

May 1963

ON TWO-STAGE NON-PARAMETRIC ESTIMATION

Elizabeth Y. H. Yen ^{1/}

Technical Report No. 26

University of Minnesota
Minneapolis, Minnesota

^{1/} Work done in part under contract Nonr-710(31), NR 042 003
of the Office of Naval Research.

TABLE OF CONTENTS

	Page
Acknowledgement	
1. Introduction and Summary	1
2. Some Basic Properties of One-stage U-statistics and Notations	3
3. Formulation of the Problem: the Two-stage Procedure and the Estimator	8
4. Asymptotic Efficiency of the Estimator	11
5. "Optimal" Choice of the Value M Relative to N	22
6. Some Examples	27
7. The Asymptotic Distribution of U'	30
8. Extension of the Two-stage Technique to Functions of More Than Two Populations	38
9. Summary of the Results on U''	40
References	43

Acknowledgement

I wish to express my deepest gratitude to Professor I. R. Savage, who as my thesis adviser, has spent numerous hours to read my rough drafts and gave valuable suggestions so that the thesis can be presented in this form.

The thesis was written during the time I resided in Raleigh, North Carolina. I am deeply indebted to Professor W. Hoeffding of the University of North Carolina for his guidance. Through his kindness and patience, I have had many instructive and illuminating discussions on the whole idea as well as details of this thesis.

Thanks are also due to Professor L. Hurwicz, who has been my adviser for several years. His constant encouragement and his confidence in me laid the foundation for pursuing such an investigation.

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 $\phi = \phi(X_1, \dots, X_r; Y_1, \dots, Y_s)$ of θ). ϕ is a function of independent observations from two populations with cumulative distribution functions $F(X)$ and $G(Y)$. (Hence, it is called a two-sample problem.) The functions $F(X)$ and $G(Y)$ will be restricted to be members of a specified class D of pairs of cumulative distribution functions, described in the context. The total number of observations from the two populations X and Y will be a fixed number N . The estimation procedure is carried out in two stages. First, take M observations from each of the populations; then, allocate the remaining $N - 2M$ observations to the same populations. The method of allocation utilizes the information from the first stage observations.

A two-stage estimator, represented by U^i , will be introduced. It is a U-statistic with random sample sizes. (See [4] on general U-statistics. U^i is defined in Section 3.) One of the main results (presented in Section 4) is that, under certain conditions, the variance of U^i approaches asymptotically a particular variance V_0 . This particular V_0 (defined in Section 2) is the minimized asymptotic variance of a one-stage estimator \bar{U} . In other words, it is computed (see Section 2) when the best one-stage allocation of N observations to the two populations is made with the help of a partial or even complete information about the distributions $F(X)$ and $G(Y)$. Such an 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. Thus, in particular, V_0 can be computed only when b_{10} and b_{01} are known. Moreover, using these parameters, it will be shown in Section 2 that V_0 is the smallest among the variances of all one-stage estimators of θ . However, no prior knowledge of b_{10} and b_{01} is required to compute $\text{Var}(U')$, and it will be proved in Section 4 that $\text{Var}(U')/V_0$ converges to unity as N approaches to infinity.

A brief review of some basic properties of one-stage U-statistics as well as some conventions on notations will be also presented in Section 2.

In Section 5, the "optimal" choice of the first stage sample size M relative to the fixed total sample size N is discussed. Three cases with different conditions on the unbiased estimator ϕ will be considered. In each case, it is found that the "optimal" choice depends on the specific conditions. (For details, see Section 5.)

Section 6 contains some examples. Here, 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 the cases that the above described two-stage estimation procedure can be applied as well as cases where it cannot be applied.

Section 7 contains a proof of the asymptotic normality of U' .

In Section 8, it is indicated that this two-stage two-sample estimation procedure can be extended to k -sample two-stage estimation with similar results for $k > 2$.

In the last section, Section 9, another two-stage two-sample estimator U'' will be introduced. It is based on the combined observations of both stages of total N observations, as compared to U' , which is based on the second stage of $N - 2M$ observations only. U'' is biased while U' is unbiased. Since U'' is of a different nature as compared to U' , the corresponding proofs are much involved. The results on U'' will be summarized, without proof, in this section.

The technique of two-stage estimation has been discussed in several papers. Stein [11] has used a two-stage procedure to determine confidence intervals of a pre-assigned length for the mean of a normal population with unknown variance. Putter [7] used such a technique to estimate the mean of a stratified normal population. Robbins [9] discussed a two-stage procedure from the point of view of the design of experiments. Later, Ghurye and Robbins [3] used a two-stage technique to estimate the difference between the means of two normal populations (or some other specified populations). Richter [8] discussed the estimation of the common mean of two normal populations. The results of the present paper, then, are to generalize 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.

2. Some Basic Properties of One-stage U-statistics and Notations

Before formulating the problem, a short review of some basic properties of U-statistics is given in this section, based on

references [4, 10]. For convenience of presentation, some specific notations are adopted here as well as throughout this paper:

(1) k will be used as a generic constant, which may represent different values according to the context.

(2) ϵ will be used as any small positive real number, its value will be specified in various situations.

(3) Vectorial notations will be used such as:

$$\begin{aligned}\bar{X}_r &= (X_{i_1}, \dots, X_{i_r}), \text{ where } r = 1, 2, \dots \\ \bar{X}_{r,j} &= (X_{i_{j+1}}, \dots, X_{i_r}) \\ \bar{X}_{i_j} &= (X_{i_1}, \dots, X_{i_j}) \\ \bar{X}_{i_{k,j}} &= (X_{i_{j+1}}, \dots, X_{i_k})\end{aligned}$$

Here, the subscripts of the coordinates are a permutation of some set of integers, which will be specified in the context.

In order to give a definition of a two-sample one-stage U-statistic, let us consider two populations X and Y with cumulative distribution functions F and G respectively. Also, let us consider a real valued estimable parameter $\theta = \theta(F, G)$.

By the statement that θ is estimable, we mean that there exists a function $\phi(\bar{X}_r; \bar{Y}_s)$ such that, with the integration taken over all values of X 's and Y 's,

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

Here, \bar{X}_r, \bar{Y}_s are r and s independent observations from population X and Y respectively. Moreover, all F 's and G 's are restricted to

be members of a specified class D of pairs of cumulative distribution functions of the populations X and Y .

Without loss of generality, the function ϕ , called the kernel, can be assumed to be symmetric in its X arguments and its Y arguments separately. (See [4, 10].) Furthermore, since any function of r X 's and s Y 's can be written as a function of $\max(r, s)$ of X 's and Y 's, we shall assume $r = s$.

Definition ([2, 4, 10])

A U -statistic associated with the parameter θ and the kernel ϕ , defined as above, in a sample of m observations on population X and n observations on population Y for $m, n \geq r$, is defined as:

$$(2.2) \quad U_{m,n} = U(\bar{X}_m; \bar{Y}_n) = \binom{m}{r}^{-1} \binom{n}{r}^{-1} \sum \phi(\bar{X}_{i_r}; \bar{Y}_{j_r})$$

where the summation is taken over all sets of integers such that

$$1 \leq i_1 < \dots < i_r \leq m; \quad 1 \leq j_1 < \dots < j_r \leq n.$$

Now, in order to write the variance of $U_{m,n}$, we define, for $c, d = 0, 1, 2, \dots, r$,

$$(2.3) \quad \phi'(\bar{X}_{i_r}; \bar{Y}_{j_r}) = \phi(\bar{X}_{i_r}; \bar{Y}_{j_r}) - \theta$$

$$\phi'_{cd}(\bar{X}_c; \bar{Y}_d) = E\phi'(\bar{x}_c, \bar{X}_{r,c}; \bar{y}_d, \bar{Y}_{r,d})$$

i.e., ϕ'_{cd} is the conditional expected value of ϕ' , given \bar{x}_c and \bar{y}_d .

Note that $\phi'_{00} = 0$. Also define

$$(2.4) \quad b_{cd} = E[\phi'_{cd}(\bar{X}_c; \bar{Y}_d)]^2.$$

It can be deduced [2, p. 224, p. 257] that

$$b_{cd} = \text{cov}[\phi'(\bar{X}_{i_r}; \bar{Y}_{j_r}), \phi'(\bar{X}_{k_r}; \bar{Y}_{t_r})]$$

where (i_1, \dots, i_r) and (k_1, \dots, k_r) are any two sets of r distinct integers from $(1, 2, \dots, m)$ and c is the number of integers common to the two sets; (j_1, \dots, j_r) and (t_1, \dots, t_r) are any two sets of r distinct integers from $(1, 2, \dots, n)$ and d is the number of integers common to the two sets. Then, the variance of $U_{m,n}$ can be expressed [2, p. 257] as:

$$(2.5) \quad \text{Var}(U_{m,n}) = \binom{m}{r}^{-1} \binom{n}{r}^{-1} \sum_{c=0}^r \sum_{d=0}^r \binom{r}{c} \binom{m-r}{r-c} \binom{r}{d} \binom{m-r}{r-d} b_{cd}.$$

Next, according to Fraser [2], the class D of pairs of cumulative distributions, $F(X)$ and $G(Y)$, for $U_{m,n}$ may be consisted of all distributions uniform within intervals. (For definition, see [2]. Particular examples are: a) a class of pairs of absolutely continuous distribution functions or b) a class of pairs of discrete distribution functions.) Then, an important theory regarding the variance of $U_{m,n}$ is also given by Fraser [2, Theorem 7.1, p. 28 and Theorem 2.1, 2.2, p. 142].

Fraser's Theorem

If the class of pairs of distribution functions includes all distributions uniform within intervals, mentioned above, then for $m, n \geq r$, $U_{m,n}$ is the unique minimum variance unbiased estimator.

Rosenblatt [10] has obtained the following Lemmas:

Rosenblatt's Lemma 2.4 For $1 \leq c < g \leq r$; $1 \leq d < h \leq r$, one has

$$(2.6) \quad g b_{co} \leq c b_{go}, \quad h b_{od} \leq d b_{oh}; \quad \text{and}$$

$$(2.7) \quad 0 \leq L_{cd} \leq \frac{cd}{gh} L_{gh}, \quad \text{where}$$

$$L_{cd} = b_{cd} - b_{co} - b_{od} = E[\phi'_{cd}(\bar{X}_c; \bar{Y}_d) - \phi'_{co}(\bar{X}_c) - \phi'_{od}(\bar{Y}_d)]^2.$$

Rosenblatt's Lemma 2.5 $\text{Var}(U_{m,n})$ has the following upper and lower bounds:

$$(2.8) \quad \text{Var}(U_{m,n}) \geq \frac{r^2}{m} b_{10} + \frac{r^2}{n} b_{01} + \frac{r^4}{mn} L_{11}$$

$$(2.9) \quad \text{Var}(U_{m,n}) \leq \frac{r}{m} b_{r0} + \frac{r}{n} b_{0r} + \frac{r^2}{mn} L_{rr}.$$

In the above discussion, concerning $U_{m,n}$ only, it is assumed that m and n are fixed numbers. Now, if m and n are not fixed but the total number of observations on populations X and Y are restricted to be a fixed number N , i.e., $m + n = N$, we shall denote such a two-sample statistic by \bar{U} instead of $U_{m,n}$. Using the quantities b_{10} , b_{01} as defined in (2.4), the following statement can be made on the lower bound of the variance of \bar{U} .

Rosenblatt's Lemma 2.6 If the ratio m/n satisfies

$$0 < a_1 \leq m/n \leq a_2 < \infty, \quad \text{as } m, n \rightarrow \infty$$

then

$$\text{Var}(\bar{U}) \geq (r^2/m)b_{10} + (r^2/n)b_{01} = V', \quad \text{say,}$$

i.e., V' is the lower bound of $\text{Var}(\bar{U})$ and V' is actually the asymptotic variance of \bar{U} .

Now, V' as defined above can be minimized by selecting the best values of m and n subject to $m + n = N$, and $m, n \geq r$. One finds that the best choices are

$$(2.10) \quad m_0 = N(b_{10})^{\frac{1}{2}} / [(b_{10})^{\frac{1}{2}} + (b_{01})^{\frac{1}{2}}] = NQ, \text{ say, and}$$

$$n_0 = N - m_0 = N(1 - Q).$$

These values for the sample sizes, represent the best allocation of N observations to the populations X and Y . They depend, however, on the unknowns b_{10} and b_{01} , which represent a partial information about the distributions $F(X)$ and $G(Y)$ and have been assumed to be positive quantities. In other words, these sample sizes can be computed and the corresponding \bar{U} statistic can be constructed only when b_{10} and b_{01} are positive and known.

The minimum value of V' , denoted by V_0 is found to be

$$(2.11) \quad V_0 = N^{-1} [r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]^2 = v'(m_0, n_0).$$

It is clear that V_0 is at least as small as the variance of any estimator of θ based on U -statistics subject to the restriction that $m + n = N$. Hence, V_0 is the minimized asymptotic variance of \bar{U} , when the best allocation of N observations to populations X and Y is made. It will be used as a basis for comparison in the remaining sections. In particular, it will be shown that there exist two-stage two-sample statistics, say U' , such that $\text{Var}(U')/V_0$ converges to unity as N approaches infinity, even though no prior knowledge of b_{10} and b_{01} is required to compute $\text{Var}(U')$.

3. Formulation of the Problem; the Two-stage Procedure and the Estimator

In this section, a two-stage statistic U' will be defined. The major result of the investigation on U' , which will be presented in Section 4, is to show that with large samples and under

certain conditions the variance of U' approaches V_0 of equation (2.11). No prior knowledge of b_{10} and b_{01} is required to obtain U' .

Definition of U'

Let the total number of observations from populations X and Y 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 shall estimate the parameters b_{10} , b_{01} . It is observed from (2.1) and (2.4) that b_{10} and b_{01} are estimable functions [4]. There exist two associated U -statistics, called T_{10} and T_{01} , which are unbiased estimators of b_{10} and b_{01} respectively. The symmetric kernels of these two statistics are functions of $2r$ X 's and $2r$ Y 's. Thus one can express T_{10} , T_{01} as follows:

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

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

where the summations are taken over all sets of integers,

$$1 \leq i_1 < \dots < i_{2r} \leq M; \quad 1 \leq j_1 < \dots < j_{2r} \leq M.$$

In analogy with (2.10), we define

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

$$Z = 0 \quad \text{otherwise.}$$

After T_{10} , T_{01} and Z are computed, the second stage is constructed by taking m' more observations on population X and n'

more observations on population Y with $m' + n' = N - 2M = N'$, where the sample sizes m' and n' are determined as follows:

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

$$\text{and } n' = N - m',$$

where $[a]$ is the largest integer contained in a .

With m' and n' so defined, the statistic U' will be defined as the estimator of θ (see equations (2.1) and (2.2)) based on m' and n' observations on populations X and Y respectively.

$$(3.5) \quad U' = \binom{m'}{r}^{-1} \binom{n'}{r}^{-1} \sum \phi(\bar{X}_{i_1 \dots i_r}; \bar{Y}_{j_1 \dots j_r})$$

where the summation is taken over all sets of integers,

$$M + 1 \leq i_1 < \dots < i_r \leq M + m'; \quad M + 1 \leq j_1 < \dots < j_r \leq M + n'.$$

In other words, U' is explicitly a function of the second stage observations only. However, the sample sizes m' and n' are in turn explicit functions of the first stage observations. Hence, implicitly, U' depends on both stages.

Finally, notice that the allocation of N' observations in (3.4) is the same as that of (2.10) with Z in place of Q . It will be shown in Lemma 4.2 that if $M \rightarrow \infty$, then $Z \rightarrow Q$ in probability. Consequently, the probability of the first case of (3.4) occurring approaches unity and the contribution of the other two cases to the variance of U' will be negligible, as $N' \rightarrow \infty$. Thus one may dispose of the other two cases and replace (3.4) by $m' = N'Z$ and

$n' = N'(1 - Z)$. (Note that the brackets of (3.4) for m' and n' will be left out hereafter, since its contribution to the variance of U' is also negligible, as $N' \rightarrow \infty$.)

REMARK: In the two-stage procedure, equal number of observations on populations X and Y are used at the first stage. Intuitively, when $r = s$ occurs in the kernel ϕ in a natural way (i.e., no argument of ϕ is identically zero), and there is no information about the relative sizes of b_{10} and b_{01} , equal size samples seem appropriate to the symmetry of the situation. When $r \neq s$, but one writes ϕ as a function of $\max(r, s)$, one might doubt the appropriateness of the equal sample sizes at the first stage.

4. Asymptotic Efficiency of the Estimator

It is mentioned in Section 3 that if $N \rightarrow \infty$ with $M \rightarrow \infty$ and $N' \rightarrow \infty$, then the second stage sample sizes (3.4) can be replaced by:

$$(4.1) \quad m' = N'Z ; \quad n' = N'(1 - Z)$$

where b_{10} and b_{01} are assumed to be positive.

In this section, it will be shown that under certain conditions, the ratio between the variance of U' (defined in Section 3) and V_0 will asymptotically approach unity. (Recall that V_0 is the smallest of the variances of any one-stage U-statistic estimator of θ subject to the restriction that $m + n = N$. V_0 can be computed only when b_{10} and b_{01} are known and the best one-stage allocation of N observations to populations X and Y are made.) The proofs are presented in Theorem 4.1. First, U' is shown to be unbiased.

Then, in Theorem 4.1, $N \text{Var}(U^i)$ is partitioned into two parts, namely, when $|Z - Q|$ is less than M^{-p} for any p within the range $0 < p < \frac{1}{4}$; and when $|Z - Q|$ is greater than M^{-p} . (Recall that Q gives the best allocation of N observations and the basis for evaluating V_0 , see (2.10)). By the results of Lemmas 4.1 and 4.2, it is concluded that the second part is of the order of magnitude of $O(M^{-2+4p} N)$. The first part of $N \text{Var}(U^i)$ is shown to be of the order of magnitude of $[r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]^2 + O(MN^{-1}) + O(M^{-p})$. The first term of this expression is equal to $N V_0$. Now, under certain assumptions (see Theorem 4.1 below) concerning the relative order of magnitude of M and N and $0 < p < \frac{1}{4}$. It will be shown that $O(M^{-2+4p} N)$, $O(M/N)$ and $O(M^{-p})$ converge to zero as N approaches infinity. Hence the ratio $\text{Var}(U^i)/V_0$ converges to unity, which is the result of Theorem 4.1.

In Section 5, it will be shown that the best choice of M (under the assumptions of Theorem 4.1) is equal to $KN^{6/7}$, where K is a non-zero unknown constant. The resulting value of p is $1/6$. Thus the ratio of $\text{Var}(U^i)$ to V_0 is equal to $1 + O(N^{-1/7})$.

Lemma 4.1 Let $\theta(F, G) = \theta$ be an estimable parameter with symmetric kernel $S = S(\bar{X}_r; \bar{Y}_r)$. Let $W = W_{MM}$ be the associated U-statistic with M observations on populations X and Y with cumulative distribution functions $F(X)$ and $G(Y)$ respectively. Assume that the $2i^{\text{th}}$ moment of the kernel is finite. Define:

a) $W^i = W - \theta$, b) $S^i = S - \theta$ and c) $S^i(\bar{X}_{rt+r,rt}; \bar{Y}_{rt+r,rt}) = S^i_t$, then for any positive integer i ,

$$\begin{aligned} E(W')^{2i} &= \frac{(2i)!}{1!2^i M^i} [E(S')^2]^i + o(M^{-1}) \\ &= o(M^{-1}) . \end{aligned}$$

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

$$\begin{aligned} W'' &= \frac{1}{k} \sum_{t=0}^{k-1} S'(\bar{X}_{rt+r,rt}; \bar{Y}_{rt+r,rt}) \\ &= \frac{1}{k} \sum_{t=0}^{k-1} S'_t, \quad \text{where } k = M/r . \end{aligned}$$

W'' is an average of k independent and identically distributed random variables with mean zero. From the work of Tchouproff [12], one has

$$\begin{aligned} E(W'')^{2i} &= \frac{1}{k^{2i}} E\left(\sum_{t=0}^{k-1} S'_t\right)^{2i} \\ &= \frac{1}{k^{2i}} \left[\frac{(2i)!}{1!2^i} k(k-1) \dots (k-i+1) E(S'_1)^2 \dots E(S'_1)^2 \right. \\ &\quad \left. + o(k^{-i-1}) \right] \\ &= \frac{(2i)!}{1!2^i k^i} [\text{Var}(S')]^i + o(k^{-i-1}) \\ &= o(M^{-1}) . \end{aligned}$$

We now prove that W' can be written in terms of W'' as follows:

$$(4.2) \quad W' = (M!)^{-2} \sum W''(\bar{X}_{h_M}; \bar{Y}_{j_M}),$$

where the summation is taken over all permutations of (h_1, \dots, h_M) , (j_1, \dots, j_M) of $(1, 2, \dots, M)$. Starting with the right side of (4.2),

$$\begin{aligned}
(M!)^{-2} \Sigma W'' &= (M!)^{-2} \Sigma \frac{1}{k} \sum_{t=0}^{k-1} S'(\bar{X}_{h_{rt+r,rt}}; \bar{Y}_{j_{rt+r,rt}}) \\
&= \frac{1}{k} \sum_{t=0}^{k-1} (M!)^{-2} \Sigma S'(\bar{X}_{h_{rt+r,rt}}; \bar{Y}_{j_{rt+r,rt}}) \\
&= \frac{1}{k} \sum_{t=0}^{k-1} \binom{M}{r}^{-2} \Sigma S'(\bar{X}_{h_{rt+r,rt}}; \bar{Y}_{j_{rt+t,rt}})
\end{aligned}$$

where Σ' is taken over all sets of integers,

$$1 \leq h_{rt+1} < \dots < h_{rt+r} \leq M; \quad 1 \leq j_{rt+1} < \dots < j_{rt+r} \leq M,$$

for all $t = 0, 1, \dots, k-1$.

Then,

$$(M!)^{-2} \Sigma W'' = \frac{1}{k} \sum_{t=0}^{k-1} W' = W'.$$

Next, since

$$(W')^{2i} = [(M!)^{-2} \Sigma W'']^{2i} \leq (W'')^{2i},$$

one has

$$\begin{aligned}
E(W')^{2i} &\leq E(W'')^{2i} = \frac{(2i)!}{i! 2^i k^i} [\text{Var}(S')^i + O(k^{-i-1})] \\
&= O(k^{-i}) = O(M^{-i}).
\end{aligned}$$

The lemma is proved.

Lemma 4.2 Let Z and Q be defined as in (3.3) and (2.10) respectively. Assume for $0 < p < \frac{i-1}{2i}$, i is an integer, $i \geq 2$, that ϕ has $4i^{\text{th}}$ finite moments. Then

$$\Pr[|Z - Q| > M^{-p}] = O(M^{-i+2ip}).$$

Proof: Write

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

Write,

$$\begin{aligned}
& \Pr\left[\frac{(T_{10})^{\frac{1}{2}}}{(T_{01})^{\frac{1}{2}} + (T_{10})^{\frac{1}{2}}} > Q + M^{-p}\right] \\
& = \Pr[(T_{10})^{\frac{1}{2}} > (Q + M^{-p})(T_{10})^{\frac{1}{2}} + (Q + M^{-p})(T_{01})^{\frac{1}{2}}] \\
& = \Pr\left\{ \left[T_{10} - \left(\frac{Q + M^{-p}}{1 - Q - M^{-p}} \right)^2 T_{01} \right] - \left[b_{10} - \left(\frac{Q + M^{-p}}{1 - Q - M^{-p}} \right)^2 b_{01} \right] \right. \\
& \quad \left. > - \left[b_{10} - \left(\frac{Q + M^{-p}}{1 - Q - M^{-p}} \right)^2 b_{01} \right] \right\}.
\end{aligned}$$

One notices that

$$\left[T_{10} - \left(\frac{Q + M^{-p}}{1 - Q - M^{-p}} \right)^2 T_{01} \right] - \left[b_{10} - \left(\frac{Q + M^{-p}}{1 - Q - M^{-p}} \right)^2 b_{01} \right]$$

is a U-statistic with mean zero and its kernel is

$$\left[h - \left(\frac{Q + M^{-p}}{1 - Q - M^{-p}} \right)^2 g \right] - \left[b_{10} - \left(\frac{Q + M^{-p}}{1 - Q - M^{-p}} \right)^2 b_{01} \right],$$

where g, h are defined in (3.1), (3.2) respectively. Since b_{10} and b_{01} are assumed positive, one has $0 < Q < 1$. Also, for M large, one can choose $M^{-p} < \min(Q, 1 - Q)$. Thus

$$\begin{aligned}
& -[b_{10} - \left(\frac{Q + M^{-P}}{1 - Q - M^{-P}}\right)^2 b_{01}] \\
& = -[b_{10} - \left(\frac{Q}{1 - Q}\right)^2 \left(1 + \frac{2M^{-P}}{Q(1 - Q)} + o(M^{-P})\right) b_{01}] \\
& = -b_{10} + b_{10} \left[1 + \frac{2M^{-P}}{Q(1 - Q)} + o(M^{-P})\right] \\
& = \frac{2M^{-P} b_{10}}{Q(1 - Q)} + o(M^{-P}).
\end{aligned}$$

Hence the last quantity is positive for small M^{-P} .

Using Tchebycheff's inequality of the form,

$\Pr\{|X| > a\} \leq a^{-2i} E(X)^{2i}$, one has

$$\begin{aligned}
& \Pr \left\{ \left[T_{10} - \left(\frac{Q + M^{-P}}{1 - Q - M^{-P}}\right)^2 T_{01} \right] - \left[b_{10} - \left(\frac{Q + M^{-P}}{1 - Q - M^{-P}}\right)^2 b_{01} \right] \right. \\
& \quad \left. > - \left[b_{10} - \left(\frac{Q + M^{-P}}{1 - Q - M^{-P}}\right)^2 b_{01} \right] \right\} \\
& \leq \frac{E \left\{ \left[T_{10} - \left(\frac{Q + M^{-P}}{1 - Q - M^{-P}}\right)^2 T_{01} \right] - \left[b_{10} - \left(\frac{Q + M^{-P}}{1 - Q - M^{-P}}\right)^2 b_{01} \right] \right\}^{2i}}{\left\{ - \left[b_{10} - \left(\frac{Q + M^{-P}}{1 - Q - M^{-P}}\right)^2 b_{01} \right] \right\}^{2i}} \\
& \leq \frac{\frac{(2i)!}{i! 2^i} r^i M^{-i} \left\{ \text{Var} \left[h - \left(\frac{Q + M^{-P}}{1 - Q - M^{-P}}\right)^2 g \right] \right\}^i}{\left\{ 2M^{-P} b_{10} [Q(1 - Q)]^{-1} + o(M^{-1}) \right\}^{2i}}
\end{aligned}$$

by Lemma 4.1, and is equal to $O(M^{-i+2iP})$ for M large.

Analogously, one gets,

$$\begin{aligned}
& \Pr \left\{ \frac{(T_{10})^{\frac{1}{2}}}{(T_{01})^{\frac{1}{2}} + (T_{10})^{\frac{1}{2}}} < Q - M^{-p} \right\} \\
&= \Pr \left\{ \left[\left(\frac{Q - M^{-p}}{1 - Q + M^{-p}} \right)^2 T_{01} - T_{10} \right] - \left[\left(\frac{Q - M^{-p}}{1 - Q + M^{-p}} \right)^2 b_{01} - b_{10} \right] \right. \\
&\quad \left. > - \left[\left(\frac{Q - M^{-p}}{1 - Q + M^{-p}} \right)^2 b_{01} - b_{10} \right] \right\} \\
&\leq \frac{\frac{(2i)!}{1!2^i} r^i M^{-i} \left\{ \left[\text{Var} \left(\frac{Q - M^{-p}}{1 - Q + M^{-p}} \right)^2 g - h \right] \right\}^i}{[2M^{-p} b_{01} [Q(1 - Q)]^{-1} + o(M^{-p})]^{2i}} \\
&= o(M^{-i+2ip}) .
\end{aligned}$$

Similarly, using Lemma 4.1,

$$\begin{aligned}
\Pr(T_{10} \leq 0) &= \Pr(T_{10} - b_{10} \leq -b_{10}) \leq \Pr(|T_{10} - b_{10}| \geq b_{10}) \\
&\leq \frac{E(T_{10} - b_{10})^{2i}}{b_{10}^{2i}} \\
&= o(M^{-i}), \quad \text{and}
\end{aligned}$$

$$\Pr(T_{01} \leq 0) = o(M^{-i}).$$

Therefore, $\Pr(|Z - Q| > M^{-p}) = o(M^{-i+2ip}) + o(M^{-i}) = o(M^{-i+2ip})$.

Lemma 4.2 is proved.

Theorem 4.1 U' is an unbiased estimator of θ , i.e.,

$E(U') = \theta$. Also, if

- (i) $\lim_{N \rightarrow \infty} \frac{N}{M^\beta}$ 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} are both positive,

then

$$\lim_{N \rightarrow \infty} \frac{\text{Var}(U^i)}{V_0} = \lim_{N \rightarrow \infty} \frac{E_{m'}[\text{Var}(U_{m'}^i)]}{V_0} = 1.$$

REMARK: In most non-parametric problems, the kernel ϕ is bounded, hence all moments exist. Therefore, the restriction (ii) is not severe. $\text{Var}(U_{m'}^i)$ denotes the conditional variance of U^i given m' and n' , and $\text{Var}(U^i)$ denotes the expected value of $\text{Var}(U_{m'}^i)$, where the expectation is over m' and n' , or $\text{Var}(U^i)$ is the unconditional variance of U^i .

Proof: Notice that m', n' are defined to be greater than r , and that they are functions of $X_1, \dots, X_M; Y_1, \dots, Y_M$ only. On the other hand, all the arguments of $\phi(\bar{X}_{1_r}; \bar{Y}_{j_r})$ in the definition of U^i (see (3.5)) are functions of $X_{M+1}, \dots, X_{M+m'}; Y_{M+1}, \dots, Y_{M+n'}$. Thus the arguments of U^i are independent of $X_1, \dots, X_M; Y_1, \dots, Y_M$. Therefore,

$$\begin{aligned} E(U^i) &= E_{m'} E(U_{m'}^i) \\ &= E_{m'} \binom{m'}{r}^{-1} \binom{n'}{r}^{-1} \sum E \phi(\bar{X}_{1_r}; \bar{Y}_{j_r}) \\ &= E_{m'} \binom{m'}{r}^{-1} \binom{n'}{r}^{-1} \sum \theta \\ &= \theta. \end{aligned}$$

Hence, U^i is unbiased.

Now, let $C =$ condition that $|Z - Q| \leq M^{-p}$

$C' =$ complement of C ,

$0 < p < \frac{1}{4}$ (see Lemma 4.2 for $i = 2$)

$$\begin{aligned} N \text{Var}(U^i) &= N E_{m'} \{ \text{Var}(U_{m'}^i) \} \\ &= N \Pr(|Z - Q| \leq M^{-p}) E_{m' \in C} \text{Var}(U_{m'}^i) \\ &\quad + N \Pr(|Z - Q| > M^{-p}) E_{m' \in C'} \text{Var}(U_{m'}^i) . \end{aligned}$$

Using the fact that $E_{m' \in C} \text{Var}(U_{m'}^i) \leq \text{Var}(U_{rr}^i) = b_{rr} = \text{Var}(\phi)$, which is bounded, by assumption (ii), and

$$\Pr(|Z - Q| > M^{-p}) = O(M^{-2+4p}), \quad 0 < p < \frac{1}{4}$$

by Lemma 4.2, one obtains

$$N \text{Var}(U^i) \leq N E_{m' \in C} \text{Var}(U_{m'}^i) + N \cdot O(M^{-2+4p}) .$$

It is easy to show that there exists a number A which is independent of m' , n' , such that

$$\text{Var}(U_{m'}^i) \leq (r^2 b_{10})/m' + (r^2 b_{01})/n' + A/\min(m'^2, n'^2) .$$

The procedure is to expand the terms of $\text{Var}(U_{m'}^i)$ and its combinatorials and then to substitute a fraction by unity. Consequently, one finds that the first two terms are less than or equal to $(r^2 b_{10})/m'$ and $(r^2 b_{01})/n'$ respectively. For the rest of the rxr terms, we substitute again a certain fraction by unity and find that each term has denominator less than or equal to $\min(m'^2, n'^2)$. Hence we find that A may be taken as the sum of all the rxr values in their numerators, which are composed of r 's and b_{cd} 's, $c, d = 1, 2, \dots, r$. One has

$$E_{m', \epsilon \epsilon} \text{Var}(U_{m'}^i) \leq E_{m', \epsilon \epsilon} \left[\frac{r^2}{m'} b_{10} + \frac{r^2}{n'} b_{01} + \frac{A}{\min(m'^2, n'^2)} \right].$$

Also, when $|Z - Q| \leq M^{-p}$ and $N \rightarrow \infty$, m' , n' can be written as $m' \geq N'(Q - M^{-p})$, $n' \geq N'(1 - Q - M^{-p})$. Thus,

$$\begin{aligned} N \text{Var}(U^i) &\leq \frac{Nr^2 b_{10}}{N'(Q - M^{-p})} + \frac{Nr^2 b_{01}}{N'(1 - Q - M^{-p})} \\ &\quad + \frac{NA}{[\min [N'(Q - M^{-p}), N'(1 - Q - M^{-p})]]^2} + o(M^{-2+4p+\beta}) \\ &= (N/N') [r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]^2 [1 + 2M^{-p} + o(M^{-p})] \\ &\quad + o(NN'^{-2}) + o(M^{-2+4p+\beta}) \\ &= [1 + 2MN^{-1} + o(MN^{-1})] [r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]^2 \\ &\quad \cdot [1 + 2M^{-p} + o(M^{-p})] + o(NN'^{-2}) + o(M^{-2+4p+\beta}) \\ &= [r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]^2 [1 + o(N^{(1-\beta)/\beta}) + o(N^{-p/\beta}) \\ &\quad + o(N^{(-2+4p+\beta)/\beta})], \end{aligned}$$

after putting $M = K(N^{1/\beta})$, where K is an unknown non-zero constant.

Since, by assumption (i), $1 < \beta < 2$, there exists p , so that

$0 < p < \frac{1}{4}$, and $(-2+4p+\beta) < 0$. Finally,

$$\begin{aligned}
\lim_{N \rightarrow \infty} \frac{\text{Var}(U^i)}{V_0} &= \lim_{N \rightarrow \infty} \frac{E_m \text{Var}(U_m^i)}{V_0} \\
&= \lim_{N \rightarrow \infty} \frac{N E_m \text{Var}(U_m^i)}{N V_0} \\
&= \lim_{N \rightarrow \infty} \frac{[r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]^2}{[r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]^2} \\
&\quad \cdot [1 + o(N^{(1-\beta)/\beta}) + o(N^{-p/\beta}) \\
&\quad + o(N^{(-2+4p+\beta)/\beta})] \\
&= 1 .
\end{aligned}$$

Hence, the theorem is proved.

In addition to U^i , other two-stage estimators of θ can be defined. For example, if θ is estimated separately at both stages, then one can combine these two estimates by weights. This paper will not include any explicit discussion of such estimators. On the other hand, the following 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), then proceed as if $b_{10} = b_{01}$. The variance of a one-stage U-statistic is minimized with respect to m , subject to $m + n = N$, when

$$m = N/2, \quad n = N/2 .$$

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,

$$\begin{aligned}
(4.3) \quad \lim_{N \rightarrow \infty} \frac{\text{Var}(U^*)}{V_0} &= \lim_{N \rightarrow \infty} \frac{N \text{Var}(U^*)}{N V_0} \\
&= \lim_{N \rightarrow \infty} \frac{2r^2(b_{10} + b_{01})}{[r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]^2} + o(N^{-1}) \\
&= \frac{2(b_{10} + b_{01})}{[(b_{10})^{\frac{1}{2}} + (b_{01})^{\frac{1}{2}}]^2} = \frac{2(1 + \rho^2)}{(1 + \rho)^2}
\end{aligned}$$

where

$$\rho = (b_{01}/b_{10})^{\frac{1}{2}}.$$

When ρ approaches 0 or ∞ , (4.3) approaches its maximum 2. Thus, comparing the results of Theorem 4.1 with (4.3), an appreciable decrease in variance can be obtained by using a two-stage procedure.

REMARK: For $s \neq r$, if we write ϕ as a function of $\max(r, s)$ X 's and Y 's, the choice of m, n shall be $Nr/(s+r)$ and $Ns/(s+r)$ respectively in order to minimize the variance of U^* assuming $b_{10} = b_{01}$. A simple computation shows that the variance ratio approaches $1 + s/r$ as ρ approaches zero and approaches $1 + r/s$ as ρ approaches infinity. Thus the variance ratio may have a maximum, for $r \neq s$, greater than 2.

5. "Optimal" Choice of the Value M Relative to N

"Optimal" choice of the value M (sample size of the first stage) relative to N (total sample size) will be studied in this section for the following three cases:

a) The first eight moments of the kernel ϕ exist

For this case, we proceed as follows. From the last step of the proof of Theorem 4.1, one has

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

A heuristic method for finding the best β and p is to find the solution of the pair of equations listed below, which are obtained by examining the exponentials in the remainder terms of the above equation.

$$(5.1) \quad \beta - 1 = p$$

$$(5.2) \quad p = 2 - 4p - \beta$$

and get $\beta = 7/6$, $p = 1/6$, thus $M = K(N^{6/7})$.

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

b) All moments of the kernel ϕ exist

By Lemma 4.2 and Theorem 4.1, for general i , $i \geq 2$, $0 < p < (i - 1)/2i$, one has

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

Similar to the above case (for $i = 2$), one solves the two equations:

$$(5.3) \quad \beta - 1 = p$$

$$(5.4) \quad p = i - 2ip - \beta.$$

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 $3/2$ and p approaches $1/2$. Therefore, $M = K(N^{2(1+i)/(3i+1)})$, where $2(i + 1)/(3i + 1)$ has $2/3$ as a lower bound. This bound, however, is not obtained. Thus when ϕ has all finite moments,

$$\text{Var}(U')/V_0 = 1 + O(N^{-1+2(1+i)/(3i+1)}) \text{ for any } i.$$

c) The kernel ϕ is bounded

First, it will be shown by the following Lemma 5.1 that $\Pr(|Z - Q| > e')$ $\leq O(e^{-e'^2 M})$, where e is the base of natural logarithm and e' is some small number. Consequently, an "optimal" choice of M can be obtained in an implicit form.

Lemma 5.1 (Hoeffding's inequality, see [5]). Let $U_{n,n}$ be a U -statistic with n observations on any two populations X and Y for estimating some parameter θ . The kernel of θ is $S(\bar{X}_r; \bar{Y}_s)$, $a \leq S \leq b$. Then for any positive number e' ,

$$\Pr(U_{n,n} - \theta > e') \leq \exp\left(\frac{-2e'^2 \left[\frac{n}{\max(r,s)}\right]}{(b-a)^2}\right)$$

$$\text{for } n \text{ large, } = \exp(-O(e'^2 n))$$

$$= O(e^{-e'^2 n}) .$$

Now, from the proof of Lemma 4.2, and neglecting the smaller order term

$$\begin{aligned}
\Pr(|Z - Q| > e') &= \Pr(Z - Q > e') + \Pr(Z - Q < -e') \\
&= \Pr \left\{ \left[\tau_{10} - \left(\frac{Q + e'}{1 - Q - e'} \right)^2 \tau_{01} \right] - \left[b_{10} - \left(\frac{Q + e'}{1 - Q - e'} \right)^2 b_{01} \right] \right. \\
&\quad \left. > - \left[b_{10} - \left(\frac{Q + e'}{1 - Q - e'} \right)^2 b_{01} \right] \right\} \\
&+ \Pr \left\{ \left[\left(\frac{Q - e'}{1 - Q + e'} \right)^2 \tau_{01} - \tau_{10} \right] - \left[\left(\frac{Q - e'}{1 - Q + e'} \right)^2 b_{01} - b_{10} \right] \right. \\
&\quad \left. > - \left[\left(\frac{Q - e'}{1 - Q + e'} \right)^2 b_{01} - b_{10} \right] \right\} .
\end{aligned}$$

Applying Lemma 5.1 and assuming that $r = s$ and

$$K_1 \leq h - \left(\frac{Q + e'}{1 - Q - e'} \right)^2 g \leq K_2, \quad K_3 \leq \left(\frac{Q - e'}{1 - Q + e'} \right)^2 g - h \leq K_4,$$

$$\begin{aligned}
\Pr(|Z - Q| > e') &\leq \exp \frac{-2 \left[b_{10} - \left(\frac{Q + e'}{1 - Q - e'} \right)^2 b_{01} \right]^2 \frac{M}{r}}{(K_2 - K_1)^2} \\
&+ \exp \frac{-2 \left[\left(\frac{Q - e'}{1 - Q + e'} \right)^2 b_{01} - b_{10} \right]^2 \frac{M}{r}}{(K_3 - K_4)^2} \\
&= O(e^{-e'^2 M}).
\end{aligned}$$

Hence, one has

$$\frac{\text{Var}(U^t)}{V_0} = 1 + O(MN^{-1}) + O(e') + O(e^{-e'^2 M}).$$

Using a similar approach as before, i.e., requiring the three terms to have the same order of magnitude, one has,

$$(5.5) \quad MN^{-1} = e'$$

$$(5.6) \quad \log e' = -e'^2 M.$$

Substitute (5.5) into (5.6),

$$\log e' = -e'^3 N, \text{ hence}$$

$$(5.7) \quad N = -\log e' / (e')^3$$

$$(5.8) \quad M = -\log e' / (e')^2 .$$

From (5.7), (5.8), $M^3 N^{-2} = -\log e' = \log(e')^{-1}$. Therefore,

$$(5.9) \quad M = N^{2/3} \log(e')^{-1/3} .$$

Taking logarithm on (5.7),

$$\begin{aligned} \log N &= \log [\log(e')^{-1}] - \log(e')^3 \\ &= \log [\log(e')^{-1}] + 3 \log(e')^{-1} . \end{aligned}$$

It is seen that for e' small,

$$(3 + \Delta) \log(e')^{-1} > \log N > 3 \log(e')^{-1} .$$

Substituting the inequalities into (5.7) and (5.9) respectively,

one has

$$(5.10) \quad \begin{aligned} M &< N^{2/3} [1/3 \log N]^{1/3} = N^{2/3} [\log N^{1/3}]^{1/3} \\ M &> N^{2/3} [1/(3+\Delta) \log N]^{1/3} = N^{2/3} [\log N^{1/(3+\Delta)}]^{1/3} . \end{aligned}$$

By (5.5), $m = Ne'$, one has

$$N^{-1/3} [\log N^{1/(3+\Delta)}]^{1/3} < e' < N^{-1/3} [\log N^{1/3}]^{1/3} .$$

Therefore, in the case with ϕ bounded,

$$\text{Var}(U')/V_0 = 1 + O(N^{-1/3} I)$$

where I is some value between $(\log N^{1/(3+\Delta)})^{1/3}$ and $(\log N^{1/3})^{1/3}$.

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

$$\begin{aligned} f(X_i, Y_j) &= 1 && \text{if } X_i > Y_j \\ &= 0 && \text{otherwise .} \end{aligned}$$

In this case, $r = s = 1$. The nuisance parameter b_{10} , b_{01} and b_{11} are

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

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

$$b_{11} = \Pr(X_1 > Y_1) - [\Pr(X_1 > Y_1)]^2 .$$

It can be shown that

$$b_{10} = 2 \Pr(X_1 > Y_1 > Y_2 > X_2)$$

$$b_{01} = 2 \Pr(Y_1 > X_1 > X_2 > Y_2)$$

$$\begin{aligned} b_{11} &= \Pr(X_1 > Y_1 > X_2 > Y_2) + \Pr(Y_1 > X_1 > Y_2 > X_2) \\ &\quad + 2 \Pr(X_1 > Y_1 > Y_2 > X_2) + 2 \Pr(Y_1 > X_1 > X_2 > Y_2) . \end{aligned}$$

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

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

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

where $h(\bar{X}_{i_2}; \bar{Y}_{j_2}) = 1/4$ if the two Y 's are ranked between the two X 's.

= 0 otherwise.

$g(\bar{X}_{i_2}; \bar{Y}_{j_2}) = 1/4$ if the two X 's are ranked between the two Y 's .

= 0 otherwise.

Here ϕ , g , h are all bounded. When θ is neither zero nor unity, only one of the b_{10} , b_{01} can be zero. Moreover, it can be shown that when $b_{10} = 0$ ($b_{01} = 0$), $b_{01} = b_{11} = \theta - \theta^2$ ($b_{10} = b_{11} = \theta - \theta^2$). If it is assumed that F , G are both strictly monotone, then both b_{10} and b_{01} are positive. Therefore when F , G are both strictly monotone, the two-stage procedure is applicable. Since in this case ϕ is bounded, one shall choose M between $N^{2/3}[\log N^{1/(3+\Delta)}]^{1/3}$ and $N^{2/3}[\log N^{1/3}]^{1/3}$.

6.2 Assume $\theta = E(X) - E(Y)$, where independent observations on populations X and Y are made with cumulative distribution functions F and G respectively. The class D contains all cumulative distribution functions with finite expectations. Then θ is estimable. The kernel is $\phi = X_i - Y_i$ and again $r = s = 1$. In this case, b_{10} and b_{01} are the population variance if they exist. The kernel of b_{10} , b_{01} are $\frac{1}{2}(X_i - X_j)^2$, $\frac{1}{2}(Y_i - Y_j)^2$, $i < j$, respectively. The corresponding U-statistic for estimating b_{10} , b_{01} are the sample variances, which can be expressed in the following form:

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

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

In this case, the kernels are not bounded, unless the distributions of X and Y are bounded. b_{10} (b_{01}) is positive if population 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, $M = N^{6/7}$.

If D contains normal distribution functions only, Ghurye and Robbins have given exact results for small samples [3].

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 such that populations X and Y have zero mean and all finite moments. Now the corresponding symmetric kernel for estimating θ will be $\theta = X_i X_j - Y_i Y_j$. Then the kernels for b_{10} , b_{01} and b_{11} are of the following forms respectively:

$$\begin{aligned} & (X_1 X_2 - Y_1 Y_2)(X_1 X_3 - Y_3 Y_4) \\ & (X_1 X_2 - Y_1 Y_2)(X_3 X_4 - Y_2 Y_3) \\ & (X_1 X_2 - Y_1 Y_2)(X_1 X_3 - Y_1 Y_3) . \end{aligned}$$

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 and one needs the kernels for b_{20} and b_{02} , which are given respectively by:

$$(X_1 X_2 - Y_1 Y_2)(X_1 X_2 - Y_3 Y_4)$$

$$(X_1 X_2 - Y_1 Y_2)(X_3 X_4 - Y_1 Y_2) .$$

Then the expected values of these kernels are:

$$E(X_1^2 X_2^2) = [\text{Var}(X)]^2 > 0$$

$$E(Y_1^2 Y_2^2) = [\text{Var}(Y)]^2 > 0 , \text{ respectively.}$$

Special attention should also be paid to the fact that in this case, the associated U-statistic may not be asymptotically normally distributed, see [10].

7. The Asymptotic Distribution of U^t

In this section, it will be shown that U^t is asymptotically normally distributed. Let us consider first, two random variables Y^t and Y^{*t} defined as the following:

$$Y^t = (U^t - \theta) / (E_m[\text{Var}(U_m^t)])^{\frac{1}{2}}$$

$$Y^{*t} = U_{N^t Q, N^t(1-Q)}^{-\theta} / [\text{Var}(U_{N^t Q, N^t(1-Q)})]^{\frac{1}{2}} .$$

It has been proved by Rosenblatt [10, Theorem 2.2] that Y^{*t} is asymptotically normal with mean zero and variance one. In what follows, Theorem 7.1 shows that Y^t is asymptotically equivalent to Y^{*t} , thus also asymptotically normally distributed with mean zero and variance one.

Theorem 7.1 Y^t and Y^{*t} are asymptotically equivalent, i.e.,

$$P\text{-}\lim_{N^t \rightarrow \infty} (Y^t - Y^{*t}) = 0 .$$

Proof: In order to show that Y^i and Y^* are asymptotically equivalent, it suffices to show that

$$E(Y^i - Y^*)^2 \rightarrow 0 \text{ as } N^i \rightarrow \infty.$$

$$\text{Now, } E(Y^i - Y^*)^2 = E(Y^i)^2 + E(Y^*)^2 - 2E(Y^i Y^*).$$

From Theorem 4.1, U^i is an unbiased estimator of θ and Y^i is its normalized form. Hence $E(Y^i)^2 = 1$. By assumption, $E(Y^*)^2 = 1$.

Also, by Theorem 4.1,

$$E_m \text{Var}(U_m^i) = (1/N^i)[r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]^2 + o(N^{i-1})$$

and

$$\text{Var}(U_{N^i Q, N^i(1-Q)}^i) = (r^2 b_{10})/(N^i Q) + (r^2 b_{01})/(N^i(1-Q)) + o(N^{i-1})$$

by Rosenblatt [10, Lemma 2.6]. Therefore,

$$(7.1) \quad E(Y^i Y^*) = E[(U^i - \theta)(U_{N^i Q, N^i(1-Q)}^i - \theta)] .$$

$$\begin{aligned} & \cdot \left\{ [E_m \text{Var}(U_m^i)] [\text{Var } U_{N^i Q, N^i(1-Q)}^i] \right\}^{-\frac{1}{2}} \\ & = N^i [r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]^{-2} E[(U^i - \theta)(U_{N^i Q, N^i(1-Q)}^i - \theta)] . \end{aligned}$$

Now let \bar{U}^i be the statistic U^i with the kernel $\phi^i(\bar{X}_{i_r}; \bar{Y}_{j_r})$, and $\bar{U}_{N^i Q, N^i(1-Q)}^i$ be the statistic $U_{N^i Q, N^i(1-Q)}^i$ with the kernel

$\phi^i(\bar{X}_{i_r}; \bar{Y}_{j_r})$. Also let:

$$C = \text{condition that } N^i(Q - M^{-P}) \leq m^i \leq N^i(Q + M^{-P}), \frac{1}{4} > p > 0 .$$

$C^i = \text{complement of } C$.

Then, one may write

$$\begin{aligned}
E(Y'Y^*) &= N' [r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]^{-2} \\
&\cdot E \left\{ E_{m' \in C} [(\bar{U}_{m'}^t)(\bar{U}_{N'Q, N'(1-Q)}) \mid m' \in C] \right\} \Pr(m' \in C) \\
&+ N' [r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]^{-2} \\
&\cdot E \left\{ E_{m' \in C'} [(\bar{U}_{m'}^t)(\bar{U}_{N'Q, N'(1-Q)}) \mid m' \in C'] \right\} \Pr(m' \in C') .
\end{aligned}$$

Now notice that $E(Y'Y^*)$ is the correlation coefficient of $Y'Y^*$ and since both Y' and Y^* are functions of random variables $X_{M+1}, \dots, X_{M+m'}, Y_{M+1}, \dots, Y_{M+n'}$, one has $1 \geq E(Y'Y^*) \geq 0$, and for any m' , $n' \geq r$,

$$E \left\{ E_{m' \in C'} (\bar{U}_{m'}^t \bar{U}_{N'Q, N'(1-Q)}) \mid m' \in C' \right\} \Pr(m' \in C') \geq 0 .$$

Consequently,

$$\begin{aligned}
E(Y'Y^*) &\geq \frac{N'}{[r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]^2} E \left\{ E_{m' \in C} (\bar{U}_{m'}^t \bar{U}_{N'Q, N'(1-Q)}) \mid m' \in C \right\} \\
&\cdot \Pr(m' \in C) \\
&= \frac{N'}{[r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]^2} E \left\{ \bar{U}_{N'(Q+\bar{e}), N'(1-Q-\bar{e})}^t \bar{U}_{N'Q, N'(1-Q)} \right\} \\
&+ o(M^{-2+4p+\beta}) ,
\end{aligned}$$

by Lemma 4.2, where \bar{e} denotes some value in the interval $(-M^{-P}, M^{-P})$. Notice,

$$\begin{aligned}
&E[\bar{U}_{N'(Q+\bar{e}), N'(1-Q-\bar{e})}^t \bar{U}_{N'Q, N'(1-Q)}] \\
&= \binom{N'(Q+\bar{e})}{r}^{-1} \binom{N'(1-Q-\bar{e})}{r}^{-1} \binom{N'Q}{r}^{-1} \binom{N'(1-Q)}{r}^{-1} \\
&\cdot \Sigma_1 \Sigma_2 \Sigma_3 \Sigma_4 E[\phi'(\bar{X}_{1_r}; \bar{Y}_{j_r}) \phi'(\bar{X}_{k_r}; \bar{Y}_{t_r})],
\end{aligned}$$

where $\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$ are sums over all sets of integers,

$$M + 1 \leq i_1 < \dots < i_r \leq N^s(Q + \bar{e})$$

$$M + 1 \leq j_1 < \dots < j_r \leq N^s(1 - Q - \bar{e})$$

$$M + 1 \leq k_1 < \dots < k_r \leq N^s Q$$

$$M + 1 \leq t_1 < \dots < t_r \leq N^s(1 - Q), \text{ respectively.}$$

The expectation of $\phi^s(\bar{X}_{i_r}; \bar{Y}_{j_r}) \phi^s(\bar{X}_{k_r}; \bar{Y}_{t_r})$ is zero when the sets $(i_1, \dots, i_r), (k_1, \dots, k_r)$ have no integer in common and the sets $(j_1, \dots, j_r), (t_1, \dots, t_r)$ have no integer in common. On the other hand, the expectation of it will become b_{cd} if there are c common integers in the former pair of sets and d common integers in the latter pair of sets. Therefore, the number of sets having (c, d) integers in common are: for all \bar{e} non-negative,

$$(7.2) \quad \binom{r}{c} \binom{N^s Q}{r} \binom{N^s(Q + \bar{e}) - c}{r - c} \binom{r}{d} \binom{N^s(1 - Q - \bar{e})}{r} \binom{N^s(1 - Q) - d}{r - d}$$

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

$$(7.3) \quad \binom{r}{c} \binom{N^s(Q + \bar{e})}{r} \binom{N^s Q - c}{r - c} \binom{r}{d} \binom{N^s(1 - Q)}{r} \binom{N^s(1 - Q - \bar{e}) - d}{r - d}.$$

Note that for \bar{e} identically zero, $Y^s \equiv Y^*$.

Consider (7.2), which, for N^s large, is

$$(7.4) \quad \frac{1}{c! [(r-c)!]^2} (N^s Q)^r [N^s(Q + \bar{e}) - c]^{r-c} \frac{1}{d! [(r-d)!]^2} \\ [N^s(1 - Q - \bar{e})]^r [N^s(1 - Q) - d]^{r-d} + \text{lower order terms} \\ = \frac{1}{c! d! [(r-c)! (r-d)!]^2} (N^s)^{4r-c-d} Q^r (Q + \bar{e})^{r-c} (1 - Q - \bar{e})^r (1 - Q)^{r-d} \\ + o[(N^s)^{4r-c-d}].$$

On the other hand, the coefficient before the summation sign is

$$(7.5) \quad \binom{N^i(Q+\bar{e})}{r}^{-1} \binom{N^i(1-Q-\bar{e})}{r}^{-1} \binom{N^i Q}{r}^{-1} \binom{N^i(1-Q)}{r}^{-1} \\ = (N^i)^{-4r} (1-Q-\bar{e})^{-r} (Q+\bar{e})^{-r} (1-Q)^{-r} Q^{-r} (r!)^4 + o[(N^i)^{-4r}] .$$

Combining (7.4), (7.5) and the above discussion, one has

$$E(\bar{U}_{N^i(Q+\bar{e}), N^i(1-Q-\bar{e})} \bar{U}_{N^i Q, N^i(1-Q)}) \\ = \sum_{\substack{c=0 \\ (c,d) \neq (0,0)}}^r \sum_{d=0}^r \left[\frac{1}{c!d![(r-c)!(r-d)!]^2} (N^i)^{4r-c-d} Q^r (1-Q)^{r-d} \right. \\ \left. (Q+\bar{e})^{r-c} (1-Q-\bar{e})^r + o(N^i)^{4r-c-d} \right] \\ \cdot [(N^i)^{4r} (1-Q-\bar{e})^r (Q+\bar{e})^r (1-Q)^r Q^r (r!)^{-4} + o(N^i)^{4r}]^{-1} \\ = \sum_{\substack{c=0 \\ (c,d) \neq (0,0)}}^r \sum_{d=0}^r (r!)^4 b_{cd} \left\{ (N^i)^{c+d} (Q+\bar{e})^c (1-Q)^d c!d! [(r-c)!(r-d)!]^2 \right\}^{-1} .$$

It is seen that for smaller $c + d$, the term is large, and as $c + d$ becomes large the term becomes small in order of magnitude. Retaining the largest order of magnitude terms $(c, d) = (1, 0)$ and $(c, d) = (0, 1)$, one has,

$$E(\bar{U}_{N^i(Q+\bar{e}), N^i(1-Q-\bar{e})} \bar{U}_{N^i Q, N^i(1-Q)}) \\ = r^2 b_{10} N^{i-1} (Q+\bar{e})^{-1} + r^2 b_{01} N^{i-1} (1-Q)^{-1} \\ = r^2 b_{10} (N^i)^{-1} Q^{-1} [1 + \bar{e}/Q]^{-1} + r^2 b_{01} [N^i(1-Q)]^{-1} + o(N^{i-1}) \\ = r^2 b_{10} (N^i)^{-1} Q^{-1} [1 + o(\bar{e})] + r^2 b_{01} [N^i(1-Q)]^{-1} + o(N^{i-1}) \\ = (N^i)^{-1} [r(b_{01})^{\frac{1}{2}} + r(b_{10})^{\frac{1}{2}}]^2 + o(N^{i-1} M^{-P}),$$

where $\bar{e} \rightarrow 0$ slowest for \bar{e} near M^{-P} .

With essentially similar steps, one will find the same result if \bar{e} is non-positive. Thus

$$1 \geq E(YY^*) \geq \frac{N^i}{[r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]^2} \left\{ \frac{[r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]^2}{N^i} + o(N^{i-1}M^{-p}) \right\} + o(M^{-2+4p+\beta})$$

$$\text{or } 1 \geq E(YY^*) \geq 1 + o(M^{-p}) + o(M^{-2+4p+\beta}) .$$

Therefore,

$$\begin{aligned} \text{Limit}_{N^i \rightarrow \infty} E(Y^i - Y^*)^2 &= \text{Limit}_{N^i \rightarrow \infty} [2 - 2 + o(N^{-p/\beta}) \\ &\quad + o(N^{-\frac{2+4p+\beta}{\beta}})] = 0 . \end{aligned}$$

Theorem 7.1 is proved.

Corollary 7.1 Y^i is also asymptotically normally distributed with mean zero and variance one, or, U^i is asymptotically normally distributed.

Next, in the expression for Y^i defined above, if one substitutes the estimated variance in terms of the values of T_{10} and T_{01} , in the variance of U^i (in terms of the values of b_{10} and b_{01}), since T_{10} and T_{01} are efficient estimators of b_{10} and b_{01} , the resulting standardized random variable Y_s^i is also asymptotically normally distributed with mean zero and variance one.

Theorem 7.2 $Y_s^i = (U^i - \theta) / \{N^{i-\frac{1}{2}}[r(T_{10})^{\frac{1}{2}} + r(T_{01})^{\frac{1}{2}}]\}$ is asymptotically equivalent to Y^i .

Proof: It suffices to show that $N^{i-\frac{1}{2}}(r(T_{10})^{\frac{1}{2}} + r(T_{01})^{\frac{1}{2}})$ is asymptotically equivalent to $N^{i-\frac{1}{2}}(r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}})$ (see [6], Theorem 5 and applications), i.e.,

$$\text{P-limit}_{N^i \rightarrow \infty} [N^{i-\frac{1}{2}} (r(T_{10})^{\frac{1}{2}} + r(T_{01})^{\frac{1}{2}}) - N^{i-\frac{1}{2}} (r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}})] = 0.$$

In other words, it is equivalent to show that for any $\epsilon^i > 0$,

$$\text{Limit}_{N^i \rightarrow \infty} \Pr \left\{ |r(T_{10})^{\frac{1}{2}} + r(T_{01})^{\frac{1}{2}} - r(b_{10})^{\frac{1}{2}} - r(b_{01})^{\frac{1}{2}}| > \epsilon^i N^{i\frac{1}{2}} \right\} = 0.$$

Now,

$$\begin{aligned} & \Pr \left\{ |r(T_{10})^{\frac{1}{2}} + r(T_{01})^{\frac{1}{2}} - r(b_{10})^{\frac{1}{2}} - r(b_{01})^{\frac{1}{2}}| > \epsilon^i N^{i\frac{1}{2}} \right\} \\ & \leq \Pr \left\{ |r(T_{10})^{\frac{1}{2}} - r(b_{10})^{\frac{1}{2}}| + |r(T_{01})^{\frac{1}{2}} - r(b_{01})^{\frac{1}{2}}| > \epsilon^i N^{i\frac{1}{2}} \right\} \\ (7.6) \quad & \leq \Pr[2|r(T_{10})^{\frac{1}{2}} - r(b_{10})^{\frac{1}{2}}| > \epsilon^i N^{i\frac{1}{2}}] \\ & \quad + \Pr[2|r(T_{01})^{\frac{1}{2}} - r(b_{01})^{\frac{1}{2}}| > \epsilon^i N^{i\frac{1}{2}}]. \end{aligned}$$

Using Tchebycheff's inequality,

$$\leq \frac{E \left\{ 2[r(T_{10})^{\frac{1}{2}} - r(b_{10})^{\frac{1}{2}}] \right\}^2}{N^i \epsilon^{i2}} + \frac{E \left\{ 2[r(T_{01})^{\frac{1}{2}} - r(b_{01})^{\frac{1}{2}}] \right\}^2}{N^i \epsilon^{i2}}.$$

Applying the identity (see [1] p. 353),

$$a^{\frac{1}{2}} - b^{\frac{1}{2}} = \frac{a - b}{2(b)^{\frac{1}{2}}} - \frac{(a - b)^2}{2(b)^{\frac{1}{2}} [a^{\frac{1}{2}} + b^{\frac{1}{2}}]^2}$$

one has

$$\begin{aligned} & E \left\{ 2[r(T_{10})^{\frac{1}{2}} - r(b_{10})^{\frac{1}{2}}] \right\}^2 \\ & = 4r^2 \left\{ E \frac{(T_{10} - b_{10})^2}{4 b_{10}} - 2E \frac{(T_{10} - b_{10})(T_{10} - b_{10})^2}{4b_{10} [(T_{10})^{\frac{1}{2}} + (b_{10})^{\frac{1}{2}}]^2} \right. \\ & \quad \left. + E \frac{(T_{10} - b_{10})^4}{4b_{10} [(T_{10})^{\frac{1}{2}} + (b_{10})^{\frac{1}{2}}]^4} \right\} \\ & \leq r^2 \left\{ E \frac{(T_{10} - b_{10})^2}{b_{10}} + 2 \left| E \left[\frac{(T_{10} - b_{10})^3}{b_{10}^2} \right] \right| + E \frac{(T_{10} - b_{10})^4}{b_{10}^3} \right\}. \end{aligned}$$

By Lemma 4.2, and by the fact that it is also easy to extend it to the case for odd moments of a U-statistic, i.e., $E(U-\theta)^{2k-1} = o(M^{-k})$, one has

$$E\left(\frac{(T_{10} - b_{10})^2}{b_{10}}\right) = o(M^{-1})$$

$$E\left(\frac{(T_{10} - b_{10})^3}{b_{10}^2}\right) = o(M^{-2})$$

$$E\left(\frac{(T_{10} - b_{10})^4}{b_{10}^3}\right) = o(M^{-2}) .$$

Consequently,

$$(7.7) \quad \frac{E\left\{2[r(T_{10})^{\frac{1}{2}} - r(b_{10})^{\frac{1}{2}}]\right\}^2}{N^i e^{i^2}} = o(N^i M e^{i^2})^{-1} \longrightarrow 0$$

with suitable choice of e^i , as $N^i \longrightarrow \infty$.

With exactly the same kind of argument, one has

$$(7.8) \quad \frac{E\left\{2[r(T_{01})^{\frac{1}{2}} - r(b_{01})^{\frac{1}{2}}]\right\}^2}{N^i e^{i^2}} \longrightarrow 0 \text{ as } N^i \longrightarrow \infty .$$

Combining (7.7), (7.8) and putting them into (7.6), one has,

$$\begin{aligned} \text{Limit}_{N^i \rightarrow \infty} \Pr[|r(T_{10})^{\frac{1}{2}} + r(T_{01})^{\frac{1}{2}} - r(b_{10})^{\frac{1}{2}} - r(b_{01})^{\frac{1}{2}}| > e^i N^{i\frac{1}{2}}] \\ \leq \text{Limit}_{N^i \rightarrow \infty} o(N^i M e^{i^2})^{-1} = 0 . \end{aligned}$$

This proves the asymptotic equivalence of $N^{i-\frac{1}{2}}[r(T_{10})^{\frac{1}{2}} + r(T_{01})^{\frac{1}{2}}]$ and $N^{i\frac{1}{2}}[r(b_{10})^{\frac{1}{2}} + r(b_{01})^{\frac{1}{2}}]$ and hence the asymptotic equivalence of Y'_S to Y^i .

Corollary 7.2 Y_s^i is asymptotically normally distributed.

8. Extension of the Two-stage Technique to Functions of More Than Two Populations

So far, the problem of two-stage estimation has been studied with estimable functions (or parameters) of two populations.

There is apparently a possibility of extending it to functions of more than two populations.

Let $X^{(1)}, \dots, X^{(k)}$ be k populations ($k \geq 2$) with cumulative distribution functions $F_1(X), \dots, F_k(X)$ respectively. Also let $\theta = \theta(F_1, \dots, F_k)$ be the functional to be estimated, with symmetric kernel $\phi(\bar{X}^{(1)}, \dots, \bar{X}^{(k)})$ where $\bar{X}^{(j)}$ represent vectors of dimension r , i.e., $\bar{X}^{(j)}$ are r independent observations on population $X^{(j)}$.

The corresponding U-statistic with n_j observations on $X^{(j)}$ will be

$$U_{n_1, \dots, n_k} = U_k = \left[\prod_{i=1}^k \binom{n_i}{r} \right]^{-1} \sum_k \phi(\bar{X}_{j_1}^{(1)}, \dots, \bar{X}_{j_k}^{(k)}),$$

where \sum_k is the k -fold summation over a set of integers such that for each vector of integers, $j_i, i = 1, 2, \dots, k, 1 \leq j_{i_1} < \dots < j_{i_r} \leq n_i$. Analogously, define

$$\phi^i(\bar{X}^{(1)}; \dots; \bar{X}^{(k)}) = \phi(\bar{X}^{(1)}; \dots; \bar{X}^{(k)}) - \theta$$

$$\phi^i_{a_1, \dots, a_k}(\bar{x}^{(1)}; \dots; \bar{x}^{(k)}) = E\phi^i(\bar{x}^{(1)} \bar{x}^{(1)}; \dots; \bar{x}^{(k)} \bar{x}^{(k)}),$$

where $\bar{x}^{(j)}$ represent vectors of dimension $a_j, a_j = 0, 1, \dots, r$ for all $j = 1, 2, \dots, k; \bar{X}^{(j)}$ represent vectors of dimension $r - a_j, (r - a_j) = (a_{j+1}, \dots, r)$ for all $j = 1, 2, \dots, k$.

$$b_{a_1, \dots, a_k} = E[\phi_{a_1, \dots, a_k}(\bar{X}^{(1)}, \dots, \bar{X}^{(k)})]^2$$

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

Then when the variance of ϕ^i exists, one can write

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

Now, when a fixed total sample size N is given $\sum_{i=1}^k n_i = N$, and if $n_i \rightarrow \infty$ in such a way that n_i/n_j are bounded away from zero and one, for all $i \neq j$, $i, j = 1, 2, \dots, k$. Then the asymptotic expression for $\text{Var}(U_k)$ is

$$\text{Var}(U_k) \leq \sum_{i=1}^k \frac{r^2}{n_i} b^{(i)} = V^i, \quad \text{say}$$

where $b^{(i)} = b_{\dots, 1, \dots, 0, \dots}$ with one of the subscript 1 at the i^{th} position and zero elsewhere.

Analogous to the case of $k = 2$, it is easy to show that V^i is minimized when

$$n_i = (b^{(i)})^{\frac{1}{2}} / \sum_{i=1}^k (b^{(i)})^{\frac{1}{2}}$$

and the minimized value of V^i is V_0 ,

$$V_0 = N^{-1} \left[\sum_{i=1}^k r (b^{(i)})^{\frac{1}{2}} \right]^2$$

when $b^{(i)}$, $i = 1, 2, \dots, k$, are all known.

The two-stage estimating procedure will be as follows:

(a) Take M observations on each of the $X^{(i)}$, $i = 1, 2, \dots, k$,

where $2rk \leq kM < N - rk$.

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

(c) Take m_i more observations on $X^{(i)}$, $i = 1, 2, \dots, k$, such that

$$m_i = N'(T^{(i)})^{\frac{1}{2}} / \sum_{i=1}^k (T^{(i)})^{\frac{1}{2}} \quad \text{where } N' = N - kM$$

for estimator using only the second stage observations.

(d) Use U'_k , the analogous two-stage estimator of U' , to estimate θ .

It can be shown, using essentially the same arguments and under the same kind of conditions as in Theorem 4.1, but replacing the condition (iii) by (iii')

$$b^{(i)} > 0, \quad i = 1, 2, \dots, k;$$

that the analogous result can be obtained. Also, the asymptotic distribution of U'_k is again normal.

9. Summary of Results on U''

In this section, results on U'' will be summarized without proof. U'' is introduced in order to utilize the data from the first-stage samples as well as the second-stage samples. (Recall that U' is constructed based on the second-stage samples only.)

The first stage for U'' is defined exactly the same as for U' (see Section 3). However, in the second stage, m'' (not m') more observations from population X and n'' (not n') more observations from population Y will be taken such that $m'' + n'' = N' = N - 2M$. The sample sizes m'' and n'' are determined as the following:

$$\begin{aligned}
 m'' &= [NZ] - M && \text{when } (M + 1)/N \leq Z \leq (N-M)/N \\
 (9.1) \quad m'' &= 0 && \text{when } Z < (M + 1)/N \\
 m'' &= N' && \text{when } Z > (N - M)/N \\
 \text{and} \quad n'' &= N' - m'' .
 \end{aligned}$$

The statistic U'' is defined as the estimator of θ based on $M + m''$ and $M + n''$ observations on populations X and Y respectively. The definition is:

$$(9.2) \quad U'' = \binom{M+m''}{r}^{-1} \binom{M+n''}{r}^{-1} \sum \phi(\bar{X}_{i_r}; \bar{Y}_{j_r})$$

where the summation is taken over sets of all integers,

$$1 \leq i_1 < \dots < i_r \leq M + m''; \quad 1 \leq j_1 < \dots < j_r \leq M + n''.$$

The statistic U'' is biased. However, by the help of two Lemmas, the following theorem is proved.

Theorem 9.1 If the conditions of Theorem 4.1 are satisfied, then $E(U'') = \theta + O(M^{-1+2p} + M^{\frac{1}{2}-\beta})$, and that

$$\lim_{N \rightarrow \infty} E(U'' - \theta)^2 / V_0 = 1 .$$

Next, it is found that there is no "optimal" choice of M relative to N so that the ratio $E(U'' - \theta)^2 / V_0$ converges to unity as quickly as possible. However, it is also found that in any case, this ratio converges to unity not slower than the ratio $\text{Var}(U') / V_0$, if the same set of values of β and p are used.

Again, although U'' is biased, it can be shown that, using similar techniques as for the case of U' , U'' is also asymptotically normally distributed, either in terms of its variance or in terms

of the estimated variance, i.e., with b_{10} (b_{01}) replaced by T_{10} (T_{01}). Finally, using analogous steps as for U' , the results on U'' are generalized to cover the situation of sampling from several populations.

REFERENCES

1. Cramér, H.: "Mathematical Methods of Statistics", Princeton University Press, 1946.
2. Fraser, D. A. S.: "Non-parametric Methods in Statistics", Wiley, 1957.
3. Ghurye, S. G. and Robbins, Herbert.: "Two-stage procedure for estimating the difference between means", *Biometrika*, 1954, (Vol. 47), 146-152.
4. Hoeffding, W.: "A class of statistics with asymptotically normal distributions", *Annals of Mathematical Statistics*, 1948, (Vol. 19), 293-325.
5. Hoeffding, W.: "Probability inequalities for sums of bounded random variables", *Journal of the American Statistical Association*, 1963, (Vol. 58), 13-30.
6. Mann, H. B. and Wald, A.: "On stochastic limit and order relationships", *Annals of Mathematical Statistics*, 1945, (Vol. 14), 217-226.
7. Putter, Joseph.: "Sur une méthode de double échantillonnage pour estimer la moyenne d'une population laplacienne stratifiée", *La Revue de L'Institut International de Statistique*, 1951: 3, 1-8.
8. Richter, Donald L.: "Two-stage experiments for estimating a common mean", *Institute of Statistics Mimeograph Series No. 231*, University of North Carolina, June 1959.
9. Robbins, H.: "Some aspect of the sequential design of experiments", *American Mathematical Society Bulletin*, 1952, (Vol. 58), 527-535.
10. Rosenblatt, Joan R.: "On a class of non-parametric tests", *Institute of Statistics Mimeograph Series No. 138*, University of North Carolina, August 1955.
11. Stein, Charles,: "A two-sample test for a linear hypothesis whose power is independent of the variance", *Annals of Mathematical Statistics*, 1945, (Vol. 16), 243-258.
12. Tchouproff, Al A.: "On the mathematical expectation of the moments of frequency distributions", *Biometrika*, 1919, (Vol. 12), 140-185.