ON OPTIMUM STRATIFICATION FOR THE OBJECTIVE VARIABLE BASED ON CONCOMITANT VARIABLES USING PRIOR INFORMATION

YASUSHI TAGA

(Received Dec. 26, 1966)

0. Summary

In estimating the population mean μ_0 of the objective variable Y (uni-variate), the optimum methods for stratification are studied which minimize the variance of the unbiased estimator $\overline{Y}$ for μ_0 based on the concomitant variable X (p-variate). Especially, in case of proportionate sample allocation to each stratum, the optimum stratification above mentioned reduces to the optimum decomposition of the distribution function $H(z)$ for the random variable $Z=\eta(X)$, where $\eta(x)$ is the regression function of Y on X. Further, a general method is shown by which such an optimum stratification can be asymptotically attained.

1. Introduction

The problem of optimum stratification was first considered by C. Hayashi [1] and T. Dalenius [2]. Their formulation for this problem is as follows.

1-1° The distribution function $F(x)$ of the objective variable X is absolutely continuous, and has the finite second moment and p.d.f. $f(x)$ which is positive over $(-\infty, \infty)$. Though $F(x)$ is discrete in the finite population, it may be considered for convenience' sake to be absolutely continuous. In the consequence, the problem of optimum stratification in the finite population is replaced by that of optimum truncations based on the interval division in the infinite population as stated below.

1-2° The population Π is classified into l strata $\Pi_1,\cdots,\Pi_l$. The ith stratum Π_i corresponds to the ith interval $I_i=[x_{i-1}, x_i)$ in the domain of X, where $-\infty=x_0<x_1<\cdots<x_{i-1}<x_i<\cdots<x_l=+\infty$, and to the p.d.f. $f(x)/w_i$, $w_i=\int_{x_{i-1}}^{x_i} f(x)dx$. (The number l of strata is predetermined.)

1-3° The method for allocation $\{n_i\}$ of total sample size n is preassigned as proportionate allocation or Neyman allocation.

1-4° As an unbiased estimator for the mean $\mu=\int_{-\infty}^{\infty} xf(x)dx$,

$$\bar{X}=\sum_{i=1}^{l} w_i\bar{X}_i, \qquad \left(\bar{X}_i=\frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}\right)$$

is used. Among all possible stratifications based on the interval division $\{I_i\}$ stated above, the one which minimizes the variance $V(\bar{X}|\{I_i\})$ of $\bar{X}$ is called "an optimum interval division (stratification)" and it is denoted by $\{I_i^*\}$, $I_i=[x_{i-1}^*, x_i^*)$, $(i=1,\cdots,l)$.

Next, we shall sketch out the results, obtained up to now, of the problem formulated as above.

1) Proportionate allocation $(n_i=w_in)$

A necessary condition that $\{I_i^*\}$ is an optimum stratification is that the relations

$$x_i=\frac{1}{2}(\mu_i+\mu_{i+1}), \qquad i=1, 2,\cdots, l-1, \tag{1.1}$$

where

$$\mu_i=\frac{1}{w_i}\int_{x_{i-1}}^{x_i} xf(x)dx \quad \text{and} \quad w_i=\int_{x_{i-1}}^{x_i} f(x)dx,$$

are satisfied simultaneously for $x_i=x_i^*$.

2) Neyman allocation $\left(n_i=\dfrac{w_i\sigma_i}{\sum_{j=1}^{l} w_j\sigma_j}n\right)$

A necessary condition that $\{I_i^*\}$ is an optimum interval division (stratification) is that the relations

$$\frac{\sigma_i^2+(x_i-\mu_i)^2}{\sigma_i}=\frac{\sigma_{i+1}^2+(x_i-\mu_{i+1})^2}{\sigma_{i+1}}, \qquad i=1, 2,\cdots, l-1, \tag{1.2}$$

where

$$\sigma_i^2=\frac{1}{w_i}\int_{x_{i-1}}^{x_i} (x-\mu_i)^2f(x)dx,$$

are satisfied simultaneously for $x_i=x_i^*$.

Though these results seem to be simple and usable, it is troublesome to obtain x_i^* numerically by the successive approximation method, since w_i, μ_i and σ_i^2 in (1.1) or (1.2) depend on x_i's and $f(x)$. Thus various computation methods, for obtaining x_i^*'s approximately at one trial, have been studied in many papers (see [5]~[12]).

By numerical investigations for various $f(x)$, W. G. Cochran [13] and V. K. Sethi [14] found that the methods given by T. Dalenius and J. L. Hodges, Jr. [9], [10] and G. Ekman [11] are satisfactory in precision for obtaining approximate solutions to x_i^*'s. In doing so, Cochran obtained an interesting result as a by-product that the optimum stratification for equal allocation ($n_i=n/l$) approaches to the one for Neyman allocation, as l becomes large enough ($l\geqq 4$). However, this fact has not yet been proved analytically.

S. P. Ghosh [15] and H. Aoyama [16] obtained the optimum stratification for $f(x)$ in two-variate case $x=(x_1, x_2)$ among all possible lattice divisions of the domain of x bounded by k and l lines parallel to x_2- and x_1-axes respectively. (k and l are predetermined.) These results leave much room for improvement to the general division of the domain of x.

Summarizing the results mentioned above, the following problems are pointed out for further studies:

1) To obtain the optimum stratification for the general division of the domain of X without restrictions of interval or lattice division. It will be called "the general optimum stratification".

2) To obtain sufficient conditions that $(l-1)$ end-points x_i^*'s of $\{I_i^*\}$ become optimum for the division of the domain of X within the limits of interval or lattice one.

3) To obtain the optimum stratification for the objective variable Y based on the concomitant variable X from practical point of view.

2. Formulation of problems

In section 3, we shall show the existence of optimum decomposition $\{F_i^*(x)\}$ of a given distribution function $F(x)$ in the sense of minimizing the variance $V(\bar{X}|\{F_i\})$ of the estimator $\bar{X}$ for the mean μ of $F(x)$, among all possible decompositions $\{F_i\}$ of F as stated below.

3-1° Let $F(x)$ be the distribution function of the objective variable X (uni-variate) and have the finite second moment.

3-2° The number l of strata is preassigned. The population Π with $F(x)$ is classified into l strata $\{\Pi_i\}$, and the non-negative and non-decreasing function $F_i(x)$ which is continuous to the right corresponds to the ith stratum Π_i such that $\sum_{i=1}^{l} F_i(x)=F(x)$ for all x. We shall call $\{F_i(x)\}$ "an l-decomposition of $F(x)$".

3-3° The total sample size n is fixed, and is allocated to each stratum proportionately to each size, i.e. $n_i=w_i n$, $(i=1,\cdots,l)$, where $w_i=F_i(+\infty)$.

3–4° To show the existence of optimum decomposition $\{F_i^*\}$, i.e. a decomposition $\{F_i^*\}$ such that

$$\mathrm{V}(\bar{X}|\{F_i^*\})=\inf_{\{F_i\}}\mathrm{V}(\bar{X}|\{F_i\}).$$

However, the uniqueness of optimum decomposition $\{F_i^*\}$ is not treated in this paper because of its complexity and partly because it will make sense to show its existence without treating its uniqueness.

In section 4, we shall treat the general optimum stratification for the objective variable Y (uni-variate) based on the concomitant variable X (p-variate) as follows:

4–1° Let $G(x)$ be the marginal distribution function of the concomitant variable X, which we assume to be known. Let $\eta(x)$ be the regression function of the objective variable Y on X, which we assume to be unknown, but to belong to $L^2(dG)$, i.e. $\int_{R^p}\eta^2(x)dG(x)<\infty$, where R^p denotes p-dimensional Euclidian space.

4–2° To consider all possible l-decompositions $\{G_i\}$ of $G(x)$, i.e.

$$\sum_{i=1}^{l} G_i(x)=G(x), \qquad \text{for all } x,$$

where $G_i(x)$ is non-negative, non-decreasing function and $G_i(+\infty,\cdots,+\infty)=w_i$. Such a decomposition $\{G_i\}$ will be called "a general stratification for Y based on X".

4–3° Let the total sample size n be fixed, and allocated to each stratum proportionately to its size, i.e. $n_i=w_i n$.

4–4° To obtain a general optimum stratification $\{G_i^*\}$ among all possible decompositions $\{G_i\}$ of G, i.e.

$$\mathrm{V}(\bar{Y}|\{G_i^*\})=\inf_{\{G_i\}}\mathrm{V}(\bar{Y}|\{G_i\}).$$

$\{G_i^*\}$ will be called "a general optimum stratification for Y based on X". This problem reduces to the one obtaining an optimum decomposition $\{H_i^*\}$ of the distribution function $H(z)$ of a new variable Z (uni-variate) defined by the relation $Z=\eta(X)$.

In section 5 a method is shown by which the general optimum stratification $\{G_i^*\}$ can be attained asymptotically, if a jointly measurable function $\hat{\eta}(x, s)$ which converges to $\eta(x)$ in the mean square is obtained by using the prior information s. In section 6 a method for construction of $\hat{\eta}(x, s)$ mentioned above will be stated, which would be called "the randomized least squares method" using an orthogonal expansion of $\eta(x)$ in $L^2(dG)$. In section 7 three examples will be shown, by which the usability of our method is certified. In section 8 conclusions are stated and unsolved problems are summarized for future studies.

3. General optimum stratification based on the objective variable itself

Let $\{F_i(x)\}$ be an l-decomposition of distribution function $F(x)$ of the objective variable X (uni-variate). $F_i(x)$'s are all non-negative and non-decreasing functions which are continuous to the right and satisfy the relations

$$\sum_{i=1}^{l} F_i(x)=F(x), \qquad \text{for all } x, \tag{3.1}$$

and

$$F_i(+\infty)=w_i, \qquad (i=1,\cdots,l).$$

Then, let us call an l-decomposition $\{F_i(x)\}$ of $F(x)$ "a general stratification based on the objective variable X itself".

Now, let us take a sample of size n, $\{(X_{i1},\cdots,X_{in_i}), i=1,\cdots,l\}$, where $(X_{i1},\cdots,X_{in_i})$ is taken from the ith stratum Π_i with the distribution function $F_i(x)/w_i$. Then, we take

$$\bar{X}=\sum_{i=1}^{l} w_i\bar{X}_i, \qquad \left(\bar{X}_i=\frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}\right), \tag{3.2}$$

as an unbiased estimator of the population mean μ of X. It is well-known that the variance of $\bar{X}$ can be obtained as

$$\mathrm{V}(\bar{X}|\{F_i\})=\sum_{i=1}^{l} w_i^2\frac{\sigma_i^2}{n_i}, \tag{3.3}$$

where

$$\sigma_i^2=\frac{1}{w_i}\int_{-\infty}^{\infty}(x-\mu_i)^2dF_i(x) \quad \text{and} \quad \mu_i=\frac{1}{w_i}\int_{-\infty}^{\infty} x\,dF_i(x).$$

If the total sample n is allocated to the ith stratum proportionately to each size, i.e. $n_i=w_in$ $(i=1,\cdots,l)$, then the relation (3.3) becomes

$$\mathrm{V}(\bar{X}|\{F_i\})=\frac{1}{n}\left\{\int_{-\infty}^{\infty} x^2dF(x)-\sum_{i=1}^{l} w_i\mu_i^2\right\}. \tag{3.4}$$

We shall call $\{F_i^*\}$ "a general optimum stratification" when $\{F_i^*\}$ minimizes $\mathrm{V}(\bar{X}|\{F_i\})$.

Now, we shall give another expression to the l-decomposition $\{F_i\}$ of F for convenience' sake to prove the existence of general optimum stratification $\{F_i^*\}$. It is easily seen from (3.1) that the measure dF_i is absolutely continuous with respect to the measure dF, i.e. $dF_i \ll dF$. Then there exists a measurable function vector $\phi=(\phi_1,\cdots,\phi_l)$ such that

$$dF_i(x)=\phi_i(x)dF(x), \qquad (i=1,\cdots,l), \tag{3.5}$$

and

$$\sum_{i=1}^{l}\phi_i(x)=1, \qquad 0\leqq\phi_i(x)\leqq 1, \quad \text{a.e. } (dF). \tag{3.6}$$

Conversely, let us define $\{F_i(x)\}$ by (3.5) for a given $\phi=(\phi_1,\cdots,\phi_l)$ satisfying (3.6). Then it is easily seen that $\{F_i(x)\}$ satisfies (3.1), 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 the one-to-one correspondence between Φ and $\mathcal{F}$ (the set of all decompositions of F). Therefore, we can consider such a function vector ϕ to be a general stratification, and the variance of $\bar{X}$ can be denoted by $\mathrm{V}(\bar{X}|\phi)$ instead of $\mathrm{V}(\bar{X}|\{F_i\})$.

Now, let Φ be the set of all function vectors ϕ satisfying (3.6), and introduce the weak topology[1] into Φ. (See, e.g., Appendix in [18].) Then we can easily prove the following two lemmas. (Proofs are omitted.)

LEMMA 1. *Φ is convex and compact with regard to the weak topology, i.e. weakly compact.*

LEMMA 2. *$\mathcal{U}$, $\mathcal{W}$ and C are all convex and compact, where*

$$\mathcal{U}=\left\{u;\ u=\int_{-\infty}^{\infty}x\phi(x)dF(x),\ \phi\in\Phi\right\}, \tag{3.7}$$

$$\mathcal{W}=\left\{w;\ w=\int_{-\infty}^{\infty}\phi(x)dF(x),\ \phi\in\Phi\right\}, \tag{3.8}$$

and

$$C=\left\{(u,w);\ u=\int_{-\infty}^{\infty}x\phi(x)dF(x),\ \ w=\int_{-\infty}^{\infty}\phi(x)dF(x),\ \phi\in\Phi\right\}.$$

Now, since the first term in the right hand side of (3.4) does not include ϕ, the general optimum stratification ϕ^* in Φ may be defined by

1) Let $\mathcal{F}$ be a family of uniformly bounded measurable functions on a measure space $(\mathcal{X},\mathcal{A},\mu)$. The sequence $\{f_n\}$, $f_n\in\mathcal{F}$, is said to converge to $f\in\mathcal{F}$ weakly if $\int f_n p\,d\mu\to\int fp\,d\mu$ for all functions $p\in L^1(\mu)$. When f is a function vector, the above definition must be considered for each component of f.

the one that attains

$$\sup_{\phi\in\Phi}\sum_{i=1}^{l} w_i\mu_i^2 \tag{3.9}$$

where

$$w_i=\int_{-\infty}^{\infty}\phi_i(x)dF(x) \quad\text{and}\quad \mu_i=\frac{1}{w_i}\int_{-\infty}^{\infty}x\phi_i(x)dF(x).$$

Let us consider the objective function

$$v(u, w)=\sum_{i=1}^{l}\frac{u_i^2}{w_i}, \qquad u_i=\int_{-\infty}^{\infty}x\phi_i(x)dF(x), \tag{3.10}$$

which is bounded and continuous except the origin on the compact subset C of $2l$-dimensional Euclidean space.

Since $\int_{-\infty}^{\infty}x^2dF(x)<\infty$, it is easily seen that $\frac{u_i^2}{w_i}\to 0$ as $w_i\to 0$ for any selection of ϕ_i. When we define $\frac{u_i^2}{w_i}=0$ for $w_i=0$, the function $v(u, w)$ is continuous on the compact set C including the origin. Therefore, $v(u, w)$ attains its supremum on C.

LEMMA 3. *$v(u, w)$ attains its supremum in the compact set C. Namely, there exists (u^*, w^*) in C such that*

$$v^*=v(u^*, w^*), \tag{3.11}$$

where v^* is defined by

$$v^*=\sup_{(u,w)\in C} v(u, w). \tag{3.12}$$

It should be noted that ϕ^* which corresponds to (u^*, w^*) is not always uniquely determined in Φ.

Next, we shall show that the general optimum stratification ϕ^* coincides with an indicator function vector $\{\chi_{I_i^*}(x), 1\leqq i\leqq l\}$ corresponding to some interval division $\{I_i^*, 1\leqq i\leqq l\}$ of the real line $(-\infty, \infty)$.

It should be noted that $w^*>0$, i.e. $w_i^*>0$ for $i=1,\cdots, l$. In fact, the assumption that $w_i^*=0$ for some i leads to a contradiction that there exists a better ϕ other than ϕ^*: We can select a ϕ_j^*, $(w_j^*>0)$, which does not degenerate at one point if $F(x)$ has increasing points more than or equal to l. Let us divide ϕ_j^* into ϕ_{j1} and ϕ_{j2} such that

$$w_{j1}+w_{j2}=w_j^*, \quad u_{j1}+u_{j2}=u_j^* \quad\text{and}\quad \frac{u_{j1}}{w_{j1}}\neq\frac{u_{j2}}{w_{j2}},$$

where

$$w_{jk}=\int_{-\infty}^{\infty}\phi_{jk}dF \quad \text{and} \quad u_{jk}=\int_{-\infty}^{\infty}x\phi_{jk}dF, \qquad (k=1,2).$$

Then, it is easily seen that

$$\frac{u_{j1}^2}{w_{j1}}+\frac{u_{j2}^2}{w_{j2}}-\frac{u_j^{*2}}{w_j^*}=\frac{w_{j1}w_{j2}}{w_j^*}\left(\frac{u_{j1}}{w_{j1}}-\frac{u_{j2}}{w_{j2}}\right)^2>0. \tag{3.13}$$

Let ϕ be defined by

$$\phi=(\phi_1^*,\cdots,\phi_{i-1}^*,\phi_{i+1}^*,\cdots,\phi_{j-1}^*,\phi_{j1},\phi_{j2},\phi_{j+1}^*,\cdots,\phi_l^*),$$

and $v(u, w)$ be the function corresponding to ϕ. From (3.13), it is seen that ϕ is better than ϕ^*, i.e. $v(u, w)>v(u^*, w^*)$, which contradicts the definition of ϕ^*. Therefore, $w^*>0$ and the supremum operation in (3.12) may be considered within the range $w>0$.

Now we shall show that the general optimum stratification ϕ_w^*, for any fixed $w>0$, will correspond to some interval division in the wide sense as stated below.

LEMMA 4. *Let Φ_w and $\mathcal{U}_w$ be defined by*

$$\Phi_w=\left\{\phi;\int_{-\infty}^{\infty}\phi(x)dF(x)=w,\ \phi\in\Phi\right\} \tag{3.14}$$

and

$$\mathcal{U}_w=\left\{u;\ u=\int_{-\infty}^{\infty}x\phi(x)dF(x),\ \phi\in\Phi_w\right\} \tag{3.15}$$

respectively for any fixed $w>0$. Then there exists a general optimum stratification ϕ_w^* in Φ_w which corresponds to an interval division such that each component ϕ_{wi}^* of ϕ_w^* coincides with the indicator function $\chi_{I_i^*}$ of the ith interval I_i^* except in its end points. Namely, $(\phi_{w1}^*,\cdots,\phi_{wl}^*)$ may have positive probability masses in common at the end points.

PROOF. It is noted that $\mathcal{U}_w$ is convex and compact like $\mathcal{U}$. The fact that ϕ_w^* corresponds to some interval division except in its end points is equivalent to the one that there exists a real number c satisfying the relation

$$\int_{c+0}^{\infty}\phi_{wi}^*(x)dF(x)=\int_{-\infty}^{c-0}\phi_{wj}^*(x)dF(x)=0, \qquad \text{for } i\neq j, \tag{3.16}$$

where ϕ_{wi}^* denotes the ith component of ϕ_w^*. (Here we suppose that $u_i^*/w_i\leqq u_j^*/w_j$.) Let us show that the assumption of non-existence of such a real number c leads to a contradiction.

Under the assumption stated above, it is easily seen that there exists a real number d $(u_i^*/w_i\leqq d\leqq u_j^*/w_j)$ satisfying the relations

$$(3.17) \qquad \int_{d+0}^{\infty} \phi_{wi}^*(x)dF(x)>0 \quad \text{and} \quad \int_{-\infty}^{d-0} \phi_{wj}^*(x)dF(x)>0.$$

Then there exist k_1 and k_2 $(0<k_1,\ k_2<1)$ such that the relation

$$(3.18) \qquad k_1\int_{d+0}^{\infty} \phi_{wi}^*(x)dF(x)=k_2\int_{-\infty}^{d-0} \phi_{wj}^*(x)dF(x)$$

is satisfied. Now let $\phi=(\phi_1,\cdots,\phi_l)$ be defined by

$$(3.19) \qquad \phi_i(x)=\begin{cases} \phi_{wi}^*(x)+k_2\phi_{wj}^*(x) & \text{if } x<d, \\ \phi_{wi}^*(x) & \text{if } x=d, \\ \phi_{wi}^*(x)-k_1\phi_{wi}^*(x) & \text{if } x>d, \end{cases}$$
$$\phi_j(x)=\begin{cases} \phi_{wj}^*(x)-k_2\phi_{wj}^*(x) & \text{if } x<d, \\ \phi_{wj}^*(x) & \text{if } x=d, \\ \phi_{wj}^*(x)+k_1\phi_{wi}^*(x) & \text{if } x>d, \end{cases}$$

and $\phi_h(x)=\phi_{wh}^*(x)$ for all x $(h\neq i, j)$.

Then it is easily seen from (3.18) that $\phi\in\Phi_w$, i.e., $\int_{-\infty}^{\infty}\phi(x)dF(x)=w$, and that $\int_{-\infty}^{\infty}x\phi(x)dF(x)=u\in\mathcal{U}_w$. On the other hand, from (3.18) and (3.19), it is easily verified that the relations

$$u_i=u_i^*-\left\{k_1\int_{d+0}^{\infty}x\phi_{wi}^*(x)dF(x)-k_2\int_{-\infty}^{d-0}x\phi_{wj}^*(x)dF(x)\right\}<u_i^*,$$
$$u_j=u_j^*+\left\{k_1\int_{d+0}^{\infty}x\phi_{wi}^*(x)dF(x)-k_2\int_{-\infty}^{d-0}x\phi_{wj}^*(x)dF(x)\right\}>u_j^*,$$

and $u_h=u_h^*$, $(h\neq i, j)$, hold. Hence it can be seen after some calculations that the relation

$$v(u, w)-v(u^*, w)=(u_j-u_j^*)\left\{\left(\frac{u_j}{w_j}-\frac{u_i}{w_i}\right)+\left(\frac{u_j^*}{w_j}-\frac{u_i^*}{w_i}\right)\right\}>0$$

holds. Since this result contradicts the definition of ϕ_w^*, there exists a real number c satisfying (3.16). Thus our assertion has been proved.

Next, let us consider the supremum of $v_w^*=v(u_w^*, w)$ corresponding to ϕ_w^*, given in lemma 4, for various $w\in\mathcal{W}$. It is easily seen that the relation

$$(3.20) \qquad v^*=\sup_{w\in\mathcal{W}} v_w^*$$

holds, where $v^*=v(u^*, w^*)=\sup_{\phi\in\Phi} v(u, w)$ is given in (3.11). From the compactness of $\mathcal{W}$, there exists a w^* which attains v^*, so there exists a general optimum stratification ϕ^* corresponding to w^*. Considering lemmas 1 to 4, together with the remark just stated above, we can show the following theorem.

THEOREM 1. *Let $F(x)$ be the distribution function of X (uni-variate) with the finite second moment and increasing points more than or equal to l. Let Φ be the set of all general stratification ϕ, and the total sample size n are allocated to each stratum proportionately to each size, i.e. $n_i = w_i n$. Then, there exists a general optimum stratification ϕ^* in Φ which minimizes the variance* $V(\bar{X}|\phi)$ *of the unbiased estimator $\bar{X}$ for the population mean μ, given by* (3.2). *Moreover, ϕ^* coincides with the indicator function vector $\{\chi_{I_i^*}\}$ of an interval division $\{I_i^*\}$. Every end point x_i^* in $\{I_i^*\}$ can be taken at a continuity point of $F(x)$ such that the condition* (1.1) *is satisfied.*

Remark. 1) We call $x=a$ "an increasing point of $F(x)$" if $F(a+\delta)-F(a-\delta)>0$ holds for any $\delta>0$.

2) The last assertion in the theorem can be verified as follows: We can assume $l=2$, $u_1+u_2=0$ and $u_1<u_2$ without loss of generality. Let the optimum division point x_1^* be a discontinuity point of $F(x)$ in a vicinity of which $F(x)$ is strictly increasing. Then the differential variation δv of $v(u, w)$, corresponding to the differential variation δw, is expressed by

$$\delta v \sim \frac{u_1^*\{2x_1^*-(\mu_1^*+\mu_2^*)\}\delta w+o\{(\delta w)^2\}}{(w_1^*-\delta w)(w_2^*+\delta w)},$$

where

$$\mu_i^* = u_i^*/w_i^*, \qquad i=1, 2.$$

Hence, we can see that x_1^* can not be an optimum (maximal) point, but that x_1^* can be optimum either if $F(x)$ is continuous and increasing at $x=x_1^*$ or if $F(x)$ is discontinuous at $x=x_1^*$ and not strictly increasing in a vicinity of x_1^*. Even in the latter case, we can take x_1^* as a continuity point of $F(x)$ for which (1.1) is satisfied.

This theorem gives a complete extension of the result, obtained by C. Hayashi and T. Dalenius for the proportionate allocation, in the sense that there exists a general optimum stratification ϕ^* for the general distribution function $F(x)$ (without assuming its absolute continuity), and ϕ^* corresponds to a "true" interval division $\{I_i^*\}$ of the real line. The latter part of theorem 1 has a meaning, since the general optimum stratification ϕ^* does not necessarily correspond to the interval division $\{I_i^*\}$ for Neyman allocation (K. Isii [19]). Moreover, K. Isii [19] shows that the general optimum stratification ϕ also plays an important role in proving the existence of optimum ϕ^* in the case when X is multi-variate.

4. General optimum stratification for objective variable based on concomitant variables

In this section, we shall give an answer to the question given in the formulations 4–1°∼4–4° in section 2. At first, we introduce a measurable function vector $\phi=(\phi_1,\cdots,\phi_l)$ corresponding to the l-decomposition $\{G_i\}$ of the marginal distribution function $G(x)$ of the concomitant variable X (p-variate), i.e. ϕ such that

$$dG_i(x)=\phi_i(x)dG(x) \qquad \text{for all } x, \quad i=1,\cdots,l \tag{4.1}$$

and

$$\sum_{i=1}^{l}\phi_i(x)=1, \qquad \phi_i(x)\geqq 0, \qquad \text{a.e. } (dG). \tag{4.2}$$

Let us call ϕ "a general stratification for the objective variable Y based on the concomitant variable X". Under the general stratification ϕ for Y based on X, let us consider the unbiased estimator $\bar{Y}$ for the population mean μ_0 of Y such that

$$\bar{Y}=\sum_{i=1}^{l} w_i\bar{Y}_i, \tag{4.3}$$

where

$$\bar{Y}_i=\frac{1}{n_i}\sum_{j=1}^{n_i}Y_{ij} \quad \text{and} \quad w_i=\int_{R^p}\phi_i(x)dG(x).$$

For the proportionate allocation ($n_i=w_in$), it is well-known that the variance $V(\bar{Y}|\phi)$ of $\bar{Y}$ under ϕ becomes

$$V(\bar{Y}|\phi)=\frac{1}{n}\left\{\sigma_0^2(1-\eta_{YX}^2)+\sum_{i=1}^{l}\int_{R^p}[\eta(x)-\mu_{0i}]^2\phi_i(x)dG(x)\right\}, \tag{4.4}$$

where σ_0^2 is the population variance of Y, η_{YX} is the multiple correlation ratio of Y on X, μ_{0i} is the mean of Y in the ith stratum, and $\eta(x)$ is the regression function of Y on X. The first term in the bracket of the right hand side of (4.4) does not depend on ϕ, and becomes small as η_{YX} approaches to 1. The second term, however, depends on ϕ. Therefore, it is sufficient to show the existence of the optimum ϕ^* which attains

$$\inf_{\phi\in\Phi}\sum_{i=1}^{l}\int_{R^p}[\eta(x)-\mu_{0i}]^2\phi_i(x)dG(x). \tag{4.5}$$

Since

$$\sum_{i=1}^{l}\int_{R^p}[\eta(x)-\mu_{0i}]^2\phi_i(x)dG(x)=\int_{R^p}\eta^2(x)\,dG(x)-\sum_{i=1}^{l}w_i\mu_{0i}^2,$$

it may be sufficient to show the existence of ϕ^* which attains

$$v^* = \sup_{\phi \in \Phi} \sum_{i=1}^{l} \frac{u_i^2}{w_i}, \tag{4.6}$$

where

$$u_i = \int_{R^p} \eta \phi_i dG \quad \text{and} \quad w_i = \int_{R^p} \phi_i dG .$$

Let us consider the transformation $Z = \eta(X)$. Let $H(z)$ be the distribution function of Z, and suppose that the number of increasing points of $H(z)$ is larger than or equal to l. Further, let $\psi = (\psi_1, \cdots, \psi_l)$ be the l-decomposition of $H(z)$ induced by ϕ through the transformation $Z = \eta(X)$. Then it is easily verified that the relations

$$u_i = \int_{-\infty}^{\infty} z\psi_i(z) dH(z) \quad \text{and} \quad w_i = \int_{-\infty}^{\infty} \psi_i(z) dH(z) \tag{4.7}$$

hold. Now, by applying theorem 1 with $H(z)$ and Ψ (the set of all possible ψ) in place of $G(x)$ and Φ, we can show the existence of ψ^* which attains

$$\sup_{\psi \in \Psi} \sum_{i=1}^{l} \frac{u_i^2}{w_i} \tag{4.8}$$

and that ψ^* coincides with the indicator function vector $\{\chi_{I_i^*}\}$ corresponding to an interval division $\{I_i^*\}$, $I_i^* = [z_{i-1}^*, z_i^*)$, of the real line $(-\infty, \infty)$. Further, division points z_i^*'s can be taken at continuity points of $H(z)$ by the remark of theorem 1. Hence, the general optimum stratification ϕ^*, corresponding to ψ^*, can be expressed in the following form:

$$\phi_i^*(x) = \begin{cases} 1 & \text{if } \eta(x) \in I_i^*, \\ 0 & \text{otherwise} \quad (1 \leqq i \leqq l) . \end{cases} \tag{4.9}$$

Therefore, we obtain the following theorem.

THEOREM 2. *Under the assumptions that $\eta(x) \in L^2(dG)$ and that the distribution function $H(z)$ has the increasing points more than or equal to l, there exists a general optimum stratification ϕ^* for Y (uni-variate) based on X (p-variate), and ϕ^* is expressed in the form* (4.9), *where $H(z)$ denotes the distribution function of $Z = \eta(X)$. (The sample allocation is supposed to be the proportionate one.)*

5. Asymptotically optimum stratification for Y based on X using prior information s

In section 4, we have seen that there exists a general optimum stratification $\phi^*=(\phi_1^*,\cdots,\phi_l^*)$ for the objective variable Y based on the concomitant variable X, and that ϕ^* is expressed in the following form:

$$\phi_i^*(x)=\begin{cases}1 & \text{if } \eta(x)\in I_i^*, \quad I_i^*=[z_{i-1}^*, z_i^*) \\ 0 & \text{otherwise}, \ (1\leqq i\leqq l)\end{cases} \tag{5.1}$$

where $\{I_i^*\}$ is an interval division of the real line $(-\infty,\infty)$. However, we can not get such a ϕ^* into our hand at once, since the regression function $\eta(x)$ is generally unknown on which ϕ^* depends.

In this section, we try to obtain an asymptotically optimum stratification $\hat{\phi}^*(x, s)$ based on the estimated function $\hat{\eta}(x, s)$ for the regression function $\eta(x)$, using prior information s. The procedure to obtain $\hat{\phi}^*$ will be stated in the following way.

5–1° Let $s=\{(X_\alpha', Y_\alpha'), 1\leqq\alpha\leqq m\}$ be the prior information (the 1st sample) of size m and $\boldsymbol{Y}=\{(Y_{i1},\cdots, Y_{in_i}), 1\leqq i\leqq l\}$, the 2nd sample, which is independent of s. Based on the second sample $\boldsymbol{Y}$, the estimator $\bar{Y}$ for μ_0 is constructed. Let us assume that there exists an estimated jointly measurable function $\hat{\eta}(x, s)$ in $L^2(dG\cdot dF)$ for $\eta(x)$, satisfying the condition

$$\lim_{m\to\infty}\iint[\hat{\eta}(x, s)-\eta(x)]^2dG(x)dF(s)=0, \tag{5.2}$$

where $F(x)$ denotes the distribution function for prior information s, which is defined by the natural conditional distribution of Y, given X, and the marginal distribution of X obtained by transforming the natural $G(x)$.

Further, let us assume that the increasing points of $\hat{H}(z, s)$ are larger than or equal to l (the preassigned number of strata) for each fixed s, where $\hat{H}(z, s)$ denotes the distribution function of the random variable $Z(s)=\hat{\eta}(X, s)$.

5–2° To obtain the asymptotically optimum stratification $\hat{\phi}^*(x, s)$ for a.a. s in accordance with the procedure stated in section 4, considering $\hat{\eta}(x, s)$ to be a regression function instead of the true regression function $\eta(x)$ for a.a. fixed s. It is possible because of the assumptions in 5–1° and theorem 2 in section 4. Here, it should be noted that $\hat{\phi}^*(x, s)$ coincides for a.a. s with an indicator function vector $\{\chi_{\hat{E}_i^*(s)}, 1\leqq i\leqq l\}$, where $\hat{E}_i^*(s)=\{x; \hat{\eta}(x, s)\in\hat{I}_i^*(s)\}$ and $\{\hat{I}_i^*(s)\}$ is a suitable interval division of the real line $(-\infty,\infty)$, and that $\hat{\phi}^*(x, s)$ is not always jointly measurable.

5–3° Let the loss $L_m(s)$, caused by using the asymptotically optimum $\hat{\phi}^*(x, s)$ instead of the optimum $\phi^*(x)$, be defined by

$$(5.3)\qquad L_m(s)=v^*-v(\hat{u}^*(s),\ \hat{w}^*(s))\geqq 0\,,$$

where

$$v(u,w)=\sum_{i=1}^{l}\frac{u_i^2}{w_i}\,,\qquad \hat{u}_i^*(s)=\int_{R^p}\eta(x)\hat{\phi}_i^*(x,s)\,dG(x)\,,$$

$$\hat{w}_i^*(s)=\int_{R^p}\hat{\phi}_i^*(x,s)dG(x)$$

and v^* is defined by (4.6). Then, we can show that "the upper expected loss" of $L_m(s)$ tends to zero as $m\to\infty$. (In this sense, $\hat{\phi}^*$ may be called "asymptotically optimum".)

LEMMA 5. *For any jointly measurable function $\zeta(x,s)$ in $L^2(dG\cdot dF)$, the inequality*

$$(5.4)\qquad \sum_{i=1}^{l}\frac{1}{w_i}\left[\int_{R^p}\zeta(x,s)\phi_i(x)\,dG(x)\right]^2\leqq\int_{R^p}\zeta^2(x,s)dG(x)$$

holds, where ϕ denotes a general stratification for Y based on X and $w_i=\int_{R^p}\phi_i(x)dG(x)$.

PROOF. Since $\zeta(x,s)$ is square integrable in x for a.a. s by the Fubini's theorem, it holds a.e. (dF) that $\infty>\int_{R^p}\zeta^2dG=\sum_{i=1}^{l}\int_{R^p}\{[\zeta-\mu_i]+\mu_i\}^2\cdot\phi_i dG\geqq\sum_{i=1}^{l}w_i\mu_i^2$, where $\mu_i=\frac{1}{w_i}\int_{R^p}\zeta\phi_i dG$. For any s where $\int_{R^p}\zeta^2dG=\infty$, the inequality (5.4) holds clearly. Thus, our assertion has been proved.

LEMMA 6. *Under the assumptions in 5-1°, for any ϕ in Φ the inequality*

$$(5.5)\qquad |\,v(\hat{u}(s),w)-v(u,w)\,|\leqq\sqrt{\gamma(s)\cdot\delta(s)}$$

holds, where

$$\hat{u}(s)=\int_{R^p}\hat{\eta}(x,s)\phi(x)dG(x)\,,\quad u=\int_{R^p}\eta(x)\phi(x)dG(x)\,,$$

$$w=\int_{R^p}\phi(x)dG(x)\,,\quad \gamma(s)=\int_{R^p}[\hat{\eta}+\eta]^2dG\quad \textit{and}\quad \delta(s)=\int_{R^p}[\hat{\eta}-\eta]^2dG\,.$$

PROOF. By the Schwarz's inequality, we obtain

$$(5.6)\qquad |\,v(\hat{u}(s),w)-v(u,w)\,|=\left|\sum_{i=1}^{l}\frac{[\hat{u}_i(s)+u_i]}{\sqrt{w_i}}\frac{[\hat{u}_i(s)-u_i]}{\sqrt{w_i}}\right|$$

$$\leqq\sqrt{\sum_{i=1}^{l}\frac{[\hat{u}_i(s)+u_i]^2}{w_i}}\sqrt{\sum_{i=1}^{l}\frac{[\hat{u}_i(s)-u_i]^2}{w_i}}\,.$$

Applying lemma 5 to (5.6) for $\zeta=\hat{\eta}\pm\eta$, we can obtain the inequality (5.5).

LEMMA 7. *Under the assumptions in* 5–1°, *there exists a constant* K, *independent of* m, *such that the inequality*

$$\int_{R^{2m}} \sqrt{\gamma(s)\cdot\delta(s)}\, dF(s) \leqq K\sqrt{\delta_m} \tag{5.7}$$

holds, where $\delta_m=\iint[\hat{\eta}-\eta]^2dGdF$.

PROOF. Applying the Schwarz's inequality, we obtain

$$\int_{R^{2m}} \sqrt{\gamma(s)\delta(s)}\, dF(s) \leqq \sqrt{\int_{R^{2m}} \gamma(s)dF(s)}\sqrt{\int_{R^{2m}} \delta(s)dF(s)}\ . \tag{5.8}$$

By the definitions of $\gamma(s)$ and $\delta(s)$ in lemma 6, we can see that

$$\int_{R^{2m}} \delta(s)dF(s)=\iint[\hat{\eta}-\eta]^2dGdF=\delta_m \tag{5.9}$$

and

$$\int_{R^{2m}} \gamma(s)dF(s)=\iint[\hat{\eta}+\eta]^2dGdF. \tag{5.10}$$

Since it follows from (5.2) that $\delta_m\to 0$ as $m\to\infty$, we can see that

$$\lim_{m\to\infty}\int_{R^{2m}} \gamma(s)dF(s)=4\int_{R^p}\eta^2dG(x). \tag{5.11}$$

Hence, the existence of a constant K in (5.7), independent of m, can be easily proved.

By lemmas 5, 6 and 7, we can evaluate the "upper expected loss" $\bar{L}_m$ of $L_m(s)$ as in the following theorem.

THEOREM 3. *Under the assumptions in* 5–1°, *the loss* $L_m(s)$, *caused by using the asymptotically optimum* $\hat{\phi}^*(x,s)$ *instead of the optimum* $\phi^*(x)$, *can be evaluated by the inequality*

$$0\leqq L_m(s)=v^*-v(\hat{u}^*(s),\hat{w}^*(s))\leqq 2\sqrt{\gamma(s)\cdot\delta(s)}\ , \tag{5.12}$$

where

$$v^*=v(u^*,w^*),\quad u^*=\int_{R^p}\eta\phi^*dG,\quad w^*=\int_{R^p}\phi^*dG,$$

$$\hat{u}^*(s)=\int_{R^p}\eta\hat{\phi}^*dG,\quad \hat{w}^*=\int_{R^p}\hat{\phi}^*dG,$$

and $\gamma(s)$ *and* $\delta(s)$ *are given in lemma* 6.

Further, the upper expected loss $\bar{L}_m$ *of* $L_m(s)$ *(see the remark* 3) *can be evaluated by the inequality*

$$0 \leqq \bar{L}_m \leqq 2K\sqrt{\delta_m}\,, \tag{5.13}$$

where K *and* δ_m *are given in lemma* 7.

PROOF. Putting $\phi=\phi^*$ and $\phi=\hat{\phi}^*$ in (5.5), we can obtain the inequalities,

$$v^* \leqq \hat{v}^*(s) + \sqrt{\gamma(s)\cdot\delta(s)}\,, \tag{5.14}$$

and

$$|\, v(\hat{u}^*(s), \hat{w}^*(s)) - \hat{v}^*(s) \,| \leqq \sqrt{\gamma(s)\delta(s)}\,, \tag{5.15}$$

where

$$\hat{v}^*(s) = v(\hat{\hat{u}}^*(s), \hat{w}^*(s)) \quad \text{and} \quad \hat{\hat{u}}^*(s) = \int_{R^p} \eta\hat{\phi}^* dG\,.$$

From (5.15), we can get easily the inequality

$$\hat{v}^*(s) \leqq v^* + \sqrt{\gamma(s)\delta(s)}\,. \tag{5.16}$$

From (5.14) and (5.16), we obtain the inequality

$$|\, \hat{v}^*(s) - v^* \,| \leqq \sqrt{\gamma(s)\cdot\delta(s)}\,. \tag{5.17}$$

Thus, we can obtain the formula (5.12), from (5.15), (5.17) and the relation that $v^* - v(\hat{u}^*(s), \hat{w}^*(s)) \geqq 0$. The inequality (5.13) is a direct result of (5.12).

Remark. 1) The asymptotically optimum stratification $\hat{\phi}^*(x, s)$ is not necessarily uniquely determined for each fixed s.

2) For any selection of $\hat{\phi}^*(x, s)$ for each s among all possible $\hat{\phi}^*$'s, the evaluation formulae (5.12) and (5.13) always hold, but $\hat{\phi}^*(x, s)$ is not necessarily jointly measurable. Therefore, $v(\hat{u}^*(s), \hat{w}^*(s))$ and $L_m(s)$ are not necessarily measurable in s. Hence we can not consider the expected loss of $L_m(s)$ in the usual sense.

3) The "upper expected loss" $\bar{L}_m$ of $L_m(s)$ in this theorem have to be considered in the following sense. The upper integral of a non-measurable function $f(x)$ is defined usually by

$$\overline{\int} f(x)d\mu(x) = \inf_{f \leqq g} \int g(x)d\mu(x)\,, \tag{5.18}$$

where $g(x)$ varies over the family of all measurable functions dominating

$f(x)$ a.e. with respect to the measure μ. $\bar{L}_m$ can be defined by (5.18) taking $L_m(s)$ for $f(x)$.

4) Under the assumption of (5.2), we can see that the upper expected loss $\bar{L}_m$ tends to zero as $m\to\infty$. In this sense, $\hat{\phi}^*$ may be called "asymptotically optimum". In practical situations, however, $\hat{\phi}^*$ is used for finite m. Therefore the above mentioned assertion can be taken as its mathematical justification, and the following remark will be important in practice.

5) We can consider δ_m, appearing in the evaluation formula (5.13), to be a measure for the degree of approximation of $\hat{\phi}^*$ to ϕ^*. So we shall show a method for construction of $\hat{\eta}(x, s)$ satisfying (5.2), and obtain $\hat{\phi}^*(x, s)$ based on this $\hat{\eta}(x, s)$ in the next section.

6) The constant K can be taken as $2\sqrt{\int_{R^p}\eta^2 dG}=2\|\eta\|$ for sufficiently large m.

In connection with theorem 3, we shall mention some theoretically interesting results. Let $\tilde{v}^*$ be defined by

$$\tilde{v}^*=\sup_{\phi\in\Phi_1}\int_{R^{2m}} v(\hat{u}(s), w(s))dF(s), \tag{5.19}$$

where Φ_1 is the set of all jointly measurable function vectors $\phi=(\phi_1,\cdots,\phi_l)$ satisfying the relation, for any fixed s

$$\sum_{i=1}^{l}\phi_i(x, s)=1, \qquad \phi_i(x, s)\geqq 0, \qquad \text{for all } x, \text{ a.e. } (dF). \tag{5.20}$$

Then the following relation can be obtained as (5.17),

$$|\tilde{v}^*-v^*|\leqq K\sqrt{\delta_m}\ . \tag{5.21}$$

Moreover, we can show that $\hat{v}^*(s)$, given in (5.15), is measurable and integrable in s, and that the inequality

$$\tilde{v}^*\leqq\int_{R^{2m}}\hat{v}^*(s)dF(s) \ \left(\leqq\iint\hat{\eta}^2 dG\,dF\right) \tag{5.22}$$

holds from the definition of $\hat{v}^*(s)$. If the equality holds in (5.22) and if there exists a $\tilde{\phi}^*$ in Φ_1 which attains $\tilde{v}^*$ [1], then the relation

$$v(\tilde{\hat{u}}^*(s),\hat{w}^*(s))=\hat{v}^*(s), \qquad \text{a.e. } (dF), \tag{5.23}$$

[1] Prof. K. Isii pointed out to the author that the equality in (5.22) always holds and that there exists a jointly measurable function $\tilde{\phi}^*(x, s)$ in Φ_1 which attains $\tilde{v}^*$ in (5.19). These results will be published in the near future.

must be satisfied, where $\tilde{\tilde{u}}^*(s)=\int\hat{\eta}\tilde{\phi}^*dG$ and $\tilde{w}^*(s)=\int\tilde{\phi}^*dG$. This relation implies that among all possible optimum stratifications for each s we can take $\hat{\phi}^*(x, s)$ so as to coincide with the jointly measurable $\tilde{\phi}^*(x, s)$ a.e. (dG) for a.a. s. In such a case, the loss function $L_m(s)$ may be considered to be measurable in s, and the expected loss $\bar{L}_m$ can be defined in the usual sense.

6. Construction of estimator for regression function

In section 5, we have shown the existence of an asymptotically optimum stratification $\hat{\phi}^*(x, s)$ under the assumption that there exists an estimated function $\hat{\eta}(x, s)$ for $\eta(x)$ satisfying the condition

$$\delta_m=\iint[\hat{\eta}(x, s)-\eta(x)]^2 dG(x)dF(s)\to 0 \qquad \text{as } m\to\infty. \tag{6.1}$$

In this section, we shall show the existence of $\hat{\eta}(x, s)$ stated above, using the orthogonal expansion in $L^2(dG)$. Of course, it is supposed that $\hat{\eta}(x, s)\in L^2(dG\cdot dF)$ is jointly measurable and $\eta(x)\in L^2(dG)$. Let $\{\psi_j(x), j=0, 1,\cdots\}$ be a C.O.N.S. in $L^2(dG)$. Then, it is wellknown that the relation

$$\lim_{r\to\infty}\int_{R^p}[\eta_r(x)-\eta(x)]^2 dG(x)=0 \tag{6.2}$$

holds, where

$$\eta_r(x)=\sum_{j=0}^{r} a_j\psi_j(x) \quad \text{and} \quad a_j=\int_{R^p}\eta\psi_j dG.$$

On the other hand, considering that $\hat{\eta}(x, s)\in L^2(dG)$ a.e. (dF), we can obtain the relation

$$\lim_{r\to\infty}\int_{R^p}[\hat{\eta}(x, s)-\eta_r(x, s)]^2 dG(x)=0, \qquad \text{a.e. } (dF), \tag{6.3}$$

where

$$\eta_r(x, s)=\sum_{j=0}^{r} a_j(s)\psi_j(x) \quad \text{and} \quad a_j(s)=\int_{R^p}\hat{\eta}(x, s)\psi_j(x)dG(x).$$

It is noted here that $a_j(s)\in L^2(dF)$, because we have

$$\int_{R^{2m}} a_j^2(s)\,dF(s)\leqq\int_{R^{2m}}\int_{R^p}\hat{\eta}^2(x, s)dG(x)dF(s)\cdot\int_{R^p}\psi_j^2(x)dG(x)<\infty.$$

Now, we shall construct an unbiased estimator $\hat{\eta}_r(x, s)$ for $\eta_r(x)$ such that

$$\hat{\eta}_r(x, s) = \sum_{j=0}^{r} \hat{a}_j(s)\phi_j(x), \tag{6.4}$$

with $\hat{a}_j(s)$ subject to the condition

$$\int_{R^{2m}} \hat{a}_j(s)dF(s) = a_j \quad \text{and} \quad \int_{R^{2m}} \hat{a}_j^2(s)dF(s) < \infty .$$

Since $\hat{\eta}_r(x, s)$, defined in (6.4), satisfies the relation

$$\iint [\hat{\eta}_r(x, s) - \eta(x)]^2 dG(x)dF(s) = \sum_{j=0}^{r} v_j + \sum_{j=r+1}^{\infty} a_j^2, \tag{6.5}$$

where

$$a_j = \int_{R^p} \eta\phi_j dG \quad \text{and} \quad v_j = \int_{R^{2m}} [\hat{a}_j(s) - a_j]^2 dF(s),$$

it is sufficient for proving (6.1) to show that there exists a $\hat{\eta}_r(x, s)$ satisfying the condition

$$\sum_{j=0}^{r} v_j \to 0 \qquad \text{as} \quad m \to \infty, \tag{6.6}$$

where m may depend on r. For constructing such a $\hat{\eta}_r(x, s)$, the following relations are useful (see, Ermakov and Zolotuxin [20]): For any $f \in L^2(dG)$ the formulae

$$\int_{R^p} \cdots \int_{R^p} \frac{\omega_f^{(i)}}{\omega} \frac{\omega^2}{(r+1)!} dG(x_0) \cdots dG(x_r) = \int_{R^p} f(x)\phi_i(x)dG(x) \tag{6.7}$$

and

$$\begin{aligned} &\int_{R^p} \cdots \int_{R^p} \left[\frac{\omega_f^{(i)}}{\omega}\right]^2 \frac{\omega^2}{(r+1)!} dG(x_0) \cdots dG(x_r) \\ &\qquad = \int_{R^p} f^2(x)dG(x) - \sum_{\substack{j=0 \\ \neq i}}^{r} \left[\int_{R^p} f(x)\phi_j(x)dG(x)\right]^2 \end{aligned} \tag{6.8}$$

hold, where

$$\omega = \omega(x_0, \cdots, x_r) = \begin{vmatrix} \phi_0(x_0) & \cdots & \phi_0(x_r) \\ \phi_1(x_0) & \cdots & \phi_1(x_r) \\ \cdots & \cdots & \cdots \\ \phi_r(x_0) & \cdots & \phi_r(x_r) \end{vmatrix}, \qquad \omega_f^{(i)}(x_0, \cdots, x_r) = \sum_{j=0}^{r} f(x_j)\Omega_{ij},$$

and Ω_{ij} is the (i, j)th cofactor of ω.

LEMMA 8. *If the random vector $(X_0', \cdots, X_r')$ is distributed according to the joint probability distribution* $\dfrac{\omega^2(x_0', \cdots, x_r')}{(r+1)!} \prod_{j=0}^{r} dG(x_j')$, *then the* *following relations hold*:

$$\text{(6.9)} \qquad \mathrm{E}\left\{\frac{\omega_\eta^{(i)}(X_0',\cdots,X_r')}{\omega(X_0',\cdots,X_r')}\right\}=\int_{R^p}\eta(x)\psi_i(x)dG(x)=a_i$$

and

$$\text{(6.10)} \qquad \mathrm{V}\left\{\frac{\omega_\eta^{(i)}(X_0',\cdots,X_r')}{\omega(X_0',\cdots,X_r')}\right\}=\int_{R^p}\eta^2(x)dG(x)-\sum_{j=0}^{r}a_j^2=\sum_{j=r+1}^{\infty}a_j^2\,.$$

This lemma is obtained directly from the formulae (6.7) and (6.8) by putting $f=\eta$. It is noted that $\Pr\{(x_0',\cdots,x_r');\ \omega(x_0',\cdots,x_r')=0\}=0$.

LEMMA 9. *If the prior information $s=\{(X_j',Y_j'),\ 0\leqq j\leqq r\}$ is distributed according to the joint probability distribution*

$$\frac{\omega^2(x_0',\cdots,x_r')}{(r+1)!}\prod_{j=0}^{r}f(y_j'\,|\,x_j')dG(x_j')\,,$$

then the estimator

$$\text{(6.11)} \qquad \hat{a}_i(s)=\frac{\omega_{Y'}^{(i)}(X_0',\cdots,X_r')}{\omega(X_0',\cdots,X_r')}$$

is unbiased for a_i, and its variance is given by

$$\text{(6.12)} \qquad \mathrm{V}(\hat{a}_i(s))=\sum_{j=0}^{r}\int_{R^p}\cdots\int_{R^p}\sigma^2(x_j')\Omega_{ij}^2\frac{1}{(r+1)!}\prod_{k=0}^{r}dG(x_k')+\sum_{j=r+1}^{\infty}a_j^2$$
$$=\int_{R^p}\sigma^2(x)dG(x)+\sum_{j=0}^{\infty}a_j^2\,,$$

where $\sigma^2(x_j')=\mathrm{V}(Y_j'\,|\,X_j'=x_j')$, $f(y'\,|\,x')$ is the conditional p.d.f. of Y' (the natural one) and $\omega_{Y'}^{(i)}(X_0',\cdots,X_r')=\sum_{j=0}^{r}Y_j'\Omega_{ij}$.

PROOF. Since $\mathrm{E}(Y_j'\,|\,X_j'=x_j')=\eta(x_j')$, it is easily seen from (6.9) that $\hat{a}_i(s)$ is unbiased for a_i. Next, we shall prove (6.12) in the following way. It is easily seen that

$$\mathrm{V}(\hat{a}_i(s))=\mathrm{E}\left\{\left(\left[\frac{\omega_{Y'}^{(i)}(X_0',\cdots,X_r')}{\omega(X_0',\cdots,X_r')}-\frac{\omega_\eta^{(i)}(X_0',\cdots,X_r')}{\omega(X_0',\cdots,X_r')}\right]+\left[\frac{\omega_\eta^{(i)}(X_0',\cdots,X_r')}{\omega(X_0',\cdots,X_r')}-a_i\right]\right)^2\right\}$$
$$=\mathrm{E}\left\{\left[\frac{\omega_{Y'-\eta}^{(i)}(X_0',\cdots,X_r')}{\omega(X_0',\cdots,X_r')}\right]^2\right\}+\mathrm{V}\left\{\frac{\omega_\eta^{(i)}(X_0',\cdots,X_r')}{\omega(X_0',\cdots,X_r')}\right\}.$$

Noting (6.10) and $\mathrm{E}\{(Y_j'-\eta(X_j'))(Y_k'-\eta(X_k'))\,|\,X_j'=x_j',\ X_k'=x_k'\}=\delta_{jk}\sigma^2(x_j')$, we can obtain (6.12), where $\delta_{jk}=1$ or 0 according to $j=k$ or $j\neq k$.

In the above discussion, we have supposed that $\{(X'_j, Y'_j), j=0, 1, \cdots, r\}$ are mutually independent.

THEOREM 4. *Let* $s^{(q)}=\{(X'_{qj}, Y'_{qj}), j=0, 1, \cdots, r\}$, $(q=1, \cdots, \rho)$, *be a set of mutually independent prior information, each of which is identically distributed according to the probability distribution of s given in lemma 9. Then the estimator*

$$\tilde{\eta}_{\rho(r+1)}(x, \boldsymbol{s})=\sum_{j=0}^{r} \tilde{a}_j(\boldsymbol{s})\phi_j(x) \tag{6.13}$$

is unbiased for $\eta_r(x)$, *and satisfies the relation*

$$\lim_{r\to\infty} \iint [\tilde{\eta}_{\rho(r+1)}(x, \boldsymbol{s})-\eta(x)]^2 dG(x)dF(s)=0 \tag{6.14}$$

for a suitable choice of $\rho=\rho_r$ *(depending on r), where*

$$\tilde{a}_i(\boldsymbol{s})=\frac{1}{\rho}\sum_{q=1}^{\rho} \hat{a}_i(s^{(q)}), \qquad \boldsymbol{s}=(s^{(1)}, \cdots, s^{(\rho)}), \tag{6.15}$$

and $\hat{a}_i(s^{(q)})$ *is given by* (6.11) *for* $s=s^{(q)}$.

PROOF. By lemma 9, it is easily seen that $\tilde{a}_i(\boldsymbol{s})$ is unbiased for a_i and its variance is given by

$$\mathrm{V}[\tilde{a}_i(\boldsymbol{s})]=\frac{1}{\rho}\left\{\int_{R^p} \sigma^2(x)dG(x)+\sum_{j=r+1}^{\infty} a_j^2\right\}. \tag{6.16}$$

From (6.5) and (6.16), we obtain at once

$$\begin{aligned} &\iint [\tilde{\eta}_{\rho(r+1)}(x, \boldsymbol{s})-\eta(x)]^2 dG(x)dF(s) \\ &=\frac{r+1}{\rho}\int_{R^p} \sigma^2(x)dG(x)+\left(1+\frac{r+1}{\rho}\right)\sum_{j=r+1}^{\infty} a_j^2 \\ &=\frac{r+1}{\rho}\sigma_0^2(1-\eta_{YX}^2)+\left(1+\frac{r+1}{\rho}\right)\sum_{j=r+1}^{\infty} a_j^2 . \end{aligned} \tag{6.17}$$

Since $\sum_{j=r+1}^{\infty} a_j^2 \to 0$ as $r\to\infty$, the relation (6.1) holds for suitable selection of ρ corresponding to r. Thus our assertions are proved.

This theorem gives a method for constructing $\hat{\eta}(x, s)$ satisfying (6.1). Moreover, it is easily verified that our method is a randomized least squares one in the case of $m=\rho(r+1)$ in the sense that $\hat{a}_i$'s are obtained as solutions which minimize $\sum_{\alpha=0}^{r}\left\{Y'_\alpha-\sum_{i=0}^{r} a_i\phi_i(X'_\alpha)\right\}^2$.

Further, it should be noted that the second term in the right hand side of (6.17) vanishes if $\eta(x)\in L_r^2(dG)$, i.e. $\eta(x)=\sum_{j=0}^{r} a_j\phi_j(x)$. In cases where $\eta(x)$ can not be represented in the finite terms, the usual (non-randomized) least squares method could not give an unbiased estimator for a_i. For simplicity, let us consider the case where the first sample size m coincides with the degree r of the orthogonal expansion $\eta_r(x)$ approximating to $\eta(x)$. Then, the usual least squares estimators $(\hat{a}_0'(s), \cdots, \hat{a}_r'(s))$ for $(a_0, \cdots, a_r)$ can be expressed in the following way.

$$\hat{a}_i'(s)=\sum_{j=0}^{r} l_{0j}\frac{L_{ij}}{L}, \tag{6.18}$$

where

$$l_{0j}=\frac{1}{r+1}\sum_{\alpha=0}^{r} Y_\alpha'\phi_j(x_\alpha), \qquad l_{ij}=\frac{1}{r+1}\sum_{\alpha=0}^{r}\phi_i(x_\alpha)\phi_j(x_\alpha),$$

$$L=\det(l_{ij}),\ (i, j=0, 1, 2, \cdots, r), \qquad s=\{(x_\alpha, Y_\alpha'), 0\leqq\alpha\leqq r\},$$

and L_{ij} is the (i, j)th cofactor of L. We can easily show that $\hat{a}_i'(s)$ is expressible in another form using the determinant ω and $\omega_{Y'}^{(i)}$ defined in (6.8):

$$\hat{a}_i'(s)=\frac{\omega_{Y'}^{(i)}(x_0, \cdots, x_r)}{\omega(x_0, \cdots, x_r)}. \tag{6.19}$$

This functional form of $\hat{a}_i'(s)$ apparently coincides with that of the unbiased estimator $\hat{a}_i(s)$ given in (6.11). However, it will be shown that $\hat{a}_i'(s)$ given by (6.19) can not be unbiased for a_i unless $\eta(x)$ is expressible in the finite terms of expansion by $\{\phi_j(x)\}$. In fact, we can obtain

$$\mathrm{E}\{\hat{a}_i'(s)\}=\frac{\omega_\eta^{(i)}(x_0, \cdots, x_r)}{\omega(x_0, \cdots, x_r)}\sim a_i+\sum_{j=r+1}^{\infty} a_j\frac{\omega_{\phi_j}^{(i)}}{\omega}. \tag{6.20}$$

Since the second term of the right hand side of (6.20) does not vanish, $\hat{a}_i'(s)$ is not unbiased. Therefore, we can not construct $\hat{\eta}_r(x, s)$ satisfying (6.1) based on the usual least squares method. On the contrary, in our method, the bias term in (6.20) can be removed by changing non-random variables $(x_0, \cdots, x_r)$ into random variables $(X_0', \cdots, X_r')$ distributed according to $\frac{\omega^2}{(r+1)!}\prod_{k=0}^{r} dG(x_k')$. It should be added that $\hat{a}_i''(s)=\frac{\omega_{Y'}^{(i)}\omega}{(r+1)!}$ is also unbiased for a_i with the natural distribution $\prod_{k=0}^{r} f(y_k'\,|\,x_k')dG(x_k')$.

In practical situations, an orthogonal polynomial system $\{P_n(x)\}$ may be recommended for $\{\phi_n(x)\}$. We shall mention below sufficient conditions under which the orthogonal polynomial system $\{P_n(x)\}$ associated with a given $G(x)$ becomes complete in p-variate case, i.e. $x=(x_1, \cdots, x_p)$. They are:

(6.21) $$\int_{R^p} |x_i|^k dG(x) < \infty, \qquad \text{for } i=1, 2, \cdots, p;\ k=1, 2, \cdots,$$

and

(6.22) $$\lim_{k\to\infty} \frac{1}{k} \left\{ \int_{R^p} \left[\sum_{i=1}^{p} x_i^2 \right]^k dG(x) \right\}^{1/2k} = 0.$$

Instead of (6.22), we may take

(6.23) $$\lim_{k\to\infty} \frac{1}{k} \left[\sum_{i=1}^{p} \beta_i^{2k} \right]^{1/2k} = 0,$$

where

$$\beta_i^{2k} = \int_{R^p} x_i^{2k} dG(x).$$

7. Examples

i) Linear regression case

We shall give two typical examples in contrast to the one in non-linear cases, though they are not so suitable for our method.

Example 1. Let Y and X be one-dimensional variates, and $\eta(x) = \beta_0 + \beta_1 x$. In this case, it is easily seen that

(7.1) $$\mathrm{V}(\bar{Y}\,|\,\phi) = \frac{1}{n} \left\{ \sigma_0^2 (1-\rho_{01}^2) + \rho_{01}^2 \frac{\sigma_0^2}{\sigma_1^2} \sum_{i=1}^{l} w_i \sigma_{1i}^2 \right\} = \frac{\sigma_0^2}{n} \left(1 - \rho_{01}^2 \frac{\sigma_{1b}^2}{\sigma_1^2} \right),$$

where

$$\sigma_0^2 = \mathrm{V}(Y), \qquad \sigma_1^2 = \mathrm{V}(X), \qquad \sigma_{1i}^2 = \frac{1}{w_i} \int_{-\infty}^{\infty} (x-\mu_{1i})^2 \phi_i(x) dG(x),$$

$$\sigma_{1b}^2 = \sum_{i=1}^{l} w_i (\mu_{1i} - \mu_1)^2, \qquad \mu_{1i} = \frac{1}{w_i} \int_{-\infty}^{\infty} x \phi_i(x) dG(x),$$

$$\mu_1 = \mathrm{E}(X) \qquad \text{and} \qquad \rho_{01} = \mathrm{Cor}(Y, X).$$

Then, we can obtain ϕ^* minimizing $\mathrm{V}(Y|\phi)$, or maximizing $\frac{\sigma_{1b}^2}{\sigma_1^2}$, among Φ. Namely, ϕ^* is nothing but an optimum stratification for $G(x)$. Therefore, the estimation of $\eta(x)$ and $\phi^*(x)$ using prior information s becomes unnecessary if $G(x)$ is known.

Example 2. Let Y be a one-dimensional variate, $X = (X_1, \cdots, X_p)'$ a p-dimensional variate distributed according to $N(\mu, \Sigma)$, $\eta(x) = \beta_0 + \sum_{j=1}^{p} \beta_j x_j$, and $\mathrm{V}(Y\,|\,X=x) = \sigma^2$ (independent of x). In this case, we can use the usual (non-randomized) least squares estimators $(\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_p)$ for $(\beta_0, \beta_1, \cdots, \beta_p)$ based on prior information $s=\{(x^{(\alpha)}, Y'_\alpha), 1\leqq\alpha\leqq m\}$, where $(x^{(1)}, \cdots, x^{(m)})$ are non-random variates representing observational levels,

$$
\begin{aligned}
&\hat{\beta}_0=\bar{Y}'-\sum_{j=1}^{p}\hat{\beta}_j\bar{x}_j, \quad \bar{Y}'=\frac{1}{m}\sum_{\alpha=1}^{m}Y'_\alpha, \quad \bar{x}_j=\frac{1}{m}\sum_{\alpha=1}^{m}x_j^{(\alpha)}, \\
(7.2)\qquad &\hat{\beta}_j=\sum_{k=1}^{p}l_{0k}\frac{L_{jk}}{L}, \qquad l_{0j}=\frac{1}{m}\sum_{\alpha=1}^{m}(Y'_\alpha-\bar{Y}')(x_j^{(\alpha)}-\bar{x}_j), \\
&l_{jk}=\frac{1}{m}\sum_{\alpha=1}^{m}(x_j^{(\alpha)}-\bar{x}_j)(x_k^{(\alpha)}-\bar{x}_k), \quad L=\det(l_{jk}),
\end{aligned}
$$

and L_{jk} is the (j, k)th cofactor of L $(1\leqq j, k\leqq p)$. Then, it is well-known that $(\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_p)$ are unbiased for $(\beta_0, \beta_1, \cdots, \beta_p)$, and their variance and covariances are given by

$$
\begin{aligned}
&\mathrm{V}(\beta_0)=\frac{\sigma^2}{m}, \qquad \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_j)=0, \qquad (1\leqq j\leqq p) \\
(7.3)\qquad & \\
&\mathrm{Cov}(\hat{\beta}_j, \hat{\beta}_k)=\frac{\sigma^2}{m}\frac{L_{jk}}{L}, \qquad (1\leqq j,\ k\leqq p).
\end{aligned}
$$

Let us put $Z=\eta(X)=\beta_0+\sum_{j=1}^{p}\beta_jX_j$ and $\hat{Z}=\hat{\eta}(X)=\hat{\beta}_0+\sum_{j=1}^{p}\hat{\beta}_jX_j$, where random variables $(X_1, \cdots, X_p)$ have to be considered implicitly in the second sample based on which the estimator $\bar{Y}$ for μ_0 is constructed. Then, it is easily seen that Z and $\hat{Z}$ are distributed according to $N(\eta(\mu), \beta'\Sigma\beta)$ and $N(\hat{\eta}(\mu), \hat{\beta}'\Sigma\hat{\beta})$ respectively, where $\beta=(\beta_1, \cdots, \beta_p)'$, $\hat{\beta}=(\hat{\beta}_1, \cdots, \hat{\beta}_p)'$, $\mu=(\mu_1, \cdots, \mu_p)'$ and $\mu_j=\mathrm{E}(X_j)$, $j=1, \cdots, p$. Further, we can see that

$$
(7.4)\qquad \mathrm{Cov}(Z, \hat{Z})=\beta'\Sigma\hat{\beta}.
$$

Now, the optimum stratification ϕ^* for Z can be defined by the interval division $\{I_i^*, 1\leqq i\leqq l\}$, $I_i^*=[z_{i-1}^*, z_i^*)$, such that

$$
(7.5)\qquad z_i^*=\eta(\mu)+\tilde{z}_i\sqrt{\beta'\Sigma\beta}, \qquad (1\leqq i\leqq l-1),
$$

where $\tilde{z}_i$'s are end-points of the interval division $\{\tilde{I}_i\}$ corresponding to an optimum stratification $\tilde{\phi}$ for $N(0, 1)$. ($\tilde{z}_i$'s are obtained numerically by Sethi [14] for $l\leqq 5$).

On the other hand, the optimum stratification $\hat{\phi}^*$ using s for $\hat{Z}$ is defined by the interval division $\{\hat{I}_i^*, 1\leqq i\leqq l\}$, $I_i^*=[z_{i-1}^*, z_i^*)$, such that

$$
(7.6)\qquad \hat{z}_i^*=\hat{\eta}(\mu)+\tilde{z}_i\sqrt{\hat{\beta}'\Sigma\hat{\beta}}, \qquad (1\leqq i\leqq l-1).
$$

Noting that $w_i^*=\hat{w}_i^*=\tilde{w}_i$ we can obtain, after some calculations,

$$
(7.7)\qquad \begin{aligned} u_i^* &= \tilde{w}_i\mu_0+\tau_i\sqrt{\beta'\Sigma\beta}\,,\\ \hat{u}_i^* &= \tilde{w}_i\mu_0+\tau_i\frac{\beta'\Sigma\hat{\beta}}{\sqrt{\hat{\beta}'\Sigma\hat{\beta}}}\,, \end{aligned}
$$

where

$$
\tau_i=\int_{\tilde{z}_{i-1}}^{\tilde{z}_i}\frac{t}{\sqrt{2\pi}}e^{-t^2/2}dt\,.
$$

Thus, we obtain

$$
(7.8)\qquad \begin{aligned} \sum_{i=1}^{l}\frac{u_i^{*2}}{w_i^*} &= \mu_0^2+K_l\beta'\Sigma\beta\,,\\ \sum_{i=1}^{l}\frac{\hat{u}_i^{*2}}{\hat{w}_i^*} &= \mu_0^2+K_l\frac{(\beta'\Sigma\hat{\beta})^2}{\hat{\beta}'\Sigma\hat{\beta}}\,, \end{aligned}
$$

where $K_l=\sum_{i=1}^{l}\frac{\tau_i^2}{\tilde{w}_i}$ represents the variance between strata for the optimum stratification $\tilde{\phi}$ in $N(0, 1)$.

Thus, we obtain:

$$
(7.9)\qquad \mathrm{V}(\bar{Y}\mid\phi^*)=\frac{1}{n}\{\sigma_0^2(1-\rho_{YX}^2)+(1-K_l)\beta'\Sigma\beta\}
$$

and

$$
(7.10)\qquad \mathrm{V}(\bar{Y}\mid\hat{\phi}^*)=\mathrm{V}(\bar{Y}\mid\phi^*)+\frac{K_l}{n}\left\{\beta'\Sigma\beta-\frac{(\beta'\Sigma\hat{\beta})^2}{\hat{\beta}'\Sigma\hat{\beta}}\right\},
$$

where ρ_{YX} denotes the multiple correlation coefficient of Y on X. The second term in the right hand side of (7.10) is non-negative and can be regarded as the loss caused by using $\hat{\phi}^*$ instead of ϕ^*.

Evaluating the expected loss, mentioned above, asymptotically, we obtain the inequality

$$
(7.11)\qquad \mathrm{E}\left\{\beta'\Sigma\beta-\frac{(\beta'\Sigma\hat{\beta})^2}{\hat{\beta}'\Sigma\hat{\beta}}\right\}\leqq\frac{\sigma_0^2}{m}(1-\rho_{YX}^2)\,\mathrm{tr}(QL^{-1})+o(m^{-1})\,,
$$

where $Q=\Sigma-\Sigma\beta\beta'\Sigma(\beta'\Sigma\beta)^{-1}$ and the constant factor K_l/n is neglected.

Finally, we shall propose a problem how to select observational levels $\boldsymbol{x}=(x^{(1)},\cdots,x^{(m)})$ optimally so that $\mathrm{tr}(QL^{-1})$ in the right hand side of (7.11) is minimized. We shall call the minimizing levels "the asymptotically optimum design" and denote it by $\boldsymbol{x}^*$. The minimum of $\mathrm{tr}(QL^{-1})$ can be obtained in the following way: Under the condition $\mathrm{tr}(L)=c$ (any positive constant),

$$\min_{x} \operatorname{tr}(QL^{-1}) = \operatorname{tr}(QL_c^{*-1}) = \frac{1}{c}[\operatorname{tr}(Q^{1/2})]^2, \tag{7.12}$$

where

$$L_c^* = cQ^{1/2}[\operatorname{tr}(Q^{1/2})]^{-1}.$$

Therefore, we can determine the asymptotically optimum design $\boldsymbol{x}^*$ for any positive constant c. However, as is seen from (7.12), $\boldsymbol{x}^*$ depends on Q which includes unknown parameters $(\beta_1, \cdots, \beta_p)$. To tackle with such a situation, more artificial means will be necessary. Detailed discussions will be done in the other chance.

ii) Non-linear regression case

Example 3. Let X be a one-dimensional variate, distributed according to $N(0, 1)$, $\eta(x) = e^{\lambda x}$ $(\lambda > 0)$, and $V(Y|X=x) = \sigma^2(x) \in L^2(dG)$, where

$$G(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt.$$

By easy calculations, we obtain the results

$$\begin{aligned} u_i &= \int_{-\infty}^{\infty} \eta(x)\phi_i(x)dG(x) = e^{\lambda^2/2}[G(x_i - \lambda) - G(x_{i-1} - \lambda)], \\ w_i &= \int_{-\infty}^{\infty} \phi_i(x)dG(x) = G(x_i) - G(x_{i-1}), \end{aligned} \tag{7.13}$$

and

$$v(u, w) = \sum_{i=1}^{l} \frac{u_i^2}{w_i} = e^{\lambda^2} \sum_{i=1}^{l} \frac{[G(x_i - \lambda) - G(x_{i-1} - \lambda)]^2}{G(x_i) - G(x_{i-1})}, \tag{7.14}$$

where ϕ denotes a stratification corresponding to an interval division $\{I_i\}$, $I_i = [x_{i-1}, x_i]$, $-\infty = x_0 < x_1 < \cdots < x_{l-1} < x_l = \infty$. Then, we can obtain an optimum ϕ^* corresponding to $\{I_i^*\}$, $I_i^* = [x_{i-1}^*, x_i^*)$, by solving numerically the equations

$$\frac{\partial v}{\partial x_i} = 0, \qquad (i = 1, 2, \cdots, l-1),$$

where v is given in (7.14).

Next, we shall state the outline of constructing the asymptotically optimum stratification $\hat{\phi}^*(x, s)$ using prior information s. At first, we mention of the orthogonal polynomials associated with $G(x)$. It is well-known that the Hermite polynomial system $\{H_n(x)\}$ is complete and orthogonal with respect to the unit normal distribution, where $\{H_n(x)\}$ is given by

$$(7.15)\qquad \begin{aligned} &H_0(x)=1,\quad H_1(x)=x,\quad H_2(x)=x^2-1,\\ &H_3(x)=x^3-3x,\quad H_4(x)=x^4-6x+3,\ \text{and so on.}\end{aligned}$$

Considering the relations

$$(7.16)\qquad \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} H_m(x)H_n(x)e^{-x^2/2}dx=\begin{cases} n! & \text{for } m=n\\ 0 & \text{for } m\neq n,\end{cases}$$

we can see that $\left\{\frac{1}{\sqrt{n!}}H_n(x)\right\}$ is the complete ortho-normal polynomial system associated with $G(x)$.

Using the C.O.N.S. $\{\psi_n(x)\}$, $\psi_n(x)=\frac{1}{\sqrt{n!}}H_n(x)$, we can obtain unbiased estimators $(\tilde{a}_0(s),\cdots,\tilde{a}_r(s))$ for $(a_0,\cdots,a_r)$ using ρ sets of the first samples, where

$$(7.17)\qquad \tilde{a}_i(s)=\frac{1}{\rho}\sum_{j=1}^{\rho}\frac{\omega^{(i)}_{Y(j)}(X_0^{(j)},\cdots,X_r^{(j)})}{\omega(X_0^{(j)},\cdots,X_r^{(j)})},\qquad a_i=\int_{-\infty}^{\infty}\eta\psi_i\,dG,$$

and $s=(s^{(1)},\cdots,s^{(\rho)})$, as in section 6. Then, we can construct an estimator $\tilde{\eta}_{\rho(r+1)}(x,s)$ for $\eta(x)$ by

$$(7.18)\qquad \tilde{\eta}_{\rho(r+1)}(x,s)=\sum_{j=0}^{r}\tilde{a}_j(s)\psi_j(x).$$

The mean square error of $\tilde{\eta}_r(x,s)$ is given by

$$(7.19)\qquad \begin{aligned}\delta_{\rho(r+1)}&=\iint[\tilde{\eta}_{\rho(r+1)}(x,s)-\eta(x)]^2\,dG(x)\,dF(s)\\ &=\frac{r+1}{\rho}\int_{-\infty}^{\infty}\sigma^2(x)dG(x)+\left(\frac{r+1}{\rho}+1\right)\sum_{j=r+1}^{\infty}a_j^2.\end{aligned}$$

Accordingly, we can calculate $\delta_{\rho(r+1)}$ explicitly or numerically only if the degree r and the functional form of $\sigma^2(x)$ are preassigned. In addition, coefficients a_n's are given explicitly by

$$(7.20)\qquad a_n=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{\lambda x-x^2/2}\psi_n(x)dx=\frac{\lambda^n}{\sqrt{n!}}e^{\lambda^2/2}.$$

8. Conclusion

In this paper, we have obtained the following results:

1) There exists a general optimum stratification ϕ^* based only on the objective variable X (uni-variate) within the limits of general stratifications, without the assumption that $F(x)$ is absolutely continuous. Especially, in the case of proportionate allocation, the general optimum stratification ϕ^* coincides with the optimum stratification within the

limits of interval divisions $\{I_i\}$.

2) In the case of proportionate allocation, the problem of general optimum stratification, for the objective variable Y (uni-variate) based on the concomitant variable X (p-variate), reduces to that of general optimum stratification for the objective variable $Z=\eta(X)$, where $\eta(x)$ is the regression function of Y on X.

3) The asymptotically optimum stratification $\hat{\phi}^*(x, s)$ using prior information (the first sample) s can be obtained in the case of proportionate allocation, only if there exists an estimated function $\hat{\eta}(x, s)$ which tends to $\eta(x)$ in the sense of $L^2(dG\cdot dF)$ as the first sample size m tends to infinity.

4) The function $\hat{\eta}(x, s)$, satisfying the condition just stated above, can be constructed by using the complete ortho-normal system $\{\psi_n(x)\}$ associated with $G(x)$. Further, the sufficient conditions are given for the completeness of $\{\psi_n(x)\}$ in the case of multivariate orthogonal polynomial system.

It should be added that our results will give some prospects to regression analysis and multivariate analysis, especially selection and discrimination problems.

Finally, we shall mention unsolved problems for optimum stratification.

1° To obtain the general optimum stratification for the objective variable Y in the case of multivariate, based on Y itself or the concomitant variable X.

2° To extend our results to the case of multi-stage sampling.

3° Considerations on cases where the empirical d.f. $G_m(x)$ is given instead of $G(x)$, or the empirical d.f. $F_m(x', y')$ for the first sample is given.

4° To find the conditions for the uniqueness of the solution satisfying the equations (1.1) or (1.2), and to find sufficient conditions for $\{x_i^*\}$ to be really optimum.

Acknowledgements

The author is deeply grateful to Professor Z. Suetuna, Director of The Institute of Statistical Mathematics, for his warm encouragement. He also wishes to express his gratitude to Professors M. Okamoto and K. Isii, Osaka University, for unstinted supports and advices. His hearty thanks are extended to the referees and Messrs. M. Sibuya, R. Shimizu, K. Takahasi and G. Suzuki for valuable advices and discussions, and Misses A. Maruyama and S. Sasaki for their help in preparing the manuscript.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] C. Hayashi and F. Maruyama, "On some method for stratification" (in Japanese), *Research Memoir*, Inst. Statist. Math., 4 (1948), 399-411.

[2] T. Dalenius, "The problem of optimum stratification," *Skand. Aktuartidskr.*, 3-4 (1950), 203-213.

[3] T. Dalenius and M. Gurney, "The problem of optimum stratification," *Skand. Aktuartidskr.*, 3-4 (1951), 133-148.

[4] C. Hayashi, F. Maruyama and M. D. Isida, "On some criteria for stratification," *Ann. Inst. Statist. Math.*, 2 (1951), 77-86.

[5] P. C. Mahalanobis, "Some aspects of the design of sample surveys," *Sankyā*, 12, Part 1 & 2 (1952), 1-7.

[6] M. H. Hansen, W. G. Hurwitz and W. G. Madow, *Sample Survey Methods and Theory*, Vol. 1 (1953), John Wiley.

[7] H. Aoyama, "A study of the stratified random sampling," *Ann. Inst. Statist. Math.*, 6 (1954), 1-36.

[8] T. Kitagawa, "Some contributions to the design of sample survey," *Sankyā*, 17, Part 1 (1956), 1-36.

[9] T. Dalenius and J. L. Hodges, Jr., "The choice of stratification points," *Skand. Aktuartidskr.*, 3-4 (1957), 198-203.

[10] T. Dalenius and J. L. Hodges, Jr., "Minimum variance stratification," *J. Amer. Statist. Ass.*, 54 (1959), 88-101.

[11] G. Ekman, "An approximation useful in univariate stratification," *Ann. Math. Statist.*, 30 (1959), 219-229.

[12] J. Durbin, Review of "Sampling in Sweden," *J. Roy. Stat. Soc.*, A, 122 (1959), 246-248.

[13] W. G. Cochran, "Comparison of methods for determining stratum boundaries," *Bull. Inst. Inter. Statist.*, 38, no. 2 (1961), 345-358.

[14] V. K. Sethi, "A note on optimum stratification of populations for estimating the population means," *Aust. J. Statist.*, 5 (1963), 20-33.

[15] S. P. Ghosh, "Optimum stratification with two characteristics," *Ann. Math. Statist.*, 34 (1963), 866-872.

[16] H. Aoyama, "Stratified random sampling with optimum allocation for multivariate population," *Ann. Inst. Statist. Math.*, 14 (1963), 251-258.

[17] K. Saito, "Theoretical considerations on the use of supplementary information in sample survey design (II)," *Sophia Economic Review*, 5 (1958), 49-72.

[18] W. A. Ericson, "Optimum stratified sampling using prior information," *J. Amer. Statist. Ass.*, 60 (1965), 750-771.

[19] K. Isii, "Mathematical programming and statistical inference" (in Japanese), *Sūgaku (Mathematics)*, 18 (1966), 85-95.

[20] S. M. Ermakov and V. G. Zolotuxin, "Polynomial approximation and Monte-Carlo method," *Probability Theory and its Applications*, 5 (1960), 473-476.

[21] E. L. Lehmann, *Testing Statistical Hypotheses*, John Wiley, 1959.

CORRECTIONS TO
"ON OPTIMUM STRATIFICATION FOR THE OBJECTIVE VARIABLE BASED ON CONCOMITANT VARIABLES USING PRIOR INFORMATION"

YASUSHI TAGA

In the above titled article (this Annals 19 (1966), 101–129) the following corrections should be made.

(i) On page 120, line 14, replace

$$\text{“}\sum_{j=0}^{\infty} a_j^2\text{”} \quad \text{by} \quad \text{“}\sum_{j=r+1}^{\infty} a_j^2\text{”}.$$

(ii) On page 121, lines 9 and 17, replace

$$\text{“}dF(s)\text{”} \quad \text{by} \quad \text{“}dF(\boldsymbol{s})\text{”}.$$

(iii) On page 127, lines 12, 13, 14 and 15, replace

$$\text{“}\tilde{\eta}_{\rho(r+1)}(x, s)\text{”} \quad \text{by} \quad \text{“}\tilde{\eta}_{\rho(r+1)}(x, \boldsymbol{s})\text{”}.$$

(iv) On page 127, line 15, replace

$$\text{“}F(\boldsymbol{s})\text{”} \quad \text{by} \quad \text{“}\boldsymbol{F}(\boldsymbol{s})\text{”}.$$