STATISTICA, anno LXIV, n. 4, 2004

EFFECT OF OPTIMUM STRATIFICATION ON SAMPLING WITH VARYING PROBABILITIES UNDER PROPORTIONAL ALLOCATION

S.E.H. Rizvi, J.P. Gupta, M. Bhargava

1. INTRODUCTION

In stratified random sampling, proper choice of strata boundaries is one of the important factors as regards to the efficiency of estimator of population characteristic under consideration. The problem of optimum stratification for univariate case was first considered by Dalenius (1950) who treated the estimation variable itself as stratification variable. Singh and Sukhatme (1969) provided some approximate solutions for the strata boundaries by using an auxiliary variable as the stratification variable under Neyman and proportional allocations. Singh and Sukhatme (1972) also considered this problem in case the units from different strata are selected with probability proportional to the value of auxiliary variable and with replacement (PPSWR) using Neyman allocation. In the sequel, Singh (1975) tackled the same problem for proportional and equal allocation procedures. In the former case they observed a gain in efficiency, however little, whereas in the latter case their study showed a remarkable decrease in the efficiency as compared to unstratified PPSWR.

Generally, information on more than one study variable is collected in any survey and in that case the theory concerning the construction of strata for one study variable can not be adopted as such. For such a situation, Rizvi *et al.* (2000) developed the theory of optimum stratification for bivariate case, on the basis of auxiliary variable, in case of simple random sampling. Thus, in the present paper the problem of optimum stratification has been considered for two study variables in case the units within a stratum are selected using PPSWR scheme. In this connection, the minimal equations have been developed under proportional allocation along with its approximate solutions. A limit expression of the generalized variance is also provided which enables us to examine the effect of stratification on sampling with PPSWR as regards to the change in number of strata. The paper concludes with numerical illustrations for three density functions.

2. VARIANCE AND COVARIANCE UNDER SUPER-POPULATION SET-UP

Let the population of size N be divided into L strata and the units within a stratum are selected using PPSWR. Let the variables under study be denoted by Y_j ($j = 1, 2$) and the known auxiliary variable by X. Let P_{hi} denote the selection probability assigned to the i-th unit in the h-th stratum, $i = 1, 2,..., N_h$; $h = 1, 2, ..., L$ so that $\sum_i P_{hi} = 1$ for each h, $P_{hi} = x_{hi}/X_h$ where x_{hi} is the i-th sampled observation in the h-th stratum and X_h is the stratum total for X.

The unbiased estimators of the population means $\overline{Y}_j$ of the study variables Y_j ($j = 1, 2$) will be the weighted means defined by

$$\bar{y}_{st.Pj} = \frac{1}{N}\sum_{h=1}^{L}\frac{1}{n_h}\sum_{i=1}^{n_h}\frac{y_{jhi}}{P_{hi}} \qquad (j=1, 2)$$

where symbols have their usual meanings.

For the proportional method of allocation, the terms for the variances of the unbiased estimators $\bar{y}_{st.Pj}$ ($j = 1, 2$) and their covariance, under repeated random sampling are given by

$$\mathrm{V}(\bar{y}_{st.Pj})_P = \frac{1}{nN^2}\sum_{h=1}^{L}\frac{1}{W_h}\left[\sum_{i=1}^{N_h}\frac{y_{jhi}^2}{P_{hi}} - Y_{jh}^2\right] \tag{1}$$

and
$$\mathrm{Cov}(\bar{y}_{st.P1}, \bar{y}_{st.P2})_P = \frac{1}{nN^2}\sum_{h=1}^{L}\frac{1}{W_h}\left[\sum_{i=1}^{N_h}\frac{y_{1hi}\cdot y_{2hi}}{P_{hi}} - Y_{1h}Y_{2h}\right] \tag{2}$$

where W_h is the proportion of units in the h-th stratum.

Let us assume that the population under consideration is finite and is a random sample of size N from an infinite superpopulation with same characteristics as those of the finite population. Let the joint density function of (X, Y_1, Y_2) in the super-population set-up be denoted by $f_s(x, y_1, y_2)$ and the marginal density function of X by $f(x)$ assuming that these functions are continuous in the range (a,b) of X. Further, let us assume that the regression of Y_j on X, in each stratum of the super-population, is linear and that these regression lines pass through origin. Thus the regression models are given by

$$Y_j = \beta_{jh}X + e_j \qquad (j = 1, 2) \tag{3}$$

where e_j is the error component such that $\mathrm{E}(e_j/X) = 0$, $\mathrm{E}(e_j e_j'/x,x') = 0$ for $x \neq x'$ and $\mathrm{V}(e_j/X) = \eta_j(x) > 0$ ($j = 1,2$) for all $x \in (a, b)$.

If we define

$$\sigma^2_{jh} = \sum_{i=1}^{N_h} \frac{y^2_{jhi}}{P_{hi}} - Y^2_{jh}, \quad (j=1, 2) \tag{4}$$

and $$\sigma_{12h} = \sum_{i=1}^{N_h} \frac{y_{1hi} \cdot y_{2hi}}{P_{hi}} - Y_{1h}Y_{2h}. \tag{5}$$

The expectations of σ^2_{1h}, σ^2_{2h} and σ_{12h} in the superpopulation become

$$E(\sigma^2_{jh}) = E_1E_2(\sigma^2_{jh}), \quad (j = 1, 2)$$

and $E(\sigma_{12h}) = E_1E_2(\sigma_{12h})$ respectively.

Here, E_1 is the expectation over all possible $x's$ and E_2 is the expectation for fixed $(x_1, x_2, ..., x_N)$. Thus, from (4) we have

$$E_2(\sigma^2_{jh}) = E_2\left[\sum_{i=1}^{N_h} \frac{y^2_{jhi}}{P_{hi}}\right] - E_2(Y^2_{jh}). \tag{6}$$

And, using model (3) we get

$$E_2(\sigma^2_{jh}) = E_2\left[\sum_{i=1}^{N_h} \frac{(\beta_{jh}x_{hi} + e_{jhi})^2}{P_{hi}}\right] - E_2\left[\sum_{i=1}^{N_h}(\beta_{jh}x_{hi} + e_{jhi})\right]^2$$

which on simplification becomes

$$E_2(\sigma^2_{jh}) = \sum_{i \neq i'}^{N_h} \phi_j(x_{hi})x_{hi'} \tag{7}$$

where $\phi_j(x_{hi}) = \eta_j(x_{hi})/x_{hi}$.

Now taking overall expectation, we get

$$\begin{aligned} E(\sigma^2_{jh}) &= E_1E_2(\sigma^2_{jh}) \\ &= E_1\sum_{i \neq i'}^{N_h} \phi_j(x_{hi})x_{hi'}. \end{aligned}$$

Since i-th and i'-th units in the h-th stratum are drawn independently, we have

$$E(\sigma^2_{jh}) = N_h(N_h - 1)\mu_{h\phi_j}\mu_{hx} \tag{8}$$

where $\mu_{h\phi_j}$ and μ_{hx} are the expected values of $\phi_j(x)$ and x, respectively, in the h-th stratum. Now, assuming the stratum size N_h to be large enough so that $N_h - 1 \cong N_h$, we get from (8)

$$\mathrm{E}(\sigma^2_{jh}) = \mathrm{N}^2_h \mu_{h\phi_j} \mu_{hx} . \tag{9}$$

And, from (5) we have

$$\mathrm{E}_2(\sigma_{12h}) = \mathrm{E}_2\left[\sum_{i=1}^{N_h} \frac{y_{1hi} \cdot y_{2hi}}{P_{hi}}\right] - \mathrm{E}_2(Y_{1h} Y_{2h}) . \tag{10}$$

Using the model (3) the above expression can be put as

$$\mathrm{E}_2(\sigma_{12h}) = X_h \mathrm{E}_2\left[\sum_{i=1}^{N_h} \{(\beta_{1h} x_{hi} + e_{1hi})(\beta_{2h} x_{hi} + e_{2hi}) / x_{hi}\}\right]$$
$$- \mathrm{E}_2\left[\sum_{i=1}^{N_h} (\beta_{1h} x_{hi} + e_{1hi})\right]\left[\sum_{i=1}^{N_h} (\beta_{2h} x_{hi} + e_{2hi})\right]$$

which, on some algebraic simplifications, gives

$$\mathrm{E}_2\,(\sigma_{12h}) = 0.$$

Therefore, $\mathrm{E}\,(\sigma_{12h}) = 0.$ (11)

Thus, the expected variances and covariance of the estimators $\overline{y}_{st.Pj}$ (j = 1, 2), under proportional method of allocation, become

$$\sigma^2_j = \mathrm{V}(\overline{y}_{st.Pj})_P = \frac{1}{n} \sum_{h=1}^{L} W_h \mu_{h\phi_j} \mu_{hx} \tag{12}$$

and $\mathrm{Cov}(\overline{y}_{st.P1}, \overline{y}_{st.P2})_P = 0$. (13)

The minimal equations and its solutions have been obtained in the next section.

3. MINIMAL EQUATIONS

For the division of the population under consideration into L strata, let us represent the set of the points of demarcation forming L strata by $\{x_h\}$, $h = 1, 2, ..., L-1$. These points are obtained by solving the minimal equations which are obtained by equating to zero the partial derivatives of G_3 w.r.t. $\{x_h\}$, where

$$G_3 = \begin{vmatrix} \sigma^2_1 & \sigma_{12} \\ \sigma_{21} & \sigma^2_2 \end{vmatrix} . \tag{14}$$

Hence, we get

$$\sigma_1^2 \frac{\partial \sigma_2^2}{\partial x_h} + \sigma_2^2 \frac{\partial \sigma_1^2}{\partial x_h} - 2\sigma_{12} \frac{\partial \sigma_{12}}{\partial x_h} = 0 \ . \tag{15}$$

Putting the values of σ_j^2 and σ_{12} from (12) and (13), respectively, in (15), we have

$$\left(\sum_{h=1}^{L} W_h \mu_{h\phi_1} \mu_{hx}\right) \frac{\partial}{\partial x_h} (W_h \mu_{h\phi_2} \mu_{hx} + W_r \mu_{r\phi_2} \mu_{rx}) + \left(\sum_{h=1}^{L} W_h \mu_{h\phi_2} \mu_{hx}\right) \frac{\partial}{\partial x_h} (W_h \mu_{h\phi_1} \mu_{hx} + W_r \mu_{r\phi_1} \mu_{rx}) = 0 \tag{16}$$

where $r = h+1,\ h = 1,2,...,L-1$.

The values of partial derivatives of the various terms involved in (16) can be easily obtained on the lines of Singh and Sukhatme (1969). Putting these values in (16), we finally get the minimal equations as follows:

$$\left(\sum_{h=1}^{L} W_h \mu_{h\phi_1} \mu_{hx}\right) \{x_h \mu_{h\phi_2} + \mu_{hx}(\phi_2(x_h) - \mu_{h\phi_2})\} + \left(\sum_{h=1}^{L} W_h \mu_{h\phi_2} \mu_{hx}\right) \{x_h \mu_{h\phi_1} + \mu_{hx}(\phi_1(x_h) - \mu_{h\phi_1})\}$$

$$= \left(\sum_{h=1}^{L} W_h \mu_{h\phi_1} \mu_{hx}\right) \{x_h \mu_{r\phi_2} + \mu_{rx}(\phi_2(x_h) - \mu_{r\phi_2})\} + \left(\sum_{h=1}^{L} W_h \mu_{h\phi_2} \mu_{hx}\right) \{x_h \mu_{r\phi_1} + \mu_{rx}(\phi_1(x_h) - \mu_{r\phi_1})\} \tag{17}$$

where $r = h+1,\ h = 1, 2,..., L-1$.

The solution of this system of minimal equations will provide the set of the optimum points of stratification $\{x_h\}$ corresponding to the minimum value of the generalized variance G_3. But, the system of equations given by (17) involves parameters which are themselves functions of points of strata boundaries indicating thereby that the exact solutions are difficult to obtain. Hence, some approximate solutions have been obtained which are given in the next section.

4. APPROXIMATE SOLUTIONS OF THE MINIMAL EQUATIONS

In order to obtain the approximate solutions of the minimal equations (17), we have to expand both the sides of the minimal equation about the point x_h, the common boundary point of the h-th and r-th strata. For this purpose, we impose certain regularity conditions on the functions $f(x)$, $\phi_j(x)$ and $\eta_j(x)$ as under:

A function $\omega(x)$ belongs to the class of function, Ω, if it satisfies the following conditions:

(i) $0 < \omega(x) < \infty$,

and (ii) $\omega(x)$, $\omega'(x)$ and $\omega''(x)$ exist and are continuous for all $x \in (a,b)$, where $(b-a) < \infty$.

Under these assumptions, we can obtain the series expansions for W_h, $\mu_{h\phi_j}$ and μ_{hx} by using Taylor's theorem about both the upper and lower boundaries of the h-th stratum. Proceeding on the lines of Singh and Sukhatme (1969), the expanded forms of different terms involved in the left hand side of the minimal equations (17) can be obtained as follows:

$$W_h = f K_h\left[1 - \frac{f'}{2f}K_h + \frac{f''}{6f}K_h^2 - \frac{f'''}{24f}K_h^3 + O(K_h^4)\right] \tag{18}$$

$$\mu_{h\phi_j} = \phi_j\left[1 - \frac{\phi_j'}{2\phi_j}K_h + \frac{f'\phi_j' + 2f\phi_j''}{12f\phi_j}K_h^2 - \frac{ff''\phi_j' + ff'\phi_j'' + f^2\phi_j''' - f'^2\phi_j'}{24f^2\phi_j}K_h^3 + O(K_h^4)\right] \tag{19}$$

where K_h is the width of h-th stratum, the function ϕ_j, f and their derivatives are evaluated at $x = x_h$.
From (19), taking $\phi_j(t) = t$, we have

$$\mu_{hx} = x\left[1 - \frac{1}{2x}K_h + \frac{f'}{12fx}K_h^2 - \frac{ff'' - f'^2}{24f^2x}K_h^3 + O(K_h^4)\right]. \tag{20}$$

Now, multiplying (18), (19) and (20) and then summing over all the strata we shall get, by proceeding on the lines of Singh and Sukhatme (1969), the following useful expression:

$$\sum_{h=1}^{L} W_h\mu_{h\phi_j}\mu_{hx} = \mu_{\eta_j} - \frac{1}{12}\sum_{h=1}^{L}K_h^2\int_{x_{h-1}}^{x_h}\mu_j(t)f(t)dt[1 + O(K_h^2)] \tag{21}$$

where $\mu_{\eta_j} = \int_a^b \eta_j(t)f(t)dt$ and $\phi_j(t) = \eta_j(t)/t$.

For finding out approximate solutions to the minimal equations (17), as obtained in the preceding section, we shall first expand the expressions on the left hand side of the minimal equation, and then the right hand side may be expanded similarly.

On using series expansions of $\mu_{h\phi_j}$ and μ_{hx} from (19) and (20), respectively, and the result (21), the left hand side of the minimal equations (17) can be put as

$$C-\frac{K_h^2\phi_2'}{4}\left[1-\frac{1}{3}\frac{f'\phi_2'+f\phi_2''}{f\phi_2'}K_h+O(K_h^2)\right]\left[\mu_{\eta_1}-\sum_{h=1}^{L}\frac{K_h^2}{12}\int_{x_{h-1}}^{x_h}\phi_1'(t)f(t)dt[1+O(K_h^2)]\right]$$

$$+\frac{K_h^2\phi_1'}{4}\left[1-\frac{1}{3}\frac{f'\phi_1'+f\phi_1''}{f\phi_1'}K_h+O(K_h^2)\right]\left[\mu_{\eta_2}-\sum_{h=1}^{L}\frac{K_h^2}{12}\int_{x_{h-1}}^{x_h}\phi_2'(t)f(t)dt[1+O(K_h^2)]\right] \quad (22)$$

where $C=x\phi_1\sum_{h=1}^{L}W_h\mu_{h\phi_2}\mu_{hx}+x\phi_2\sum_{h=1}^{L}W_h\mu_{h\phi_1}\mu_{hx}$.

On simplifying (22) and neglecting the terms of $O(m^4)$, m = Sup K_h, we can put the left hand side of (17), after some algebraic simplifications, as

$$C-\frac{[f(\mu_{\eta_1}\phi_2'+\mu_{\eta_2}\phi_1')]^{1/3}}{4f}\left[K_h^2\int_{x_{h-1}}^{x_h}l_3(t)f(t)dt[1+O(K_h^2)]\right]^{2/3} \quad (23)$$

where $\quad l_3(t)=\mu_{\eta_1}\phi_2'(t)+\mu_{\eta_2}\phi_1'(t)$. (24)

Similarly the right hand side of the minimal equation can be shown to be equal to

$$C-\frac{[f(\mu_{\eta_1}\phi_2'+\mu_{\eta_2}\phi_1')]^{1/3}}{4f}\left[K_r^2\int_{x_{h-1}}^{x_h}l_3(t)f(t)dt[1+O(K_r^2)]\right]^{2/3}. \quad (25)$$

Substituting (23) and (25) in (17) and then equating both the sides, we get the following equalities

$$\frac{[f(\mu_{\eta_1}\phi_2'+\mu_{\eta_2}\phi_1')]^{1/3}}{4f}\left[K_h^2\int_{x_{h-1}}^{x_h}l_3(t)f(t)dt[1+O(K_h^2)]\right]^{2/3}$$

$$=\frac{[f(\mu_{\eta_1}\phi_2'+\mu_{\eta_2}\phi_1')]^{1/3}}{4f}\left[K_r^2\int_{x_{h-1}}^{x_h}l_3(t)f(t)dt[1+O(K_r^2)]\right]^{2/3} \quad (26)$$

where $l_3(t)$ is as given by (24), $r = h+1$, $h = 1,2,...,L-1$ and we assume that the function $l_3(t)f(t) \in \Omega$, the class of all functions $f(x)$, such that $x \in (a, b)$ where $(b-a) < \infty$ and the first two derivatives of $f(x)$ exist for all x in the interval (a, b).

Now, if we assume that the number of strata is large enough so that the strata widths K_h are small and higher powers of K_h in the expansion can be neglected, then the system of minimal equations (17) and equivalently the equations (26) can approximately be put as

$$\left[K_h^2 \int_{x_{h-1}}^{x_h} l_3(t)f(t)dt\right]^{2/3} = \left[K_r^2 \int_{x_h}^{x_{h+1}} l_3(t)f(t)dt\right]^{2/3} \tag{27}$$

which gives $\quad K_h^2 \int_{x_{h-1}}^{x_h} l_3(t)f(t)dt$ = constant, $h=1, 2,...,L$ (28)

or equivalently by

$Q_3(x_{h-1},x_h)$ = constant (29)

where $Q_3(x_{h-1},x_h)$ is a function of order $O(m^3)$, $m = \underset{(a,b)}{\text{Sup}}(K_h)$ such that

$$K_h^2 \int_{x_{h-1}}^{x_h} l_3(t)f(t)dt = Q_3(x_{h-1}, x_h)[1 + O(K_h^2)]\,. \tag{30}$$

Various methods of finding approximate solutions to the minimal equations (17) can be established through the system of equations (28). Singh and Sukhatme (1969) developed different forms of $Q(x_{h-1},x_h)$ corresponding to univariate case under Neyman allocation. Proceeding on the same lines, one such form of the function $Q_3(x_{h-1},x_h)$ can be obtained as

$$\left[\int_{x_{h-1}}^{x_h} \sqrt[3]{l_3(t)f(t)}dt\right]^3 = \text{constant (say, C)} \tag{31}$$

where $\quad C = \dfrac{1}{L^3}\left[\int_a^b \sqrt[3]{l_3(t)f(t)}dt\right]^3$.

We now propose the following rule of finding out approximately optimum strata boundaries (*AOSB*), in case of two study variables, when the method of varying probabilities of selection is adopted.

Cum $\sqrt[3]{R_3(x)}$ rule:

If the function $R_3(x) = l_3(x)f(x)$, is bounded and possesses its first two derivatives, then for a given value of L taking equal intervals on the cumulative cube root of $R_3(x)$, that is

$$\int_{x_{h-1}}^{x_h} \sqrt[3]{R_3(x)}dx = \frac{1}{L}\int_a^b \sqrt[3]{R_3(x)}dx$$

will give the set $\{x_h\}$ of *AOSB*.

5. EFFECT OF OPTIMUM STRATIFICATION

In this section, we shall discuss as to whether stratification provides any gain on using *PPSWR* method of selecting samples from different strata, under proportional method of allocation. For this purpose a limiting expression of generalized variance G_3 as defined by (14) is to be obtained. Now, under the regression model (3), the generalized variance G_3 can be expressed as

$$n^2 G_3 = \left(\sum_{h=1}^{L} W_h \, \mu_{h\phi_1} \mu_{hx} \right) \left(\sum_{h=1}^{L} W_h \, \mu_{h\phi_2} \mu_{hx} \right). \tag{32}$$

Further, on using result 21, the above expression can be put in its expanded form as

$$n^2 G_3 = \left[\mu_{\eta_j} - \frac{1}{12} \sum_{h=1}^{L} K_h^2 \int_{x_{h-1}}^{x_h} \phi'_1(t) f(t) dt [1 + O(K_h^2)] \right]$$

$$\times \left[\mu_{\eta_2} - \frac{1}{12} \sum_{h=1}^{L} K_h^2 \int_{x_{h-1}}^{x_h} \varphi'_2(t) f(t) dt [1 + O(K_h^2)] \right] \tag{33}$$

which on neglecting the terms of order $O(m^4)$, $m = \underset{(a,b)}{\text{Sup}} (K_h)$, can be obtained as

$$n^2 G_3 = \mu_{\eta_1} \mu_{\eta_2} - \frac{1}{12} \sum_{h=1}^{L} K_h^2 \int_{x_{h-1}}^{x_h} l_3(t) f(t) dt [1 + O(K_h^2)] \,. \tag{34}$$

For further simplifying (34), we have the following lemma due to Singh and Sukhatme (1969), which can be proved by expanding $\sqrt[\lambda]{f(t)}$ about the point $t = x_h$ on making use of Taylor's theorem.

Lemma 5.1.
$$\left[\int_{x_{h-1}}^{x_h} \sqrt[\lambda]{f(t)} \; dt \right]^{\lambda} = K_h^{\lambda-1} \int_{x_{h-1}}^{x_h} f(t) \; dt [1 + O(K_h^2)] \,. \tag{35}$$

Now, by using lemma 5.1 in (34), we get

$$n^2 G_3 = \mu_{\eta_1} \mu_{\eta_2} - \frac{1}{12} \sum_{h=1}^{L} \left[\int_{x_{h-1}}^{x_h} \sqrt[3]{l_3(t) f(t)} dt \right]^3 . \tag{36}$$

In case the strata boundaries are determined by making use of the proposed $\sqrt[3]{R_3(x)}$ rule, the equation (36) will reduce to

$$n^2 G_3 = \alpha - \frac{\beta}{12L^2}$$

or, $$G_3 = \frac{1}{n^2}\left[\alpha - \frac{\beta}{12L^2}\right] \tag{37}$$

where $\alpha = \mu_{\eta_1}\mu_{\eta_2}$ and $\beta = \left[\int_a^b \sqrt[3]{l_3(t)f(t)}\,dt\right]^3$.

Taking limit on both the sides of (37) as L→∞, we get

$$\lim_{L\to\infty} \mathrm{G}_3 = \frac{\alpha}{\mathrm{n}^2}. \tag{38}$$

From relation (38) we infer that the generalized variance G_3 will be less than α/n^2, which goes on increasing with the increase in number of strata and approaches its maximum value α/n^2. This result leads to the following remarks:

1. The limiting expression (38) is exactly same as the result (38) of Rizvi *et al.* (2000), indicating thereby that the stratified simple random sampling is equally efficient as PPSWR when the number of strata becomes large, under proportional method of allocation.
2. Here, it may be pointed out that the result (38) is also supported by Sukhatme *et al.* (1984) wherein it is observed that the efficiency of stratified PPSWR would decrease as the allocation departs from Neyman allocation.
3. It is interesting to note that the similar observation have also been made by Singh (1975), for univariate case, under proportional allocation.

The empirical investigations in this regard have been made in the next section.

6. EMPIRICAL STUDY

The effectiveness of the proposed method of finding the set $\{x_h\}$ of *AOSB* has been demonstrated empirically. For this purpose the following three density functions of the stratification variable *X* have been considered.

1. Uniform distribution: $f(x) = 1$ $1 \le x \le 2$
2. Right triangular distribution: $f(x) = 2(2-x)$ $1 \le x \le 2$
3. Exponential distribution: $f(x) = e^{-x+1}$ $1 \le x \le \infty$

These densities to some extent represent those usually encountered in practice. For the purpose of comparison of the proposed method with stratified simple random sampling (SRS), we have taken the same specifications and procedures as given by Rizvi *et al.* (2000). The point of truncation as well as the form of conditional variance i.e. $\eta_j(x) = A_j x^{g_j}$, (j = 1, 2), where $A_j > 0$ and g_j = 0, 1, 2, are also the same.

In the present investigation, only the two combinations of g_1 and g_2 viz. $g_1, g_2 =$

1, 2 and $g_1 = g_2 = 2$ could be taken as the other combinations have no significance. For instance, for $g_1 = g_2 = 0$ or $g_1 = g_2 = 1$, the function $l_3(t)$, as defined by (4.7), becomes zero, and for such cases any set of boundaries is optimum.

For the aforesaid combinations, values of *AOSB*, n^2G_3 and per cent relative efficiency of proposed method with unstratified PPSWR (% RE_1) and per cent relative efficiency of stratified PPSWR with stratified SRS (% RE_2) have been presented in tables 1, 2 and 3 for uniform, right triangular and exponential distributions, respectively. These values indicate that the percent relative efficiency goes on decreasing with the increase in number of strata, however slowly, as compared to unstratified PPSWR sampling scheme for all the three distributions considered. This trend is obvious from the relation (36) which is also supported by the numerical investigations made by Singh (1975) for univariate case. Also in each case the relative efficiency goes on decreasing with the increase in the value of g_1 or g_2 and attains minimum value for $g_1 = g_2 = 2$. Similar finding has also been reported by Singh (1975). Through perusal of the following tables it may be inferred that the % RE_1, in case of right triangular distribution, is slightly higher than that for the uniform distribution. In the former case it varies from 97.79 to 94.17 and that for the latter case it ranges between 97.27 to 93.13. When the stratification variable follows exponential distribution, the percent increase in the relative efficiency was found to be less than the other two distributions.

TABLE 1

AOSB and relative efficiencies of optimum stratification (Uniform distribution)

L	AOSB				n^2G_3	$\%RE_1$	$\%RE_2$
	$g_1, g_2 = 1, 2$						
1	1.00000	2.00000			.0066944	100.00	623.13
2	1.00000	1.49653	2.00000		.0068822	97.27	226.19
3	1.00000	1.32990	1.65981	2.00000	.0069177	96.77	155.61
4	1.00000	1.24649	1.49298	1.73947	.0069287	96.62	131.74
	2.00000						
5	1.00000	1.19648	1.39296	1.58945	.0069361	96.51	120.33
	1.78593	2.00000					
6	1.00000	1.16318	1.32635	1.48953	.0069388	96.48	114.21
	1.65271	1.81588	2.00000				
	$g_1, g_2 = 2$						
1	1.00000	2.00000			.0064554	100.00	646.21
2	1.00000	1.49723	2.00000		.0068200	94.65	228.26
3	1.00000	1.33050	1.66101	2.00000	.0068898	93.69	158.09
4	1.00000	1.24719	1.49438	1.74157	.0069128	93.38	132.04
	2.00000						
5	1.00000	1.19718	1.39436	1.59155	.0069258	93.21	120.56
	1.78873	2.00000					
6	1.00000	1.16388	1.32775	1.49163	.0069314	93.13	114.33
	1.65551	1.81939	2.00000				

If we look at $\%RE_2$ values, it is found to be as high as 862.59% for exponential distribution followed by right triangular distribution (638.10%) and uniform distribution (646.21%). A further comparison of proposed method with that of Rizvi *et al.* (2000) reveals that $\%RE_2$ goes on decreasing as the number of strata increases. And, it is remarkable to note that for larger number of strata the stratified *SRS* would found to be equally efficient as PPSWR when proportional method of allocation is envisaged. This fact is also supported by remark 2 under section 5.

TABLE 2

AOSB and relative efficiencies of optimum stratification (Right triangular distribution)

L	AOSB				n^2G_3	%RE_1	%RE_2
	g_1 , g_2 = 1, 2						
1	1.00000	2.00000			.0029920	100.00	618.75
2	1.00000	1.40142	2.00000		.0030597	97.79	242.49
3	1.00000	1.25849	1.55249	2.00000	.0030741	97.33	165.55
4	1.00000	1.19048	1.39747	1.63235	.0030797	97.15	137.55
	2.00000						
5	1.00000	1.15057	1.31065	1.48443	.0030825	97.06	124.29
	1.68111	2.00000					
6	1.00000	1.12427	1.25484	1.39351	.0030842	97.01	117.00
	1.54379	1.71337	2.00000				
	g_1 , g_2 = 2						
1	1.00000	2.00000			.0029013	100.00	638.10
2	1.00000	1.40222	2.00000		.0030337	95.63	244.57
3	1.00000	1.25929	1.55439	2.00000	.0030621	94.75	166.20
4	1.00000	1.19118	1.39907	1.63525	.0030727	94.42	137.87
	2.00000						
5	1.00000	1.15127	1.31225	1.48703	.0030780	94.26	124.47
	1.68521	2.00000					
6	1.00000	1.12507	1.25644	1.39602	.0030810	94.17	117.12
	1.54749	1.71877	2.00000				

TABLE 3

AOSB and relative efficiencies of optimum stratification (Exponential distribution)

L	AOSB				n^2G_3	%RE_1	%RE_2
	g_1 , g_2 = 1, 2						
1	1.00000	6.00000			.5944715	100.00	710.35
2	1.00000	2.55953	6.00000		.6812319	87.26	270.39
3	1.00000	1.94507	3.33092	6.00000	.7027103	84.60	178.33
4	1.00000	1.67921	2.55797	3.80547	.7108992	83.62	144.73
	6.00000						
5	1.00000	1.53028	2.17468	2.99579	.7148522	83.16	128.86
	4.12998	6.00000					
6	1.00000	1.43499	1.94388	2.55654	.7170472	82.90	120.13
	3.32717	4.36679	6.00000				
	g_1 , g_2 = 2						
1	1.00000	6.00000			.4895521	100.00	862.59
2	1.00000	2.55991	6.00000		.6428539	76.15	286.53
3	1.00000	1.94532	3.33180	6.00000	.6840050	71.57	183.21
4	1.00000	1.67946	2.55872	3.80710	.7000186	69.93	146.98
	6.00000						
5	1.00000	1.53053	2.17531	2.99692	.7077993	69.16	130.14
	4.13223	6.00000					
6	1.00000	1.43511	1.94426	2.55729	.7121312	68.74	120.96
	3.32855	4.36941	6.00000				

Division of Agricultural Economics & Statistics SYED EJAZ HUSAIN RIZVI
S.K. University of Agricultural Sciences and Technology
Main Campus, Chatha, Jamnu, India

Meerut City (U.P.) India JAI P. GUPTA

Department of Agronomy MANOJ BHARGAVA
HPKV, Palampur, HP, India

ACKNOWLEDGEMENT

The authors wish to thank the referee for his constructive suggestions and valuable comments.

REFERENCES

T. DALENIUS, (1950), *The problem of optimum stratification*, "Skandinavisk Aktuaritidskrift", 33, pp. 203-213.

S.E.H. RIZVI, J.P. GUPTA, R. SINGH (2000), *Approximately optimum stratification for two study variables using auxiliary information*, "Journal of the Indian Society of Agricultural Statistics", 53, pp. 287-298.

R. SINGH (1975), *A note on optimum stratification in sampling with varying probabilities*, "Australian Journal of Statistics", 17, pp. 12-21.

R. SINGH, B.V. SUKHATME (1969), *Optimum stratification*, "Annals of the Institute of Statistical Mathematics", 21, pp. 515-528.

R. SINGH, B.V. SUKHATME (1972), *Optimum stratification in sampling with varying probabilities*, "Annals of the Institute of Statistical Mathematics", 24, pp. 485-494.

P.V. SUKHATME, B.V. SUKHATME, S. SUKHATME, C. ASHOK (1984), *Sampling theory of surveys with applications*, 3rd edition, Indian Society of Agricultural Statistics, New Delhi, India, pp. 175-176.

RIASSUNTO

Effetto della stratificazione nel campionamento a probabilità variabile nel caso di allocazione proporzionale

Il problema della stratificazione ottima di una variabile ausiliare quando le unità dei differenti strati sono scelte con probabilità proporzionale al valore della variabile ausiliare è stata considerata da Singh (1975) nel contesto univariato. In questo lavoro viene esteso il metodo al caso di due variabili, proponendo una regola per ottenere un'approssimazione dei limiti degli strati. Attraverso sviluppi teorici e applicativi viene mostrato che l'uso della stratificazione ha effetti inversi sulla efficienza relativa rispetto alla non-stratificazione.

SUMMARY

Effect of optimum stratification on sampling with varying probabilities under proportional allocation

The problem of optimum stratification on an auxiliary variable when the units from different strata are selected with probability proportional to the value of auxiliary variable (PPSWR) was considered by Singh (1975) for univariate case. In this paper we have extended the same problem, for proportional allocation, when two variates are under study. A cum. $\sqrt[3]{R_3(x)}$ rule for obtaining approximately optimum strata boundaries has been provided. It has been shown theoretically as well as empirically that the use of stratification has inverse effect on the relative efficiency of PPSWR as compared to unstratified PPSWR method when proportional method of allocation is envisaged. Further comparison showed that with increase in number of strata the stratified simple random sampling is equally efficient as PPSWR.