STRATIFIED RANDOM SAMPLING ; RANK CORRELATION COEFFICIENTS, TESTS OF INDEPENDENCE AND RANDOM CONFIDENCE INTERVALS

Yanagawa, Takashi
Department of Mathematics, Kyushu University

https://doi.org/10.5109/13069

出版情報：統計数理研究. 15 (3/4), pp.21-30, 1973-03. Research Association of Statistical Sciences

STRATIFIED RANDOM SAMPLING; RANK CORRELATION COEFFICIENTS, TESTS OF INDEPENDENCE AND CONFIDENCE INTERVALS

By

Takashi Yanagawa*

(Received November 25, 1971)

§ 1. Introduction

The problem of giving reasonable measures of association when a population is stratified was first investigated by Aoyama [1]. Recently Wakimoto [3] considered the problem more extensively. He gave an estimator of the correlation coefficient based on a stratified random sample and showed it to be superior to the one given by Aoyama.

The purpose of this paper is to propose new measures of association, test of independence and confidence intervals based on a stratified random sample. These measures are stratified version of Kendall and Speaman rank correlation coefficients. Throughout this paper we assume that each size of stratum is sufficiently large compared with that of sample taken from it so that the finite correction term may be neglected. In section 2 measures of association, tests of independence and confidence intervals based on a stratified random sample is given. A stratified version of Kendall rank correlation coefficient is discussed in section 2.1 and then in section 2.2 the one of Speaman type is discussed. In section 3 gains in efficiency due to stratification is demonstrated in the case of proportional allocation by comparing proposed measures with respect to Kendall and Speaman rank correlation coefficient based on a simple random sample.

§ 2. Measures of association, tests of independence and confidence intervals

Suppose that the population π with two-dimensional distribution function $F(x)$ is classified into L strata $\{\pi_1, \cdots, \pi_L\}$, which may be overlapping, in such a way that the two-dimensional distribution function $F_i(x)$ corresponds to the i-th stratum π_i satisfies the relation

$$F(x) = \sum_{i=1}^{L} w_i F_i(x) \qquad \text{for all} \quad x, \tag{2.1}$$

where w_i is a known weight of F_i such that $0 \leqq w_i \leqq 1$, $\sum_{i=1}^{L} w_i = 1$.

* Department of Mathematics, Kyushu University, Fukuoka.

Now suppose we have a sampling plan to take a two-dimensional random sample $(X_{i1}^{(1)}, X_{i1}^{(2)}), \cdots, (X_{in_i}^{(1)}, X_{in_i}^{(2)})$ from the i-th stratum π_i $(i=1\sim L)$. We assume $n=\sum_{i=1}^{L} n_i$ the total sample size. Let $(X_1^{(1)}, X_1^{(2)}), \cdots, (X_n^{(1)}, X_n^{(2)})$ be a two-dimensional simple random sample of size n from the population π.

We need the following definitions and notations in the sequel.

$$S(u)=\begin{cases} -1 & \text{if } u<0 \\ 0 & \text{if } u=0 \\ 1 & \text{if } u>0, \end{cases}$$

R_α ; rank of $X_\alpha^{(1)}$ among $X_1^{(1)}, \cdots, X_n^{(1)}$,

S_α ; rank of $X_\alpha^{(2)}$ among $X_1^{(2)}, \cdots, X_n^{(2)}$,

$R_{i\alpha}$; rank of $X_{i\alpha}^{(1)}$ among $X_{i1}^{(1)}, \cdots, X_{in_i}^{(1)}$,

$S_{i\alpha}$; rank of $X_{i\alpha}^{(2)}$ among $X_{i1}^{(2)}, \cdots, X_{in_i}^{(2)}$,

$R_{i\alpha}^{ij}$; rank of $X_{i\alpha}^{(1)}$ among $X_{i1}^{(1)}, \cdots, X_{in_i}^{(1)}, X_{j1}^{(1)}, \cdots, X_{jn_j}^{(1)}$,

$S_{i\alpha}^{ij}$; rank of $X_{i\alpha}^{(2)}$ among $X_{i1}^{(2)}, \cdots, X_{in_i}^{(2)}, X_{j1}^{(2)}, \cdots, X_{jn_j}^{(2)}$.

We get following equalities.

(2.2) $$R_\alpha=\frac{n+1}{2}+\frac{1}{2}\sum_{\beta=1}^{n} S(X_\alpha^{(1)}-X_\beta^{(1)})$$

(2.3) $$S_\alpha=\frac{n+1}{2}+\frac{1}{2}\sum_{\beta=1}^{n} S(X_\alpha^{(2)}-X_\beta^{(2)})$$

(2.4) $$R_{i\alpha}=\frac{n_i+1}{2}+\frac{1}{2}\sum_{\beta=1}^{n_i} S(X_{i\alpha}^{(1)}-X_{i\beta}^{(1)})$$

(2.5) $$S_{i\alpha}=\frac{n_i+1}{2}+\frac{1}{2}\sum_{\beta=1}^{n_i} S(X_{i\alpha}^{(2)}-X_{i\beta}^{(2)})$$

(2.6) $$R_{i\alpha}^{ij}-R_{i\alpha}=\frac{n_j}{2}+\frac{1}{2}\sum_{\beta=1}^{n_j} S(X_{i\alpha}^{(1)}-X_{j\beta}^{(1)})$$

(2.7) $$S_{i\alpha}^{ij}-S_{i\alpha}=\frac{n_j}{2}+\frac{1}{2}\sum_{\beta=1}^{n_j} S(X_{i\alpha}^{(2)}-X_{j\beta}^{(2)}).$$

2.1. Kendall's type

Kendall's rank correlation coefficient in the simple random sampling is defined as

$$r_K=\frac{1}{n(n-1)}\sum_{\alpha\neq\beta} S(X_\alpha^{(1)}-X_\beta^{(1)})S(X_\alpha^{(2)}-X_\beta^{(2)}).$$

r_K is an unbiased estimator of

$$\tau=\int s(x_1^{(1)}-x_2^{(1)})S(x_1^{(2)}-x_2^{(2)})dF(x_1)dF(x_2), \qquad x_i=(x_1^{(i)}, x_2^{(i)}) \qquad i=1, 2.$$

In the stratified case we propose following r_K^* as a measure of association, which belongs to the class of U-statistics considered by Yanagawa, T. and Wakimoto, K. [5].

$$r_K^* = \sum_{i=1}^{L} \sum_{j=1}^{L} w_i w_j t_{ij} , \tag{2.8}$$

where

$$t_{ii} = \frac{1}{n_i(n_i-1)} \sum_{\alpha \neq \beta}^{n_i} S(X_{i\alpha}^{(1)} - X_{i\beta}^{(1)}) S(X_{i\alpha}^{(2)} - X_{i\beta}^{(2)}) , \tag{2.9}$$

$$t_{ij} = \frac{1}{n_i n_j} \sum_{\alpha=1}^{n_i} \sum_{\beta=1}^{n_j} S(X_{i\alpha}^{(1)} - X_{j\beta}^{(1)}) S(X_{i\alpha}^{(2)} - X_{j\beta}^{(2)}) \qquad (i \neq j) . \tag{2.10}$$

Let us put

$$\begin{aligned} &\varphi_K(x_1, x_2) \equiv S(x_1^{(1)} - x_2^{(1)}) S(x_1^{(2)} - x_2^{(2)}) , \\ &\varphi_{K1}(x) = \int S(x^{(1)} - x_2^{(1)}) S(x^{(2)} - x_2^{(2)}) dF(x_2) , \\ &\qquad x = (x^{(1)}, x^{(2)}) , \qquad x_i = (x_i^{(1)}, x_i^{(2)}) . \end{aligned} \tag{2.11}$$

THEOREM 1. (i) *r_K^* is an unbiased estimator of τ and its variance is given by*

$$\mathrm{Var}\,[r_K^*] = 4 \sum_{i=1}^{L} \frac{w_i^2}{n_i} \mathrm{Var}\,[\varphi_{K1}(X_{i1})] + 2 \sum_{i=1}^{L} \frac{w_i^4}{n_i(n_i-1)} a_{ii} + 2 \sum_{i \neq j}^{L} \frac{w_i^2 w_j^2}{n_i n_j} a_{ij} , \tag{2.12}$$

where

$$a_{ij} = \mathrm{Var}\,[\varphi_K(X_{i1}, X_{j2})] - 2\,\mathrm{Cov}\,[\varphi_K(X_{i1}, X_{j2}), \varphi_K(X_{i1}, X_{j3})] .$$

(ii) *Let $n \to \infty$ in such a way that $n_i/n \to \lambda_i$, where λ_i is a constant such that $0 < \lambda_i < 1$ $(i = 1 \sim L)$ and $\sum_{i=1}^{L} \lambda_i = 1$. Then $\sqrt{n}\,(r_K^* - \tau)$ has a normal limiting distribution with mean 0 and variance*

$$\sigma_K^{*2} = 4 \sum_{i=1}^{L} \frac{w_i^2}{\lambda_i} \mathrm{Var}\,[\varphi_{K1}(X_{i1})] . \tag{2.13}$$

PROOF. (i) is obtained directly by simple but some lengthy computations and (ii) comes from a slight modification of the proof of Lemma 1.2 of [5], and so we shall omit the proof.

From the theorem we immediately get following corollaries.

COROLLARY 1. (*Asymptotic optimum allocation*) *The infimum value of σ_K^{*2} over all λ_i such that $0 < \lambda_i < 1$ and $\sum_{i=1}^{L} \lambda_i = 1$ is*

$$\sigma_{K.\mathrm{opt.}}^{*2} = \Big(\sum_{i=1}^{L} w_i A_i \Big)^2 , \tag{2.14}$$

where

$$A_i^2 = \mathrm{Var}\,[\varphi_{K1}(X_{i1})] . \tag{2.15}$$

Further the optimum λ_i^* which attains the infimum value is given by

$$\lambda_i^* = w_i A_i / \sum_{j=1}^{L} w_j A_j . \tag{2.16}$$

COROLLARY 2. (*Asymptotic proportional allocation*) *When $\lambda_i = w_i$ for all $i = 1 \sim L$ we get a proportional allocation and σ_K^{*2} becomes*

(2.17) $$\sigma^{*2}_{K.\mathrm{prop.}}=4\Big\{\int\varphi^2_{K1}dF-\sum_{i=1}^{L}w_i\Big(\int\varphi_{K1}dF_i\Big)^2\Big\}.$$

Since F and F_i are assumed to be unknown and from (2.15)

$$A_i^2=\int\Big(\varphi_{K1}-\int\varphi_{K1}dF_i\Big)^2dF_i\,,$$

then A_i^2 is also unknown. Thus for the problem of testing and confidence interval we need its consistent estimator, which is given by the following Lemma.

LEMMA 1. *Put*

(2.18) $$g(x_{i\alpha})=\frac{w_i}{n_i-1}\sum_{\beta\neq\alpha}^{n_j}\varphi(x_{i\alpha},\,x_{i\beta})+\sum_{j\neq i}^{L}\frac{w_j}{n_j}\sum_{\beta=1}^{n_j}\varphi(x_{i\alpha},\,x_{j\beta})\,,$$

(2.19) $$\bar{g}_i=\sum_{\alpha=1}^{n_i}g(x_{i\alpha})/n_i\,,$$

and

(2.20) $$S_i^2=\sum_{\alpha=1}^{n_i}[g(x_{i\alpha})-\bar{g}_i]^2/(n_i-1)\,,$$

then for A_i^2, *given in* (2.15), *we get*

$$\text{(i)}\qquad E[S_i^2]\longrightarrow A_i^2\quad as\quad n\longrightarrow\infty\qquad(i=1\sim L)$$

and

$$\text{(iii)}\qquad S_i^2\longrightarrow A_i^2\quad in\ prob.\quad as\quad n\longrightarrow\infty\qquad(i=1\sim L)\,.$$

PROOF. See Yanagawa, T. [4] Lemma 2.

COROLLARY 3. *Put*

(2.21) $$T^*_{K,n}=\sqrt{n}\,(r^*_K-\tau)/2\Big(\sum_{i=1}^{L}w_i^2S_i^2/\lambda_i\Big)^{1/2}.$$

then under the assumption of Theorem 1 $T^*_{K,n}$ *has a standard normal limiting distribution.*

PROOF. Proof is immediately obtained from Theorem 1 and Lemma 1.

From the lemma and corollaries we get following applications.

APPLICATION 1. An estimator of optimum λ_i^* is given as $\hat{\lambda}_i^*=w_iS_i/\sum_{i=1}^{L}w_iS_i$.

APPLICATION 2. Consider the test of independence such that

$$H;\quad F(x^{(1)},\,x^{(2)})=F(x^{(1)},\,\infty)\cdot F(\infty,\,x^{(2)})$$
$$AH;\quad F(x^{(1)},\,x^{(2)})\neq F(x^{(1)},\,\infty)\cdot F(\infty,\,x^{(2)})\,.$$

Since under the hypothesis we get $\tau=0$, $T^*_{K,n}$ given by (2.21) with $\tau=0$ is proposed as a test statistic of the hypothesis based on the stratified random sample. Especially in the case of proportional allocation since we get $\tau=0$ and $\varphi_{K1}(x)=[2F(x^{(1)},\infty)-1][2F(\infty,x^{(2)})-1]$ under the hypothesis, then $\sigma^{*2}_{K.\mathrm{prop.}}$ given by (2.17) becomes

$$\sigma^{*2}_{K.\mathrm{prop.}}=4\Big[1/9-\sum_{i=1}^{L}w_i\Big(\int\varphi_{K1}dF_i\Big)^2\Big].$$

Thus instead of $T^*_{K,n}$, following $T^*_{n.\mathrm{prop.}}$ is proposed as a test statistic.

$$T^*_{n.\mathrm{prop.}}=\sqrt{n}\,r^*_K/2\Big[1/9-\sum_{i=1}^{L}w_i\bar{g}_i^2\Big]^{1/2},$$

where $\bar{g}_i$ is given by (2.19).

APPLICATION 3. Approximate confidence intervals of confidence coefficient $(1-\alpha)$ is given as follows.

$$\sqrt{n}\, r_K^* - 2k(\alpha/2)\Big(\sum_{i=1}^{L} w_i^2 S_i^2/\lambda_i\Big)^{1/2} \leqq \tau$$

$$\leqq \sqrt{n}\, r_K^* - 2k(\alpha/2)\Big(\sum_{i=1}^{L} w_i^2 S_i^2/\lambda_i\Big)^{1/2},$$

where $k(\alpha)$ is an upper α percentile point of a standard normal distribution.

2.2. Speaman's type in the case of proportional allocation

In the simple random sampling Speaman's rank correlation coefficient is

$$r_S = \frac{12}{n^3-n}\sum_{\alpha=1}^{n}\Big(R_\alpha - \frac{n+1}{2}\Big)\Big(S_\alpha - \frac{n+1}{2}\Big). \tag{2.22}$$

It is an unbiased estimator of

$$\rho = \frac{3}{n+1}\Big[(n-2)\int S(x_1^{(1)}-x_2^{(1)})S(x_1^{(2)}-x_3^{(2)})\prod_{i=1}^{3} dF(x_i) \tag{2.23}$$

$$+\int S(x_1^{(1)}-x_2^{(1)})S(x_1^{(2)}-x_2^{(2)})\prod_{i=1}^{2} dF(x_i)\Big].$$

In this section we shall give a stratified version of Speaman's rank correlation coefficient. It is quite complicated and that we shall consider it only the case of proportional allocation. Through similar discussions as previous section we can get an unbiased estimator of ρ based on U-statistics. But here we shall concentrate our considerations to estimators based on ranks and that we shall propose following r_s^* as a measure of association in the case of proportional allocation.

$$r_s^* = \frac{12}{n^2(n+1)}\sum_{i=1}^{L}\sum_{\alpha=1}^{n_i}\Big[\sum_{j\neq i}^{L} R_{i\alpha}^{ij} - (L-2)R_{i\alpha} - \frac{n+1}{2}\Big]\Big[\sum_{K\neq i}^{L} S_{i\alpha}^{ik} - (L-2)S_{i\alpha} - \frac{n+1}{2}\Big]. \tag{2.24}$$

Put

$$\varphi_s(x_1, x_2, x_3) = S(x_1^{(1)}-x_2^{(1)})S(x_1^{(2)}-x_3^{(2)}), \tag{2.25}$$

$$\phi(x_1, x_2, x_3) = \sum{}' \varphi_s(x_{i_1}, x_{i_2}, x_{i_3})/2, \tag{2.26}$$

where the sum $\sum'$ is taken over all permutation (i_1, i_2, i_3) of (1, 2, 3). Then we get the following theorem.

THEOREM 2. (i) *$E(r_s^*) = \rho + B/(n+1)$, where*

$$B = 3\sum_{i=1}^{L} w_i\Big\{\int \phi(x_1, x_2, x_3)\prod_{j=1}^{2} dF_i(x_j)dF(x_3) - 2\int \phi(x_1, x_2, x_3)\prod_{j=1}^{3} dF(x_j)\Big\}.$$

(ii) *Let $n\to\infty$ in such a way that $n_i/n \to w_i$ for all $i = 1\sim L$, then $\sqrt{n}(r_s^*-\rho)$ has a normal limiting distribution with mean 0 and variance*

$$\sigma_s^{*2} = 9\sum_{i=1}^{L} w_i \int [\phi_s(x) - \rho_{i\cdot}]^2 dF_i(x), \tag{2.27}$$

where

(2.28) $$\psi_s(x)=\int \psi(x, x_2, x_3)\prod_{i=2}^{3} dF(x_i)$$

and

(2.29) $$\rho_{i\cdot}=\int \psi_s(x)dF_i(x)\,.$$

PROOF. As the proof of the theorem is complicated, we shall give it in the appendix.

Now let us consider the test of independence. Under the hypothesis we get $\rho=0$ and $\psi_s(x)=[2F(x^{(1)}, \infty)-1][2F(\infty, x^{(2)}-1]$, then it follows that $\sigma_s^{*2}=9\sigma_{K,\text{prop}}^{*2}/4$, where $\sigma_{K.\text{prop}.}^{*2}$ is given by (2.17). Thus we get the following test statistic for the test of independence.

$$T_{s,n}^{*}=\sqrt{n}\, r_s^*/3\Big[1/9-\sum_{i=1}^{L} w_i \bar{g}_i^2\Big]^{1/2},$$

where $\bar{g}_i$ is given by (2.19).

§ 3. Gain in efficiency due to stratification

In this section we shall consider the gain in efficiency due to stratification. Let $X_\alpha=(X_\alpha^{(1)}, X_\alpha^{(2)})$, $\alpha=1, 2, 3$ be i. i. d. random variables with distribution function $F(x)$, $X_{i\beta}=(X_{i\beta}^{(1)}, X_{i\beta}^{(2)})$ be i. i. d. random variables with distribution function $F_i(x)$, $i=1\sim L$, and let us put

$$\varphi_K^j(x)=\int \varphi_K(x, y)dF_j(y)\,, \qquad j=1\sim L\,.$$

LEMMA 2.

(i) $$\text{Var}\,[\varphi_K(X_1, X_2)]\geqq 2\,\text{Var}\,[\varphi_{K1}(X_1)]\,.$$

(ii) $$\sum_{i,j}^{L} w_i w_j \{\text{Cov}\,[\varphi_K(X_{i1}, X_{j2}), \varphi_K(X_{i1}, X_{j3})]-\text{Var}\,[\varphi_{K1}(X_{i1})]\}$$
$$=\sum_{i,j}^{L} w_i w_j E\{[\varphi_K^j(X_{i1})-E\varphi_K^j(X_{i1})]-[\varphi_{K1}(X_{i1})-E\varphi_{K1}(X_{i1})]\}^2\,.$$

PROOF. (i) See Fraser [2], p. 227, Theorem 5.2. (ii) Since

$$\text{Cov}\,[\varphi_K(X_{i1}, X_{j2}), \varphi_K(X_{i1}, X_{j3})]-\text{Var}\,[\varphi_{K1}(X_{i1})]$$
$$=E[\varphi_K^j(X_{i1})-E\varphi_K^j(X_{i1})]^2-E[\varphi_{K1}(X_{i1})-E\varphi_{K1}(X_{i1})]^2\,,$$

and

$$\sum_{i,j} w_i w_j E\{\varphi_K^j(X_{i1})-E\varphi_K^j(X_{i1})][\varphi_{K1}(X_{i1})-E\varphi_{K1}(X_{i1})]$$
$$=\sum_i w_i E[\varphi_{K1}(X_{i1})-E\varphi_{K1}(X_{i1})]^2\,,$$

then we get

$$\sum_{i,j} w_i w_j \{\text{Cov}\,[\varphi_K(X_{i1}, X_{j2}), \varphi_K(X_{i1}, X_{j3})]-\text{Var}\,[\varphi_{K1}(X_{i1})]\}$$
$$=\sum_{i,j} w_i w_j E\{[\varphi_K^j(X_{i1})-E\varphi_K^j(X_{i1})]-[\varphi_{K1}(X_{i1})-E\varphi_{K1}(X_{i1})]\}^2\,.$$

THEOREM 3. *Suppose the proportional allocation.* (i) *Put*

$$P=\sum_{i=1}^{L} w_i [E\varphi_{K1}(X_{i1})-E\varphi_{K1}(X_1)]^2 ,$$

$$Q=\sum_{i,j}^{L} w_i w_j [E\varphi_K(X_{i1}, X_{j2})-E\varphi_K(X_1, X_2)]^2-2\sum_{i=1}^{L} w_i [E\varphi_{K1}(X_{i1})-E\varphi_{K1}(X_1)]^2$$
$$+2\sum_{i,j}^{L} w_i w_j \{\mathrm{Cov}\,[\varphi_K(X_{i1}, X_{j2}), \varphi_K(X_{i1}, X_{j3})]-\mathrm{Var}\,[\varphi_{K1}(X_{i1})]\} ,$$

then we get

$$\mathrm{Var}\,[r_K]-\mathrm{Var}\,[r_K^*]=\frac{4}{n} P+\frac{2}{n(n-1)} Q+0\left(\frac{1}{n^3}\right)$$

and

$$\frac{4}{n} P+\frac{2}{n(n-1)} Q \geqq 0 .$$

(ii) $$E[r_s-\rho]^2-E[r_s^*-\rho]^2=\frac{9}{n}\sum_{i=1}^{L} w_i\left[\int\psi_s(x)dF_i(x)-\rho\right]^2+0\left(\frac{1}{n^2}\right) .$$

PROOF. (i) Variance of Kendall's rank correlation coefficient r_K is given by

$$\mathrm{Var}\,[r_K]=\frac{4(n-2)}{n(n-1)}\,\mathrm{Var}\,[\varphi_{K1}(X_1)]+\frac{2}{n(n-1)}\,\mathrm{Var}\,[\varphi_K(X_1, X_2)] .$$

Substituting (2.1), it follows

$$\mathrm{Var}\,[r_K]=\frac{4}{n}\left\{\sum_{i=1}^{L} w_i\,\mathrm{Var}\,[\varphi_{K1}(X_{i1})]+\sum_{i=1}^{L} w_i [E\varphi_{K1}(X_{i1})-E\varphi_{K1}(X_1)]^2\right\}$$
$$+\frac{2}{n(n-1)}\left\{\sum_{i,j}^{L} w_i w_j\,\mathrm{Var}\,[\varphi_K(X_{i1}, X_{j2})]\right.$$
$$+\sum_{i,j}^{L} w_i w_j [E\varphi_K(X_{i1}, X_{j2})-E\varphi_K(X_1, X_2)]^2$$
$$\left.-2\sum_{i=1}^{L} w_i \{\mathrm{Var}\,[\varphi_{K1}(X_{i1})]+[E\varphi_{K1}(X_{i1})-E\varphi_{K1}(X_1)]^2\}\right\} .$$

Thus from (2.12) we get

$$\mathrm{Var}\,[r_K]-\mathrm{Var}\,[r_K^*]=\frac{4}{n}\sum_{i=1}^{L} w_i [E\varphi_{K1}(X_{i1})-E\varphi_{K1}(X_1)]^2$$
$$+\frac{2}{n(n-1)}\left\{\sum_{i,j}^{L} w_i w_j [E\varphi_K(X_{i1}, X_{j2})-E\varphi_K(X_1, X_2)]^2\right.$$
$$-2\sum_{i=1}^{L} w_i [E\varphi_{K1}(X_{i1})-E\varphi_{K1}(X_1)]^2$$
$$\left.+2\sum_{i,j}^{L} w_i w_j (\mathrm{Cov}\,[\varphi_K(X_{i1}, X_{j2}), \varphi_K(X_{i1}, X_{j3})]-\mathrm{Var}\,[\varphi_1(X_{i1})])\right\}$$
$$+\frac{2}{n^2}\left\{\sum_{i,j}^{L}\frac{w_i w_j}{n-1}(\mathrm{Var}\,[\varphi_K(X_{i1}, X_{j2})]-2\,\mathrm{Cov}\,[\varphi_K(X_{i1}, X_{j2}), \varphi_K(X_{i1}, X_{j3})])\right.$$
$$\left.-\sum_{i=1}^{L}\frac{w_i^2}{n_i-1}(\mathrm{Var}\,[\varphi_K(X_{i1}, X_{j2})]-2\,\mathrm{Cov}\,[\varphi_K(X_{i1}, X_{j2}), \varphi_K(X_{i1}, X_{j3})])\right\}$$
$$=\frac{4}{n} P+\frac{2}{n(n-2)} Q+0\left(\frac{1}{n^3}\right) .$$

From Lemma 2, $\frac{4}{n}P+\frac{2}{n(n-1)}Q\geqq 0$ is clear. Thus we get (i).

(ii) From Theorem 2

$$E[r_s-\rho]^2-E[r_s^*-\rho]^2=\text{Var}\,[r_s]-\text{Var}\,[r_s^*]+0(1/n^2)$$
$$=\text{Var}\,[K]-\text{Var}\,[K^*]+0(1/n^2)\,,$$

where $K=6\sum''\psi(X_\alpha, X_\beta, X_\gamma)/n(n-1)(n-2)$, the summation $\sum''$ being over all combinations (α, β, γ) such that $1\leqq\alpha<\beta<\gamma\leqq n$ and K^* is given by (2) in the appendix. Since K and K^* are U-statistics, and through some similar discussions as (i), we get from the proof of Theorem 2 that

$$\text{Var}\,[K]=\frac{9}{n}\,\text{Var}\,[\psi_s(X_1)]+0\left(\frac{1}{n^2}\right)$$

and

$$\text{Var}\,[K^*]=\frac{9}{n}\sum_{i=1}^{L}w_i\int[\psi_s(x)-\rho_{i\cdot}]^2dF_i(x)\,,$$

where

$$\rho_{i\cdot}=\int\psi_s(x)dF_i(x)\,,\qquad i=1\sim L\,.$$

Thus

$$\text{Var}\,[K]-\text{Var}\,[K^*]=\frac{9}{n}\left\{\sum_{i=1}^{L}w_i\rho_{i\cdot}^2-\rho^2\right\}+0\left(\frac{1}{n^2}\right)$$
$$=\frac{9}{n}\sum_{i=1}^{L}w_i\left[\int\psi_s(x)dF_i(x)-\rho\right]^2+0\left(\frac{1}{n^2}\right).$$

Thus we complete the proof of the theorem.

NOTE. From Theorem 3 we get inequalities such that $\text{Var}\,(r_k)\geqq\text{Var}\,(r_k^*)$ and $E[r_s-\rho]^2\geqq E[r_s^*-\rho]^2$ asymptotically in the case of proportional allocation, which show the gain in efficiency due to stratification. Unfortunately these relations do not hold even asymptotically for arbitrary allocation. For example suppose $F_1=\cdots=F_L$, which is considered as a "bad stratification", then we get the inverse inequality $\sigma_K^2\leqq\sigma_{K,\text{prop.}}^{*2}$, where equality holds only when $\lambda_i=w_i$ for all $i=1\sim L$.

References

[1] AOYAMA, H. (1954): *A study of the stratified random sampling,* Ann. Inst. Stat. Math., **6**, 1-36.

[2] FRASER, D. A. S. (1957): *Nonparpmetric Methods in Statistic.* John Wiley & Sons, Inc.

[3] WAKIMOTO, K. (1971): *Stratified random samaling (III); Estimation of the correlation coefficient.* Ann. Inst. Stat. Math., **23**, pp. 339-353.

[4] YANAGAWA, T. (1972): *Stratified random sampling; Testing hypotheses based on U-statistics.* Memoirs of the Faculty of Science, Kyushu University. Series A, Vol. **26**, No. 2, pp. 361-379.

[5] YANAGAWA, T. and WAKIMOTO, K. (1972): *Estimation of some functional of population distribution based on a stratified random sampling.* Ann. Inst. Stat. Math., Vol. **24**, No. 1, pp. 137-151.

Appendix. PROOF OF THEOREM 2. (i) Put

$$\rho_1 = 3\int \varphi(x_1, x_2, x_3)\prod_{i=1}^{3} dF(x_i)\,,$$

then ρ can be written as follows

(1) $$\rho = [(n-2)\rho_1 + 3\tau]/(n+1)\,,$$

where τ is given in section 2.1. From Yanagawa, T. and Wakimoto, K. [5] U.M.V. unbiased estimator of ρ_1 based on stratified random sample is given as following K^*.

(2) $$K^* = A_1 + A_2 + A_3\,,$$

where

$$A_1 = 6\sum_{i=1}^{L} w_i^3 \sum_{\alpha<\beta<\gamma}^{n_i} \phi(x_{i\alpha}, x_{i\beta}, x_{i\gamma})/n_i(n_i-1)(n_i-2)\,,$$

$$A_2 = 6\sum_{i=1}^{L} w_i^2 w_j \sum_{\alpha<\beta}^{n_i}\sum_{\gamma=1}^{n_j} \phi(x_{i\alpha}, x_{i\beta}, x_{j\gamma})/n_i(n_i-1)n_j$$

and

$$A_3 = \sum_{\substack{i\neq j\neq k\\ i\neq k}}^{L} w_i w_j w_k \sum_{\alpha=1}^{n_i}\sum_{\beta=1}^{n_j}\sum_{\gamma=1}^{n_k} \phi(x_{i\alpha}, x_{j\beta}, x_{k\gamma})\,.$$

By (2.25), (2.26), (2.2), (2.3), (2.4), (2.5), (2.6) and (2.7) and using the relation $n_i = w_i n$ for all $i = 1 \sim L$ (proportional allication), we get

$$(n-2)A_1 = 3\sum_{i=1}^{L} \frac{w_i^2}{n_i(n_i-1)} \sum_{\substack{\alpha\neq\beta\neq\gamma\\ \alpha\neq\beta}}^{n_i} \varphi(x_{i\alpha}, x_{i\beta}, x_{i\gamma}) + A_{11}\,,$$

$$(n-2)A_2 = 3\sum_{i\neq j}^{L} w_i w_j \sum_{\alpha=1}^{n_j}\sum_{\beta\neq\gamma}^{n_j} \varphi(x_{i\alpha}, x_{j\beta}, x_{j\gamma})/n_i n_j$$

$$+3\sum_{i\neq j}^{L} w_i w_j \sum_{\alpha=1}^{n_j} \{[2R_{i\alpha} - (n_i+1)][2(S_{i\alpha}^{ij} - S_{i\alpha}) - n_j]$$

$$+[2(R_{i\alpha}^{ij} - R_{i\alpha}) - n_j][2S_{i\alpha} - (n_i+1)]\}/n_i n_j + A_{21}\,.$$

and

$$(n-2)A_3 = 3\sum_{\substack{i\neq j\neq k\\ i\neq k}}^{L} \frac{w_i w_j}{n_i n_j} \sum_{\alpha=1}^{n_i} [2(R_{i\alpha}^{ij} - R_{i\alpha}) - n_j][2(S_{i\alpha}^{ij} - S_{i\alpha}) - n_k] - A_{31}\,,$$

where

$$A_{11} = 6\sum_{i=1}^{L} w_i^2(1-w_i) \sum_{\substack{\alpha\neq\beta\neq\gamma\\ \alpha\neq\gamma}}^{n_i} \varphi(x_{i\alpha}, x_{i\beta}, x_{i\gamma})/n_i(n_i-1)(n_i-2)\,,$$

$$A_{21} = 3\sum_{i\neq j}^{L} w_i w_j (1-2w_i) \sum_{\gamma=1}^{n_j}\sum_{\alpha\neq\beta}^{n_i} [\varphi(x_{i\alpha}, x_{i\beta}, x_{i\gamma})$$

$$+\varphi(x_{i\beta}, x_{j\gamma}, x_{i\alpha}) + \varphi(x_{j\gamma}, x_{i\alpha}, x_{i\beta})]/n_i(n_i-1)n_j$$

and

$$A_{31} = 6\sum_{\substack{i\neq j\neq k\\ j\neq k}}^{L} w_i w_j w_k \sum_{\alpha=1}^{n_i}\sum_{\beta=1}^{n_j}\sum_{\gamma=1}^{n_k} \varphi(x_{i\alpha}, x_{j\beta}, x_{k\gamma})/n_i n_j n_k\,.$$

Note that $[(n-2)K^* + 3r_k^*]/(n+1)$ is an unbiased estimator of ρ.

Now

$$(n-2)K^*+3r_K^* = 3\sum_{i=1}^{L}\frac{w_i^2}{n_i(n_i-1)}\sum_{\alpha=1}^{n_i}[2R_{i\alpha}-(n_i+1)][2S_{i\alpha}-n_i(n_i+1)]$$

$$+3\sum_{i\neq j}^{L} w_iw_j\sum_{\alpha=1}^{n_i}[2(R_{i\alpha}^{ij}-R_{i\alpha})-n_j][2(S_{i\alpha}^{ij}-S_{i\alpha})-n_j]/n_in_j$$

$$+3\sum_{i\neq j}^{L} w_iw_j\sum_{\alpha=1}^{n_j}\{[2R_{i\alpha}-(n_i+1)][2(S_{i\alpha}^{ij}-S_{i\alpha})-n_j]$$

$$+[2(R_{i\alpha}^{ij}-R_{i\alpha})-n_j][2S_{i\alpha}-(n_i+1)]\}/n_in_j$$

$$+\sum_{\substack{i\neq j\neq k\\ i\neq k}}^{L} w_iw_j\sum_{\alpha=1}^{n_i}[2(R_{i\alpha}^{ij}-R_{i\alpha})-n_j][2(S_{i\alpha}^{ik}-S_{i\alpha})-n_k]/n_in_j$$

$$+A_{11}+A_{21}-A_{31}$$

$$=\frac{12}{n^2}\sum_{i=1}^{L}\sum_{\alpha=1}^{n_i}\Big\{\sum_{j\neq i}^{L}R_{i\alpha}^{ij}-(L-2)R_{i\alpha}-\frac{n+1}{2}\Big\}\Big\{\sum_{K\neq i}^{L}S_{i\alpha}^{ik}-(L-2)S_{i\alpha}-\frac{n+1}{2}\Big\}$$

$$+A_{11}+A_{21}-A_{31}+A_{41},$$

where

$$A_{41}=3\sum_{i=1}^{L}w_i^2\sum_{\alpha=1}^{n_i}[2R_{i\alpha}-(n_i+1)][2S_{i\alpha}-(n_i+1)]/n_i^2(n_i-1).$$

Thus we finally get the equation

$$(2)\qquad r_s^*=[(n-2)K^*+3r_K^*-A_{11}-A_{21}+A_{31}-A_{41}]/(n+1)$$

where r_s^* is given by (2.24).

Now

$$E(r_s^*)=\rho-E[A_{11}+A_{21}-A_{31}+A_{41}]/(n+1),$$

and we get easily the relation $E[A_{11}+A_{21}-A_{31}+A_{41}]=B$, where B is given in Theorem 2. Thus the proof of (i) is completed. (ii) Since it is easily shown that

$$A_{11}+A_{21}-A_{31}+A_{41}\longrightarrow 0\quad \text{in prob. as}\quad n\longrightarrow\infty$$

and from Theorem 1, $\sqrt{n}\ (K^*-\rho_1)$ has a normal limiting distribution, then from (2) the asymptotic distribution of $\sqrt{n}\ (r_s^*-\rho)$ is equivalent to that of $\sqrt{n}\ (K^*-\rho_1)$. But it is known from Yanagawa, T. and Wakimoto, K. [5] $\sqrt{n}\ (K^*-\rho_1)$ has, as $n\to\infty$, a normal limiting distribution with mean 0 and variance σ_s^{*2} given by (2.27). Thus we complete the proof of the theorem.