응용통계연구 제 6권 1호
1993년 3월 pp.145 - 162

A Study on Unequal Probability Sampling over Two Successive Occasions in Time Series[4]

Hong Nai Park[5] and Kay O Lee[6]

Abstract

We review sampling schemes on successive occasions with partial replacement of units and propose a Rao-Hartley-Cochran(RHC) type's sampling scheme over two successive occasions with probability proportionate to observations on the previous occasion. For comparison of the reviewed and proposed sampling schemes, optimal estimator of population mean on second occasion and its variance are derived. The relative efficiency of the proposed sampling scheme is compared with other equal and unequal probability sampling scheme by theoretical and numerical simulation study. For simulation study, three artificial populations are generated by a time series model. It is observed that RHC type's sampling scheme has small variance and deviation in general.

1. Introducton

Repeated sampling survey should be executed when the characteristic of populaton is changed in the course of time or when statistic is produced by the periodic sampling survey. In accomplishing the repeated sampling survey in the same population, three sampling schemes can be considered as an appropriate methods ; such as the repeated sample survey on every occasion, independent sample survey on each occasion and the sample survey combining the above two methods. And to increase the precision of estimate and to attain the purpose of

[4] The present study was supported by the Basic Science Research Institute Program, Ministry of Education, 1991, Project NO. 108

[5] Department of Computer Science and Statistics, Seoul National University, Seoul 151-742, Korea.

[6] Department of Computer Science, Korea Air Force Academy, Chongju, Chungbuk 363-849

survey, the following sampling schemes can be considered.

(1) When the variation of the characteristic of population is to be estimated at different points of time, the same sample is surveyed over all occasions.
(2) In estimating the mean of the over all survey period, an independent sample is taken and investigated on each occasion.
(3) When estimating the population mean on each occasion, a part of the sample is retained while the rest is drawn afresh.

Many statisticians have been concerned about the sampling method that a part of the sample is replaced on each occasion and the replaced sample is taken by the equal probability sampling. However, there are few methods for the unequal probability sampling. Especially, in the repeated sampling survey, probability proportional sampling method can increase the precision by using the observed results at the previous occasion as the measure of probability. This study on the process of unequal probability sampling is limited on two successive occasions survey.

First of all, simple random sampling with replacement is reviewed as an equal probability sampling and it is used with the criterion of comparison. The second method is a partial probability proportional sampling(PPPS) which is to select sample on the first occasion and the independent sample on the second occasion by unequal probability selection, and then to select matched sample on the second occasion by equal probability sampling like SRSWOR. The third method is the complete probability proportional sampling(CPPS) which is the same as the first method, except the selection of the matched sample on the second occasion through using the observed data on the first occasion as the measure of the probability proportional sampling. In each scheme the optimal estimator and variance of estimator are derived and the relative efficiencies are compared. Finally, with the assumption of a special time series model, numerical comparison is made by doing computer simulation for unequal probability sampling.

2. Simple Random Sampling with Replacement(Scheme I)

Assume that, for the simplicity of study, the population size is 'N' and the sample size is 'n' at every time of investigation. In practice, sampling with replacement will be applied to simplify the comparison analysis about unequal probability sampling and the results of the comparison analysis will be easily enlarged to sampling without replacement.

On the first occasion, sample(s_1) is composed of n units selected from the population(U) of size N by SRSWR, and on the second occasion the matched sample(s_{2m}) and unmatched sample(s_{2u}) are composed of m units from s_1 by SRSWOR and of u units from subpopulation (U-s_1) by SRSWR, respectively.

Here it can be easily known that sample(s_2) on the second occasion is composed of the union of s_{2m} and s_{2u}, and the selection of sample on the third or more occasions is not different from that on the second occasion. Therefore, let's deal with the sampling scheme only on the two successive occasions.

On the second occasion, an estimator of population mean (μ_2) is given by the weighted mean of two independent estimators on the basis of matched sample(s_{2m}) and unmatched sample(s_{2u}).

$$\hat{\mu}_{2I} = a\ \hat{\hat{Y}}_{2u} + (1-a)\ \hat{\hat{Y}}_{2m} \tag{2.1}$$

where $\hat{\hat{Y}}_{2u}$ is an estimator of μ_2 based on s_{2u} and $\hat{\hat{Y}}_{2m}$ is a regression estimator of μ_2 based on s_{2m}.

Regression estimator of μ_2 and variance of it are given as follows.

$$\hat{\hat{Y}}_{2m} = \bar{y}_{2m} + b_{21}(\ \bar{y}_{1.} - \bar{y}_{1m}) \tag{2.2}$$

where $\bar{y}_{2m}$ is the average of matched sample of size m on the second occasion,

$\bar{y}_{1.}$ is the average of sample of size n on the first occasion,

$\bar{y}_{1m}$ is the average of the matched sample of size m on the first occasion and b_{21} is the regression coefficient (=$\rho_{21}S_2/S_1$) of the variate of the second occasion on the variate of the first occasion.

$$Var(\ \hat{\hat{Y}}_{2m}) = (\frac{1}{m} - \frac{1}{n})\ S_2^2(1-\rho_{21}^2) + \frac{S_2^2}{n} \tag{2.3}$$

Since an estimator of μ_2 observed on s_{2u} is the simple average of sample, it can be easily shown that

$$\hat{\bar{Y}} = \frac{1}{u}\sum_{s_{2u}} Y_{2i} \qquad (2.4)$$

$$Var(\hat{\bar{Y}}_{2u}) = \frac{S_2^2}{u}$$

In (2.1), since the weighted values of $\hat{\bar{Y}}_{2u}$ and $\hat{\bar{Y}}_{2m}$ are inverse numbers of variances, the estimator in (2.1) can be represented as

$$\hat{\mu}_{2I} = \frac{Var(\hat{\bar{Y}}_{2m})\hat{\bar{Y}}_{2u} + Var(\hat{\bar{Y}}_{2u})\hat{\bar{Y}}_{2m}}{Var(\hat{\bar{Y}}_{2u}) + Var(\hat{\bar{Y}}_{2m})} \qquad (2.5)$$

and variance of estimator is as follows.

$$Var(\hat{\mu}_{2I}) = \frac{S_2^2(n - u\rho_{21}^2)}{n^2 - u^2\rho_{21}^2} \qquad (2.6)$$

3. Unequal Probability Sampling Schemes

Suppose that the characteristic of population changes on different occasions and has a stochastic probability model. Then it is natural to use the unequal probability proportional sampling on the basis of the information obtained in the previous sample survey. For example, when the Ministry of Construction declares the quarterly fluctuation index of land price which is produced from the data obtained by taking a census of all over the country in the beginning of the year and by the sampling survey in each quarter, the efficiency of the sampling strategy may be increased by using the over all surveyed data for selected sample at the second quarter as the measure of probability proportional sampling. Though the various types of unequal probability sampling can be considered, we will describe PPPS and CPPS in detail, so as not to fall into complexity and to lose the generality of comparison analysis among sampling schemes.

3.1 Partial Probability Proportional Sampling(PPPS : Scheme II)

When the previous information about each unit of population is known, the probability proportional sampling is an efficient sampling scheme. If the census or the similar sample survey had been executed in the same population, then the investigated data can be used as probability measures in probability proportional sampling. Let's deal with a sampling scheme which applied Raj's procedure(1965) to the successive investigations. In estimating the mean, PPS will be adopted.

Suppose that the measure of each unit employed in probability proportional sampling is stochastically similar to the value of interested variable. Let the measure of the ith unit be described as follows. The sample(s_1) on the first occasion is composed of n units selected from the population of size N and PPS is used to select this sample with the probability p_i for ith unit.

The sample(s_2) on the second occasion is formed by two samples : the matched sample(s_{2m}) and the unmatched sample(s_{2u}). s_{2m} is composed of m units selected from s_1 by SRSWOR and s_{2u} is composed of u(u=n-m) units selected from population by PPS.

An estimator of μ_2 on the second occasion is given by the weighted average of two independent estimators derived from s_{2u}, s_{2m} and s_1, respectively.

$$\hat{\mu}_{2U} = \beta\ \hat{\hat{Y}}_{2u} + (1-\beta)\ \hat{\hat{Y}}_{2m} \tag{3.1}$$

where $\hat{\hat{Y}}_{2u} = \frac{1}{Nu}\sum_{s_{2u}}\frac{Y_{2j}}{P_j}$, $\quad \hat{\hat{Y}}_{2m} = \frac{1}{Nm}\sum_{s_{2m}}\frac{Y_{2j}-Y_{1j}}{P_j} + \frac{1}{Nn}\sum_{s_1}\frac{Y_{1j}}{P_j}$.

The variance of estimator based on the independent sample on the second occasion is given by

$$Var(\ \hat{\hat{Y}}_{2m}) = \frac{1}{u}\sum_{i=1}^{N} P_1\left(\frac{Y_{2i}}{NP_i} - \mu_2\right)^2 \tag{3.2}$$

and the variance of estimator based on the matched sample can be represented as

$$Var(\ \hat{\hat{Y}}_{2m}) = \left(\frac{1}{m} - \frac{1}{n}\right)\left[\ \sum_{i=1}^{N} P_i\left(\frac{Y_{2i}}{NP_i} - \mu_2\right)^2 + \sum_{i=1}^{N} P_i\left(\frac{Y_{1i}}{NP_i} - \mu_1\right)^2\right.$$

$$-2\sum_{i=1}^{N} P_i\left(\frac{Y_{2i}}{NP_i}-\mu_2\right)\left(\frac{Y_{1i}}{NP_i}-\mu_1\right)\Big] \tag{3.3}$$

In (3.3), since the optimal values of weighted number are proportional to the inverse of vaiances, the equation (3.1) can be written as

$$\hat{\mu}_{2II}=\frac{Var(\hat{\hat{Y}}_{2m})\hat{\hat{Y}}_{2u}+Var(\hat{\hat{Y}}_{2u})\hat{\hat{Y}}_{2m}}{Var(\hat{\hat{Y}}_{2u})+Var(\hat{\hat{Y}}_{2m})} \tag{3.4}$$

Therefore, the variance of the estimator of μ_2 in equation (3.4) is

$$Var(\hat{\mu}_{II})=\frac{\frac{1}{mn}V_{pps}(Y_2)\{V_{pps}(Y_2)+V_{pps}(Y_1)-2COV_{pps}(Y_1,Y_2)\}}{\frac{1}{u}V_{pps}(Y_2)+\frac{m-n}{mn}\{V_{pps}(Y_2)+V_{pps}(Y_1)-2COV_{pps}(Y_1,Y_2)\}}$$

where $V_{pps}(Y_k)=\sum_{i=1}^{N}P_i\left(\frac{Y_{ki}}{NP_i}-\mu_k\right)^2,\quad k=1,\ 2.$

3.2 Complete Probability Proportional Sampling(SchemeⅢ)

Rao-Hartley-Cochran(RHC)'s PPS is largely quotated as unequal probability sampling which is made use of the all obtainable information about the interested variations in sample design. In this section, consider the sample selection scheme which applies RHC's PPS to the repeated sample survey. Suppose that the information of the measure for PPS is given and is closely related with the observed value of sample unit. For example, if the value of each unit in population is known, it can be used as a measure of PPS. Assuming that, for simplicity of sample selection, N/n, n/m and N/u are integer values, we propose the following sampling scheme.

Divide the population of size N at random into n groups of size k(k = N/n) and draw a sample of size one with the measure $Z_{oi}(=Y_{oi}/\sum_{i=1}^{N}Y_{oi})$ from each of these n groups by PPS. Sample(s_1) on the first occasion is composed of these n units. The sample(s_2) on the second occasion is the union of the matched sample(s_{2m}) and the independent sample(s_{2u}). Split s_1 into m groups of size k_1(= n/m) and select a sample of size one with measure $Z_{1i}(=Y_{1i}/\sum_{i=1}^{n}Y_{1i})$ from each of these m groups by PPS and then s_{2m} is composed of these m units. The sample s_{2u} is selected as follows. Divide the population of size N into u groups of size k_2(= N/u) and draw a sample of size one with measure $Z_{oi}(=Y_{oi}/\sum_{i=1}^{N}Y_{oi})$ by PPS.

On the second occasion, an estimator ($\hat{\mu}_{2m}$) of mean is evaluated as

$$\hat{\mu}_2=\phi\ \hat{\hat{Y}}_{2u}+(1-\phi)\ \hat{\hat{Y}}_{2m} \tag{3.6}$$

where $\hat{\hat{Y}}_{2u}=\frac{1}{N}\sum_{S_{2u}}Y_{2i}\frac{Z^*_{oi}}{Z_{oi}}$, and

$$\hat{\hat{Y}}_{2m}=\frac{1}{N}\sum_{S_{2m}}\frac{(Y_{2i}-Y_{1i})\ Z^*_{1i}}{Z_{1i}}\cdot\frac{Z^*_{oi}}{Z_{oi}}+\frac{1}{N}\sum_{S_1}\frac{Y_{1i}\ Z^*_{oi}}{Z_{oi}},$$

Z^*_{oi} is the sum of Z_{oi} in ith group
Z^*_{1i} is the sum of Z_{1i} in ith group.

In (3.6), the optimal value of weighted number is proportional to the inverse value of its variance and μ_{2m} can be represented as

$$\hat{\mu}_{2III}=\frac{Var(\ \hat{\hat{Y}}_{2m})\ \hat{\hat{Y}}_{2u}+Var(\ \hat{\hat{Y}}_{2u})\ \hat{\hat{Y}}_{2m}}{Var(\ \hat{\hat{Y}}_{2u})+Var(\ \hat{\hat{Y}}_{2m})} \tag{3.7}$$

The variance of each estimator is easily calculalted as follows.

$$Var(\ \hat{\hat{Y}}_{2u})=\frac{N-u}{N-1}\cdot\frac{1}{u}\sum_{i}^{N}Z_{oi}(\ \frac{Y_{2i}}{NZ_{oi}}-\mu_2)^2 \tag{3.8}$$

$$Var(\ \hat{\hat{Y}}_{2m})=E_1\ V_2(\ \hat{\hat{Y}}_{2m})+V_1\ E_2(\ \hat{\hat{Y}}_{2m})$$

$$E_2(\ \hat{\hat{Y}}_{2m})=\frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{k_1}\frac{Z_{1j}}{Z^*_{1j}}\cdot\frac{(Y_{2i}-Y_{1i})\ Z^*_{1i}}{Z_{1i}}\cdot\frac{Z^*_{oi}}{Z_{oi}}$$

$$+\frac{1}{N}\sum_{i=1}^{n}\frac{Y_{1i}\, Z^{*}_{oi}}{Z_{oi}}$$

$$V_1 E_2(\hat{\hat{Y}}_{2m}) = V_1\left(\frac{1}{N}\sum_{i=1}^{n}\frac{Y_{1i}\, Z^{*}_{oi}}{Z_{oi}}\right)$$

$$= \frac{N-n}{n(N-1)}\cdot\sum_{t=1}^{N} Z_{\alpha t}\left(\frac{Y_{2t}}{NZ_{\alpha t}}-\mu_2\right)^2$$

$$V_2(\hat{\hat{Y}}_{2m}) = \frac{1}{N^2}\cdot\frac{(n-m)}{mnN(N-1)}\sum_{i=1}^{n}\left(\frac{Z^{*}_{oi}}{Z_{oi}}\right)^2\left(\frac{W_{2i}^{\ 2}}{Z_{1i}}-W_2^{\ 2}\right)$$

where $W_{2i}=y_{2i}-y_{1i}$.

$$E_1 V_2(\hat{\hat{Y}}_{2m}) = \frac{(n-m)}{mnN(N-1)}\sum_{t=1}^{N}\frac{Z^{*}_{oi}}{Z_{oi}}\left(\frac{W_{2t}^{\ 2}}{Z_{1t}}-W_2^{\ 2}\right)$$

$$Var(\hat{\hat{Y}}_{2m}) = \frac{N(n-m)}{mn(N-1)}\sum_{t=1}^{N}\frac{Z^{*}_{oi}}{Z_{oi}}\left(\frac{W_{2t}^{\ 2}}{Z_{1t}\cdot N^2}-W_2^{\ 2}\right)$$

$$+\frac{N-n}{n(N-1)}\cdot\sum_{t=1}^{N} Z_{\alpha t}\left(\frac{Y_{2t}}{NZ_{\alpha t}}-\mu_2\right)^2 \qquad (3.9)$$

Since variance of $\hat{\mu}_{2III}$ is the sum of the inverse of variance for each estimator, it can be written as

$$Var(\hat{\mu}_{2III}) = \frac{Var(\hat{\hat{Y}}_{2u})Var(\hat{\hat{Y}}_{2m})}{Var(\hat{\hat{Y}}_{2u})+Var(\hat{\hat{Y}}_{2m})}$$

Applying the results of calculation for variance into the above equation, the variance of the estimator ($\hat{\mu}_{2III}$)can be represented as

$$Var(\hat{\mu}_{2III}) = \frac{\frac{N(N-u)}{m(N-1)}V_{ppz}(Y_2)V_{ppz}(W_2)+\frac{(N-u)(N-n)}{u(N-1)}V_{ppz}^{\ 2}(Y_2)}{\frac{N(u+n)-2un}{u}V_{ppz}(Y_2)+\frac{Nu}{n}V_{ppz}(W_2)} \qquad (3.10)$$

where $V_{ppz}(Y_k)=\sum_{t=1}^{N} Z_{\alpha t}\left(\frac{Y_{kt}}{Z_{\alpha t}}-\mu_k\right)^2.$

4. Comparison of Sampling Schemes

The proposed unequal probability sampling in the previous section makes use of the observed results on the previous occasion in estimation of parameters as well as in sample selection process. In order to determine the optimal sample selection strategy by the comparison of three sampling schemes mentioned in the previous sections, it is natural to compare and analyze them on the basis of economical efficiency, convenience and efficiency, but variances of the estimator are only compared with each other for the purpose of simplicity.

4.1 Comparison between Sampling Scheme II and Sampling Scheme I

In sampling strategy for successive occasion investigations, where the unequal probability sampling scheme is compared with equal probability sampling scheme on the basis of the variance, direct comparison can not be made easily. Therefore, they should be compared under some restricted assumptions. Assume that $V_{pps}(Y_1) = V_{pps}(Y_2)$ and $\delta = \frac{COV_{pps}(Y_1, Y_2)}{V_{pps}(Y_2)}$ in (3.5). Then the variance of the estimator($\hat{\mu}_{2II}$) for the sampling scheme II can be represented as

$$Var(\hat{\mu}_{2II}) = \frac{2uV_{pps}(Y_2)(1-\delta)}{2u^2+mn-2u^2\delta} \tag{4.1}$$

Let the proportion of the sample replacement be 0.5 on each occasion. Then m=u=n/2 ,and (2.6) and (4.1) can be written as follows.

$$Var(\hat{\mu}_{2II}) = \frac{V_{pps}(Y_2)(1-\delta)}{n(1-\frac{\delta}{2})} \tag{4.2}$$

$$Var(\hat{\mu}_{2I}) = \frac{S_2^2(1-\frac{\rho_{21}^2}{2})}{n(1-\frac{\rho_{21}^2}{4})} \tag{4.3}$$

If $P_i = \frac{1}{N}$, then $V_{pps}(Y_2) \approx S_2^2$. And the correlation coefficient between Y_{2i} and Y_{1i} is considered as positive value.

Also if $\delta = \rho_{21}$ and the correlation of coefficient is positive,sampling scheme II is efficient to sampling scheme I since the following facts are easily verified.

$$RE(II,I) = \frac{(1-\delta)(1-\frac{\rho_{21}^2}{4})}{(1-\frac{\delta}{2})(1-\frac{\rho_{21}^2}{2})}$$

$$= 1 - \frac{\delta}{2-\delta^2}, \quad \rho_{21} = \delta \tag{4.4}$$

If $0 < \delta < 1$, then $Var(\mu_{2II}) < Var(\mu_{2I})$.

4.2 Comparison between Sampling SchemeIII and Sampling Scheme I

Sampling schemeIII is an unequal probability sampling scheme which applies the observed results in the previous occasions to updated sampling selection process, and is the strategy which use the maximum information available. If $V_{ppz}(Y_1) = V_{ppz}(Y_2)$ and $V_{ppz}(W_2) = 2V_{ppz}(1-\rho)$, then (3.10) is represented as

$$Var(\hat{\mu}_{2III}) = \frac{\frac{(N-u)V_{ppz}(Y_2)}{(N-1)}\left\{\frac{2N}{m}(1-\rho) + \frac{(N-n)}{u}\right\}}{\frac{1}{un}\{nN(u+n) - 2un^2 + 2u^2N(1-\rho)\}} \tag{4.5}$$

And if the proportion of sample replacement is 0.5 on each occasion, then m=u=n/2 . Applying these to (4.5), it is easily shown that

$$Var(\hat{\mu}_{2III}) = \frac{(2N-n)\{N(3-2\rho)-n\}V_{ppz}(Y_2)}{(N-1)\{N(4-\rho)-2n\}n} \tag{4.6}$$

To find the relative efficiency of sampling schemeIII to sampling scheme I, divide the equation (4.6) by the equation (4.3) and apply the following conditions : $V_{ppz} = S_2^2$ and $\rho = \rho_{21}$, then the following equation is obtained easily. The result relation is given in the following term.

$$RE(III,I) = \frac{(2-f)(3-f-2\rho)(1-\frac{\rho^2}{4})}{(4-2f-\rho)(1-\frac{\rho^2}{2})} \tag{4.7}$$

When sampling fraction(f) is not small and correlation cofficient (ρ) is sufficiently large, it can be observed from the numerical calculation result of (4.7) that sampling schemeIII is efficient to sampling scheme I, but it is difficult to draw the general conclusions from the result directly. There is an important problem to be studied deeply all the more : evaluation of the optimal value of sample replacement proportion when correlation coefficient ρ is varied. It is desirable to compare the relative efficiency on the basis of the optimal proportion of sample replacement.

4.3 Comparison between Sampling SchemeIII and Sampling Scheme II

Sampling schemeIII and sampling scheme II are PPS in common. Sampling scheme II is applied to sampling selection on the successive occasion survey without the change of probability proportional measure(P_i). But sampling schemeIII adopts the changing probability proportional measure(Z_{1i}) since the observed result on the previous occasion is reflected to the matched sample selection process.

By using the result evaluated on the previous occasion, the relative efficiency is obtained as follows.

$$RE(III,II) = \frac{Var(\hat{\mu}_{2III})}{Var(\hat{\mu}_{2II})}$$

$$= \frac{(2N-n)(N(3-2\rho)-n)V_{ppz}(Y_2)}{(N-1)(N(4-\rho)-2n)n} \cdot \frac{n(2-\delta)}{V_{pps}(Y_2)(2-2\delta)} \tag{4.8}$$

If it is assumed that $V_{ppz}(Y_2) = V_{pps}(Y_2)$ and $\rho = \delta$, then the relative efficiency is given by the function of sampling fraction (f = n/N) and ρ .

$$RE(III,II)=\frac{(2-f)(3-f-2\rho)}{(4-2f-\rho)}\cdot\frac{(2-\rho)}{(2-2\rho)} \tag{4.9}$$

In (4.9), RE(III, II) is less than one whenever f and ρ are moderately large or when f goes to zero and $0<\rho<2-\sqrt{2}$. So it is difficult to say generally that sampling schemeIII is more efficient than sampling scheme II. It can be considered that such results do not come from the application of the previous observations to the process of the matched sample selection, but the properties of Rao-Hartley-Cochran scheme.

4.4 Numerical Comparisons

When the discussed sampling schemes are applied to time series data, the optimal sampling strategy may not be determined by analytical comparison and practical conclusion cannot be obtained. So numerical comparison of sampling schemes will be made in the simulated population which are generated by a simple time series model.

The procedures of numerical comparison are followed in three steps ; generation of population, implementation of three sampling schemes and computations of estimates and variances for each sampling scheme. Such sampling strategies are operated 20 times as follows;

(1) Three populations of size 64 units which are generated by the following time series model.

$$Y_{ti}=0.9Y_{(t-1)}+e_{ti},\quad i=1,\ 2,\ \ldots,\ 64;\quad t=1,\ 2 \tag{4.10}$$

where Y_{oi} ~ n(5,4) and e_{ti} ~ n(0,1).

The population consisting of Y_{oi} is used to probability measure in unequal probability sampling, the population on the first occasion is composed of Y_{1i} and the other population at the second occasion is composed of Y_{2i}.

(2) On the first occasion, a sample s_1 of size 16 is selected by the corresponding sampling scheme. On the second occasion, the matched sample s_{2m} of size 8 units is selected by the following methods ; under

the schemes I and II, s_{2m} is selected by SRSWOR from s_1, for the scheme III s_{2m} is obtained by the probability proportional to Y_{1i}(observations in s_1), and the independent sample s_{2u} is selected by the given sampling scheme in the previous subsection. For each sampling strategy, estimate and variance are computed.

(3) Repeat the process of (2) 20 times and compute average of estimates and variance. Next, comparison of sampling schemes is done by comparing estimates and average with population mean on second occasion(μ_2)

The generated populations and the measure of probability proportional selection are presented in table 1. For each sampling strategy, 20 estimates are obtained by the above procedure. The results of simulation study are given in table 2. Average of 20 estimates and variance are summarized as follows ;

method	average	variance	deviation
scheme I	3.857	0.8027	-0.206
scheme II	4.225	0.1632	0.162
scheme III	3.964	0.1544	-0.099

From Tables 1 ~ 2 the followings can be observed :

(1) Scheme I leads estimates to have the largest variance among the schemes. The range of estimstes in scheme I is [1.822, 5.341], but them of schemes II and III are[3.601, 5.002] and [2.793, 4.749] respectively.

(2) Scheme II has a tendency to overestimate slightly. The average of 20 estimates is 4.225 with the positive deviation 0.162. Estimates of scheme II have the shortest range, but variance is larger than scheme III .

(3) Scheme III leads the most efficient estimate, since the variance of eatimator and the absolute deviation of the average of eatimates are the smallest among the three schemes respectively.

(4) Since schemes II and III make good use of the previous observations and the probability measure which are highly correlated with the study variables, they can be performed well on any case of time series data.

5. Conclusions

In this paper, when sampling survey is periodically executed on the population consisting of time series data, unequal probability sampling schemes are reviewed and proposed. The proposed sampling scheme makes use of the previous observations as a probability measure on the current occasion sample selection.

The performance of the proposed scheme is compared with the reviewed ones by the theoretical relative efficiency and the simulated study. The deviation and variance of the estimates of the schemeIII are the smallest among three schemes. The proposed schemeIII is more efficient than the schemeII, but the advantage of schemeIII with respect to schemeII is not so great as expectation. Such results may come from randomly dividing the population into n groups. Since schemeIII is a type of Rao-Hartley-Cochran's sampling method, the estimates of variance for the population mean are always nonnegative. In computing the estimator, we can use the estimate of variance instead of the unknown variance.

For each sampling scheme, it is of interest to study the estimaion of the variance for estimator, because the optimal estimator contains the unknown variance.

However the studies of estimation of variance and optimal proportion of sample replacement are not dealt with in this paper. We leave these problems for further research.

Table 1 Simulated Populations and Probability measures

unit	$P_0(i)$	U_I	$P_1(i)$	U_{II}
1	0.0152	6.0832	0.0188	5.5905
2	0.0101	2.9126	0.0098	1.2733
3	0.0174	6.4739	0.0199	4.4071
4	0.0231	7.0713	0.0217	7.7477
5	0.0088	3.7027	0.0121	2.5479
6	0.0158	5.1618	0.0162	4.6473
7	0.0208	7.1040	0.0220	4.7156
8	0.0152	5.6665	0.0176	5.4750
9	0.0123	3.1515	0.0104	0.0899
10	0.0202	5.2585	0.0165	5.4284
11	0.0180	5.7067	0.0179	4.8307
12	0.0196	6.5480	0.0202	5.7908
13	0.0076	2.5991	0.0087	1.8012
14	0.0145	5.5286	0.0173	5.2477

U_I : Population on the first occasion
U_{II} : Population on the second occasion

15	0.0095	0.2611	0.0020	0.4888
16	0.0019	1.2300	0.0049	0.1598
17	0.0104	0.3501	0.0023	0.3103
18	0.0180	8.0653	0.0246	5.8440
19	0.0133	4.0313	0.0130	2.4202
20	0.0164	4.4721	0.0142	3.2010
21	0.0145	5.2785	0.0165	3.2567
22	0.0227	7.8734	0.0240	6.3833
23	0.0136	4.6918	0.0147	4.7241
24	0.0136	4.6095	0.0147	2.3790
25	0.0231	7.1738	0.0220	5.4991
26	0.0164	5.5575	0.0173	2.9221
27	0.0009	0.1845	0.0017	0.2239
28	0.0145	3.7838	0.0121	2.3567
29	0.0177	4.2659	0.0136	3.7771
30	0.0142	4.9957	0.0156	5.0470
31	0.0133	4.7646	0.0150	2.5708
32	0.0120	3.4059	0.0113	4.4248
33	0.0268	8.3519	0.0254	7.2729
34	0.0284	8.8020	0.0269	9.2805
35	0.0139	4.3405	0.0139	3.5083
36	0.0152	4.8376	0.0153	4.7538
37	0.0240	7.9876	0.0243	7.1979
38	0.0139	3.1040	0.0104	1.0359
39	0.0079	3.2340	0.0107	2.6468
40	0.0126	3.1444	0.0104	1.7335
41	0.0199	4.3035	0.0139	4.1459
42	0.0224	7.5857	0.0237	8.8999
43	0.0243	8.1490	0.0248	8.0590
44	0.0120	5.3491	0.0168	2.6634
45	0.0189	6.0064	0.0188	3.9596
46	0.0073	1.9927	0.0069	1.0759
47	0.0186	5.6218	0.0176	5.1007
48	0.0057	1.0948	0.0043	0.0315
49	0.0303	9.5467	0.0315	9.1222
50	0.0167	5.7337	0.0179	3.9809
51	0.0145	4.0795	0.0130	4.9937
52	0.0129	2.7081	0.0092	1.8603
53	0.0142	4.6570	0.0147	3.3647
54	0.0111	2.2296	0.0078	0.9136
55	0.0133	3.1744	0.0104	3.2967
56	0.0164	5.8853	0.0182	4.9354
57	0.0158	5.5585	0.0173	7.8297
58	0.0028	1.5688	0.0058	0.5351
59	0.0224	4.9224	0.0156	4.2800
60	0.0287	12.0603	0.0361	12.0127
61	0.0249	7.5884	0.0231	6.7748
62	0.0073	4.5460	0.0144	2.9042
63	0.0120	4.4963	0.0142	3.0123
64	0.0202	5.8421	0.0182	6.2632

Table 2 Estimates of μ_2 and Deviations ($D_i = \mu_{2i} - \mu_2$) obtained by Simulation Study (N = 64, n = 16, m = 8)

Sample	μ_{2I}	D_I	μ_{2I}	D_{II}	μ_{3III}	D_{III}
1	5.319	1.256	4.528	0.465	3.919	-0.144
2	4.050	-0.013	4.038	-0.025	3.772	-0.291
3	4.543	0.480	4.065	0.002	3.892	-0.171
4	4.359	0.296	4.352	0.289	4.060	-0.003
5	4.326	0.263	4.647	0.584	4.189	0.126
6	3.235	-0.828	4.213	0.150	3.708	-0.355
7	5.000	0.937	4.874	0.811	4.749	0.686
8	5.341	1.278	4.062	-0.001	4.124	0.061
9	3.999	-0.064	4.235	0.172	3.952	-0.111
10	4.256	0.193	5.002	0.939	4.655	0.592
11	3.563	-0.500	4.717	0.654	3.061	-0.002
12	2.962	-1.101	4.503	0.440	3.839	-0.224
13	3.696	-0.367	3.646	-0.417	4.034	-0.029
14	3.545	-0.518	3.601	-0.462	4.307	0.244
15	1.822	-2.241	3.975	-0.088	3.907	-0.156
16	2.838	-1.225	4.457	0.394	3.903	-0.160
17	3.106	-0.957	3.655	-0.408	3.961	-0.102
18	2.575	-1.488	4.332	0.269	3.736	-0.327
19	4.134	0.071	3.779	-0.284	3.703	-0.360
20	4.477	0.414	3.811	-0.252	2.793	-1.270
Average	3.857	-0.206	4.225	0.162	3.963	-0.099
Variance	0.8027	.	0.1632	.	0.1544	.

References

[1] Avadhani, M.S. and Sukhatme(1970), "A Comparison of Two Sampling Procedures with an Application to Successive Sampling," *Applied Statistics*, 19, 251-259.

[2] Chaudhuri, A. and Arnab, R.(1977), "On the Relative Efficiencies of a Few Strategies of Sampling with Varying Probabilities on Two Occasions," *Calcutta Statistics Association Bulletine*, 26, 25-38.

[3] Chotai, J.(1974), "A Note on Rao-Hartley-Cochran method for PPS Sampling over Two Occasions," *Sankhya, Ser.C*, 36, 173-180.

[4] Raj, Des(1965), "On Sampling over Two Occasions with Probability Proportionate to Size," *Annals of Mathematical Statistics*, 36, 327-330.

[5] Ghangurde, P.D. and Rao, J.N.K.(1969), "Some Results on Sampling over Two Occasions," *Sankhya, Ser.A*, 31, 463-472.

[6] Rao, J.N.K., Hartley, H.O., and Cochran, W.G.(1962), "On a Simple Procedure of Unequal Probability Sampling without Replacement," *Journal of the Royal Statistics Society, Ser.B,* 24, 482-491.

[7] Patterson, H.D.(1950), "Sampling on Successive Occasions with Partial Replacement of Units," *Journal of the Royal Statistics Society, Ser.B,* 12, 241-255.

[8] Saxena, R.R., Singh, P., and Srivastava, A.K.(1986), "An Unequal Probability Sampling Scheme," *Biometrika,* 73, 761-763.

[9] Scott, A.J. and Smith, T.M.F.(1974), "Analysis of Repeated Surveys Using Time Series Methods," *Journal of the American Statistical Association,* 69, 674-678.

[10] Tripathi, T.P. and Srivastava, O.P.(1979), "Estimation on Successive Occasions Using PPSWR Sampling," *Sankhya, Ser.C,* 41, 84-91.

[11] 박홍래(1967). *Comparison of Estimators for two Multistage Designs when Sampling on Successive Occasions,* Unpublished Ph.D. thesis, Department of Experimental Statistics, North Carolina State University, North Carolina.

[12] 박홍래(1973), "Rotation Sampling in Time Series," *통계학 연구,* 제 2 권, 17-23.

[13] 박홍래(1989). *통계조사론,* 영지문화사.

[14] 이계오(1985). *Comparison of Some Successive Occasions Sampling Schemes from Spatially Correlated Process,* Unpublished Ph.D. thesis, Department of Statistics, North Carolina State University, North Carolina.

시계열 계속 표본조사에서 불균등확률 추출법 연구[1]

박홍래[2], 이계오[3]

요 약

본 논문에서는 반복적 계속 표본조사에서 일부의 표본을 교체하는 2회 계속조사의 표본 추출법들을 요약하고 앞 조사시기의 관찰값을 확률측도로 이용한 RHC (Rao-Hartley-Cochran) 유형의 불균등 확률추출법을 제안하였다. 제안된 추출법과 기존의 확률추출법의 비교를 위하여 둘째 조사시기의 모평균 추정량과 그의 분산을 유도하였으며, 제안된 추출법의 상대 효율은 이론적인 측면과 수치적 시뮬레이션 방법으로 비교 분석되었다. 시뮬레이션 비교를 위하여 한 특별한 시계열 모형을 가정하고 이를 사용하여 인위적인 모집단을 생성하였으며 이 모집단에서 각 추출법에 해당되는 표본을 컴퓨터로 추출하여 각각의 추정치를 계산하여 비교한 결과에서 RHC 유형의 새로 제안된 추출법의 분산과 편차가 일반적으로 적음을 보였다.

[1] 이 연구는 1991년도 교육부 기초과학육성연구비의 지원에 의한것임.
[2] (151-742) 서울시 관악구 신림2동 산 56-1
서울대학교 자연과학대학 계산통계학과
[3] (363-849) 충북 청원군 남일면 쌍수리 사서함 335-2호
공군사관학교 전산학과