*The Annals of Statistics*
1995, Vol. 23, No. 2, 598–607

# ADMISSIBILITY AND MINIMAXITY OF THE UMVU ESTIMATOR OF $P\{X < Y\}$

BY QIQING YU[1] AND Z. GOVINDARAJULU

*State University of New York, Stony Brook and University of Kentucky*

Suppose that $X_1, \dots, X_m$ are i.i.d. from a continuous distribution function $F$ and $Y_1, \dots, Y_n$ are i.i.d. from a continuous distribution function $G$; $X$'s and $Y$'s are independent. A minimum variance unbiased estimator of $P\{X < Y\}$ is the Mann–Whitney statistic. We show that the Mann–Whitney statistic is admissible under a class of weighted squared error losses and is minimax under a proper weighted squared error loss.

**1. Introduction.** Suppose that $X_1, \dots, X_m$ are i.i.d. from a continuous distribution function $F$ and $Y_1, \dots, Y_n$ are i.i.d. from a continuous distribution function $G$; $X$'s and $Y$'s are independent. The Mann–Whitney statistic [Mann and Whitney (1947)], $\hat{p}_0$, is a minimum variance unbiased estimator of $P\{X < Y\}$, where

$$\hat{p}_0 = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} 1[X_j \le Y_i], \tag{1.1}$$

and $1[A]$ is the indicator function of a set $A$. $\hat{p}_0$ is consistent and asymptotically normal [Govindarajulu (1967, 1976)].

In Section 2, we consider the admissibility and minimaxity issue. We prove that $\hat{p}_0$ is admissible under a wide class of loss functions, with the form

$$(\hat{p} - p)^2 h(F, G) \quad \text{where } h(F,G) \text{ is a positive function.} \tag{1.2}$$

Furthermore, $\hat{p}_0$ is minimax under the loss function $(\hat{p}_0 - p)^2/\sigma^2(F, G)$, where $\sigma^2(F, G) = \mathrm{Var}(\hat{p}_0)$, that is,

$$\sigma^2(F,G) = \Big\{ p - p^2 + (m-1)\int \big(F^2(t) - p^2\big)\, dG(t) + (n-1)\int \Big[(1 - G(t))^2 - p^2\Big]\, dF(t) \Big\} \Big/ (mn).$$

It is worth noting the following well-known results about the binomial distribution.

*Suppose that $X$ has a binomial distribution with sample size $n$ and probability of success $\theta$. The estimator $\hat{\theta} = X/n$ is admissible under the loss*

Received October 1992; revised September 1993.

[1]Partially supported by NSF Grant DMS-94-02561 and DAMD17-94-J-4332.

*AMS* 1991 *subject classifications*. Primary 62C15; secondary 62C10.

*Key words and phrases.* Admissibility, minimaxity, multinomial distribution, conditional Bayes estimator.

$(\hat{\theta} - \theta)^2 h(\theta)$, *where* $h(\theta)$ *is a positive weight function. Furthermore,* $X/n$ *is minimax under the loss* $(\hat{\theta} - \theta)^2/[\theta(1-\theta)]$. [*See, e.g., Yu and Kuo* (1992)].

It should be pointed out that $\mathrm{Var}(X/n) = \theta(1-\theta)/n$ and $(\hat{p}_0 - p)^2/\mathrm{Var}(\hat{p}_0) \neq (\hat{p}_0 - p)^2/[p(1-p)/n]$. Thus these two problems are similar but not exactly parallel.

It is natural to consider using a traditional Bayes approach in proving admissibility and minimaxity. A difficulty arises since the family of distribution functions we considered is one of all *continuous* distribution functions. The well-known Dirichlet prior [Ferguson (1973)] is a prior which takes discrete distribution functions with probability 1. So we do not have a suitable prior on the family of continuous distribution functions.

Yu and Kuo (1992) showed that the empirical distribution function $\hat{F}(t) = (\sum_{i=1}^{m} 1[t \geq X_i])/m$ is admissible and minimax under the loss function

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

where $dW$ is a finite measure and $F(t)$ is an unknown continuous distribution function. Thus they faced the same difficulty as we do here. In our original approach, we used an idea similar to that of Yu and Kuo (1992). Due to a referee's suggestion and the special feature of this problem, the current proof uses a different approach to overcome this difficulty.

**2. Main results.** We first address the admissibility issue. There are two steps in our proof in order to overcome this difficulty that there is no proper prior over the family of all continuous distribution functions. The first step is to consider the discrete version of the problem, that is, to assume that $G(t)$ is a discrete distribution function. By making a transformation, the problem can be reduced to one with the family of multinomial distribution functions. Thus we can make use of the result on the multinomial distribution function. The second step is to convert the continuous problem to the discrete version mentioned in step 1.

Since we need to transfer the nonparametric problem to a multinomial distribution problem, we first introduce the latter problem: The multinomial distribution has a probability density function (pdf) of the form $f_m(\boldsymbol{\eta}) = \binom{m}{\boldsymbol{\eta}} \prod_{i=1}^{n+1} \pi_i^{\eta_i}$, where $\boldsymbol{\eta} = (\eta_1, \dots, \eta_{n+1})$, the $\eta_i$'s are nonnegative integers, $\sum_{i=1}^{n+1} \eta_i = m$, $\pi_1, \dots, \pi_{n+1} \geq 0$, $\pi_1 + \cdots + \pi_{n+1} = 1$ and $\binom{m}{\boldsymbol{\eta}}$ is the multinomial coefficient. The parameter of interest is $p = \sum_{j=1}^{n} a_j \pi_j$, where the $a_i$'s are constants. The loss function is $(\hat{p} - p)^2 h(\boldsymbol{\pi})$, where $h$ is a positive function. An estimator of $p$ is $\delta_\tau(\boldsymbol{\eta}) = \sum_{i=1}^{n} a_i \eta_i / m$. We have the following result.

LEMMA 1. *Suppose that* $\mathbf{M}$ *has a multinomial distribution, with the pdf* $f_m(\boldsymbol{\eta})$. *Let* $\tilde{p}$ *be an estimator of* $p$ *which does not equal* $\delta_\tau$. *Let* $k$ *be the smallest positive integer* $(1 \leq k \leq n)$ *such that* $\tilde{p}(\boldsymbol{\eta}) \neq \sum_{i=1}^{n} a_i \eta_i / m$ *for at least one* $\boldsymbol{\eta}$ *which has* $k$ *entries of* $\boldsymbol{\eta}$, $\eta_{i_1}, \dots, \eta_{i_k} \neq 0$, *and the remaining entries* $\eta_i = 0$,

*$i \neq i_1, \dots, i_k$. For these $i_1, \dots, i_k$, let*

$$d\tau = \begin{cases} (h(\boldsymbol{\pi}))^{-1} \prod_{j=1}^{k} \pi_{i_j}^{-1} 1[\pi_{i_1} + \cdots + \pi_{i_k} = 1]\, d\pi_{i_1} \cdots d\pi_{i_{k-1}}, & \text{if } k \geq 2, \\ (h(\boldsymbol{\pi}))^{-1} \pi_{i_1}^{-1} \pi_1^{-1} 1[\pi_{i_1} + \pi_1 = 1]\, d\pi_{i_1}, & \text{if } i_1 \neq 1 \text{ and } k = 1, \\ (h(\boldsymbol{\pi}))^{-1} \pi_{i_1}^{-1} \pi_2^{-1} 1[\pi_{i_1} + \pi_2 = 1]\, d\pi_{i_1}, & \text{if } i_1 = 1 \text{ and } k = 1, \end{cases}$$

*be a prior on $\boldsymbol{\pi} = (\pi_1, \dots, \pi_n)$. Then*

$$\int [R(p, \delta_\tau) - R(p, \tilde{p})]\, d\tau < 0. \tag{2.1}$$

Note that the Bayes risks may not be finite, thus, (2.1) cannot be replaced by

$$\int R(p, \delta_\tau)\, d\tau < \int R(p, \tilde{p})\, d\tau.$$

PROOF. Under the assumption given in the lemma, the risk of an estimator $\tilde{p}$ is

$$R(p, \tilde{p}) = \sum_{\boldsymbol{\eta}} (p - \tilde{p}(\boldsymbol{\eta}))^2 h(\boldsymbol{\pi}) \binom{m}{\boldsymbol{\eta}} \prod_{i=1}^{n+1} \pi_i^{\eta_i}. \tag{2.2}$$

In the following, we use the notation as in Lemma 1. If $k \leq 2$, the prior $d\tau$ is proportional to a binomial distribution, and the lemma is trivial. Thus without loss of generality (w.l.o.g.), we can assume $k > 2$. Let $V = \{\boldsymbol{\eta};\ \tilde{p}(\boldsymbol{\eta}) \neq \sum_{i=1}^{n} a_i \eta_i / m\}$ and $V_\tau = \{\boldsymbol{\eta} \in V;\ \eta_i = 0 \text{ if } i \notin \{i_1, \dots, i_k\}\}$. Then due to the definition of $k$, $V_\tau$ is a subset of $V$ such that $\boldsymbol{\eta}$ has nonzero entries of $\eta_{i_1}, \dots, \eta_{i_k}$ and zero entries otherwise.

The reason that we restrict our attention to $V_\tau$ is as follows. If $\boldsymbol{\eta}^* \notin V_\tau$ but $\boldsymbol{\eta}^* \in V$, by definition, there is an entry of $\boldsymbol{\eta}^*$, say $\eta_{i_0}^* \neq 0$, where $i_0 \notin \{i_1, \dots, i_k\}$. Under $d\tau$, $\pi_{i_0} = 0$, and then $\pi_{i_0}^{\eta_{i_0}^*} = 0$. Consequently, the corresponding summand in (2.2),

$$(p - \tilde{p}(\boldsymbol{\eta}))^2 h(\boldsymbol{\pi}) \binom{m}{\boldsymbol{\eta}^*} \prod_{i=1}^{n+1} \pi_i^{\eta_i^*} = 0,$$

for any estimator $\tilde{p}(\boldsymbol{\eta})$, including $\delta_\tau$. It follows that we have

$$\begin{aligned} &\int [R(p, \delta_\tau) - R(p, \tilde{p})]\, d\tau \\ &\quad = \int E\{(p - \delta_\tau(\mathbf{M}))^2 h(\boldsymbol{\pi}) 1[\mathbf{M} \in V_\tau]\}\, d\tau \\ &\qquad - \int E\{(p - \tilde{p}(\mathbf{M}))^2 h(\boldsymbol{\pi}) 1[\mathbf{M} \in V_\tau]\}\, d\tau \\ &\quad = \int \left\{ \sum_{\boldsymbol{\eta} \in V_\tau} \left[(p - \delta_\tau(\boldsymbol{\eta}))^2 - (p - \tilde{p}(\boldsymbol{\eta}))^2\right] h(\boldsymbol{\pi}) \binom{m}{\boldsymbol{\eta}} \prod_{i=1}^{n+1} \pi_i^{\eta_i} \right\} d\tau. \end{aligned} \tag{2.3}$$

Note under the prior $d\tau$, $\pi_i = 0$ for $i \notin \{i_1, \dots, i_k\}$, and consequently, if $\boldsymbol{\eta} \in V_\tau$, then $\eta_i = 0$ (a.s.) for $i \notin \{i_1, \dots, i_k\}$. We define $(\pi_i^{\eta_i} =)\ 0^0 = 1$ if $i \notin \{i_1, \dots, i_k\}$. Integrating each summand corresponding to $\boldsymbol{\eta} \in V_\tau$ in expression (2.2) yields

$$
\begin{aligned}
0 &\le \int \cdots \int \binom{m}{\boldsymbol{\eta}} \left( \sum_{i=1}^{n} a_i \pi_i - \tilde{p}(\boldsymbol{\eta}) \right)^2 \\
&\quad \times \prod_{j=1}^{k} \pi_{i_j}^{\eta_{i_j}} h(\boldsymbol{\pi}) (h(\boldsymbol{\pi}))^{-1} \pi_{i_1}^{-1} \cdots \pi_{i_k}^{-1}\, d\pi_{i_1} \cdots d\pi_{i_{k-1}} \qquad (2.4) \\
&= \int \cdots \int \binom{m}{\boldsymbol{\eta}} \left( \sum_{h=1}^{k} a_{i_h} \pi_{i_h} - \tilde{p}(\boldsymbol{\eta}) \right)^2 \prod_{j=1}^{k} \pi_{i_j}^{\eta_{i_j} - 1}\, d\pi_{i_1} \cdots d\pi_{i_{k-1}} < \infty,
\end{aligned}
$$

since $\eta_{i_j} - 1 \ge 0$ for $j = 1, \dots, k$, and $\pi_i = 0$ if $i \notin \{i_1, \dots, i_k\}$. Then the multiple integral in (2.4) can be evaluated for each $\boldsymbol{\eta} \in V_\tau$. Note that $h(\boldsymbol{\pi}) h^{-1}(\boldsymbol{\pi}) = 1$, so w.l.o.g. we can assume that $h(\boldsymbol{\pi}) = 1$. It follows that the Bayes estimator $\delta_b$ is

$$
\delta_b = E_\tau(p \mid \boldsymbol{\eta}) / E_\tau(1 \mid \boldsymbol{\eta}) = \sum_{j=1}^{k} a_{i_j} E_\tau\left(\pi_{i_j} \mid \boldsymbol{\eta}\right) / E_\tau(1 \mid \boldsymbol{\eta}).
$$

It can be shown that

$$
E_\tau\left(\pi_{i_j} \mid \boldsymbol{\eta}\right) / E_\tau(1 \mid \boldsymbol{\eta}) = \eta_{i_j} / m, \qquad j = 1, \dots, k \tag{2.5}
$$

[the proof is similar to that in Brown (1988), page 1576]. Then

$$
\begin{aligned}
\delta_b &= \sum_{j=1}^{k} a_{i_j} E_\tau\left(\pi_{i_j} \mid \boldsymbol{\eta}\right) / E_\tau(1 \mid \boldsymbol{\eta}) \\
&= \sum_{j=1}^{k} a_{i_j} \eta_{i_j} / m = \sum_{i=1}^{n} a_i \eta_i / m = \delta_\tau;
\end{aligned} \tag{2.6}
$$

the third equality holds since $\eta_i = 0$ if $i \notin \{i_1, \dots, i_k\}$ and if $\boldsymbol{\eta} \in V_\tau$. It is important to note that

$$
\tilde{p} \ne \delta_\tau = \delta_b \quad \text{if } \boldsymbol{\eta} \in V_\tau. \tag{2.7}
$$

Note that integrating each term with $\boldsymbol{\eta} \in V_\tau$ over $d\tau$ in (2.2) yields a finite value. In view of this together with (2.5), (2.6) and (2.7), (2.1) follows. This completes the proof of the lemma. □

To the best of our knowledge, in the literature of the multinomial problem [see, e.g., Brown (1988)], the case of estimating an arbitrary linear combination of $\pi_1, \dots, \pi_n$ has not been considered. The proof of the lemma is not trivial either. Let $Y_{(1)}, \dots, Y_{(n)}$ be the order statistics of $Y_1, \dots, Y_n$, $Y_{(0)} = -\infty$ and $Y_{(n+1)} = \infty$. Denote $Z_1 = \sum_{i=1}^m 1[X_i \le Y_{(1)}]$, $Z_j = \sum_{i=1}^m 1[X_i \le Y_{(j)}] - \sum_{i=1}^m 1[X_i \le Y_{(j-1)}]$, $j = 2, \dots, n$, and $Z_{n+1} = m - Z_1 - \cdots - Z_n$. To take advantage of the result in Lemma 1, we need the following lemma, which enables us to reduce the family of all continuous distribution functions to the family of multinomial distribution functions.

LEMMA 2. *Given* $Y_{(i)} = y_i$, $i = 0, \dots, n, n+1$, $\mathbf{Z} = (Z_1, \dots, Z_{n+1})$ *has the conditional density function* $P\{\mathbf{Z} = \boldsymbol{\eta} \mid \mathbf{Y} = \mathbf{y}\} = f_m(\boldsymbol{\eta}) = \binom{n}{\boldsymbol{\eta}} \prod_{i=1}^{n+1} \pi_i^{\eta_i}$, *where* $\pi_i = F(y_i) - F(y_{i-1})$, $i = 1, 2, \dots, n+1$.

The proof of the lemma is trivial and is omitted.

Now we make use of the result based on the multinomial distribution. Let $N$ be the cardinality of the support set $\{\xi_1, \dots, \xi_N\}$ of $G$ and let $\mathscr{M}_N$ be the family of all discrete distributions $G(t)$ with the cardinality of the support set less than or equal to $N$. In the following theorem we first consider a discrete version of our problem.

THEOREM 1. *Suppose that* $X_1, \dots, X_m$ *are i.i.d. from a distribution function* $F$, $Y_1, \dots, Y_n$ *are i.i.d. from an arbitrary discrete distribution function* $G$ *and* $X$'*s and* $Y$'*s are independent. Suppose that the loss function is as in* (1.2). *Suppose that* $\{\xi_1, \dots, \xi_N\}$ *is the support set of* $G$, *where* $N$ *is an integer. If there is an estimator* $\tilde{p}$ *and* $\mathbf{y}$ *such that* $\tilde{p}(\mathbf{x}, \mathbf{y}) \neq \hat{p}_0(\mathbf{x}, \mathbf{y})$ *on a set of* $\mathbf{x}$ *with positive measure and* $y_1, \dots, y_n \in \{\xi_1, \dots, \xi_N\}$, *then there exists a prior* $d\tau$ *on* $F$ *such that*

$$\int R(p(F,G), \hat{p}_0)\, d\tau - \int R(p(F,G), \tilde{p})\, d\tau < 0. \tag{2.8}$$

PROOF. Since the order statistic is a sufficient statistic, without loss of generality, we can restrict ourselves to the class of estimators which are functions of the order statistic. With a slight abuse of the notation, we let $(Y_1, \dots, Y_n)$ be the order statistic [of $(Y_1, \dots, Y_n)$]. As a consequence of Lemma 2, there is a transformation from $(X_1, \dots, X_m, Y_1, \dots, Y_n)$ [$= (\mathbf{X}, \mathbf{Y})$] to $(\mathbf{Z}, \mathbf{Y})$, where $\mathbf{Z} = (Z_1, \dots, Z_{n+1})$ is a multinomial random vector. An estimator of $p$ can be written as $\tilde{p} = \tilde{p}(\boldsymbol{\eta}, \mathbf{y})$ [write $\mathbf{Z} = (\eta_1, \dots, \eta_{n+1})$]. It can be shown that

$$\hat{p}_0 = \frac{n}{mn}\eta_1 + \frac{n-1}{mn}\eta_2 + \cdots + \frac{1}{mn}\eta_n. \tag{2.9}$$

That is, $\hat{p}_0$ is a linear combination of $\eta_i$'s (or $Z_i$'s). Write $\hat{p}_0 = \hat{p}_0(\boldsymbol{\eta}, \mathbf{y})$. Furthermore, it can be verified that

$$p = \sum_{\mathbf{y}} \left\{ \frac{n}{n}\pi_1 + \cdots + \frac{1}{n}\pi_n \right\} P\{\mathbf{Y} = \mathbf{y}\}, \tag{2.10}$$

where $\pi_i = F(y_i) - F(y_{i-1})$, $i = 1, \dots, n+1$, and that the random vector $(\mathbf{Z}, \mathbf{Y})$ has the distribution $dP = \binom{n}{\boldsymbol{\eta}} \prod_{i=1}^{n+1} \pi_i^{\eta_i} P\{\mathbf{Y} = \mathbf{y}\}$. If $G$ is discrete, then the risk of $\tilde{p}$ is

$$R(p(F,G),\tilde{p}) = \sum_{\mathbf{y}} \sum_{\boldsymbol{\eta}} (p - \tilde{p}(\boldsymbol{\eta}))^2 h(F,G) \binom{n}{\boldsymbol{\eta}} \prod_{i=1}^{n+1} \pi_i^{\eta_i} P\{\mathbf{Y} = \mathbf{y}\}. \tag{2.11}$$

We proceed to prove the theorem by induction on $N$.

$N = 1$. Then $Y_i = y_i = \xi_1$ with probability 1, $\eta_2 = \cdots = \eta_n = 0$. Further, $Z_2 = \cdots = Z_n = 0$ and $p = F(\xi_1)\ (= \pi_1)$ and $\pi_2 = \cdots = \pi_n = 0$. Thus,

$$\begin{aligned} R(p(F,G),\tilde{p}) &= \sum_{\boldsymbol{\eta}} (p - \tilde{p}(\boldsymbol{\eta}))^2 h(F,G) \binom{n}{\boldsymbol{\eta}} \prod_{i=1}^{n+1} \pi_i^{\eta_i} \\ &= \sum_{\boldsymbol{\eta}} (p - \tilde{p}(\boldsymbol{\eta}))^2 h(F,G) \binom{n}{\boldsymbol{\eta}} \pi_1^{\eta_1} \pi_{n+1}^{\eta_{n+1}}. \end{aligned}$$

Note that the risk is the same as one in the multinomial problem, $p = \pi_1$ and $\hat{p}_0 = \eta_1/m$. It follows from Lemma 1 that if $\tilde{p} \neq \hat{p}_0$, then we have (2.1) or (2.8).

Now assume that the theorem is true for $N - 1$. We need to show that the theorem is true also for $N$. Without loss of generality, we can assume that for any proper subset of $\{\xi_1, \dots, \xi_N\}$, say, $\{\xi_{i_1}, \dots, \xi_{i_k}\}$ $(k < N)$, we have $\tilde{p} = \hat{p}_0$ when $y_1, \dots, y_n \in \{\xi_{i_1}, \dots, \xi_{i_k}\}$. Otherwise, $k < N$ implies (2.8) by the induction assumption and the proof is completed. Let $V$ be the set of $(\boldsymbol{\eta}, \mathbf{y})$ such that $\tilde{p} \neq \hat{p}_0$. Then for any $(\boldsymbol{\eta}, \mathbf{y}) \in V$, the set of distinct points in $\{y_1, \dots, y_n\}$ is the same as the set $\{\xi_1, \dots, \xi_N\}$. Without loss of generality, we can further assume that $n \geq N$. Given $\mathbf{y} = (y_1, \dots, y_n)$, let $\psi_i = \psi_i(\mathbf{y}) = \sum_{j=1}^n 1[y_j = \xi_i]$, that is, the number of $y_j = \xi_i$. Then $\psi_i(\mathbf{y}) \geq 1$ if $(\boldsymbol{\eta}, \mathbf{y}) \in V$.

Let $d\mu = (h(F,G))^{-1} \rho_1^{-1} \cdots \rho_N^{-1}\, d\rho_1 \cdots d\rho_{N-1}$, where $\rho_1 = F(\xi_1)$, $\rho_2 = F(\xi_2) - F(\xi_1), \dots$ and $\rho_N = F(\xi_N) - F(\xi_{N-1})$. It can be checked that for any $(\boldsymbol{\eta}, \mathbf{y}) \in V$, either $y_i - y_{i-1} = 0$ and thus $\pi_i = 0$, or $y_i - y_{i-1} = \xi_j - \xi_{j-1}$ and thus

$$\pi_i = \rho_j \quad \text{where } i = \psi_1 + \cdots + \psi_{j-1} + 1. \tag{2.12}$$

In the latter case [i.e., (2.12) holds], denote

$$i_j = i_j(\mathbf{y}) = i, \qquad j = 1, \dots, N. \tag{2.13}$$

If $\mathbf{y}$ and $k = N$ are given, then $d\mu = d\tau$, where $d\tau$ is defined as in Lemma 1. For any $(\boldsymbol{\eta}, \mathbf{y}) \in V$, if $\pi_i = 0$, then $\eta_i = 0$ (with probability 1). In this case, define $\pi_i^{\eta_i} = \pi_i^0 = 1$. Given $G \in \mathscr{M}_N$, the measure $d\mu$ is a measure on $F$ only and is independent of $P\{\mathbf{Y} = \mathbf{y}\}$. Thus, $p\ [= P\{X \leq Y\}$; see (2.10)] is a linear combination of $\rho_1, \dots, \rho_N$. That is,

$$p = \sum_{\mathbf{y}} \sum_{i=1}^{n} P\{\mathbf{Y} = \mathbf{y}\} \left( \frac{n - i + 1}{n} \right) \pi_i(\mathbf{y}) = \sum_{\mathbf{y}} \sum_{j=1}^{N} P\{\mathbf{Y} = \mathbf{y}\} \left( \frac{n - i_j + 1}{n} \right) \rho_j.$$

where given $\mathbf{y}$, $i_j$ satisfies (2.13) [note that implicitly, $i$ and $j$ in (2.13) depend on **y**]. In view of (2.9), $\hat{p}_0$ can be written as

$$\begin{aligned}\hat{p}_0 &= \left[\frac{n}{mn}\eta_1 + \frac{n-1}{mn}\eta_2 + \cdots + \frac{1}{mn}\eta_n\right]\sum_{\mathbf{y}} P\{\mathbf{Y}=\mathbf{y}\}\\ &= \sum_{\mathbf{y}}\sum_{i=1}^{n} P\{\mathbf{Y}=\mathbf{y}\}\left(\frac{n-i+1}{n}\right)\frac{\eta_i}{m}\\ &= \sum_{\mathbf{y}}\sum_{j=1}^{N} P\{\mathbf{Y}=\mathbf{y}\}\left(\frac{n-i_j+1}{n}\right)\frac{\eta_{i_j}}{m}.\end{aligned}$$

Note that

$$\begin{aligned}&\int R(p(F,G),\tilde{p})\,d\mu\\ &\quad= \int \sum_{\mathbf{y}}\sum_{\boldsymbol{\eta}} (p-\tilde{p}(\boldsymbol{\eta}))^2 h(F,G)\binom{n}{\boldsymbol{\eta}}\prod_{i=1}^{n+1}\pi_i^{\eta_i} P\{\mathbf{Y}=\mathbf{y}\}\,d\mu\\ &\quad= \sum_{\mathbf{y}}\sum_{\boldsymbol{\eta}} \int (p-\tilde{p}(\boldsymbol{\eta}))^2 h(F,G)\binom{n}{\boldsymbol{\eta}}\prod_{j=1}^{N}\rho_j^{\eta_{i_j}}\,d\mu\, P\{\mathbf{Y}=\mathbf{y}\} \qquad (2.14)\\ &\quad= \sum_{\mathbf{y}}\sum_{\boldsymbol{\eta}} \int\cdots\int (p-\tilde{p}(\boldsymbol{\eta}))^2 h(F,G)\binom{n}{\boldsymbol{\eta}}\\ &\qquad\times\prod_{j=1}^{N}\rho_j^{\eta_{i_j}-1}\,d\rho_1\cdots d\rho_{N-1}\,P\{\mathbf{Y}=\mathbf{y}\}.\end{aligned}$$

It is important to note that the integral of the summand corresponding to $(\boldsymbol{\eta},\mathbf{y}) \in V$ in (2.14) exists, since $\eta_{i_j} - 1 \geq 0$ by the induction assumption, and given $(\boldsymbol{\eta},\mathbf{y}) \in V$, the Bayes rule with respect to $d\mu$ is $\hat{p}_0$. The derivation is the same as in the proof of Lemma 1. It follows that

$$\begin{aligned}&\int\cdots\int (p-\hat{p}_0(\boldsymbol{\eta}))^2 h(F,G)\binom{n}{\boldsymbol{\eta}}\prod_{j=1}^{N}\rho_j^{\eta_{i_j}-1}\,d\rho_1\cdots d\rho_{N-1}\\ &\qquad\left\{\begin{matrix}<\\=\end{matrix}\right\}\int\cdots\int (p-\tilde{p}(\boldsymbol{\eta}))^2 h(F,G)\binom{n}{\boldsymbol{\eta}}\\ &\qquad\qquad\times\prod_{j=1}^{N}\rho_j^{\eta_{i_j}-1}\,d\rho_1\cdots d\rho_{N-1}\begin{cases}\text{if } (\boldsymbol{\eta},\mathbf{y}) \in V,\\ \text{otherwise.}\end{cases}\end{aligned}$$

Consequently, (2.8) holds. Thus the theorem holds for $N$. This completes the induction argument and the proof of the theorem. □

The theorem implies that $\hat{p}_0$ is admissible in the discrete case. In the following theorem we will remove the discreteness-assumption on $G$.

THEOREM 2. *Suppose that $X_1,\ldots,X_m$ are i.i.d. from a continuous distribution function $F$, $Y_1,\ldots,Y_n$ are i.i.d. from a continuous distribution function*

*G and X's and Y's are independent. Suppose that the loss function is* $(\hat{p} - p)^2 h(F, G)$. *Then* $\hat{p}_0$ *is admissible.*

PROOF. We will show the following:

(C.1) If there is an estimator $\tilde{p}$ such that its risk is smaller than and equal to the risk of $\hat{p}_0$ for all $F$ and $G$ in the family of all continuous distribution functions, then

$$\tilde{p} = \hat{p}_0 \quad \text{with probability 1 for all } F \text{ and } G. \tag{2.15}$$

We will assume that (C.1) does not imply (2.15) and reach a contradiction. Due to sufficiency, we restrict ourselves to the class of estimators which are functions of order statistics $X_1 \leq \cdots \leq X_m$ and $Y_1 \leq \cdots \leq Y_n$. Suppose that $\tilde{p}$ satisfies (C.1) and $\tilde{p} \neq \hat{p}_0$ on a set $B$ of positive measure. W.l.o.g., we can assume that the measure is the Lebesgue measure (in $R^{m+n}$). Then using Littlewood's three principles [see, e.g., Royden (1968), page 71], it can be shown that there is a product set $I$ such that $B \cap I$ is of positive measure, where $I = [a_1, b_1) \times \cdots \times [a_m, b_m) \times [u_1, v_1) \times \cdots \times [u_n, v_n)$ and $[a_i, b_i)$'s and $[u_i, v_i)$'s are disjoint intervals. Furthermore, $b_i < a_{i+1}$ and $v_i < u_{i+1}$ for all $i$ (due to order statistics).

Without loss of generality, we can further assume that $[a_i, b_i) = [2i, 2i + 1)$ and $[u_i, v_i) = [2i - 1, 2i)$ for all possible $i$. Define a subclass of the family of all continuous distribution functions $F$ and $G$ as follows: $F$ and $G$ have densities of the forms

$$\begin{aligned} f(x) &= \sum_{i=0}^{N} \alpha_i 1[x \in [2i, 2i + 1)] \quad \text{and} \\ g(y) &= \sum_{i=1}^{N} \beta_i 1[y \in [2i - 1, 2i)], \end{aligned} \tag{2.16}$$

respectively, where $\alpha_i, \beta_j \geq 0$ and $\sum_{i=0}^{N} \alpha_i = \sum_{i=1}^{N} \beta_i = 1$. Then $\tilde{p} \neq \hat{p}_0$ on a set $B$ of positive measure for some $F$ and $G$ satisfying (2.16). Define the transformations

$$U_j = \sum_{i=1}^{m} 1[X_i \in [2j - 1, 2j)], \qquad j = 0, \ldots, N$$

and

$$W_j = \sum_{i=1}^{n} 1[Y_i \in [2j, 2j + 1)], \qquad j = 1, \ldots, N.$$

Let $\mathbf{U} = (U_0, \ldots, U_N)$, $\mathbf{W} = (W_1, \ldots, W_N)$ and denote in an obvious way $\mathbf{u}$, $\mathbf{w}$, etc. Then $\mathbf{U}$ and $\mathbf{W}$ are random vectors with discrete density functions $f_1(\mathbf{u}) = P\{\mathbf{U}(\mathbf{X}) = \mathbf{u}\}$ and $g_1(\mathbf{w}) = P\{\mathbf{W}(\mathbf{Y}) = \mathbf{w}\}$, respectively.

Since we want to take advantage of Theorem 1 (the discrete version of Theorem 2), we define a class of discrete distribution functions $F_d$ and $G_d$

parallel to those in (2.16) with densities as follows:

$$f_d(i) = \alpha_i, \quad i = 0, 1, \dots, N, \quad \text{and} \tag{2.17}$$
$$g_d(i) = \beta_i, \quad i = 1, \dots, N.$$

Let $U_{id} = \sum_{j=0}^{m} 1[X_j = i]$ and $W_{id} = \sum_{j=1}^{n} 1[Y_j = i]$ for all possible $i$, and denote in an obvious way $\mathbf{U}_d$ and $\mathbf{W}_d$.

It can be verified that:

(f.1) $\mathbf{U}$ and $\mathbf{U}_d$ have the same multinomial distribution and so do $\mathbf{W}$ and $\mathbf{W}_d$.
(f.2) $p(F,G) = p(F_d, G_d)$.
(f.3) $(p(F,G) - \hat{p}_0(\mathbf{x},\mathbf{y}))^2$ is constant on the set $B_{\mathbf{u},\mathbf{w}} = \{(\mathbf{x},\mathbf{y});\ \mathbf{U}(\mathbf{x}) = \mathbf{u}, \mathbf{W}(\mathbf{y}) = \mathbf{w}\}$.

On the latter set, we write $\hat{p}(\mathbf{u},\mathbf{w}) = \hat{p}(\mathbf{x},\mathbf{y})$. Replacing $(F,G)$ by $(F_d, G_d)$, the risk of the estimator $\hat{p}_0$ remains the same, since

$$\begin{aligned} &R(p(F,G), \hat{p}_0) \\ &= \sum_{\mathbf{u},\mathbf{w}} \int \cdots \int_{B_{\mathbf{u},\mathbf{w}}} (p(F,G) - \hat{p}_0(\mathbf{x},\mathbf{y}))^2 \prod_{i=1}^{m} g(x_i) \prod_{i=1}^{n} f(y_i)\, d\mathbf{x}\, d\mathbf{y} \\ &= \sum_{\mathbf{u},\mathbf{w}} (p(F,G) - \hat{p}_0(\mathbf{u},\mathbf{w}))^2 g_1(\mathbf{u}) f_1(\mathbf{w}) \\ &= R(p(F_d, G_d), \hat{p}_0) \quad \text{by (f.1), (f.2) and (f.3).} \end{aligned} \tag{2.18}$$

Note that (f.3) may not be true for $\tilde{p}$. However, $\tilde{p}$ satisfies

$$\begin{aligned} &R(p(F,G), \tilde{p}) \\ &= \sum_{\mathbf{u},\mathbf{w}} \int \cdots \int_{B_{\mathbf{u},\mathbf{w}}} (p(F,G) - \tilde{p}(\mathbf{x},\mathbf{y}))^2 \prod_{i=1}^{m} g(x_i) \prod_{i=1}^{n} f(y_i)\, d\mathbf{x}\, d\mathbf{y} \\ &\geq \sum_{\mathbf{u},\mathbf{w}} (p(F,G) - \tilde{p}_c(\mathbf{u},\mathbf{w}))^2 g_1(\mathbf{u}) f_1(\mathbf{w}) \\ &= R(p(F_d, G_d), \tilde{p}_c), \end{aligned} \tag{2.19}$$

where

$$\tilde{p}_c(\mathbf{u},\mathbf{w}) = \frac{\int \cdots \int_{\mathbf{U}(\mathbf{x})=\mathbf{u}, \mathbf{W}(\mathbf{y})=\mathbf{w}} \tilde{p}(\mathbf{x},\mathbf{y}) \prod_{i=1}^{m} g(x_i) \prod_{i=1}^{n} f(y_i)\, d\mathbf{x}\, d\mathbf{y}}{g_1(\mathbf{u}) f_1(\mathbf{w})}.$$

Furthermore, the strict inequality holds unless $\tilde{p} = \tilde{p}_c$ w.p.1. [This follows from $E(X-p)^2 = \mathrm{Var}(X) + [E(X) - p]^2 \geq [E(X) - p]^2$ for any random variable $X$ and any constant $p$.]

If $\tilde{p} = \tilde{p}_c$ w.p.1 for any $F$ and $G$ satisfying (2.16), then $\hat{p}_0 \neq \tilde{p}_c$ since $\hat{p}_0 \neq \tilde{p}$. Otherwise, we have a strict inequality in (2.19) for at least one pair of $F$ and $G$ satisfying (2.16). Thus, it follows from (2.18), (2.19) and (C.1) that

$$R(p, \tilde{p}_c) \leq R(p, \hat{p}_0) \quad \text{for all possible } p = p(F_d, G_d) \tag{2.20}$$

and $R(p_1, \tilde{p}_c) < R(p_1, \hat{p}_0)$ for some $p_1 = p_1(F_d, G_d)$. It is obvious that $\hat{p}_0 \neq \tilde{p}_c$. Thus in any case, $\hat{p}_0 \neq \tilde{p}_c$ and (2.20) holds.

Since $F_d$ and $G_d$ are discrete, Theorem 1 applies to $F_d$ and $G_d$. It follows that (2.8) holds replacing $\tilde{p}$ by $\tilde{p}_c$. However, (2.8) contradicts (2.20). It follows that (C.1) implies (2.15). □

THEOREM 3. *Suppose that* $X_1, \dots, X_m$ *are i.i.d. from a continuous distribution function* $F$, $Y_1, \dots, Y_n$ *are i.i.d. from a continuous distribution function* $G$ *and* $X$*'s and* $Y$*'s are independent. Let the loss function be* $(\hat{p} - p)^2/\sigma^2(F, G)$. *Then* $\hat{p}_0$ *is minimax.*

PROOF. Note that $\hat{p}_0$ is the UMVU estimator of $p$ and it is easy to derive

$$R(p, \hat{p}_0) = E\left(\frac{(\hat{p}_0(\mathbf{X}, \mathbf{Y}) - p)^2}{\operatorname{Var}(\hat{p}_0)}\right) = 1.$$

Since $\hat{p}_0$ is admissible and is an equalizer rule under the loss function, it is minimax. □

**Acknowledgment.** The authors are grateful to the referee for a very useful suggestion which simplified and replaced our original version of the proof of Theorem 2.

## REFERENCES

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

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

GOVINDARAJULU, Z. (1967). Distribution-free confidence bounds for $P(X < Y)$. *Ann. Inst. Statist. Math.* **20** 229–238.

GOVINDARAJULU, Z. (1976). A note on distribution-free confidence bounds for $P(X < Y)$ when $X$ and $Y$ are dependent. *Ann. Inst. Statist. Math.* **28** 307–308.

MANN, H. B. and WHITNEY, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.* **18** 50–60.

ROYDEN, H. L. (1968). *Real Analysis*, 2nd ed. Macmillan, New York.

YU, Q. and KUO, L. (1992). An analogy between nonparametric problems of estimating a distribution function and their parametric versions. *Sankyā Ser. A*. To appear.

DEPARTMENT OF APPLIED MATHEMATICS
AND STATISTICS
SUNY AT STONY BROOK
STONY BROOK, NEW YORK 11794-3600

DEPARTMENT OF STATISTICS
UNIVERSITY OF KENTUCKY
LEXINGTON, KENTUCKY 40506-0027