ON TWO-STAGE NON-PARAMETRIC ESTIMATION[1]

By Elizabeth H. Yen[2]

University of Minnesota

1. Introduction and summary. In this paper, a two-sample, two-stage non-parametric estimation problem will be studied. The parameter $\theta = \theta(F, G)$ under consideration is estimable (i.e., there exists an unbiased estimator $\phi = \phi(X_1, \cdots, X_r; Y_1, \cdots, Y_s)$ of θ). ϕ is a function of independent observations from two populations with cumulative distribution functions $F(X)$ and $G(Y)$. The functions $F(X)$ and $G(Y)$ belong to a specified class D, such that a U-statistic based on ϕ is the unique minimum variance unbiased estimator of θ. The total number of observations on populations X and Y will be a fixed number N. The sampling procedure is carried out in two stages. First, take M observations from each of the populations; then allocate the remaining $N - 2M$ observations between the populations. The method of allocation utilizes the information from the first stage observations.

Two kinds of two-stage estimators, represented by U', and U'' will be introduced in this paper. Both U' and U'' are U-statistics with random sample sizes. U' is based essentially on the second stage observations only. U'' is defined on all N observations. Intuitively, the statistic U'' is more appealing. The first stage observations are used not only to determine the allocation of the second stage observations, but also to estimate the parameter θ. (see Section 3) One of the main results (Section 4) is that U' is unbiased and under certain conditions, the variance of U' approaches asymptotically to a particular variance V_0. (Here we shall consider the cases that both the variances of U' and U'', are finite.) U'' is in general biased. However, under the same conditions the value $E(U'' - \theta)^2$ approaches asymptotically to the same value V_0. This value V_0 is the smallest variance of any one-stage U-statistic estimator of θ, subject to the restriction that the total number of observations on X and on Y is N. V_0 is computed (see Section 2) when the best one-stage allocation of N observations to the two populations is made with the help of partial or even complete information about the distributions $F(X)$ and $G(Y)$. Such information about F and G is represented by the "nuisance parameters" $b_{10} = b_{10}(F, G)$, $b_{01} = b_{01}(F, G)$, etc., defined in Section 2. No prior knowledge of b_{10} and b_{01} is required to compute $\text{Var}(U')$, and $E(U'' - \theta)^2$.

In Section 5, the "optimal" choice of the first stage sample size M relative to the fixed total sample size N is discussed. The term "optimal" is in the sense

Received 6 June and 18 July 1963; revised 28 February 1964.

[1] Part of this paper is a thesis submitted to the Faculty of the Graduate School of the University of Minnesota in partial fulfillment of the requirements for the degree of Doctor of Philosophy. This work is supported in part by the Office of Naval Research, contract Nonr-710(31) NR042003.

[2] Now at Catholic University of America.

that the particular choices of M in relative order of magnitude of N, such that as N goes to infinity, the ratios $\text{Var}(U')/V_0$ and $E(U'' - \theta)^2/V_0$ approach unity as fast as possible in order of magnitude of N. Three cases with different conditions on ϕ are considered. It is found that the "optimal" choices depend on the specific conditions.

Section 6 contains some examples. To each $\theta(F, G)$, the corresponding estimators for b_{10} and b_{01} together with their behavior under different conditions on F and G, will be given. The examples include cases where the proposed procedures can be applied as well as cases where it cannot be applied.

Section 7 shows the asymptotic normality of U' and U''.

Section 8 indicates that the proposed procedures can be extended to k-sample case, for $k > 2$, with similar results.

The technique of two-stage estimation has been used in several papers. Stein [12] used it to determine confidence interval of a pre-assigned length for the mean of a normal population with unknown variance. Putter [8] used it to estimate the mean of a stratified normal population, Robbins [10] discussed such a technic for the design of experiments. Later, Ghurye and Robbins [4] used it to estimate the difference between the means of two normal populations (or some other specified populations). Richter [9] discussed the estimation of the common mean of two normal populations. During the preparation of the present paper, Alam [1] discussed the estimation of the common mean of $k \geqq 2$ normal populations. This paper generalizes these two-stage procedures in two ways. First, the underlying cumulative distributions F, G are members of a larger class of distributions. Secondly, the underlying parameters $\theta(F, G)$ are not restricted to population means or functions of means. Consequently, in such a general setup the question of "the best" estimator of any particular parameter $\theta(F, G)$ is not considered in this paper.

2. Some notations and the smallest variance of any one-stage U-statistic.

For convenience of presentation, some specific notations are adopted in this paper:

(1) K is used as a specified generic constant.

(2) e' is used as any given positive real number.

(3) Vectorial notations will be used such as: $(r = 1, 2, \cdots)$, $\bar{X}_r = (X_1, \cdots, X_r)$, $\bar{X}_{r,j} = (X_{j+1}, \cdots, X_r)$; $\bar{X}_{i_j} = (X_{i_1}, \cdots, X_{i_j})$, $\bar{X}_{i_k,j} = (X_{i_{j+1}}, \cdots, X_{i_k})$.

(4) $i_{t,s} : (C_1, C_2); j_{t,s} : (C_3, C_4)$ represents the statement "The summation is taken over all sets of integers, $C_1 \leqq i_{s+1} < \cdots < i_t \leqq C_2$; $C_3 \leqq j_{s+1} < \cdots < j_t \leqq C_4$."

(5) The notations ψ_{cd} and ξ_{cd} used by Hoeffding [5] and Rosenblatt [11] will be replaced by ϕ_{cd}' and b_{cd} respectively, for $c, d = 1, 2, \cdots, r$.

Consider two populations X and Y with cumulative distribution functions $F(X)$ and $G(Y)$ respectively, and a real valued estimable parameter $\theta = \theta(F, G)$. By Fraser ([3], Chapters 1 and 3), if F, G belong to a specific class D of cumulative distribution functions, (e.g. class of pairs of absolutely continuous distribution functions) then there exists a U-statistic which is the unique minimum variance unbiased estimator of θ. Let $\phi(\bar{X}_r;\bar{Y}_s)$ be the symmetric kernel of θ(see [5], [11]) such that,

$$(2.1)\quad \theta(F, G) = \int \cdots \int \phi(\bar{X}_r;\ \bar{Y}_s)\, dF(X_1)\cdots dF(X_r)dG(Y_1)\cdots dG(Y_s).$$

Since any function of rX's and sY's can be written as a function of $\max(r, s)$ of X's and Y's, we shall assume $r = s \neq 0$.

Denote $\phi'(\bar{X}_{i_r};\ \bar{Y}_{j_r}) = \phi(\bar{X}_{i_r};\ \bar{Y}_{j_r}) - \theta$ and $\phi'_{cd}(\bar{X}_c;\ \bar{Y}_d) = E\phi'(\bar{x}_c, \bar{X}_{r,c};\ \bar{y}_d, \bar{Y}_{r,d})$, the conditional expected value of ϕ' given $\bar{x}_c$ and $\bar{y}_d$, where $c, d = 0, 1, 2, \cdots r$. The "nuisance parameters" can be expressed as

$$(2.2)\qquad b_{cd} = E[\phi'_{cd}(\bar{X}_c;\ \bar{Y}_d)]^2 \quad \text{for} \quad c, d = 0, 1, 2, \cdots, r,$$

where $(i_1, \cdots, i_r)$ and $(k_1, \cdots, k_r)$ are any two sets of r distinct integers from $(1, 2, \cdots, m)$ and c is the number of integers common to the two sets; $(j_1, \cdots, j_r)$ and $(t_1, \cdots, t_r)$ are any two sets of r distinct integers from $(1, 2, \cdots, n)$ and d is the number of integers common to the two sets.

Now in a fixed sample of mX's and nY's the lower bound of the variance of the associated U-statistic of θ is by Rosenblatt ([11], Lemma 2.4 and Lemma 2.5)

$$(2.3)\qquad \operatorname{Var}(U_{m,n}) \geqq (r^2/m)b_{10} + (r^2/n)b_{01} + (r^4/mn)\eta_{11}, \quad \text{where} \quad \eta_{11} = b_{11} - b_{10} - b_{01} \geqq 0.$$

Denote $\bar{U}$ as the U-statistic of θ with $m + n = N$, and N is fixed. When m, n satisfy the condition $0 < a_1 \leqq m/n \leqq a_2 < \infty$ as $m, n \to \infty$, then by Rosenblatt ([11], Lemma 2.6),

$$(2.4)\qquad \operatorname{Var}(\bar{U}) \geqq (r^2/m)b_{10} + (r^2/n)b_{01} = V', \text{ say.}$$

V' can be minimized, if b_{10}, b_{01} are known and both bounded away from zero and infinity, by selecting the best values of m, n, subject to $m + n = N$, such that

$$(2.5)\quad m_0 = N(b_{10})^{\frac{1}{2}}/[(b_{10})^{\frac{1}{2}} + (b_{01})^{\frac{1}{2}}] = NQ, \quad \text{say}, \quad n_0 = N - m_0 = N(1 - Q).$$

The minimum value of V', denoted by V_0, is

$$(2.6)\qquad V_0 = N^{-1}[r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]^2 = V'(m_0, n_0).$$

Clearly, V_0 is the smallest variance of any estimator of θ based on U-statistics subject to the restriction that $m + n = N$. It will be used as a basis for comparison in the remaining sections.

3. The two-stage procedures and the estimators.

DEFINITION of U'. Let the total observations be fixed at N where $N > 6r$. At the first stage, M observations are made on each of the two populations, where $M > 2r$ and $2M < N - 2r$. From these $2M$ observed values, we estimate the parameters b_{10}, b_{01}. Clearly, b_{10} and b_{01} are estimable functions [5]. The associated U-statistics, called T_{10} and T_{01}, can be expressed as follows:

$$(3.1)\qquad T_{10} = \binom{M}{2r}^{-1}\binom{M}{2r}^{-1}\sum[\phi_{10}'(X_1)]^2 = \binom{M}{2r}^{-2}\sum h(\bar{X}_{i_{2r}}; \bar{Y}_{j_{2r}})$$

$$(3.2)\qquad T_{01} = \binom{M}{2r}^{-1}\binom{M}{2r}^{-1}\sum[\phi_{01}'(Y_1)]^2 = \binom{M}{2r}^{-2}\sum g(\bar{X}_{i_{2r}}; \bar{Y}_{j_{2r}}),$$

where $i_{2r,0} : (1, M); j_{2r,0} : (1, M)$.

In analogy with (2.4), we define

$$(3.3)\quad Z = (T_{10})^{\frac{1}{2}}[(T_{01})^{\frac{1}{2}} + (T_{10})^{\frac{1}{2}}]^{-1} \text{ for } T_{10}, T_{01} \text{ both positive, and 0 otherwise.}$$

The second stage is constructed by taking m' and n' more observations on X and on Y respectively. With $m' + n' = N - 2M = N'$, we define,

$$(3.4)\qquad \begin{aligned} m' &= [N'Z] && \text{when } r/N' \leqq Z \leqq (N' - r)/N' \\ m' &= r && \text{when } Z < r/N' \\ m' &= N' - r && \text{when } Z > (N' - r)/N' \end{aligned}$$

and $n' = N' - m'$, where $[a]$ is the largest integer contained in a. (Note: In general r is the minimum number of observations required from X and Y such that ϕ is an unbiased estimator of θ); and

$$(3.5)\qquad U' = \binom{m'}{r}^{-1}\binom{n'}{r}^{-1}\sum\phi(\bar{X}_{i_r}; \bar{Y}_{j_r}), \quad \text{where} \quad i_{r,0}:(M + 1, M + m'); \quad j_{r,0}:(M + 1, M + n').$$

Hence, U' is explicitly a function of the second stage observations only. It depends implicitly also on the first stage, through m' and n'.

Definition of U''. Its first stage procedure is the same as U'. The second-stage $N - 2M$ observations are combined, after allocation, with the first-stage $2M$ observations to form a U-statistic of all N observations.

The second stage is constructed by taking m'' and n'' more observations on X and on Y respectively. With $m'' + n'' = N'$, we define,

$$(3.6)\qquad \begin{aligned} m'' &= [NZ] - M && \text{when } (M + 1)/N \leqq Z \leqq (N - M)/N \\ m'' &= 0 && \text{when } Z < (M + 1)/N \\ m'' &= N' && \text{when } Z > (N - M)/N \\ n'' &= N' - m''; \text{ and} \end{aligned}$$

$$(3.7)\qquad U'' = \binom{M + m''}{r}^{-1}\binom{M + n''}{r}^{-1}\sum\phi(\bar{X}_{i_r}; \bar{Y}_{j_r}),$$

$$\text{where} \quad i_{r,0}:(1, M + m''); \qquad j_{r,0}:(1, M + n'').$$

Here U'' is an explicit function of all N observations.

4. Asymptotic efficiency of the estimators. Before we present the main theorems, some lemmas will be given first.

LEMMA 4.1. *Let $\theta(F, G) = \theta$ be an estimable parameter with symmetric kernel $S = S(\bar{X}_r ; \bar{Y}_s)$, where 2ith moment of S is finite. Let $W = W_{MM}$ be the associated U-statistic. Define*: (a) $W' = W - \theta$, (b) $S' = S - \theta$ *and* (c) $S'(\bar{X}_{rt+r,rt} ; \bar{Y}_{rt+r,rt}) = S'_t$, *then for any positive integer i, $E(W)^{2i} = O(M^{-i})$.*

PROOF. For convenience, again let $r = s$. Also define:

$$W'' = (1/k)\sum_{t=0}^{k-1} S'_t, \qquad \text{where } k = M/r.$$

W'' is an average of k independent and identically distributed random variables with mean zero and finite variances. From the work of Tchouproff [13], $E(W'')^{2i} = O(M^{-i})$; and by Hoeffding [6], $W' = (M!)^{-2}\sum W''(\bar{X}_{h_M} ; \bar{Y}_{j_M})$, where the summation is taken over all permutations of $(h_1, \cdots, h_M)$, $(j_1, \cdots, j_M)$ of $(1, 2, \cdots, M)$. Next, since $(W')^{2i} = [(M!)^{-2}\sum (W'')^{2i}] \leqq (W'')^{2i}$, one has $E(W')^{2i} \leqq E(W'')^{2i} = O(M^{-i})$, and the lemma is proved.

LEMMA 4.2. *Let Z and Q be defined as in* (3.3) *and* (2.4) *respectively. Assume for $0 < p < (i - 1)/2i$, $i = 2, 3, \cdots$, that ϕ has 4ith finite moments, and b_{10}, $b_{01} \geqq a > 0$ for any positive constant a. Then* $\Pr[|Z - Q| > M^{-p}] = O(M^{-i+2ip})$.

PROOF

$$\begin{aligned}
\Pr[|Z - Q| > M^{-p}] &= \Pr[|Z - Q| > M^{-p}; T_{10}, T_{01} > 0] \\
&\quad + \Pr[|Z - Q| > M^{-p}; T_{10}, T_{01} \text{ not both positive}] \\
&\leqq \Pr\{(T_{10})^{\frac{1}{2}}[(T_{01})^{\frac{1}{2}} + (T_{10})^{\frac{1}{2}}]^{-1} > Q + M^{-p}\} \\
&\quad + \Pr\{(T_{01})^{\frac{1}{2}}[(T_{01})^{\frac{1}{2}} + (T_{10})^{\frac{1}{2}}]^{-1} < Q - M^{-p}\} \\
&\quad + \Pr[T_{10} \leqq 0] + \Pr[T_{01} \leqq 0].
\end{aligned} \tag{4.1}$$

The first term on the right side of (4.1) can be expressed in the form of $\Pr(U - EU > -EU)$.

Using the generalized Chebyshev's inequality of the form, $\Pr[|X| > a] \leqq a^{-2i}E(X)^{2i}$, and applying Lemma 4.1, one finds that it is a term of order of $O(M^{-i+2ip})$, for large M. Similarly, the second term on the right side of (4.1) is also of order of $O(M^{-i+2ip})$.

Using Lemma 4.1, the last two terms of (4.1) is of order of $O(M^{-i})$. Therefore, $\Pr(|Z - Q| > M^{-p}) = O(M^{-i+2ip}) + O(M^{-i}) = O(M^{-i+2ip})$ is proved.

LEMMA 4.3. *Let X be any random variable with cumulative distribution function $F(X)$ and* $\Pr(X < 0) = 0$. *Let K be any real number. Then $E(X \mid X \leqq K) \leqq E(X)$.*

The proof is elementary and will be omitted.

LEMMA 4.4. *Define*

$$T'_{10} = \binom{M}{2r}^{-1}\binom{M}{2r}^{-1}\sum h(\bar{X}_{i_r} ; \bar{Y}_{j_r}),$$

where $i_{2r,0} : (r + 1, M)$; $j_{2r,0} : (r + 1, M)$. *Let* T'_{01} *be defined in an analogous way. Define* Z' *analogously as* Z *(see (3.3)) with* T_{10}, T_{01} *replaced by* T'_{10}, T'_{01}, *respectively. Let* $b_{10}, b_{01} \geqq \alpha > 0$ *and* ϕ *has finite eighth moment. Then* $E|Z - Z'| = O(M^{-1})$.

Proof. The procedure by Cramér ([2] pages 353, 356) for treating functions of moments will be followed. Define,

$$T''_{10} = VT'_{10}(T''_{01} = VT'_{01}), \quad \text{where} \quad V = \binom{M}{2r}\binom{M}{2r}\Big/\binom{M-r}{2r}\binom{M-r}{2r}.$$

Then, $T''_{10}(T''_{01})$ is a U-statistic with $M - r$ observations on both X and Y. For T_{10}, $T_{01}(T''_{10}, T''_{01})$ within small intervals $b_{10} \pm e'$, $b_{01} \pm e'$ respectively, one may represent $Z(Z')$ by a Taylor's series around the point (b_{10}, b_{01}). This assertion follows from the facts that Q is bounded away from zero and unity and that for small e' all order of derivations of $Z(Z')$ are bounded. Let R be the condition that $|T_{10} - b_{10}| < e'$, $|T_{01} - b_{01}| < e'$, $|T''_{10} - b_{10}| < e'$ and $|T''_{01} - b_{01}| < e'$ are all true, and let R' = complement of R. The expectation is broken down according to whether R or R' occurred. Since $Z(Z')$ are bounded, so is their difference; and by the generalized Chebyshev's inequality, for $e' = M^{-p}$, $0 < p < (i - 1)/2i$. $i = 2, 3, \cdots,$. The expectation when R' occurred is of order of $O(M^{-i+2ip}) = o(M^{-1})$. While the expectation when R occurred, after applying Lemma 4.3, is less or equal to

$$\begin{aligned} KE\{&|T_{10} - T''_{10}| + |T_{01} - T''_{01}| + (T_{10} - b_{10})^2 \\ &+ |(T_{10} - b_{10})(T_{01} - b_{01})| + (T_{01} - b_{01})^2 + (T''_{10} - b_{10})^2 \\ &+ |(T''_{10} - b_{10})(T''_{01} - b_{01})| + (T''_{01} - b_{01})^2\}. \end{aligned} \tag{4.2}$$

With or without first applying the Schwarz inequality, we use the definition of V, and the result of Lemma 4.1. Then each term in (4.2) is found to be of order of $O(M^{-1})$. Therefore $E|Z - Z'| = O(M^{-1})$, and the lemma is proved.

Theorem 4.1. $E(U') = \theta$. *Also, if* $M \to \infty$, *as* $N \to \infty$, *and*

(i) $\text{Limit}_{N\to\infty} N/M^{\beta}$ *exists and is finite for some* β, *such that* $1 < \beta < 2$,

(ii) *the eighth moment of* ϕ *is finite, and*

(iii) $b_{10}, b_{01} \geqq \alpha > 0$ *for any positive constant* α; *then*

$$\text{Limit}_{N\to\infty} \text{Var}(U')/V_0 = \text{Limit}_{N\to\infty} E_{m'}[\text{Var}(U'_{m'})]/V_0 = 1.$$

Remark. In most non-parametric problems ϕ is bounded, hence all moments exist. Therefore, the restriction (ii) is not severe. (Note that (ii) insures b_{10}, b_{01} both finite). Also, Var $(U'_{m'})$ denotes the conditional variance of U' given m' and n', and Var (U') denotes the expected value of Var $(U'_{m'})$, where the expectation is over m' and n'.

Proof. Notice that m', n' are functions of $X_1, \cdots, X_M$; $Y_1, \cdots, Y_M$ only. While all the arguments of $\phi(\bar{X}_{i_r}; \bar{Y}_{j_r})$ in the definition of U' are functions of $X_{M+1}, \cdots, X_{M+m'}$; $Y_{m+1}, \cdots, Y_{M+n'}$. Thus the arguments of U' are independent of $X_1, \cdots, X_M; Y_1, \cdots, Y_M$. Therefore

$$E(U') = E_{m'}E(U'_{m'}) = E_{m'}\binom{m}{r}^{-1}\binom{n'}{r}^{-1}\sum E\phi(\bar{X}_{i_r}; \bar{Y}_{j_r}) = \theta.$$

Now, let C be the condition that $|Z - Q| < M^{-p}$; C' = complement of C, for $0 < p < \frac{1}{4}$.

$$\begin{aligned} N \operatorname{Var}(U') &= NE_{m'}[\operatorname{Var}(U'_{m'})] \\ &= N \Pr(C)E_{m'\varepsilon C}\operatorname{Var}(U'_{m'}) + N\Pr(C')E_{m'\varepsilon C'}\operatorname{Var}(U'_{m'}). \end{aligned} \tag{4.3}$$

Noticing that $E_{m'\varepsilon C'} \operatorname{Var}(U'_{m'}) \leqq \operatorname{Var}(U'_{rr}) = b_{rr} = \operatorname{Var}(\phi)$ is bounded, by assumption (ii); and by using the result of Lemma 4.2 for $i = 2$; we find the second term on the right of (4.3) is of order of $O(M^{-2+4p+\beta})$.

It is easy to show that there exists a number A which is independent of m', n', such that $\operatorname{Var}(U'_{m'}) \leqq (r^2b_{10})/m' + (r^2b_{01})/n' + A/\min(m'^2, n'^2)$. Also, under condition C and for sufficiently large N and M, m', n' can be written as $m' \geqq N'(Q - M^{-p})$, $n' \geqq N'(1 - Q - M^{-p})$. Thus,

$$\begin{aligned} &N \operatorname{Var}(U') \\ &\quad = [r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]^2[1 + O(N^{(1-\beta)/\beta}) + O(N^{-p/\beta}) + O(N^{(-2+4p+\beta)/\beta})], \end{aligned}$$

after some computation and simplification, and by putting $M = K(N^{1/\beta})$, where K is an unknown non-zero constant. By assumption (i), $1 < \beta < 2$, there exists p, so that $0 < p < \frac{1}{4}$, and $(-2 + 4p + \beta) < 0$. Hence

$$\begin{aligned} \operatorname{Limit}_{N\to\infty} \operatorname{Var}(U')/V_0 &= \operatorname{Limit}_{N\to\infty} NE_{m'}\operatorname{Var}(U'_{m'})/NV_0 \\ &= \operatorname{Limit}_{N\to\infty}[1 + O(N^{(1-\beta)/\beta}) + O(N^{-p/\beta}) + O(N^{(-2+4p+\beta)/\beta})] = 1, \end{aligned}$$

which completes the proof.

THEOREM 4.2. *If the conditions in Theorem* 2.1 *are satisfied, then,* $E(U'') = \theta + O(M^{-1+2p} + M^{\frac{1}{2}-\beta})$, *and* $\operatorname{Limit}_{N\to\infty} E(U'' - \theta)^2/V_0 = 1$.

PROOF. Let

$$p(m, n) = \binom{M+m}{r}\binom{M+n}{r},$$

$$q(c, d, m, n) = \binom{M}{c}\binom{m}{r-c}\binom{M}{d}\binom{n}{r-d}.$$

Using Z' in Lemma 4.4, we denote

$$m''' = NZ' - M, \qquad n''' = NZ' - M. \tag{4.4}$$

For $c, d = 0, 1, \cdots, r$, let $\sum_{cd}$ be the summation over all $(i_1, \cdots, i_r)$, $(j_1, \cdots, j_r)$, such that

$$1 \leqq i_1 < \cdots < i_c \leqq M < i_{c+1} < \cdots < i_r \leqq M + m'',$$

$$1 \leqq j_1 < \cdots < j_d \leqq M < j_{d+1} < \cdots < j_r \leqq M + n''.$$

Then

$$(4.5)\qquad |E(U'') - \theta| = \left| E_{m''} \left\{ p(m'', n'')^{-1} \sum_{c=0}^{r} \sum_{d=0}^{r} \sum_{cd} [E\phi'(\bar{X}_{i_r} ; \bar{Y}_{j_r}) \mid m''] \right\} \right| .$$

Notice that U'' is in general biased. For some of its kernels are functions of observations from both stages. The conditional expectation of such kernels given first stage observations fixed, does not (in general), equal to the unconditional expectation. Since the number of terms under $\sum_{cd}$ is $q(c, d, m'', n'')$, (4.5) is less or equal to the following expression:

$$(4.6)\qquad \sum_{c=0}^{r} \sum_{d=0}^{r} \left| E_{m''} \left\{ \left[\frac{q(c, d, m'', n'')}{p(m'', n'')} - \frac{q(c, d, m''', n''')}{p(m''', n''')} \right] \phi'_{cd}(\bar{X}_c ; \bar{Y}_d) \right\} + E_{m''} \frac{q(c, d, m''', n''')}{p(m''', n''')} \phi'_{cd}(\bar{X}_c , \bar{Y}_d) \right| .$$

By (4.4) m''' and n''' are independent of $X_1, \cdots, X_r ; Y_1, \cdots Y_r$, and $q(c, d, m''', n''')/p(m''', n''')$ is bounded. The last term in (4.6) is zero for all $c, d = 0, 1, \cdots, r$. Also,

$$E_{m''} \left\{ \left[\frac{\sum (c, d, m'', n'')}{p(m'', n'')} - \frac{\sum (c, d, m''', n''')}{p(m''', n''')} \right] \phi'_{00} \right\} = 0.$$

Let conditions C, C' as defined in Theorem 4.1. Apply the Schwarz inequality on each non-zero term, then break down the expectation according to whether C or C' occurred. Using Lemma 4.2 for $i = 2$, and noticing that the coefficient before each kernel is bounded, the expectation when C' occurred is of order of $O(M^{-2+4p})$. While the expectation when C occurred can be written as:

$$(4.7)\qquad K(MN^{-1})^{2c+2d} E_{m''\varepsilon C}[(1 - Z')^d Z'^c - (1 - Z)^d Z^c]^2.$$

Since the higher order terms are for $c + d$ small, we shall retain only the terms $(c, d) = (0, 1)$ and $(c, d) = (1, 0)$. Now, for $0 \leqq Z, Z' \leqq 1$, $(Z' - Z)^2 \leqq |Z - Z'|$, and by Lemma 4.3 and Lemma 4.4, $E_{m''\varepsilon C}(|Z - Z'|) = O(M^{-1})$. Therefore, it is found,

$$|E(U'') - \theta| = O(M^{\frac{1}{2}-\beta} + M^{-1+2p}).$$

In Section 5, we shall find that the "optimal" choices of β and p lead to $|E(U'') - \theta| = O(N^{-4/7})$. But when ϕ is bounded, from (4.5) we can have $|E(U'') - \theta| = O(N^{-1}) + O(M^{-2+4p}) = O(N^{-1})$, for $-2 + 4p + \beta < 0$.

$E(U'' - \theta)^2$ is computed by the truncation method. Let $\bar{\phi}(\bar{X}_r ; \bar{Y}_r) = \phi(\bar{X}_r ; \bar{Y}_r) - \theta = \phi'(\bar{X}_r ; \bar{Y}_r)$ if $|\phi'(\bar{X}_r ; \bar{Y}_r)| < M^{\frac{1}{2}}$ (for brevity, call this situation T); and 0 otherwise (call it T'). Add and subtract $\bar{\phi}$ from the kernel of U'' and expand the square, we have,

$$(4.8)\qquad \begin{aligned} NE(U'' - \theta)^2 = {} & NE[p(m'', n'')^{-1} \sum \bar{\phi}(\bar{X}_{i_r} ; \bar{Y}_{j_r})]^2 + 2NE\{p(m'', n'')^{-2} \\ & \cdot \sum \bar{\phi}(\bar{X}_{i_r} ; \bar{Y}_{j_r}) \sum [\phi'(\bar{X}_{i_r} ; \bar{Y}_{j_r}) - \bar{\phi}(\bar{X}_{i_r} ; \bar{Y}_{j_r})]\} \\ & + NE\{p(m'', n'')^{-1} \sum [\phi'(\bar{X}_{i_r} ; \bar{Y}_{j_r}) - \bar{\phi}(\bar{X}_{i_r} ; \bar{Y}_{j_r})]\}^2 \end{aligned}$$

where $i_{r,0} : (1, M + m'')$; $j_{r,0} : (1, M + n'')$. Now the second term on the right of (4.8) is of order $O(NM^{-2})$. It is seen by first applying the Schwarz inequality then writing $\bar{\phi}(\bar{X}_{i_r} ; \bar{Y}_{j_r})$ in terms of $\phi'(\bar{X}_{i_r} ; \bar{Y}_{j_r})$, and finally using the generalized Chebyshev's inequality, for ϕ have finite eighth moment. Similarly, the last term of (4.8) is of order of $O(NM^{-4})$. Now the first term of (4.8) can be written as

$$\begin{aligned} &NE_{m''}\{p(m'', n'')^{-2} \sum_{c=0}^{r} \sum_{d=0}^{r} \binom{M + m''}{r} \binom{r}{c} \\ (4.9) \qquad &\cdot \binom{M + m'' - r}{r - c} \binom{M + n''}{r} \binom{r}{d} \binom{M + n'' - r}{r - d} \\ &\cdot E\{\bar{\phi}(\bar{X}_{i_r} ; \bar{Y}_{j_r}) \cdot \bar{\phi}(\bar{X}_{k_r} ; \bar{Y}_{t_r}) \mid m''\}, \end{aligned}$$

where $(i_1, \cdots, i_r)$ and $(k_1, \cdots, k_r)$ $((j_1, \cdots, j_r)$ and $(t_1, \cdots, t_r))$ have exactly $c(d)$ common integers. Any $X(Y)$ having subscripts less than or equal to M is taken as constant under the second expectation sign. For $(c, d) = (0, 0)$ in (4.9), the term is less or equal to

$$(4.10) \quad NE_{m''}E\{[\phi(\bar{X}_{i_r} ; \bar{Y}_{j_r})\bar{\phi}(\bar{X}_{k_r} ; \bar{Y}_{t_r}) \mid m''\} \leqq N\{|E\phi(\bar{X}_{i_r} ; \bar{Y}_{j_r})|\}^2.$$

Since $E[\phi'(\bar{X}_{i_r} ; \bar{Y}_{j_r})] = 0$, we find the absolute value of the expectation under the situation T is equal to the absolute value of the expectation under its complement. Using the generalized Chebyshev's inequality, $|E\bar{\phi}(\bar{X}_{i_r} ; \bar{Y}_{j_r})| = O(M^{-7/2})$. Therefore, (4.10) is a term of order of $O(NM^{-7})$. Consider the term $(c, d) = (0, 1)$ in (4.9), the expectation is broken down according to whether C or C' occurred. (C, C' defined in Theorem 4.1). The part when C' occurred is of order of $O(M^{-1+4p})$, and the part C occurred is computed first by applying Lemma 4.3 then adding and subtracting a term $\phi'(\bar{X}_{i_r} ; \bar{Y}_{j_r})\phi'(\bar{X}_{k_r} ; \bar{Y}_{t_r})$. We obtain,

$$\begin{aligned} &E[\bar{\phi}(\bar{X}_{i_r} ; \bar{Y}_{j_r})\bar{\phi}(\bar{X}_{k_r} ; \bar{Y}_{t_r}) \mid C] \Pr(C) \\ &\qquad \leqq b_{01} + E|\bar{\phi}(\bar{X}_{i_r} ; \bar{Y}_{j_r})[\bar{\phi}(\bar{X}_{k_r} ; \bar{Y}_{t_r}) - \phi'(\bar{X}_{k_r} ; \bar{Y}_{t_r})]| \\ (4.11) \qquad &\qquad + E|[\bar{\phi}(\bar{X}_{i_r} ; \bar{Y}_{j_r}) - \phi'(\bar{X}_{i_r} ; \bar{Y}_{j_r})]\phi'(\bar{X}_{k_r} ; \bar{Y}_{t_r})| \\ &\qquad \leqq b_{01} + M^{\frac{1}{2}}E[|\phi'(\bar{X}_{k_r} ; \bar{Y}_{t_r})| \mid T'] \Pr(T') \\ &\qquad\qquad + E[\phi'(\bar{X}_{k_r} ; \bar{Y}_{t_r})^2 \mid T'] \Pr(T') = b_{01} + O(M^{-3}), \end{aligned}$$

by the generalized Chebyshev's inequality. Combine with the coefficient, the term $(c, d) = (0, 1)$ in (4.9) is

$$\begin{aligned} &r^2b_{01}/(1 - Q) + O(M^{-p}) + O(M^{-1+4p}) + r^2b_{r1}N/M \cdot O(M^{-1+4p}) \\ (4.12) \qquad &\qquad + \text{lower order terms.} \\ &\qquad = r^2(b_{01})^{\frac{1}{2}}[(b_{01})^{\frac{1}{2}} + (b_{10})^{\frac{1}{2}}] + O(M^{-p}) + O(M^{-2+4p+\beta}). \end{aligned}$$

Similarly, it can be shown that the term $(c, d) = (1, 0)$ in (4.9) is

$$r^2(b_{10})^{\frac{1}{2}}[(b_{01})^{\frac{1}{2}} + (b_{10})^{\frac{1}{2}}] + O(M^{-p}) + O(M^{-2+4p+\beta}). \tag{4.13}$$

Combining results (4.10), (4.11), (4.12), and (4.13) with the main expressions (4.9), and (4.8) one obtains,

$$\begin{aligned} &NE(U'' - \theta)^2 \\ &\quad = O(NM^{-2}) + O(M^{-4}N) + O(NM^{-7}) + r^2(b_{01})^{\frac{1}{2}}[(b_{01})^{\frac{1}{2}} + (b_{10})^{\frac{1}{2}}] \\ &\qquad + r^2(b_{10})^{\frac{1}{2}}[(b_{01})^{\frac{1}{2}} + (b_{10})^{\frac{1}{2}}] + O(M^{-p}) + O(M^{-2+4p+\beta}). \end{aligned} \tag{4.14}$$

Therefore,

$$\begin{aligned} \text{Limit}_{N\to\infty} \frac{E(U'' - \theta)^2}{V_0} &= \text{Limit}_{N\to\infty} \frac{NE(U'' - \theta)^2}{NV_0} \\ &= \text{Limit}_{N\to\infty} \frac{[r(b_{01})^{\frac{1}{2}} + r(b_{10})^{\frac{1}{2}}]^2}{[r(b_{01})^{\frac{1}{2}} + r(b_{10})^{\frac{1}{2}}]^2}[1 + O(M^{-p}) + O(M^{-2+4p+\beta})] = 1, \end{aligned}$$

which completes the proof.

In addition to U', one may like to estimate θ separately at both stages, then combine these two estimates by weights. The determination of the proper weights under some criterion are quite complicated in such a general set-up, hence such procedure is not included in this paper. However, another one-stage statistic will be discussed.

Assume that N observations are to be made, and that the b_{cd}'s are unknown, (except that b_{10}, b_{01} are positive and finite), then proceed as if $b_{10} = b_{01}$. The variance of such a one-stage U-statistic is minimized with respect to m, subject to $m + n = N$, when $m = N/2$, $n = N/2$. Let the statistic be denoted by U^*, then its variance is given by

$$\text{Var}(U^*) = N^{-1}2r^2(b_{10} + b_{01}) + O(N^{-2}).$$

Hence,

$$\begin{aligned} \text{Limit}_{N\to\infty} \text{Var}(U^*)/V_0 &= \text{Limit}_{N\to\infty} N\,\text{Var}(U^*)/NV_0 \\ &= 2(1 + \rho^2)/(1 + \rho)^2, \quad \text{where} \quad \rho = (b_{01}/b_{10})^{\frac{1}{2}}. \end{aligned} \tag{4.15}$$

When ρ nears 0 or ∞, (4.15) nears its maximum 2. Thus, an appreciable decrease in variance can be obtained by using a two-stage procedure.

Remark. For $s \neq r$, but we write ϕ as a function of max (r, s) X's and Y's, Var (U^*) is minimized when $m = Nr/(s + r)$ and $n = Ns/(s + r)$. Consequently, the variance ratio approaches $1 + s/r(1 + r/s)$ as ρ approaches zero (infinity). Thus the variance ratio may have a maximum greater than 2.

5. "Optimal" choice of the value M relative to N.

For the ratio $\text{Var}(U')/V_0$.

(1) *The first eight moments of the kernel ϕ exist.* From the last step of the proof

of Theorem 4.1, one has

$$\mathrm{Var}\,(U')/V_0 = 1 + O(M^{-(\beta-1)}) + O(M^{-p}) + O(M^{-2+4p+\beta}).$$

Heuristically, the best β and p is the solution of the pair of equations, which are obtained by equating the exponentials in the remainder terms of the above equation, and get $\beta = \frac{7}{6}$, $p = \frac{1}{6}$, thus $M = K(N^{6/7})$.

Actually, this pair of values is the "optimal" solution, because any other choices will make one of the three terms have a larger order of magnitude than $O(M^{-\frac{1}{6}})$ (or equivalently, $O(N^{-\frac{1}{7}})$). Therefore, $\mathrm{Var}\,(U')/V_0 = 1 + O(N^{-\frac{1}{7}})$.

(2) *All moments of the kernel ϕ exist.* By Lemma 4.2 and Theorem 4.1, for $i = 2, 3, \cdots, 0 < p < (i-1)/2i$, one has

$$\mathrm{Var}\,(U')/V_0 = 1 + O(M^{-(\beta-1)}) + O(M^{-p}) + O(M^{-i+2ip+\beta}).$$

Similar to (1) above, it is found that $\beta = (3i+1)/2(1+i)$ and $p = (i-1)/2(1+i)$ is the set of solutions. When i approaches infinity, β approaches $\frac{3}{2}$ and p approaches $\frac{1}{2}$. Therefore, $M = K(N^{2(1+i)/(3i+1)})$, where $2(i+1)/(3i+1)$ has $\frac{2}{3}$ as a lower bound. This bound, however, is not obtained. Thus,

$$\mathrm{Var}\,(U')/V_0 = 1 + O(N^{-1+2(1+i)/(3i+1)})$$

for any i.

(3) *The kernel ϕ is bounded.* Referring to the proof in Lemma 4.2, and applying Hoeffding's inequality ([6], Theorem 2) and Lemma 5.1, we have $\Pr\,(|Z - Q| > e') = O(e^{-e'^2M})$. Hence, from the proof of Theorem 4.1, one obtains,

$$\mathrm{Var}\,(U')/V_0 = 1 + O(MN^{-1}) + O(e') + O(Ne^{-e'^2M}).$$

After taking logarithm and some computation, it amounts to solve the equations

$$M = N^{\frac{2}{3}}(\log N/e')^{\frac{1}{3}}, \tag{5.1}$$

and

$$\log\,(e')^{-1} = \log N^{\frac{1}{3}} - \tfrac{1}{3}\log\,\{\log N + \log\,(e')^{-1}\}. \tag{5.2}$$

Since $(e')^{-1} < N$, $\log\,(e')^{-1} < \log N$, (5.2) leads to

$$\log\,(e')^{-1} \geqq \log N^{\frac{1}{3}} - \tfrac{1}{3}\log\,[\log N^2]. \tag{5.3}$$

On the other hand, $\log\,(e')^{-1}$ is positive, $\log N + \log\,(e')^{-1} \geqq \log N$, thus (5.2) leads to

$$\log\,(e')^{-1} \leqq \log N^{\frac{1}{3}} - \tfrac{1}{3}\log\,(\log N). \tag{5.4}$$

Comparing with (5.1), one has

$$N^{\frac{2}{3}}[\log N^{\frac{4}{3}} - \tfrac{1}{3}\log\,(\log N)]^{\frac{1}{3}} > M > N^{\frac{2}{3}}[\log N^{\frac{4}{3}} - \tfrac{1}{3}\log\,(\log N^2)]^{\frac{1}{3}}. \tag{5.5}$$

Therefore, $\mathrm{Var}\,(U')/V_0 = 1 + O(N^{-\frac{1}{3}}I)$, where I is some value in the interval,

$$([\log N^{\frac{4}{3}} - \tfrac{1}{3}\log\,(\log N^2)]^{\frac{1}{3}}, [\log N^{\frac{4}{3}} - \tfrac{1}{3}\log\,(\log N)]^{\frac{1}{3}}).$$

For the ratio $E(U'' - \theta)^2/V_0$. In all three cases, there are no "optimal" choices of M relative to N in order of magnitude. Because, for any fixed p, one shall choose β as small as possible subject to $\beta > 1$. However, there exist, for the Case (1), at least one set of values, $\beta = \frac{7}{6}$, $\rho = \frac{1}{6}$, such that the remainder term of the ratio $E(U'' - \theta)^2/V_0$ is of order of $O(N^{-\frac{1}{7}})$. If one chooses $\beta = 1 + \Delta$, for some $\Delta > 0$, the ratio is expressed as $1 + O(N^{-(1-\Delta)/[5(1+\Delta)]})$, which may converge to unity faster than $1 + O(N^{-\frac{1}{7}})$. The limit of the ratio is $1 + O(N^{-\frac{1}{5}})$ which is, however, never reached. In Cases (2) and (3), the situations are analogous.

6. Some examples.

6.1. *Consider the Wilcoxon Statistic.* The class D contains all pairs of cumulative distribution functions F, G which are continuous. $\theta = \Pr(X > Y)$ with the kernel: $f(X_i, Y_j) = 1$, if $X_i > Y_j$, and 0 otherwise.

Here, $r = s = 1$, and b_{10}, b_{01} can be shown as

$$b_{10} = \Pr(X_1 > Y_1, Y_2) - [\Pr(X_1 > Y_1)]^2 = 2\Pr(X_1 > Y_1 > Y_2 > X_2)$$

$$b_{01} = \Pr(X_1, X_2 > Y_1) - [\Pr(X_1 > Y_1)]^2 = 2\Pr(Y_1 > X_1 > X_2 > Y_2).$$

The estimators of b_{10}, b_{01} are respectively,

$$T_{10} = \binom{M}{2}^{-1}\binom{M}{2}^{-1} \sum_{1 \leqq i_1 < i_2 < M} \sum_{1 \leqq j_1 < j_2 \leqq M} 2h(\bar{X}_{i_2}; \bar{Y}_{j_2})$$

$$T_{01} = \binom{M}{2}^{-1}\binom{M}{2}^{-1} \sum_{1 \leqq i_1 < i_2 \leqq M} \sum_{i \leqq j_1 < j_2 \leqq M} 2g(\bar{X}_{i_2}; \bar{Y}_{j_2})$$

where $h(\bar{X}_{i_2}; \bar{Y}_{j_2}) = \frac{1}{4}$ if the two Y's are ranked between the two X's, and 0 otherwise; $g(\bar{X}_{i_2}; \bar{Y}_{j_2}) = \frac{1}{4}$ if the two X's are ranked between the two Y's, and 0 otherwise. Here ϕ, g, h are all bounded. When F, G are both strictly monotone, both b_{10} and b_{01} are positive. In that case, the two-stage procedure is applicable, and one shall choose M in order of magnitude, between

$$O\{N^{\frac{2}{3}}[\log N^{\frac{4}{3}} - \tfrac{1}{3}\log(\log N^2)]^{\frac{1}{3}}\} \quad \text{and} \quad O\{N^{\frac{2}{3}}[\log N^{\frac{4}{3}} - \tfrac{1}{3}\log(\log N)]^{\frac{1}{3}}\}.$$

6.2. *Assume* $\theta = E(X) - E(Y)$, *where* X *and* Y *have cumulative distribution functions* F *and* G *respectively*. Here $\phi = X_i - Y_i$ and again $r = s = 1$. In this case, b_{10} and b_{01} are the population variances. The corresponding U-statistic for estimating b_{10}, b_{01} are, respectively.

$$T_{10} = \binom{M}{2}^{-1} \sum_{i<j} (X_i - X_j)^2/2 = S_X^2,$$

$$T_{01} = \binom{M}{2}^{-1} \sum_{i<j} (Y_i - Y_j)^2/2 = S_Y^2.$$

The kernels are not bounded, unless the distributions of X and Y are bounded. $b_{10}(b_{01})$ is positive if populations $X(Y)$ is not a constant with probability one.

To apply the theorems of this paper, the distributions of X and Y must have finite eighth moments. One may choose, say $M = O(N^{6/7})$.

If D contains normal distribution functions only, Ghurye and Robbins [4] have given exact results for small samples.

6.3. *An example where the theorems of this paper do not apply.* Let the parameter be $\theta = [E(X)]^2 - [E(Y)]^2$, and let F, G belong to any class D of cumulative distribution functions having zero mean and all finite moments. Here, $r = s = 2$, and $\phi = X_i X_j - Y_i Y_j$. The kernels for b_{10}, b_{01} are respectively $(X_1X_2 - Y_1Y_2) \cdot (X_1X_3 - Y_3Y_4)$, and $(X_1X_2 - Y_1Y_2)(X_3X_4 - Y_1Y_3)$.

Since it can be shown that each of these has zero expected value, one cannot use any of the results of this paper. However, the theory of U-statistic is applicable. The kernels for b_{20} and b_{02} are $(X_1X_2 - Y_1Y_2)(X_1X_2 - Y_3Y_4)$, and $(X_1X_2 - Y_1Y_2)(X_3X_4 - Y_1Y_2)$ respectively. Their expected values are $E(X_1^2X_2^2) = [\text{Var}(X)]^2 > 0$, and $E(Y_1^2Y_2^2) = [\text{Var}(Y)]^2 > 0$, respectively.

Special attention should also be paid to the fact that in this case, the associated U-statistic may not be asymptotically normally distributed, see [11].

7. The asymptotic distribution of U' and U''. Consider two random variables Y' and Y^* defined as the following:

$$Y' = (U' - \theta)/(E_{m'}[\text{Var}(U'_{m'})])^{\frac{1}{2}}, \quad Y^* = U_{N'Q,N'(1-Q)} - \theta/[\text{Var}(U_{N'Q,N'(1-Q)})]^{\frac{1}{2}}.$$

Rosenblatt ([11], Theorem 2.2) has proved that Y^* is asymptotically normal with mean zero and variance one.

THEOREM 7.1. *Y' and Y^* are asymptotically equivalent, therefore, Y' is asymptotically normal with mean zero and variance one.*

PROOF. It suffices to show that

$$E(Y' - Y^*)^2 = E(Y')^2 + E(Y^*)^2 - 2E(Y'Y^*) \to 0, \quad \text{as} \quad N' \to \infty.$$

From Theorem 4.1, $E(Y')^2 = 1$. By assumption, $E(Y^*)^2 = 1$. Now let $\bar{U}'$ be the statistic U' with the kernel $\phi'(\bar{X}_{i_r}; \bar{Y}_{j_r})$, and $\bar{U}_{N'Q,N'(1-Q)}$ be the statistic $U_{N'Q,N'(1-Q)}$ with the kernel $\phi'(\bar{X}_{i_r}; \bar{Y}_{j_r})$. Write the expectation into two parts, namely when C occurred or when C' occurred (defined as in Theorem 4.1). Since $E(Y'Y^*)$ is the correlation coefficient of $Y'Y^*$ and since both Y' and Y^* are functions of random variables $X_{M+1}, \cdots, X_{M+m'}; Y_{M+1}, \cdots, Y_{M+n'}$, one has $1 \geqq E(Y'Y^*) \geqq 0$, and for any m', $n' \geqq r$, both parts are non-negative. Consequently,

$$\begin{aligned} &E(Y'Y^*) \\ &\geqq \{N'/[r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]^2\} \\ &\qquad \cdot E\{E_{m' \varepsilon C}(\bar{U}'_{m'} \bar{U}_{N'Q,N'(1-Q)} \mid m' \varepsilon C)\} \Pr(m' \varepsilon C) \qquad (7.1) \\ &\geqq \{N'/[r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]^2\} E\{\bar{U}_{N'(Q+\bar{\varepsilon}),N'(1-Q-\bar{\varepsilon})} \bar{U}_{N'Q,N'(1-Q)}\} \\ &\quad + O(M^{-2+4p+\beta}), \end{aligned}$$

by Lemma 4.2, where $\bar{e}$ denotes some value in the interval $(-M^{-p}, M^{-p})$. Note that for $\bar{e}$ identically zero, $Y' \equiv Y^*$. Notice,

$$E[\bar{U}'_{N'(Q+\bar{e}),N'(1-Q-\bar{e})}\bar{U}_{N'Q,N'(1-Q)}] = \binom{N'(Q+\bar{e})}{r}^{-1}\binom{N'(1-Q-\bar{e})}{r}^{-1}\binom{N'Q}{r}^{-1}\binom{N'(1-Q)}{r}^{-1} \cdot \sum_1 \sum_2 \sum_3 \sum_4 E[\phi'(\bar{X}_{i_r}; \bar{Y}_{j_r})\phi'(\bar{X}_{k_r}; \bar{Y}_{t_r})],$$

where $i_{r,0} : (M+1, N'(Q+\bar{e})), j_{r,0} : (M+1, (1-Q-\bar{e})); k_{r,0} : (M+1, N'Q)$, and $t_{r,0} : (M+1, N'(1-Q))$. The number of sets having (c, d) integers in common are: for all $\bar{e}$ non-negative,

$$\binom{r}{c}\binom{N'Q}{r}\binom{N(Q+\bar{e})-c}{r-c}\binom{r}{d}\binom{N'(1-Q-\bar{e})}{r}\binom{N'(1-Q)-d}{r-d}; \tag{7.2}$$

and for all $\bar{e}$ non-positive,

$$\binom{r}{c}\binom{N'(Q+\bar{e})}{r}\binom{N'Q-c}{r-c}\binom{r}{d}\binom{N'(1-Q)}{r}\binom{N'(1-Q-\bar{e})-d}{r-d}. \tag{7.3}$$

Consider (7.2), and simplify, one has

$$E(\bar{U}'_{N'(Q+\bar{e}),N'(1-Q-\bar{e})}\bar{U}_{N'(1-Q)})$$
$$= \sum_{\substack{c=0\\ c,d\neq 0,0}}^{r} \sum_{d=0}^{r} (r!)^4 b_{cd}/\{(N')^{c+d}(Q+\bar{e})^c(1-Q)^d c!\,d![(r-c)!(r-d)!]^2\}$$
$$= (N')^{-1}[r(b_{01})^{\frac{1}{2}} + r(b_{10})^{\frac{1}{2}}]^2 + O(N'^{-1}M^{-p}) + \text{lower order terms},$$

where $\bar{e} \to 0$ is the slowest for $\bar{e}$ near M^{-p}.

Similarly, one finds the same results if $\bar{e}$ is non-positive. Thus, combine (7.1) and (7.2), it is found, $1 \geqq E(Y'Y^*) \geqq 1 + O(M^{-p}) + O(M^{-2+4p+\beta})$. Therefore,

$$\text{Limit}_{N'\to\infty} E(Y' - Y^*)^2$$
$$= \text{Limit}_{N'\to\infty} [2 - 2 + O(N^{-p/\beta}) + O(N(-2+4p+\beta)/\beta)] = 0,$$

which completes the proof.

THEOREM 7.2. *$Y'_s = (U' - \theta)/\{N'^{-\frac{1}{2}}(r(T_{10})^{\frac{1}{2}} + r(T_{01})^{\frac{1}{2}})\}$ is asymptotically equivalent to Y'. Consequently, Y'_s is asymptotically normally distributed with mean zero and variance one.*

PROOF. It suffices to show that $N'^{-\frac{1}{2}}[r(T_{10})^{\frac{1}{2}} + r(T_{01})^{\frac{1}{2}}]$ is asymptotically equivalent to $N'^{-\frac{1}{2}}[r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]$ (see [7], Theorem 5 and applications).

In other words, it is sufficient to show that for any $e' > 0$,

$$\text{Limit}_{N'\to\infty} \Pr\{|r(T_{10})^{\frac{1}{2}} + r(T_{01})^{\frac{1}{2}} - r(b_{10})^{\frac{1}{2}} - r(b_{01})^{\frac{1}{2}}| > e'N'^{\frac{1}{2}}\} = 0. \tag{7.4}$$

Apply the Chebyshev inequality and evaluate the expectation by applying the identity on page 353 of [2]. Using the result of Lemma 4.2 for odd moments, one finds that the probability in (7.4) is of order of $O(NMe'^2)^{-1}$, which approach zero as $N' \to \infty$ for suitable e'. Hence, Theorem 7.2 is proved.

Next, consider the random variables,

$$Y'' = \frac{U'' - \theta}{[E_{m''} \operatorname{Var}(U''_{m''})]^{\frac{1}{2}}}, \quad \text{and} \quad Y^{**} = \frac{U_{NQ,N(1-Q)} - \theta}{[\operatorname{Var}(U_{NQ,N(1-Q)})]^{\frac{1}{2}}},$$

where Y^{**} is asymptotically normally distributed with mean zero and variance one (same as Y^*). U'' is biased, and the bias is of negligible order of magnitude comparing to the order of magnitude of the value $[E_{m''} \operatorname{Var}(U''_{m''})]^{\frac{1}{2}} = [E(U'' - \theta)^2 - (\text{Bias of } U'')^2]^{\frac{1}{2}}$ (see Theorem 4.2). Hence, by showing that Y'' is asymptotically equivalent to Y^{**}, one concludes that U'' is also asymptotically normally distributed. Along the similar arguments as for the case of U', we have the analogous Theorems in terms of Y'', Y^{**}, and Y''_s.

8. Extension of the two-stage technique to k-sample case, $k > 2$. Let $X^{(1)}, \cdots, X^{(k)}$ be k populations ($k > 2$) with cumulative distribution functions $F_1(X), \cdots, F_k(X)$ respectively, and $\theta = \theta(F_1, \cdots, F_k)$ be the functional to be estimated. Let the symmetric kernel be $\phi(\bar{X}^{(1)}, \cdots, \bar{X}^{(k)})$, where $\bar{X}^{(j)}$ are r independent observations on population $X^{(j)}$. Analogously, define

$$b_{a_1, \cdots, a_k} = E[\phi'_{a_1, \cdots, a_k}(\bar{\bar{X}}^{(1)}, \cdots, \bar{\bar{X}}^{(k)})]^2$$

for vectors $\bar{\bar{X}}^{(j)}$ of dimension a_j, $a_j = 0, 1, \cdots, r$ for all $j = 1, \cdots, k$; and

$$\operatorname{Var}(U_k) = \left[\prod_{i=1}^{k} \binom{n_i}{r}\right]^{-1} \sum_{a_1=0}^{r} \cdots \sum_{a_k=0}^{r} \binom{r}{a_1}\binom{n_1 - r}{r - a_1} \cdots \binom{r}{a_k}\binom{n_k - r}{r - a_k} b_{a_1 \cdots a_k}.$$

Let N be fixed, and $\sum_{i=1}^{k} n_i = N$. If $n_i \to \infty$ in such a way that n_i/n_j are bounded away from zero and one, for all $i \neq j$, $i, j = 1, 2, \cdots, k$, then the asymptotic expression for Var (U_k) is $\operatorname{Var}(U_k) \geqq \sum_{i=1}^{k} (r^2/n_i) b^{(i)} = V'_k$, say, where $b^{(i)} = 0, \cdots, 1, \cdots, 0, \cdots$, with unity at the ith subscript and zero elsewhere.

When $b^{(i)}$, $i = 1, 2, \cdots, k$, are known; it is easy to show that V' is minimized when $n_i = (b^{(i)})^{\frac{1}{2}}/\sum_{i=1}^{k} (b^{(i)})^{\frac{1}{2}}$, and its minimum value is $V_0 = N^{-1}[\sum_{i=1}^{k} r(b^{(i)})^{\frac{1}{2}}]^2$.

The two-stage estimating procedure will be as follows:

(a) Take M observations on each of the $X^{(i)}$, $i = 1, 2, \cdots, k$, where $2rk \leqq kM < N - rk$.

(b) Estimate the k unknowns $b^{(i)}$, $i = 1, 2, \cdots, k$, by, say, $T^{(i)}$.

(c) Take m_i more observations on $X^{(i)}$, $i = 1, 2, \cdots, k$, such that $m_i = N'(T^{(i)})^{\frac{1}{2}}/\sum_{i=1}^{k} (T^{(i)})^{\frac{1}{2}}$ where $N' = N - kM$, (or $M + m_i = N(T^{(i)})^{\frac{1}{2}}/\sum_{i=1}^{k} (T^{(i)})^{\frac{1}{2}}$ for using U'').

(d) Use U'_k (or U''_k) the analogous two-stage k-sample estimator to estimate θ.

Under the same kind of conditions as in Theorem 4.1, but replacing the condition (iii), by: $b^{(i)} > 0$, $i = 1, 2, \cdots, k$, the analogous result can be obtained. Also, the asymptotic distribution of U'_k (or U''_k) is again normal.

Acknowledgment. The author wishes to express her deepest gratitude to Professor I. R. Savage, who has patiently read the rough draft and gave many valuable suggestions. The author is also deeply indebted to Professor W. Hoeffding of the University of North Carolina for his many instructive and illuminating discussions during the writing of this paper.

REFERENCES

[1] ALAM, KHURSHEED (1963). A two-stage sampling procedure for estimating a common mean. Technical Report No. 25. Department of Statistics, Univ. of Minnesota.

[2] CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

[3] FRASER, D. A. S. (1957). *Non-Parametric Methods in Statistics*. Wiley, New York.

[4] GHURYE, S. G. and ROBBINS, HERBERT (1954). Two-stage procedure for estimating the difference between means. *Biometrika* **47** 146–152.

[5] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distributions. *Ann. Math. Statist.* **19** 293–325.

[6] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

[7] MANN, H. B. and WALD, A. (1945). On stochastic limit and order relationships. *Ann. Math. Statist.* **14** 217–226.

[8] PUTTER, JOSEPH. (1951). Sur une méthode de double echantillonnage pour estimer la moyenne d'une population laplacienne stratifiée. *Rev. Inst. Internat. Statist.* **3** 1–8.

[9] RICHTER, DONALD L. (1960). Two-stage experiments for estimating a common mean. *Ann. Math. Statist.* **31** 1164–1173.

[10] ROBBINS, H. (1952). Some aspect of the sequential design of experiments. *Amer. Math. Soc. Bull.* **58** 527–535.

[11] ROSENBLATT, JOAN R. (1955). On a class of non-parametric tests. Institute of Statistics Mimeograph Series No. 138. Univ. of North Carolina.

[12] STEIN, CHARLES. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16** 243–258.

[13] TCHOUPROFF, A. A. (1919). On the mathematical expectation of the moments of frequency distributions. *Biometrika* **12** 140–185.