \in V : \xi_J < x_1 \leq x_n \leq t \leq \xi_{m-1}\} \\ V_0 &= V \setminus (V_m \cup V_1). \end{aligned} \quad (4.5)$$

We first show that

$$d_1 = d_L \text{ for } (\vec{x}, t) \in (\{\xi_j\}_{j>J})^{n+1}, \quad (4.6)$$

and then we conclude the proof of the lemma by showing

$$d_1 = d_L \text{ for } (\vec{x}, t) \in (\{\xi_j\}_{j\geq 1})^{n+1}. \quad (4.7)$$

Indeed (4.7) contradicts the assumption $d_1 \neq d_L$.

To show (4.6), by defining a distribution with support in $\{\xi_{J+1}, \xi_{J+2}, \ldots\}$, WLOG, we can assume that $J = 0$, *i.e.*,

(1) $d_1(\vec{x}, t) = l_0$ if $x_1, \ldots, x_n, t \in \{\xi_1, \xi_2, \ldots\}$ and $t > x_n$;

(2) $d_1(\vec{x}, x_n) = l'_0$ if $x_1, \ldots, x_n \in \{\xi_1, \xi_2, \ldots\}$.

Recall that when $\beta > 0$, (3.12) holds for $(\vec{x}, t) \in V$. Thus (3.13) holds and we conclude the proof since (3.13) contradicts (3.6). When $\beta = 0$, with V replaced by $V_0 \cup V_1$, (3.12) holds, *i.e.*, there exists an $\epsilon > 0$ such that

$$I(d_L(\vec{x}, t), \epsilon)/I(d_1(\vec{x}, t), \epsilon) < 1 \text{ for all } (\vec{x}, t) \in V_0 \cup V_1 \text{ and for all } m. \quad (4.8)$$

If V_m is empty, we are done, so that we can assume that V_m is not empty. Then (4.8) or (3.12) does not hold for $(\vec{x}, t) \in V_m$ and for all $\epsilon > 0$. It can be shown (see Lemma 3) that a variant of (4.8) or (3.12) holds, namely, for any $\epsilon > 0$ such that (4.8) holds, we can choose m large enough so that

$$\frac{\sum_{(\vec{x}, \xi_j) \in V_1 \cup V_m} I(d_L(\vec{x}, \xi_j), \epsilon)}{\sum_{(\vec{x}, \xi_j) \in V_1 \cup V_m} I(d_1(\vec{x}, \xi_j), \epsilon)} < 1. \quad (4.9)$$

Since $V = V_0 \cup (V_1 \cup V_m)$, (4.8) and (4.9) imply (3.13), which contradicts (3.6). This completes the proof of (4.6).

We now give the proof of (4.7). It follows from (4.6) that V_1 is empty, $V = V_0 \cup V_m$ (see (4.5)) and $V_m = \{(\vec{x}, \xi_m) \in V : x_1 \in \{\xi_1, \ldots, \xi_J\}\}$. Note that there exists an $\epsilon > 0$ such that (4.8) holds for all $(\vec{x}, t) \in V_0$. It follows from Lemma 4 below that for any $\epsilon > 0$ such that (4.8) holds, there exists m such that for all $\vec{x}$ for which $(\vec{x}, \xi_m) \in V_m$ and $x_1 \le \xi_J$,

$$\frac{\sum_{j>J}^{m} I(d_L(\vec{x}, \xi_j), \epsilon)}{\sum_{j>J}^{m} I(d_1(\vec{x}, \xi_j), \epsilon)} < 1. \tag{4.10}$$

Then (4.8) and (4.10) yield (3.13), which contradicts (3.6). The contradiction shows that (4.7) holds, which completes the proof of the lemma.

In the above proof, the set $\{(\vec{x}, \xi_j) \in V_1 \cup V_m\}$ is the B_1 in Proposition 1 and the remaining B_i are singleton sets $\{(\vec{x}, \xi_j)\}$, which are subsets of V_0.

Lemma 3. *Inequality* (4.9) *holds.*

Proof. Inequality (4.9) is equivalent to

$$\frac{-\sum_{(\vec{x},t)\in V_1}[I(d_L(\vec{x}, t), \epsilon) - I(d_1(\vec{x}, t), \epsilon)]}{\sum_{(\vec{x},t)\in V_m}[I(d_L(\vec{x}, t), \epsilon) - I(d_1(\vec{x}, t), \epsilon)]} > 1. \tag{4.11}$$

Thus, it suffices to show that for any $\epsilon > 0$ such that (4.8) holds for $(\vec{x}, \xi_j) \in V_1$, we can choose m so that

$$\frac{-\sum_{(\vec{x},t)\in V_1}[I(d_L(\vec{x}, t), \epsilon) - I(d_1(\vec{x}, t), \epsilon)]}{\sum_{(\vec{x},t)\in V_m}[I(d_L(\vec{x}, t), \epsilon) - I(d_1(\vec{x}, t), \epsilon)]} > 1 \text{ if } l_0 = l_0' \text{ and } V_m \text{ is not empty;} \tag{4.12}$$

$$\begin{cases} \frac{-\sum_{(\vec{x},x_n)\in V_1}[I(d_L(\vec{x},x_n),\epsilon)-I(d_1(\vec{x},x_n),\epsilon)]}{\sum_{(\vec{x},x_n)\in V_m}[I(d_L(\vec{x},x_n),\epsilon)-I(d_1(\vec{x},x_n),\epsilon)]} > 1 \ \text{ if } l_0 \ne l_0' \text{ and } l_0' \ne \frac{n+s}{n+r+s}; \\ \frac{-\sum_{(\vec{x},t)\in V_1, t>x_n}[I(d_L(\vec{x},t),\epsilon)-I(d_1(\vec{x},t),\epsilon)]}{\sum_{(\vec{x},t)\in V_m, t>x_n}[I(d_L(\vec{x},t),\epsilon)-I(d_1(\vec{x},t),\epsilon)]} > 1 \ \text{ if } l_0 \ne l_0' \text{ and } l_0 \ne \frac{n+s}{n+r+s}. \end{cases} \tag{4.13}$$

For simplicity, we only prove (4.12). The idea of proving (4.13) is the same. By assumption, V is not empty, (4.3) and (4.4) hold, $J = 0$ and $l_0 = l_0'$, *i.e.*, d_1 is constant for $t \ge x_{(n)}$. If an estimate $d(\vec{x}, t)$ is constant for $t \ge x_{(n)}$, say, equal to $d(\vec{x}_{(n)})$, then (3.4) yields

$$\begin{aligned} &\sum_{(\vec{x},\xi_j)\in V_1} I(d(\vec{x}, \xi_j), \epsilon) \\ &= \sum_{j=1}^{m-1} \int_{P_\epsilon} p_j \Big(\sum_{i=1}^{j} p_i - d(x_{(n)})\Big)^2 (\sum_{i=1}^{j} p_i)^n \prod_{k=1}^{m} p_k^{-1} p_1^{s-1-\alpha} p_m^{r-\beta} dp \end{aligned}$$

$$= \sum_{j=1}^{m-1} \Big[\prod_{k=1}^{j-2} \int_{\frac{(k/m)\epsilon}{1-[(m-k-1)/m]\epsilon}}^{\frac{1-[(m-k)/m]\epsilon}{1-[(m-k-1)/m]\epsilon}} q_k^{s-\alpha-2}(1-q_k)^{-1} dq_k \Big] \int_{\frac{[(j-1)/m]\epsilon}{1-[(m-j)/m]\epsilon}}^{\frac{1-[(m-j+1)/m]\epsilon}{1-[(m-j)/m]\epsilon}} q_{j-1}^{s-\alpha-2} dq_{j-1}$$
$$\times \int_{\frac{j}{m}\epsilon}^{1-\frac{(m-j)}{m}\epsilon} (u_j - d(x_{(n)}))^2 u_j^{n+1+s-2}(1-u_j)^{r-1} du_j$$
$$\times \prod_{k=j+1}^{m-1} \int_{\frac{\epsilon/m}{1-[(k-1)/m]\epsilon}}^{\frac{1-[(m-1)/m]\epsilon}{1-[(k-1)/m]\epsilon}} (1-r_k)^{r-\beta-1}(r_k)^{-1} dr_k, \tag{4.14}$$

and

$$\sum_{(\vec{x},\xi_j)\in V_m} I(d(\vec{x},\xi_j),\epsilon) = \int_{\frac{1}{m}\epsilon}^{1-\epsilon(m-2)/m} p_1^{s-\alpha-2}(1-p_1)^{r-\beta} dp_1$$
$$\times \Big[\prod_{k=2}^{m-1} \int_{\frac{\epsilon/m}{1-[(k-1)/m]\epsilon}}^{\frac{1-[(m-1)/m]\epsilon}{1-[(k-1)/m]\epsilon}} r_k^{-1}(1-r_k)^{r-\beta} dr_k \Big] (1-d(x_{(n)}))^2. \tag{4.15}$$

Note that (4.14) and (4.15) hold for $d = d_1$ and for $d = d_L$.

Note that by assumption, $l_0 \neq l_0'$ and for $t \geq x_{(n)}$ (thus for all $(\vec{x},t) \in V_1 \cup V_m$),

$$d_1(\vec{x},t) = l_0 \text{ (see (4.2)) and } d_L(\vec{x},t) = \frac{n+s}{n+r+s} (\neq l_0) \text{ (see (1.5) and (4.4))}. \tag{4.16}$$

Denote

$$\triangle = \int_0^1 [(u_j - d_L(\vec{x},\xi_j))^2 - (u_j - d_1(\vec{x},\xi_j))^2] u_j^{s+n-1}(1-u_j)^{r-1} du_j. \tag{4.17}$$

Then $\triangle$ is a constant for $(\vec{x},\xi_j) \in V_1$ and $\triangle < 0$ by (3.9). If $l_0 < \frac{n+r}{n+r+s}$, in view of (3.14) and (4.16), (3.10) holds. Thus we can assume that $l_0 > \frac{n+r}{n+r+s}$. Denoting the third integral in (4.14) by $\int_{\frac{j}{m}\epsilon}^{1-\frac{(m-j)}{m}\epsilon} G(d)$, we obtain

$$\Big|\Big(\int_{\frac{j}{m}\epsilon}^{1-\frac{(m-j)}{m}\epsilon} [G(d_L) - G(d_1)] \Big) - \triangle \Big| \leq \int_{1-\epsilon}^1 |G(d_1) - G(d_L)| + \int_0^\epsilon |G(d_1) - G(d_L)| = o(\epsilon)$$

uniformly in m. In view of (4.14), (4.17) and the above inequality, for small enough ϵ, uniformly in m,

$$- \sum_{(\vec{x},t)\in V_1} [I(d_L(\vec{x},t),\epsilon) - I(d_1(\vec{x},t),\epsilon)]$$
$$= \sum_{j=1}^{m-1} -[\triangle + o(\epsilon)] \Big\{ \Big[\prod_{k=1}^{j-2} \int_{\frac{(k/m)\epsilon}{1-[(m-k-1)/m]\epsilon}}^{\frac{1-[(m-k)/m]\epsilon}{1-[(m-k-1)/m]\epsilon}} q_k^{s-\alpha-2}(1-q_k)^{-1} dq_k \Big]$$
$$\times \int_{\frac{[(j-1)/m]\epsilon}{1-[(m-j)/m]\epsilon}}^{\frac{1-[(m-j+1)/m]\epsilon}{1-[(m-j)/m]\epsilon}} q_{j-1}^{s-\alpha-2} dq_{j-1} \Big[\prod_{k=j+1}^{m-1} \int_{\frac{\epsilon/m}{1-[(k-1)/m]\epsilon}}^{\frac{1-[(m-1)/m]\epsilon}{1-[(k-1)/m]\epsilon}} (1-r_k)^{r-\beta-1} r_k^{-1} dr_k \Big] \Big\}. \tag{4.18}$$

In view of (4.15) and (4.16),

$$\sum_{(\vec{x},t)\in V_m}[I(d_L(\vec{x},t),\epsilon)-I(d_1(\vec{x},t),\epsilon)]$$
$$=\int_{\frac{1}{m}\epsilon}^{1-\epsilon(m-2)/m} p_1^{s-\alpha-2}(1-p_1)^{r-\beta}dp_1\Big[\prod_{k=2}^{m-1}\int_{\frac{\epsilon/m}{1-[(k-1)/m]\epsilon}}^{\frac{1-[(m-1)/m]\epsilon}{1-[(k-1)/m]\epsilon}} r_k^{-1}(1-r_k)^{r-\beta}dr_k\Big]$$
$$\times\Big[(1-\frac{n+s}{n+r+s})^2-(1-l_0)^2\Big]. \tag{4.19}$$

In the following, we evaluate the order of integrals in (4.18) and (4.19), which depends on α,β,r and s. When $\beta=0$ and $r>0$, the case

$$\alpha=\beta=0 \text{ and } r=s=1 \tag{4.20}$$

is the most interesting and important one (see (1.3) and (1.4)). Assume first that (4.20) is true. Then for a large enough m (for a given small positive number ϵ), one has, for the factors in (4.18) and (4.19),

$$\int_{\frac{(k/m)\epsilon}{1-[(m-k-1)/m]\epsilon}}^{\frac{1-[(m-k)/m]\epsilon}{1-[(m-k-1)/m]\epsilon}} q_k^{s-\alpha-2}(1-q_k)^{-1}dq_k$$
$$=-\text{ln}k-2\text{ln}\epsilon+2\ln m+\ln(1-\frac{m-k}{m}\epsilon)+\ln(1-\frac{m-1}{m}\epsilon) \tag{4.21}$$
$$\approx 2\ln m-\text{ln}k-2\text{ln}\epsilon,\ k\geq 1.$$

Note that the approximation is uniform for all m. Similarly, uniformly in m,

$$\int_{\frac{[(j-1)/m]\epsilon}{1-[(m-j)/m]\epsilon}}^{\frac{1-[(m-j+1)/m]\epsilon}{1-[(m-j)/m]\epsilon}} q_{j-1}^{s-\alpha-2}dq_{j-1}\approx \text{ln}m-\ln(j-1)-\text{ln}\epsilon,\ j\geq 2$$
$$\int_{\frac{\epsilon/m}{1-[(k-1)/m]\epsilon}}^{\frac{1-[(m-1)/m]\epsilon}{1-[(k-1)/m]\epsilon}} (1-r_k)^{r-\beta-1}r_k^{-1}dr_k\approx \text{ln}m-\text{ln}\epsilon k\geq 1. \tag{4.22}$$

Thus (4.18) through (4.22) yields (for the given $\epsilon>0$ and for a large enough m),

$$\frac{-\sum_{(\vec{x},\xi_j)\in V_1}[I(d_L(\vec{x},\xi_j),\epsilon)-I(d_1(\vec{x},\xi_j),\epsilon)]}{\sum_{(\vec{x},\xi_j)\in V_m}[I(d_L(\vec{x},\xi_j),\epsilon)-I(d_1(\vec{x},\xi_j),\epsilon)]}$$
$$\approx\frac{-\sum_{j=1}^{m-1}(\triangle)\prod_{k=1}^{j-2}(2\text{ln}m-\text{ln}k-2\text{ln}\epsilon)(\text{ln}m-\ln(j-1)-\text{ln}\epsilon)^{1(j\geq 2)}\prod_{k=j+1}^{m-1}(\text{ln}m-\text{ln}\epsilon)}{[(\frac{1}{n+2})^2-(1-l_0)^2](\text{ln}m-\text{ln}\epsilon)^{m-1}}$$
$$\geq\sum_{j=3}^{m-1}\frac{-(\triangle)(-\text{ln}\epsilon)}{[(\frac{1}{n+2})^2-(1-l_0)^2](\text{ln}m-\text{ln}\epsilon)^2}$$
$$=(m-3)\frac{-(\triangle)(-\text{ln}\epsilon)}{[(\frac{1}{n+2})^2-(1-l_0)^2](\text{ln}m-\text{ln}\epsilon)^2}\geq 2. \tag{4.23}$$

The last inequality holds since $\lim_{m\to\infty}(m/(\ln m)^2) = \infty$. This completes the proof of (4.12) under the assumption (4.20). If (4.20) is not true, only (4.21) through (4.23) need to be modified. We skip the details; (4.13) can be shown in a similar manner.

Lemma 4. *Inequality* (4.10) *holds.*

Proof. The idea to prove (4.10) for each possible $\vec{x}$ is similar to the arguments along the lines (4.14) through (4.23). For simplicity, we only give the proof for the case $x_n \le \xi_J$. Note that both $d_1(\vec{x}, t)$ and $d_L(\vec{x}, t)$ are constant for $t > x_n$ (see (4.1)). Furthermore, $\eta_k = 0$ for $k > J$ since $x_n \le \xi_J$. It can be shown as in (4.14) that the sum of the first $m-1-J$ terms in the numerator (or denominator) of (4.10) is (for $d = d_L$ (or $d = d_1$))

$$\begin{aligned}
&\sum_{j>J}^{m-1} I(d(\vec{x}, \xi_j), \epsilon) \\
&= \sum_{j>J}^{m-1} \Big[\prod_{k=1}^{J} \int_{\frac{(k/m)\epsilon}{1-[(m-k-1)/m]\epsilon}}^{\frac{1-[(m-k)/m]\epsilon}{1-[(m-k-1)/m]\epsilon}} q_k^{\sigma_k+s-\alpha-2}(1-q_k)^{\eta_{k+1}-1} dq_k \Big] \\
&\times \Big[\prod_{k>J}^{j-2} \int_{\frac{(k/m)\epsilon}{1-[(m-k-1)/m]\epsilon}}^{\frac{1-[(m-k)/m]\epsilon}{1-[(m-k-1)/m]\epsilon}} q_k^{n+s-\alpha-2}(1-q_k)^{-1} dq_k \Big] \int_{\frac{[(j-1)/m]\epsilon}{1-[(m-j)/m]\epsilon}}^{\frac{1-[(m-j+1)/m]\epsilon}{1-[(m-j)/m]\epsilon}} q_{j-1}^{n+s-\alpha-2} dq_{j-1} \\
&\times \int_{\frac{j}{m}\epsilon}^{1-\frac{(m-j)}{m}\epsilon} (u_j - d(x_n))^2 u_j^{n+s-1}(1-u_j)^{r-1} du_j \\
&\quad \prod_{k=j+1}^{m-1} \int_{\frac{\epsilon/m}{1-[(k-1)/m]\epsilon}}^{\frac{1-[(m-1)/m]\epsilon}{1-[(k-1)/m]\epsilon}} (1-r_k)^{r-\beta-1} r_k^{-1} dr_k
\end{aligned} \tag{4.24}$$

and the last summand in the numerator (or denominator) of (4.23) is

$$\begin{aligned}
I(d(\vec{x}, \xi_m), \epsilon) = &\int_{\epsilon/m}^{1-\epsilon(m-2)/m} p_1^{\eta_1+s-\alpha-2}(1-p_1)^{n-\sigma_1+r-\beta} dp_1 \\
&\times \prod_{k>1}^{J} \int_{\frac{\epsilon/m}{1-[(k-1)/m]\epsilon}}^{\frac{1-[(m-1)/m]\epsilon}{1-[(k-1)/m]\epsilon}} (1-r_k)^{n-\sigma_k+r-\beta} r_k^{-1+\eta_k} dr_k \\
&\times \prod_{k>J}^{m-1} \int_{\frac{\epsilon/m}{1-[(k-1)/m]\epsilon}}^{\frac{1-[(m-1)/m]\epsilon}{1-[(k-1)/m]\epsilon}} (1-r_k)^{r-\beta} r_k^{-1} dr_k (1-d(x_n))^2.
\end{aligned} \tag{4.25}$$

To evaluate the order of the expressions in (4.24) and (4.25), WLOG, we can assume that (4.20) is true. Then, for a large enough m (and fixed $\epsilon > 0$), $k > J$,

$$\int_{\frac{(k/m)\epsilon}{1-[(m-k-1)/m]\epsilon}}^{\frac{1-[(m-k)/m]\epsilon}{1-[(m-k-1)/m]\epsilon}} q_k^{n+s-\alpha-2}(1-q_k)^{-1} dq_k \approx -\ln\epsilon - \ln(1-\epsilon) + \ln m,$$

$$\int_{\frac{[(j-1)/m]\epsilon}{1-[(m-j)/m]\epsilon}}^{\frac{1-[(m-j+1)/m]\epsilon}{1-[(m-j)/m]\epsilon}} q_{j-1}^{n+s-\alpha-2} dq_{j-1} \approx \frac{1}{n+s-\alpha-1} \text{ (denoted by } c_1 \text{)},$$

$$\int_{\frac{\epsilon/m}{1-[(k-1)/m]\epsilon}}^{\frac{1-[(m-1)/m]\epsilon}{1-[(k-1)/m]\epsilon}} (1-r_k)^{r-\beta-1}(r_k)^{-1} dr_k \approx \ln m - \ln\epsilon. \tag{4.26}$$

$$\int_{\frac{m-1}{m}\epsilon}^{1-\frac{1}{m}\epsilon} u_{m-1}^{n+s-\alpha-2}(1-u_{m-1})^{r-\beta} du_j \approx \int_0^1 u_{m-1}^{n+s-\alpha-2}(1-u_{m-1})^{r-\beta} du_j \; (= c_2).$$

The approximation in (4.26) is uniform in m. For a small $\epsilon > 0$, the substitution $d = d_L$ and $d = d_1$ into (4.24) and (4.25) yields

$$\begin{aligned}
&\frac{-\sum_{j>J}^{m-1}[I(d_L(\vec{x},\xi_j),\epsilon) - I(d_1(\vec{x},\xi_j),\epsilon)]}{[I(d_L(\vec{x},\xi_m),\epsilon) - I(d_1(\vec{x},\xi_m),\epsilon)]} \\
&\approx \frac{-\sum_{j>J}^{m-1}(\triangle)\prod_{k>J}^{j-2}(\ln m - \ln\epsilon)(c_1)\prod_{k=j+1}^{m-1}(\ln m - \ln\epsilon)}{[(\frac{1}{n+2})^2 - (1-l_0)^2]\prod_{k>J}^{m-1}(\ln m - \ln\epsilon)(c_2)} \\
&= \frac{-(m-1-J)(\triangle)c_1}{[(\frac{1}{n+2})^2 - (1-l_0)^2](\ln m - \ln\epsilon)^2(c_2)} \geq 2 \text{ if } m \text{ is large enough.}
\end{aligned} \tag{4.27}$$

Formula (4.27) yields (4.10) for $x_n \leq \xi_J$. Similarly, we can show (4.10) for each $\vec{x}$ such that $(\vec{x},\xi_m) \in V_m$ and $x_1 \leq \xi_J < x_n$.

Proof of Theorem 1 for case (2). Given an arbitrary estimator $d \in U$, we can assume $d \neq d_L$ on a subset of positive Lebesgue measure. In view of Lemma 1, by taking subsets and taking limits, we can assume that there exists an infinite sequence of points $\xi_1 < \xi_2 < \cdots$ and a finite integer $J > 0$ such that

(1) $d \neq d_L$ for some $(\vec{x},t) \in \{\xi_1,\xi_2,\ldots,\xi_J\}^{n+1}$;

(2) $d = d_n$ for $(\vec{x},t) \in \{\xi_1,\xi_2,\ldots\}^{n+1}$, where d_n takes on finitely many values.

Statement (2) implies that for $(\vec{x},t) \in \{\xi_1,\xi_2,\ldots\}^{n+1}$, d takes on finitely many values.

It follows from Lemma 2 that $d = d_L$ on $\{\xi_1,\xi_2,\ldots\}^{n+1}$. The latter equality contradicts (1). This concludes the proof of the theorem for the case (2).

5. Comment

The proof of the theorem for case (3) is very similar to that for case (2), except that it is more technical. We refer to a technical report (Yu (1994)). Finally, we give some comments on the new approach for proving the admissibility used in this paper. The method is summarized in Proposition 1. Note that if the B_k's in the inequality (1.6) are all singletons and if the τ_j in (1.6) is replaced by τ, then Proposition 1 reduces to the standard (generalized) Bayes argument. Thus, like Blyth's Lemma (1951), Proposition 1 is a refinement of the standard (generalized) Bayes argument. However, they are different, since Blyth's Lemma requires $\frac{r_j(\delta_0)-r_k}{\int_a^b d\tau_j} \to 0$, where $r_k(\delta_0)$ and r_k are the Bayes risks (with respect to τ_j) of δ_0 and the Bayes rule, respectively. We were not able to evaluate these two Bayes risks directly in our problem.

Under the conditions of Proposition 1, we do not require that the Bayes risk $r(\tau, \delta_0)$ is finite or $\delta_0(\vec{x})$ is a limit of posterior Bayes actions. Recall that the step-wise Bayes arguments have been widely used in proving admissibility when $r(\tau, \delta_0) = \infty$. Using the stepwise Bayes argument, the sample space is partitioned into several subsets. Then on each subset S_j, a proper prior π_j is selected so that $E_{\pi_j}(L(\theta, \delta_0(\vec{x}))|\vec{x}) < \infty$, where $\delta_0(\vec{x})$ is a Bayes estimate with respect to π_j for all $\vec{x} \in S_j$.

There are two differences between the new approach proposed in Proposition 1 and the stepwise Bayes approach. The new approach is simpler when both approaches are applicable to the same problem, such as in the special case considered in this paper, i.e., $\alpha = \beta = -1$ (which is solved by Brown (1988) using a step-wise Bayes argument). While the step-wise Bayes approach needs to find several priors for the different subsets of the sample space, the new approach essentially needs one prior τ^* (as in (3.3)) and the priors τ^*_ϵ, $\epsilon > 0$ as in (3.3), are truncated versions of τ^*. Another difference is that the new approach no longer requires that for each $\vec{x}$, $\delta_0(\vec{x})$ minimizes $E_\pi(L(\theta, \delta(\vec{x}))|\vec{x})$ w.r.t. some prior π. That is, it does not requires that $\delta_0(\vec{x})$ be a posterior Bayes action (or limit of posterior Bayes actions), whereas the stepwise Bayes argument or the others does (see Remark 1).

Proof of Proposition 1. We first assume that there is a rule δ which improves on δ_0, that is,

$$0 \leq R(\theta, \delta_0) - R(\theta, \delta) = E(\sum_k 1[\vec{X} \in B_k][L(\theta, \delta_0(\vec{X})) - L(\theta, \delta(\vec{X}))]), \quad (5.1)$$

where $\{B_k; k \in K\}$ is a partition of the subset $\{\vec{x} \in \mathcal{X}; \delta_0(\vec{x}) \neq \delta(\vec{x})\}$. Then we reach a contradiction.

By (5.1), we have $E(\sum_k 1[\vec{X} \in B_k]L(\theta, \delta_0(\vec{X})) \geq E(\sum_k 1[\vec{X} \in B_k]L(\theta, \delta(\vec{X}))$. Integrating expressions on both sides of the inequality over the prior $d\tau_j$ and changing the order of integration and expectation yield

$$E(\sum_k 1[\vec{X} \in B_k]E_{\tau_j}(L(\theta, \delta_0(\vec{X}))|\vec{X})) \geq E(\sum_k 1[\vec{X} \in B_k]E_{\tau_j}(L(\theta, \delta(\vec{X}))|\vec{X})).$$

It further yields

$$\frac{E(\sum_k 1[\vec{X} \in B_k]E_{\tau_j}(L(\theta, \delta_0(\vec{X}))|\vec{X}))}{E(\sum_k 1[\vec{X} \in B_k]E_{\tau_j}(L(\theta, \delta(\vec{X}))|\vec{X}))} \geq 1. \quad (5.2)$$

Assumption (1.6) in Proposition 1 yields

$$\lim_{j\to\infty} \frac{E(\sum_k 1[\vec{X}\in B_k]E_{\tau_j}(L(\theta,\delta_0(\vec{X}))|\vec{X}))}{E(\sum_k 1[\vec{X}\in B_k]E_{\tau_j}(L(\theta,\delta(\vec{X}))|\vec{X}))} < 1,$$

which contradicts to (5.2). This completes the proof of Proposition 1.

Acknowledgement

The author gratefully acknowledges the valuable opinions and suggestions from a referee. Partially supported by NSF Grants DMS 9202070 and DMS-9402561.

References

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

Berger, J. O. (1985). *Statistical Decision theory and Bayes Analysis. 2nd ed.* Spring-verlag. New York. p.159.

Blyth, C. R. (1951). On minimax statistical decision procedures and their admissibility. *Ann. Math. Statist.* **22**, 22-42.

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

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

Ferguson, T. S. (1967). Mathematical Statistics, A Decision Theoretic Approach. Academic Press, New York. p.697.

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

Yu, Q. (1989b). Admissibility of the empirical distribution function in the invariant problem. *Statist. and Decisions* **7**, 383-398.

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

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

Yu, Q. (1992). Minimaxity of the empirical distribution function in discrete invariant estimation of a distribution function. *Statist. and Decisions* **10**, 25-38.

Yu, Q. (1993). Admissibility of the empirical distribution function in discrete nonparametric invariant problems. *Statist. and Prob. Let.* **18**, 337-343.

Yu, Q. and Phadia, E. G. (1993). Admissibility of the best invariant estimator of a distribution function under the Kolmogorov-Smirnov Loss. *Comm. Statist. Theory Methods* **22**, 2103-2124.

Yu, Q. (1994). Technical report on "Admissibility of the best invariant estimator of a discrete distribution function". Dept. of Math, SUNY at Binghamton.

Yu, Q. and Kuo, L. (1995). An analogy between nonparametric problems of estimating a distribution function and their parametric versions. *Sankhya.* **57**, 472-485.

Department of Mathematical Sciences, State University of New York at Binghamton, Box 6000, Binghamton, NY 13902-6000, U.S.A.

E-mail: qyu@math.binghamton.edu

(Received June 1995; accepted April 1997)