ESTIMATION OF SOME FUNCTIONAL OF THE POPULATION DISTRIBUTION BASED ON A STRATIFIED RANDOM SAMPLE

TAKASHI YANAGAWA AND KAZUMASA WAKIMOTO

(Received April 21, 1971)

Summary

In the preceding papers ([7], [8] and [9]), one of the authors discussed about the estimation of variances, covariances and correlation coefficients of the population based on a stratified random sample. In this paper we consider more general problem; estimating some functional $\theta(F)$ of the population distribution F based on a stratified random sample, which include our previous papers as special cases. We propose an unbiased estimator of $\theta(F)$ based on a stratified random sample and give an asymptotic expression of the gain in precision due to stratification in the case of proportional allocation. Furthermore, we present the general form of the optimum stratification in the proportional allocation for the estimation of $\theta(F)$.

1. One sample case

1.1 *Estimators of population characteristic*

Let p-dimensional vectors $X_i=(X_i^{(1)},\cdots,X_i^{(p)})$, $i=1,2,\cdots,N$, be a random sample from the population Π with p-variate distribution function $F(x)$, and let $\theta(F)$ be a univariate population characteristic of Π. Suppose $\theta(F)$ is a regular (estimable) functional, that is, there exists a kernel $\Phi(x_1,\cdots,x_m)$ symmetric in $m(\leqq N)$ vectors $x_\gamma=(x_\gamma^{(1)},\cdots,x_\gamma^{(p)})$, $\gamma=1,\cdots,m$ such that

$$(1.1)\qquad \theta(F)=\int\Phi(x_1,\cdots,x_m)dF(x_1),\cdots,dF(x_m)\ .$$

The statistic of the form

$$(1.2)\qquad U=\Sigma'\Phi(X_{\alpha_1},\cdots,X_{\alpha_m})\Big/\binom{N}{m}\ ,$$

where the sum Σ' is extended over all combinations $(\alpha_1,\cdots,\alpha_m)$ of m different integers $1\leqq\alpha_1<\alpha_2<\cdots<\alpha_m\leqq N$, is called a U statistic. U is easily shown to be an unbiased estimator of $\theta(F)$.

Now suppose that the population Π with the distribution function F is classified into l strata $\{\Pi_1,\cdots,\Pi_l\}$, which may be overlapping, in the following way: the distribution function $F_i(x)$ corresponds to the ith stratum Π_i and these F_i satisfy the relation

$$(1.3)\qquad F(x)=\sum_{i=1}^{l} w_i F_i(x) \qquad \text{for all } x\ ,$$

where w_i is a weight of F_i such that $0\leqq w_i\leqq 1$, $\sum_{i=1}^{l} w_i=1$. We shall call $\{F_i(x)\}$ "l-decomposition of $F(x)$." If for each $i=1,2,\cdots,l$, we have a sampling plan to take a random sample of size n_i from the ith stratum Π_i, then as an estimator of $\theta(F)$ we can apply the following U^*.

$$(1.4)\qquad U^*=\Sigma^*\binom{m}{r_1,r_2,\cdots,r_l}\prod_{k=1}^{l} w_k^{r_k} U^*_{1(r_1)2(r_2)\cdots l(r_l)}\ ,$$

where

$$\binom{m}{r_1,r_2,\cdots,r_l}=\frac{m!}{r_1!r_2!\cdots r_l!}\ ,$$

$$U^*_{1(r_1)2(r_2)\cdots l(r_l)}=\Sigma^{**}\Phi(X_{1\alpha_{11}},\cdots,X_{1\alpha_{1r_1}},\cdots,X_{l\alpha_{l1}},\cdots,X_{l\alpha_{lr_l}})\Big/\prod_{k=1}^{l}\binom{n_k}{r_k}\ ,$$

the sum Σ^* is extended over all combinations $(r_1,r_2,\cdots,r_l)$ such that $r_i\geqq 0$ $r_1+r_2+\cdots+r_l=m$, the sum Σ^{**} is extended over all combinations $(\alpha_{i1},\cdots,\alpha_{ir_i})$, $1\leqq\alpha_{i1}<\cdots<\alpha_{ir_i}\leqq n_i$, $i=1,2,\cdots,l$, and $X_{i\alpha}=(X_{i\alpha}^{(1)},\cdots,X_{i\alpha}^{(p)})$, $\alpha=1,2,\cdots,n_i$, is a random sample drawn from Π_i, $i=1,2,\cdots,l$.

Since

$$\begin{aligned}U^*&=\Sigma^*\binom{m}{r_1,\cdots,r_l}\prod_{k=1}^{l} w_k^{r_k}\prod_{j=1}^{l}\binom{n_j-r_j}{m-r_j}r_j!(m-r_j)!\\&\quad\times\Sigma^{**}\Phi(X_{1\alpha_{11}},\cdots,X_{1\alpha_{1r_1}},\cdots,X_{l\alpha_{l1}},\cdots,X_{l\alpha_{lr_l}})\\&=\Sigma^*\binom{m}{r_1,\cdots,r_l}\prod_{k=1}^{l} w_k^{r_k}\\&\quad\times\Sigma''\Phi_{r,_l\cdots,r_1}(X_{1\alpha_{11}},X_{1\alpha_{1m}},\cdots,X_{l\alpha_{l1}},\cdots,X_{l\alpha_{lm}})\Big/\prod_{j=1}^{l}{}_{n_j}P_m\ ,\end{aligned}$$

where $\Phi_{r_1\cdots r_l}(X_{11},\cdots,X_{1m},\cdots,X_{l1},\cdots,X_{lm})=\Phi(X_{11},\cdots,X_{1r_1},\cdots,X_{l1},\cdots,X_{lr_l})$ and the sum Σ'' is extended over all permutations $(\alpha_{i1},\cdots,\alpha_{im})$: $1\leqq\alpha_{iu}\leqq n_i$, $u=1,\cdots,n_i$, $i=1,\cdots,l$, then U^* can be rewritten as a generalized U statistic as follows.

$$U^*=\Sigma''\Phi^*(X_{1\alpha_{11}},\cdots,X_{1\alpha_{1m}},\cdots,X_{l\alpha_{l1}},\cdots,X_{l\alpha_{lm}})\Big/\prod_{k=1}^{l}{}_{n_k}P_m\ ,$$

where

$$\Phi^*(x_{11},\cdots,x_{1m},\cdots,x_{l1},\cdots,x_{lm})$$
$$=\Sigma^*\binom{m}{r_1,\cdots,r_l}\prod_{k=1}^{l} w_k^{r_k}\Phi_{r_1\cdots r_l}(x_{11},\cdots,x_{1m},\cdots,x_{l1},\cdots,x_{lm}) .$$

LEMMA 1.1. *Suppose the relation* (1.3) *hold. Then U^* is an unbiased estimator of $\theta(F)$.*

PROOF. Substituting (1.3) into (1.1), we can get

$$\theta(F)=\Sigma^*\binom{m}{r_1,\cdots,r_l}\prod_{k=1}^{l} w_k^{r_k}\theta_{1(r_1)\cdots l(r_l)}(F) , \tag{1.5}$$

where

$$\theta_{1(r_1)\cdots l(r_l)}(F)=\int \Phi(x_{11},\cdots,x_{1r_1},\cdots,x_{l1},\cdots,x_{lr_l})\prod_{j=1}^{l}\prod_{i=1}^{r_j} dF_j(x_{ji}) .$$

Since $U^*_{1(r_1)\cdots l(r_l)}$ is an unbiased estimator of $\theta_{1(r_1)\cdots l(r_l)}(F)$, from (1.4) and (1.5) U^* is easily shown to be an unbiased estimator of $\theta(F)$.

Remark. By putting $m=1$ and $\Phi(x)=x$, we can solve the problem of estimation of the population mean. Also, by putting $m=2$ and $\Phi(x_1, x_2)=(x_1^{(1)}-x_2^{(1)})^2/2$, $(x_1^{(1)}-x_2^{(1)})(x_1^{(2)}-x_2^{(2)})/2$ where $x_1=(x_1^{(1)}, x_1^{(2)})$, $x_2=(x_2^{(1)}, x_2^{(2)})$, we can solve the problem of estimation of the population variance and covariance, respectively.

1.2 *Comparison of the estimators*

In this section we shall compare two estimators U and U^* of $\theta(F)$ defined in the preceding section in the case of proportionate allocation, and demonstrate an effect of stratification.

Suppose the total sample size N is fixed, and it is allocated to each stratum proportionately to each size, i.e. $n_i=w_iN$, $(i=1,\cdots,l)$, where w_i is the weight of F_i given in (1.3).

ASSUMPTION 1.1. $E\{\Phi^2(X_1,\cdots,X_m)\}<\infty$.

Under the above assumption, the variances of U and U^* exist. Let

$$V_\infty[U]=\lim_{N\to\infty}\mathrm{Var}\,[\sqrt{N}(U-\theta(F)] ,$$
$$V_\infty[U^*]=\lim_{N\to\infty}\mathrm{Var}\,[\sqrt{N}(U^*-\theta(F))]$$

and

$$\Phi_1(x)=\begin{cases}\int \Phi(x_1,\cdots,x_m)\prod_{i=2}^{m} dF(x_i) , & \text{if } m\geqq 2 ,\\ \Phi(x) , & \text{if } m=1 .\end{cases} \tag{1.6}$$

LEMMA 1.2. *Under the Assumption* 1.1,

(i) $\sqrt{N}[U-\theta(F)]$ *has, as* $N\to\infty$, *a normal limiting distribution with mean* 0 *and variance*

$$V_\infty[U]=m^2\left\{\int \Phi_1^2(x)dF(x)-\theta^2(F)\right\} .$$

(ii) $\sqrt{N}[U^*-\theta(F)]$ *has, as* $N\to\infty$, *a normal limiting distribution with mean* 0 *and variance*

$$V_\infty[U^*]=m^2\left\{\int \Phi_1^2(x)dF(x)-\sum_{i=1}^{l} w_i\left[\int \Phi_1(x)dF_i(x)\right]^2\right\} .$$

PROOF.

(i) It is well known that the U has a normal limiting distribution, its variance is given by

$$m^2 \operatorname{Cov}[\varphi(X_1, X_2,\cdots, X_m), \varphi(X_1, X_2',\cdots, X_m')]=m^2\left\{\int \Phi_1^2(x)dF(x)-\theta^2(F)\right\} ,$$

and the variance coincides with $V_\infty[U]$ (see Hoeffding [2]). So we shall omit the proof of (i).

(ii) The asymptotic normality of U^* is obtained from the fact that U^* can be written as a form of generalized U statistic (see the previous section). But here we shall give the direct proof. Put $a_i=w_i^m$,

$$b(r)=\binom{m}{r_1,\cdots, r_l}\prod_{k=1}^{l} w_k^{r_k} ,$$

$$U_{N,k}=\sqrt{N}\left[U^*_{1(0),\cdots,k-1(0),k(m),k+1(0),\cdots,l(0)}-\theta_{1(0),\cdots,k-1(0),k(m),k+1(0),\cdots,l(0)}(F)\right] ,$$

$$U_{N,1(r_1)\cdots l(r_l)}=\sqrt{N}[U^*_{1(r_1)\cdots l(r_l)}-\theta_{1(r_1)\cdots l(r_l)}(F)] ,$$

and

$$U_N^*=\sqrt{N}[U^*-\theta(F)] . \tag{1.7}$$

Then U_N^* can be written as

$$U_N^*=\sum_{k=1}^{l} a_i U_{N,k}^*+\sum_{(r)}{}^{\#} b(r)U^*_{N,1(r_1)\cdots l(r_l)} ,$$

where the sum $\sum_{(r)}^{\#}$ is extended over all combinations of $(r_1, r_2,\cdots, r_l)$ such that $0\leqq r_i\leqq m-1$, $r_1+r_2+\cdots+r_l=m$. It is noted that the statistics $\{U_{N,k}^*\}_{k=1,\cdots,l}$ are independent of each other but the statistics $\{U^*_{N,1(r_1)\cdots l(r_l)}\}$ are not independent. Put

$$A_i(x_{i\alpha})=E[\varphi(X_{11},\cdots,X_{1r_1},\cdots,x_{i\alpha},X_{i2},\cdots,X_{ir_i},\cdots,X_{l1},\cdots,X_{lr_l})]\ ,$$

$$B_i(x_{i\alpha})=E[\varphi(x_{i\alpha},X_{i2},\cdots,X_{im})]\ ,\qquad i=1,\cdots,l\ ,$$

$$C(r)=\frac{1}{\sqrt{N}}\sum_{k=1}^{l}\frac{r_k}{w_k}\sum_{\alpha=1}^{n_k}[A_k(X_{k\alpha})-\theta_{1(r_1)\cdots l(r_l)}(F)]\ ,$$

$$D_k=m\sum_{\alpha=1}^{n_k}[B_k(X_{k\alpha})-\theta_{1(0)\cdots k-1(0)\ k(m)\ k+1(0)\cdots l(0)}(F)]\Big/w_k\sqrt{N}$$

and

$$T=\sum_{i=1}^{l}a_iD_i+\sum_{(r)}{}^{\#}\,b(r)C(r)\ . \tag{1.8}$$

The same arguments as Sugiura [4] lead us to get

$$\lim_{N\to\infty}E[U^*_{N,1(r_1)\cdots l(r_l)}-C(r)]^2=0 \tag{1.9}$$

and

$$\lim_{N\to\infty}E[U^*_{N,k}-D_k]^2=0\ ,\qquad k=1,\cdots,l\ . \tag{1.10}$$

Now from (1.7) and (1.8) we get

$$\begin{aligned}E[U^*_N-T]^2=&\sum_{k=1}^{l}a_k^2E[U^*_{N,k}-D_k]^2\\&+\sum_{(r)}{}^{\#}\sum_{(s)}{}^{\#}\,b(r)b(s)E[U^*_{N,1(r_1)\cdots l(r_l)}-C(r)]\\&\times[U^*_{N,1(s_1)\cdots l(s_l)}-C(s)]\\&+2\sum_{k=1}^{l}\sum_{(r)}{}^{\#}\,a_kb(r)E[U^*_{N,k}-D_k][U^*_{N,1(r_1)\cdots l(r_l)}-C(r)]\ .\end{aligned} \tag{1.11}$$

Since

$$\begin{aligned}&|E[U^*_{N,1(r_1)\cdots l(r_l)}-C(r)][U^*_{N,1(s_1)\cdots l(s_l)}-C(s)]|\\&\qquad\leqq\{E[U^*_{N,1(r_1)\cdots l(r_l)}-C(r)]^2E[U^*_{N,1(s_1)\cdots l(s_l)}-C(s)]^2\}^{1/2}\ ,\end{aligned}$$

we get from (1.9)

$$\lim_{N\to\infty}E[U^*_{N,1(r_1)\cdots l(r_l)}-C(r)][U^*_{N,1(s_l)\cdots l(s_l)}-C(s)]=0\ . \tag{1.12}$$

Similarly from (1.9) and (1.10) we get

$$\lim_{N\to\infty}E[U^*_{N,k}-D_k][U^*_{N,1(r_1)\cdots l(r_l)}-C(r)]=0\ . \tag{1.13}$$

Thus, substituting (1.10), (1.12) and (1.13) into (1.11), we get

$$(1.14)\qquad \lim_{N\to\infty} E[U_N^* - T]^2 = 0 .$$

Hence the statistic U_N^* has the same limiting distribution as T (see Cramér [1], p. 254). Put

$$(1.15)\qquad Q_k(x_{k\alpha}) = \frac{m}{w_k} a_k B_k(x_{k\alpha}) + \sum\nolimits^{*} \frac{r_k}{w_k} b(r) A_k(x_{k\alpha}) ,$$

and

$$(1.16)\qquad R_k = \frac{m}{w_k} a_k \theta_{1(0)\cdots k-1(0)\ k(m)\ k+1(0)\cdots l(0)} + \sum\nolimits^{*} \frac{r_k}{w_k} b(r) \theta_{1(r_1)\cdots l(r_l)}(F) .$$

Then T can be rewritten as the sum of independent random variables as follows.

$$(1.17)\qquad T = \frac{1}{\sqrt{N}} \sum_{k=1}^{l} \sum_{\alpha=1}^{n_k} [Q_k(X_{k\alpha}) - R_k] .$$

Since $E(\Phi^2) < \infty$, we get $E[Q_k(X_{k\alpha}) - R_k]^2 < \infty$. Thus by the central limit theorem T has, as the limiting distribution when $N \to \infty$, the normal whose mean is 0 and whose variance is given as follows.

$$(1.18)\qquad \begin{aligned} \mathrm{Var}\,(T) &= \frac{1}{N} \sum_{k=1}^{l} \sum_{\alpha=1}^{n_k} \mathrm{Var}\,[Q_k(X_{k\alpha}) - R_k] \\ &= \frac{1}{N} \sum_{k=1}^{l} \sum_{\alpha=1}^{n_k} E[Q_k(X_{k\alpha})]^2 - \frac{1}{N} \sum_{k=1}^{l} n_k R_k^2 . \end{aligned}$$

Now since from (1.15)

$$\begin{aligned} Q_k(x_{k\alpha}) &= \frac{m}{w_k} a_k B_k(x_{k\alpha}) + \sum\nolimits^{*} \frac{r_k}{w_k} b(r) A_k(x_{k\alpha}) \\ &= \frac{m w_k^m}{w_k} E\{\Phi(x_{k\alpha}, X_{k2}, \cdots, X_{km})\} + \sum_{(r)}\nolimits^{*} \frac{r_k}{w_k} \binom{m}{r_1, \cdots, r_l} \\ &\quad \times \prod_{i=1}^{l} w_i^{r_i} E\{\Phi(X_{11}, \cdots, X_{1r_1}, \cdots, x_{k\alpha}, X_{k2}, \cdots, X_{kr_k}, \cdots, X_{l1}, \cdots, X_{lr_l})\} \\ &= \sum_{(r\cdot r)}\nolimits^{*} m \binom{m-1}{r_1, \cdots, r_{k-1}, r_k - 1, r_{k+1}, \cdots, r_l} \\ &\quad \times \prod_{i=1}^{l} \frac{w_i^{r_i}}{w_k} E\{\Phi(x_{k\alpha}, X_{11}, \cdots, X_{1r_1}, \cdots, X_{k1}, \cdots, X_{kr_k - 1}, \\ &\qquad \cdots, X_{l1}, \cdots, X_{lr_l})\} , \end{aligned}$$

where the sum $\sum_{(r\cdot r)}^{*}$ is extended over all combinations $(r_1, r_2, \cdots, r_l)$ such that $r_1 + r_2 + \cdots + r_l = m$, $1 \leq r_k \leq m$, $0 \leq r_i \leq m-1$, $i \neq k$, thus we get

$$Q_k(x_{k\alpha}) = m \int \Phi(x_{k\alpha}, y_1, \cdots, y_{m-1}) \prod_{j=1}^{m-1} dF(y_j) = m \Phi_1(x_{k\alpha}) .$$

Hence

$$\frac{1}{N}\sum_{k=1}^{l}\sum_{j=1}^{n_k} E[Q_k(X_{k\alpha})]^2 = m^2 \sum_{k=1}^{l} w_k \int \Phi_1^2(x)dF_k(x) = m^2 \int \Phi_1^2(x)dF(x) \ . \tag{1.19}$$

Similarly we get from (1.6)

$$\begin{aligned} R_k &= \sum_{(r\cdot r)}{}^{*} m \binom{m-1}{r_1,\cdots,r_{k-1},r_k-1,r_{k+1},\cdots,r_l} \\ &\quad \times \prod_{i=1}^{l} \frac{w_i^{r_i}}{w_i} E\{\Phi(X_{11},\cdots,X_{1r_1},\cdots,X_{l1},\cdots,X_{lr_l})\} \\ &= m\int \Phi(x_1,\cdots,x_m)\prod_{i=2}^{m} dF(x_i)dF_k(x) \\ &= m\int \Phi_1(x)dF_k(x) \ . \end{aligned}$$

Thus

$$\frac{1}{N}\sum_{k=1}^{l} n_k R_k^2 = m^2 \sum_{k=1}^{l} w_k \left[\int \Phi_1(x)dF_k(x)\right]^2 . \tag{1.20}$$

Hence substituting (1.19), (1.20) into (1.18) we get

$$\mathrm{Var}\,(T) = m^2\left\{\int \Phi_1^2(x)dF(x) - \sum_{k=1}^{l} w_k\left[\int \Phi_1(x)dF_k(x)\right]^2\right\} . \tag{1.21}$$

Finally we get

$$\begin{aligned} |\mathrm{Var}\,(T)-\mathrm{Var}\,(U_N^*)| &= |E(T+U_N^*)(T-U_N^*)| \\ &\leqq \{E(T+U_N^*)^2E(T-U_N^*)^2\}^{1/2} . \end{aligned} \tag{1.22}$$

Since $E(\Phi^2)<\infty$, there exists some constant $M(<\infty)$ independent of N such that $E(T+U_N^*)^2<M$. Thus from (1.14) and (1.22) we get

$$V_\infty[U^*] = \lim_{N\to\infty} \mathrm{Var}\,(U_N^*) = \mathrm{Var}\,(T) \ . \tag{1.23}$$

Hence combining (1.21) and (1.23) we complete the proof.

THEOREM 1. *Under the Assumption* 1.1, *we get*

$$V_\infty[U]-V_\infty[U^*] = m^2\sum_{i=1}^{l} w_i\left[\int \Phi_1(x)dF_i(x)-\theta(F)\right]^2 .$$

Hence for any underlying distribution F the limiting variance of the estimator U^* is smaller than that of U.

PROOF. From Lemma 1.2,

$$V_\infty[U]-V_\infty[U^*] = m^2\left\{\sum_{i=1}^{l} w_i\left[\int \Phi_1(x)dF_i(x)\right]^2-\theta^2(F)\right\}$$

$$=m^2\sum_{i=1}^{l} w_i\left[\int \Phi_1(x)dF_i(x)-\theta(F)\right]^2 ,$$

thus we get the theorem.

Remark. From the above theorem, it is seen that the precision of the estimator of $\theta(F)$ is asymptotically higher in the case of proportional allocation than in the case of simple random sampling.

1.3 *Optimum stratification*

Let $\{F_i(x)\}$ be an l-decomposition of the distribution function $F(x)$. $F_i(x)$ is the distribution function and satisfy the relation

$$F(x)=\sum_{i=1}^{l} w_i F_i(x) , \qquad \text{for all } x . \tag{1.24}$$

In this section we shall show the existence of an optimum l-decomposition $\{F_i^*\}$, i.e. l-decomposition which satisfy (1.24) and attains the infimum value of $V_\infty[U^*]$ for fixed F, N and l. Applying the same technique as Taga [5], and Isii and Taga [6] we can solve the problem of this section as follows. It is easily seen from (1.24) that the measure dF_i is absolutely continuous with respect to the measure dF. Then there exists a measurable function vector $\chi=(\chi_1,\cdots,\chi_l)$ such that

$$w_i dF_i(x)=\chi_i(x)dF(x) , \qquad i=1,\cdots,l \tag{1.25}$$

and

$$\sum_{i=1}^{l}\chi_i(x)=1 , \qquad 0\leqq\chi_i(x)\leqq 1 , \qquad \text{a.e. } (dF) . \tag{1.26}$$

Conversely, let us define $\{F_i(x)\}$ by (1.25) for a given $\chi=(\chi_1,\cdots,\chi_l)$ satisfying (1.26). Then it is easily seen that $\{F_i(x)\}$ satisfies (1.24), and that $\{F_i(x)\}$ becomes an l-decomposition of $F(x)$. Let two measurable function vectors be identified if they coincide elementwise except on sets of dF-measure zero. Then, there is a one to one correspondence between the set of all measurable function vectors χ which satisfy (1.25) and (1.26) (we shall denote it by $\mathscr{H}$) and the set of the all decomposition of F. Using (1.25) and (1.26), $V_\infty(U^*)$ given in the Lemma 1.2 can be rewritten as follows.

$$V_\infty[U^*|\chi]=V_\infty[U^*]=m^2\left\{\int \Phi_1^2(x)dF(x)-\sum_{i=1}^{l}\frac{\left[\int \Phi_1(x)\chi_i(x)dF(x)\right]^2}{\int \chi_i(x)dF(x)}\right\} . \tag{1.27}$$

Now we shall show the existence of $\chi\in\mathscr{H}$ which attains the infimum of $V_\infty[U^*|\chi]$. Put

$$w_i = w_i(\chi) = \int \chi_i(x)dF(x) \ ,$$

$$u_i = u_i(\chi) = \int \Phi_1(x)\chi_i(x)dF(x) \ ,$$

$$V_\infty[U^* | \chi] = G(u_1, \cdots, u_k, w_1, \cdots, w_k) = \int \Phi^2(x)dF(x) - \frac{\sum_{i=1}^{l} u_i^2}{w_i} \ ,$$

$$\mathcal{U} = \left\{ u ; \ u = \int \Phi_1(x)\chi(x)dF(x), \ \chi \in \mathcal{H} \right\} \ ,$$

$$\mathcal{W} = \left\{ w ; \ w = \int \chi(x)dF(x), \ \chi \in \mathcal{H} \right\}$$

and

$$C = \left\{ (u, w) ; \ u = \int \Phi_1(x)\chi(x)dF(x), \ w = \int \chi(x)dF(x), \ \chi \in \mathcal{H} \right\} \ .$$

We shall state the following lemmas which are given by Taga [5] and Isii and Taga [6] without proof.

LEMMA 1.3.

(i) $\mathcal{H}, \mathcal{U}, \mathcal{W}, C$ *are all convex and compact where the topology is defined by the weak convergence (for example, see Lehmann* [3], *Appendix).*
(ii) $G(u_1, \cdots, u_k, w_1, \cdots, w_k)$ *is a continuous and concave mapping from* C *into* R^1.

From the above Lemma (i) we can easily get the existence of $\chi \in \mathcal{H}$ which attains the infimum of $V_\infty[U^* | \chi]$. It should be noted that such a χ is not always uniquely determined. Let us denote by $\mathcal{H}^*$ the set of all $\chi \in \mathcal{H}$ which attains the infimum of $V_\infty[U^* | \chi]$.

ASSUMPTION 1.2. The support of the measure F contains at least l points.

LEMMA 1.4. *Under the Assumption* 1.2, *it holds for* w^* *which corresponds to* $\chi^* \in \mathcal{H}^*$ *that* $w_i^* > 0, \ i = 1, \cdots, l$.

LEMMA 1.5. *For any fixed* $\chi^* \in \mathcal{H}^*$, χ^* *attains the minimal point of the set* $\left\{ \int \sum_{i=1}^{l} (\Phi_1(x) - \mu_i^*)^2 \chi_i(x)dF(x) - \int \Phi_1^2(x)dF(x), \ \chi \in \mathcal{H} \right\}$, *where* $\mu_j^* = u_j^* / w_j^*$ *and* u_j^*, w_j^* *corresponds to* χ_j^* $(j = 1, \cdots, l)$.

PROOF. The proof is easily obtained from the above Lemma 1.3 (ii) and the Lemma 2.2 of Isii and Taga [6], since G is continuously differentiable on C.

From the above lemma we can easily get the χ^* which attains the infimum of $V_\infty[U^*|\chi]$.

THEOREM 2. *In the proportional allocation the χ^* which attains the infimum of $V_\infty[U^*|\chi]$ is given by*

$$(1.28)\qquad \chi_i^*(x)=\begin{cases} 1 & \text{if } (\Phi_1(x)-\mu_i^*)^2<(\Phi_1(x)-\mu_j^*)^2 \quad \text{for any } j\neq i\ , \\ 0 & \text{if } (\Phi_1(x)-\mu_i^*)^2>(\Phi_1(x)-\mu_j^*)^2 \quad \text{for some } j\neq i\ . \end{cases}$$

2. Two-sample case

2.1 *Estimators of a population characteristic*

Let p-dimensional vectors $X_i=(X_i^{(1)},\cdots,X_i^{(p)})$, $i=1,2,\cdots,N_1$ be a random sample from the population Π^X with p-variate distribution function $F(x)$, and $Y_j=(Y_j^{(1)},\cdots,Y_j^{(p)})$, $j=1,2,\cdots,N_2$, be a random sample from the population Π^Y with p-variate distribution function $G(y)$. Let $\theta(F;G)$ be a real valued population characteristic of both populations. Suppose $\theta(F;G)$ is a regular (estimable) functional. Then there exists a kernel $\Phi(x;y)$ such that

$$(2.1)\qquad \theta(F;G)=\int\Phi(x;y)dF(x)dG(y)\ .$$

As an unbiased estimator of $\theta(F;G)$ we shall apply the following U-statistic

$$(2.2)\qquad U=\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\Phi(X_i;Y_j)\Big/N_1N_2\ .$$

Suppose that the population Π^X with distribution function $F(x)$ and the population Π^Y with $G(y)$ are classified into l_1 and l_2 strata $\{\Pi_1^X,\cdots,\Pi_{l_1}^X\}$ and $\{\Pi_1^Y,\cdots,\Pi_{l_2}^Y\}$, respectively, in such a way that the distribution function F_i and G_j respectively corresponds to the ith stratum Π_i^X and jth stratum Π_j^Y satisfy the relations

$$(2.3)\qquad \begin{aligned} F(x)&=\sum_{i=1}^{l_1}v_iF_i(x)\ , \qquad 0\leqq v_i\leqq 1\ ,\quad \sum_{i=1}^{l_1}v_i=1\ ,\\ G(y)&=\sum_{j=1}^{l_2}w_jG_j(y)\ , \qquad 0\leqq w_j\leqq 1\ ,\quad \sum_{j=1}^{l_2}w_j=1\ ,\end{aligned}$$

for all x and y.

If for each $i=1,2,\cdots,l_1$ and $j=1,2,\cdots,l_2$, we have a sampling plan to take a random samples of size m_i and n_j from the ith stratum Π_i^X and jth stratum Π_j^Y, respectively, then as an estimator of $\theta(F;G)$ we can apply the following statistic U^*.

$$(2.4)\qquad U^*=\sum_{i=1}^{l_1}\sum_{j=1}^{l_2}w_iv_jU_{ij}^*$$

where

$$U_{ij}^{*}=\sum_{\alpha=1}^{m_i}\sum_{\beta=1}^{n_j}\Phi(X_{i\alpha};Y_{j\beta})\Big/m_in_j$$

and $X_{i\alpha}=(X_{i\alpha}^{(1)},\cdots,X_{i\alpha}^{(p)})$, $\alpha=1,2,\cdots,m_i$ is a random sample drawn from Π_i^X, $i=1,2,\cdots,l_1$ and $Y_{j\beta}=(Y_{j\beta}^{(1)},\cdots,Y_{j\beta}^{(p)})$, $\beta=1,2,\cdots,n_j$ is one from Π_j^Y, $j=1,2,\cdots,l_2$.

Substituting (2.3) into (2.1) we can get

$$\theta(F;G)=\sum_{i=1}^{l}\sum_{j=1}^{l}w_iv_j\int\Phi(x;y)dF_i(x)dG_j(y)\ .$$

Thus U^* defined in (2.4) is easily shown to be an unbiased estimator of $\theta(F;G)$. We shall compare two unbiased estimators U and U^* of $\theta(F;G)$ in the next section.

Remark. By putting $\Phi(x,y)=1$ if $x<y$ and 0 otherwise, we can solve the problem of estimation of the probability $P[X<Y]$.

2.2 *Comparison of the estimators*

Suppose that the total sample sizes N_1 and N_2 of both populations are fixed, and they are allocated to each stratum proportionately to each size, i.e. $n_i=v_iN_1$, $(i=1,2,\cdots,l_1)$, and $m_j=w_jN_2$, $(j=1,2,\cdots,l_2)$, where v_i and w_j are weights of F_i and G_j, respectively, given in (2.3).

ASSUMPTION 2.1. $E[\Phi(X;Y)]<\infty$.

Under the Assumption 2.1 the variances of both estimators U and U^* exist. Let N be such that $N_i=\rho_iN$, $0<\rho_i<1$, $\rho_1+\rho_2=1$, $(i=1,2)$, and let

$$V_\infty[U]=\lim_{N\to\infty}\mathrm{Var}\,[\sqrt{N}(U-\theta(F;G))]\ ,$$

and

$$V_\infty[U^*]=\lim_{N\to\infty}\mathrm{Var}\,[\sqrt{N}(U^*-\theta(F;G))]\ .$$

LEMMA 2.1. *Under the Assumption* 2.1,

(i) $\sqrt{N}(U-\theta(F;G))$ *has, as* $N\to\infty$, *a normal limiting distribution with mean zero and variance*

$$V_\infty[U]=\rho_1^{-1}\int\Phi(x;y_1)\Phi(x;y_2)dF(x)dG(y_1)dG(y_2)$$
$$+\rho_2^{-1}\int\Phi(x_1;y)\Phi(x_2;y)dF(x_1)dF(x_2)dG(y)-(\rho_1\rho_2)^{-1}\theta^2(F;G)\ .$$

(ii) $\sqrt{N}(U^*-\theta(F;G))$ *has, as* $N\to\infty$, *a normal limiting distribution with mean zero and variance*

$$(2.5)\qquad \begin{aligned} V_\infty[U^*]=&\rho_1^{-1}\int \Phi(x;y_1)\Phi(x;y_2)dF(x)dG(y_1)dG(y_2)\\ &+\rho_2^{-1}\int \Phi(x_1;y)\Phi(x_2;y)dF(x_1)dF(x_2)dG(y)\\ &-\sum_{k=1}^{l_1}\rho_1^{-1}v_k\left[\int \Phi(x;y)dF_k(x)dG(y)\right]^2\\ &-\sum_{k=1}^{l_2}\rho_2^{-1}w_k\left[\int \Phi(x;y)dF(x)dG_k(y)\right]^2 . \end{aligned}$$

PROOF. (i) is immediately obtained from the theory of U-statistics. (ii) Put

$$A_j(x_{i\alpha})=E\{\Phi(x_{i\alpha};Y_{j1})\}-\int \Phi(x;y)dF_i(x)dG_j(y) ,$$

$$B_i(y_{j\beta})=E\{\Phi(X_{i1};y_{j\beta})\}-\int \Phi(x;y)dF_i(x)dG_j(y) ,$$

$$C_{Nij}=\frac{\sqrt{N}}{m_i}\sum_{\alpha=1}^{m_i}A_j(X_{i\alpha})+\frac{\sqrt{N}}{n_j}\sum_{\beta=1}^{n_j}B_i(Y_{j\beta}) ,$$

$$T_N=\sum_{i=1}^{l_1}\sum_{j=1}^{l_2}v_iw_jC_{Nij} ,$$

$$U_{Nij}^*=\sqrt{N}\left(U_{ij}^*-\int \Phi(x, y)dF_i(x)dG_j(y)\right) ,$$

and

$$U_N^*=\sqrt{N}(U^*-\theta(F;G)) .$$

Then the proof is given through the following steps.

(a) $\lim_{N\to\infty} E[U_{Nij}^*-C_{Nij}]^2=0$,

(b) $\lim_{N\to\infty} E[U_N^*-T_N^*]^2=0$,

(c) U_N^* has the same asymptotic distribution as T_N, and

(d) T_N has, as $N\to\infty$, a normal limiting distribution with mean zero and whose variance is given by (2.5). Since each step (a), (b), (c), and (d) is shown similarly as the corresponding one in the proof of Lemma 1.2, so we shall omit the details.

From Lemma 2.1 we can get the theorem which shows an effect of stratification.

THEOREM 3. *Suppose the Assumption* 2.1. *Then we get*

$$V_\infty[U]-V_\infty[U^*]=\sum_{i=1}^{l_1}\rho_1^{-1}v_i\left[\int \Phi(x;y)dF_i(x)dG(y)-\theta(F;G)\right]^2$$

$$+\sum_{j=1}^{l_2}\rho_2^{-1}w_j\left[\int\Phi(x;y)dF(x)dG_j(y)-\theta(F;G)\right]^2.$$

Hence the limiting variance of U^* is smaller than that of U for any underlying distribution F.

PROOF. From Lemma 2.1 we get

$$\begin{aligned}V_\infty[U]-V_\infty[U^*]&=\sum_{i=1}^{l_1}\rho_1^{-1}v_i\left[\int\Phi(x;y)dF_i(x)dG(y)\right]^2\\&\quad+\sum_{j=1}^{l_2}\rho_2^{-1}w_j\left[\int\Phi(x;y)dF(x)dG_j(y)\right]^2-(\rho_1\rho_2)^{-1}\theta^2(F;G)\\&=\sum_{i=1}^{l_1}\rho_1^{-1}v_i\left[\int\Phi(x;y)dF_i(x)dG(y)-\theta(F;G)\right]^2\\&\quad+\sum_{j=1}^{l_2}\rho_2^{-1}w_j\left[\int\Phi(x;y)dF(x)dG_j(y)-\theta(F;G)\right]^2.\end{aligned}$$

2.3 *Optimum stratification*

Let $\{F_i(x)\}$ and $\{G_j(y)\}$ be an l_1 and l_2-decomposition of the distribution function $F(x)$ and $G(y)$, respectively. $F_i(x)$ and $G_j(y)$ are the distributions which satisfy the relations

$$(2.6)\qquad \begin{aligned}F(x)&=\sum_{i=1}^{l_1}v_iF_i(x)\,, &&0\leqq v_i\leqq 1\,, \quad \sum_{i=1}^{l_1}v_i=1\,,\\ G(y)&=\sum_{j=1}^{l_2}w_jG_j(y)\,, &&0\leqq w_j\leqq 1\,, \quad \sum_{j=1}^{l_2}w_j=1\,.\end{aligned}$$

In this section we shall consider the existence and its concrete form of optimum decomposition $\{F_i^*(x)\}$ and $\{G_j^*(y)\}$; i.e. l_1 and l_2- decomposition which satisfy (2.6) and attain the infimum of $V_\infty[U^*]$ for fixed F, G, N_1, N_2, l_1 and l_2.

$$(2.7)\qquad \begin{aligned}V_\infty[U^*]&=\rho_1^{-1}\left\{\int\Phi(x;y_1)\Phi(x;y_2)dF(x)dG(y_1)dG(y_2)\right.\\&\qquad\left.-\sum_{i=1}^{l_1}v_i\left[\int\Phi(x;y)dF_i(x)dG(y)\right]^2\right\}\\&\quad+\rho_2^{-1}\left\{\int\Phi(x_1;y)\Phi(x_2;y)dF(x_1)dF(x_2)dG(y)\right.\\&\qquad\left.-\sum_{j=1}^{l_2}w_j\left[\int\Phi(x;y)dF(x)dG_j(y)\right]^2\right\}.\end{aligned}$$

The first { } of the right-hand side of (2.7) depends only on l_1-decomposition $\{F_i(x)\}$ and the second { } depends on l_2-decomposition $\{G_j(y)\}$ alone. Thus without loss of generality, our problem is reduced to showing the existence and its form of l_1-decomposition $\{F_i(x)\}$ which satisfy (2.6) and attains the infimum of the first { } of the right-hand side of (2.7).

Hence we can also apply the same arguments stated in Section 1.3 to this problem and get the following theorem.

THEOREM 4. *Let measurable function vectors* $\chi_f=(\chi_{f1},\cdots,\chi_{fl_1})$ *and* $\chi_g=(\chi_{g1},\cdots,\chi_{gl_2})$, *be respectively, such that*

$$v_i dF_i(x)=\chi_{fi}(x)dF(x)\ ,\qquad i=1,\cdots,l_1\ , \tag{2.8}$$

$$\sum_{i=1}^{l_1}\chi_{fi}(x)=1\ ,\qquad 0\leqq\chi_{fi}(x)\leqq 1\ ,\quad \text{a.e.}\ (dF)\ ,$$

and

$$w_j dG_j(y)=\chi_{gj}(y)dG(y)\ ,\qquad j=1,\cdots,l_2\ , \tag{2.9}$$

$$\sum_{j=1}^{l_2}\chi_{gj}(y)=1\ ,\qquad 0\leqq\chi_{gj}(y)\leqq 1\ ,\quad \text{a.e.}\ (dG)\ .$$

Let $\mathcal{H}_f(\mathcal{H}_g)$ *be the set of all* $\chi_f(\chi_g)$ *which satisfy* (2.8) ((2.9)) *and let* $\mathcal{H}_f^*$ *and* $\mathcal{H}_g^*$, *be respectively, the set of all* $\chi_f\in\mathcal{H}_f$ *and* $\chi_g\in\mathcal{H}_g$ *which attain the infimum of* $V_\infty[U^*]$. *For any* $\chi_f^*\in\mathcal{H}_f^*$ *and* $\chi_g^*\in\mathcal{H}_g^*$ *let*

$$\mu_{fi}^*=\int\Phi_f(x)\chi_{fi}^*(x)dF(x)\Big/\int\chi_{fi}^*(x)dF(x)\ ,\qquad (i=1,\cdots,l_1)\ ,$$

and

$$\mu_{gj}^*=\int\Phi_g(y)\chi_{gj}^*(y)dG(y)\Big/\int\chi_{gj}^*(y)dG(y)\ ,\qquad (j=1,\cdots,l_2)\ ,$$

where

$$\Phi_f(x)=\int\Phi(x\,;y)dG(y)$$

and

$$\Phi_g(y)=\int\Phi(x\,;y)dF(x)\ ,$$

Then, in the two-sample proportionate allocation, $\chi_f^*\in\mathcal{H}_f^*$ *and* $\chi_g^*\in\mathcal{H}_g^*$ *are given by*

$$\chi_{fi}^*(x)=\begin{cases}1 & \text{if}\ \ (\Phi_f(x)-\mu_{fi}^*)^2<(\Phi_f(x)-\mu_{fj}^*)^2 \quad \textit{for any}\ j\neq i\ ,\\ 0 & \text{if}\ \ (\Phi_f(x)-\mu_{fi}^*)^2>(\Phi_f(x)-\mu_{fj}^*)^2 \quad \textit{for some}\ j\neq i\ ,\end{cases}$$

and

$$\chi_{gj}^*(y)=\begin{cases}1 & \text{if}\ \ (\Phi_g(y)-\mu_{gj}^*)^2<(\Phi_g(y)-\mu_{gk}^*)^2 \quad \textit{for any}\ k\neq j\ ,\\ 0 & \text{if}\ \ (\Phi_g(y)-\mu_{gj}^*)^2>(\Phi_g(y)-\mu_{gk}^*)^2 \quad \textit{for some}\ k\neq j\ .\end{cases}$$

References

[1] Cramér, H. (1946). *Mathematical Methods of Statistics*, Princeton University Press.

[2] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution, *Ann. Math. Statist.*, **19**, 293-325.

[3] Lehmann, E. L. (1959). *Testing Statistical Hypothesis*, John Wiley and Sons Inc.

[4] Sugiura, N. (1965). Multisample and multivariate nonparametric tests based on U statistics and their asymptotic efficiencies, *Osaka J. Math.*, **2**, 385-426.

[5] Taga, Y. (1967). On optimum stratification for the objective variable based on concomitant variables using prior information, *Ann. Inst. Statist. Math.*, **19**, 101-129.

[6] Isii, K. and Taga, Y. (1969). On optimum stratification for multivariate distributions, *Shand. Aktuarietidskr.*, **52**, 24-38.

[7] Wakimoto, K. (1971). Stratified random sampling (I); Estimation of the population variance, *Ann. Inst. Statist. Math.*, **23**, 233-252.

[8] Wakimoto, K. (1971). Stratified random sampling (II); Estimation of the population covariance, *Ann. Inst. Statist. Math.*, **23**, 327-337.

[9] Wakimoto, K. (1971). Stratified random sampling (III); Estimation of the correlation coefficient, *Ann. Inst. Statist. Math.*, **23**, 339-353.