The Korean Communications
in Statistics Vol. 3, No. 3, 1996
pp. 161-168

A Study of Circular Sampling in Finite Population[1)]

Hae-Yong Lee[2)]

Abstract

This paper describes a sampling method, which can be used instead of the simple random sampling without replacement(SRSWOR). This method, circular sampling, assumes that the sampling units of the population are arranged in circular format, and randomly selects as many as samples of contiguous units. Therefore this method gathers information quicker and easier than SRSWOR. In certain circumstances, the reliability of this method is better than that of SRSWOR. And if circular sampling would be applied to nonprobability sampling methods, the reliability of the sample results in terms of probability could be determined.

1. Introduction

There are many cases to survey a group of people, animal and plant in special place in order to get public opinions, to check the state of national health, and to protect certain animals and plants. When a population of a group is small and distrbuted in small area, complete survey is doable. But most of cases, the population is too large and distribution of the population is too wide to do complete survey. Complete survey is much expensive and the relability is not much than a sample survey(Cochran, 1977)

If the residence places of the sampling units to be surveyed are distinct like persons and plants, it is easy to survey the sampling units which are defined as sample. But if the residence places of sampling units are not fixed like animals, it is difficult to survey every one of them.

When the residence places of sampling units are unstable, is there any better sampling method which is less expensive without compromising the reliability of estimators of population ?. Circular sampling is designed to overcome the problem.

The circular sampling is applied under the assumption of that sampling units are arranged in circular format. So this sampling method takes samples without replacement and assumes

1) This paper was supported by Sungshin Women's University Research Fund, 1995.
2) Professor, Department of Statistics, Sungshin Women's University, Seoul 136-742, Korea.

that the first sampling unit and the last one are consirdered as contiguous unit each other. Though the fact that population is arranged in circular format has been taken into account by Godambe(1970), Royall(1970), Hartley and Rao(1971), Rao(1975), Cochran(1977), and so on, they could not conect the idea with sample selection methodology.

Hedayet, Rao and Stufkan(1988) present a sampling method which selects samples avoiding pairs of contiguous units. They assume that population is arranged in circular format within the order of each unit size, so contiguous units are anticipated to provide simliar data.

For example, denote the units by 1,2,..., N, with i and (i+1) mod N as contiguous units. All samples in the support will be fixed size n. The following samples are example obtained by the sampling method without contiguous units for N =9, $i_1 - i_2 \geq 2$, and n =3.

{1, 3, 6} {1, 4, 8} {1, 5, 7} {2, 4, 7} {2, 5, 9}

{2, 6, 8} {3, 5, 8} {3, 7, 9} {4, 6, 9}.

In this example we see that the first and second order inclusion probabilities are given by

$$\pi_i = 1/3, \quad i \subseteq \{1,2,3,...,9\}$$

and, with $i_1 \neq i_2$,

$$\pi_{i_1 i_2} = \begin{cases} 1/9 & \text{if } i_1 \text{ and } i_2 \text{ are non-contiguous,} \\ 0 & \text{otherwise,} \end{cases}$$

respectively. This method works better for ordered population.

However the most elements of the population are randomly arranged in natural states rather than in order of unit value. In comparison to the above method, the circular sampling works better for natural states of arrangement. Because this sampling selects the sampling units under the assumption that the elements of the population are arranged in random. Also this technique is easer to select sampling units and more efficient than SRSWOR in any circumstances.

If we use the sampling technique presented by Hedayat et al.(1988) to estimate population mean, we should know the order of each wildlife animal weight. If we don't know the order of each weight of N wildlife animals, we have to do the complete survey to know the order of them by amount.

But the circular sampling technique assumes that the numbering of each sampling unit was done randomly without considering the amount of weight. Therefore this method is easier and have more applicable than method presented by Hedayat et al..

Let N be a population size, n be a sample size, i be the first sampling unit selected from population size N, and X_i denote the value of the characteristic attached to the first selected sampling unit, then the elements of the sample size n are { i, i+1, ..., $i+n-1$}. The set $S(i)$ of the sample values can be shown the following equation (1-1).

$$S(i) = \begin{cases} \{X_i, X_{i+1}, \ldots, X_{i+n-1}\} & \text{for } (i+n-1) \le N \\ \{X_i, X_{i+1}, \ldots, X_N, X_1, X_2, \ldots, X_{i+n-1-N}\} & \text{for } (i+n-1) > N \end{cases} \tag{1-1}$$

And let $\overline{X_c}$ be the sample mean , then, the mean can be calculated by following equation (1-2)

$$\overline{X_c} = \begin{cases} \frac{1}{n} \sum_{j=i}^{i+n-1} X_j, & \text{for } (i+n-1) \le N \\ \frac{1}{n} \{ \sum_{j=i}^{N} X_j + \sum_{j=1}^{i+n-1-N} X_j \}, & \text{for } (i+n-1) > N \end{cases} \tag{1-2}$$

2. Sample Selection Procedure

When a given population with size $N\{X_1, X_2, \ldots, X_N\}$ is arranged in circular format, the method to select samples with size $n(\ n \le N\ )$, as following procedures, is named circular sampling.

The first, randomly select an element from the population $\{X_1, X_2, \ldots, X_N\}$ which is assumed to be in circular format.

The second, if the number of the first selection is i, sequentialy select n sampling units from starting point i with satisfying the equation (1-1).

The third, the total number of sample support obtained from one arrangement of population size N is N as followings;

$\{X_1, X_2, \ldots, X_n\}$, $\{X_2, X_3, \ldots, X_{n+1}\}$,...,$\{X_i, X_{i+1}, \ldots, X_{i+n-1}\}$, ... $\{X_N, X_1, X_2, \ldots, X_{n-1}\}$.

As an example let us suppose that we wish to estimate the population mean with size 3 sample selected from the population by using circular sampling, when a hypothetical population is given like {1,2,4,6,7} and the arrangement is same order. For solving this problem, if the first selected value is "2", the desired sample is {2,4,6}. The total number of sample support $S(i)$, i=1,2,..., N from starting point "2" is as follows;

$S(1)$ = {2,4,6}, $S(2)$= {4,6,7}, $S(3)$ = {6,7,1}, $S(4)$ = {7,1,2}, $S(5)$ = {1,2,4}.

But the arrangement of circular form with the population composed by size N is same as circular permutation of N, and the number of circular arrangement of N is $(N-1)!$.

As stated above even if the population size is fixed with N, there are many types of population according to how to arrange the population with size N. Therefore this circular sampling is the sampling method based on superpopulation rather than a single population. This is essential difference between the SRSWOR and the circular sampling.

This circular sampling has the following distinctive features.

(1) This sampling is a sampling method to be taken into account the total circular permutation $(N-1)!$ of population size N.

(2) The number of supports obtained from a arrangement of population size N are N without regard to sample size n.

(3) This sampling is balanced sampling, because the total appearence number of each unit in the total supports N obtained from one arrangement is same as n.

(4) The total number of supports obtained from population size N by this sampling is $N!$, because the circular permutation of N is $(N-1)!$ and the number of support for one arrangement is N.

(5) The appearence number of each unit in the possible supports number $N!$ is same as $n(N-1)!$, so the circular sampling considered all possible arrangements is also balanced sampling same as one arrangement case.

3. Properties of Circular Sampling

As mentioned the preceding chapter the circular sampling is balanced sampling. Thus the sample mean $\overline{X_c}$ satisfies the following theorem (3-1).

<Theorem 3-1> : The sample mean $\overline{X_c}$ obtained from the circular sampling is an unbiased estimator of population mean μ.

Proof.
$$E(\overline{X_c}) = \frac{1}{N!}[\overline{X_1}+\overline{X_2}+,\ldots,\overline{X_{N!}}]$$
$$= \frac{1}{N!}[\frac{1}{n}n(N-1)!(X_1+X_2+,\ldots,+X_N)]$$
$$= \mu$$
▨

The sample mean variance $Var(\overline{X_c})$ of the circular sampling $\overline{X_c}$ satisfies the following theorem (3-2), when a sample with size n is selected from a population with size N by circular sampling.

<Therom 3-2> : The sample mean variance $Var(\overline{X_c})$ obtained from circular sampling is presented as the following equation (3-1).

$$Var(\overline{X_c}) = \frac{N-n}{N-1}\frac{\sigma^2}{n} \qquad (3\text{-}1)$$

Proof.

$$Var(\overline{X}_c) = \frac{1}{N!}\sum_{i=1}^{N!}(\overline{X}_i-\mu)^2 = \frac{1}{N!}\sum_{i=1}^{N!}\left(\frac{1}{n}\sum_{j=1}^{n}X_{ij}-\mu\right)^2$$

$$= \frac{1}{N!}\frac{1}{n^2}\left\{\sum_{i=1}^{N!}\sum_{j=1}^{n}(X_{ij}-\mu)^2+\sum_{i=1}^{N!}\sum_{j\neq j'}^{n}(X_{ij}-\mu)(X_{ij'}-\mu)\right\}$$

$$= \frac{n}{n^2}\frac{(N-1)!}{N!}\sum_{k=1}^{N}(X_k-\mu)^2 + \frac{n(n-1)}{n^2}\frac{(N-2)!}{N!}\sum_{k\neq k'}^{N}(X_k-\mu)(X_{k'}-\mu)$$

$$= \frac{\sigma^2}{n} + \frac{n-1}{n}\frac{1}{N(N-1)}\sum_{k\neq k'}^{N}(X_k-\mu)(X_{k'}-\mu)$$

$$= \frac{\sigma^2}{n} - \frac{n-1}{n}\frac{1}{(N-1)N}\sum_{k=1}^{N}(X_k-\mu)^2$$

$$= \frac{N-n}{N-1}\frac{\sigma^2}{n}$$

▨

And the estimaror $\widehat{Var}(\overline{X}_c)$ of the sample mean variance $Var(\overline{X}_c)$ is as following equation (3-2).

$$\widehat{Var}(\overline{X}_c) = \frac{N-n}{N}\frac{s_i^2}{n} \tag{3-2}$$

where , $s_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_{ij}-\overline{X}_i)^2$, $i=1,2,\ldots,N!$.

It can be shown that $\widehat{Var}(\overline{X}_c)$ is the unbiased estimator of $Var(\overline{X}_c)$. As the above results it is known that the mean variance $Var(\overline{X}_c)$ of circular sampling is equal to the mean variance $Var(\overline{X}_r) = \frac{N-n}{N-1}\frac{\sigma^2}{n}$ of SRSWOR (Cochran 1977, Hansen et. al.1953).

But the sample mean variance $Var(\overline{X}_i)$ obtained from i-th arrangement of population is not same as $Var(\overline{X}_r)$, because this sample is not balanced sample. For an example let the i-th arrangement be one population, the variance $Var(\overline{X}_i)$ of sample mean $\overline{X}_i$ obtained from this arrangement is the following equation (3-3).

$$Var(\overline{X}_i) = E(\overline{X}_i-\mu)^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{n}\sum_{j=1}^{n}X_{ij}-\mu\right)^2 \tag{3-3}$$

Here let the first order circular serial correlation coefficient between the contiguous units of i-th arrangement be ρ as equation (3-4),

$$\rho = \frac{1}{n(n-1)N}\frac{1}{\sigma^2}\sum_{i=1}^{N}\sum_{j\neq j'}^{n}(X_{ij}-\mu)(X_{ij'}-\mu) \tag{3-4}$$

the equation (3-3) becomes equation (3-5a) or (3-5b).

$$Var(\overline{X_i}) = \frac{\sigma^2}{n} + \frac{1}{Nn^2}\sum_{i=1}^{N}\sum_{j\neq j'}^{n}(X_{ij}-\mu)(X_{ij'}-\mu) \qquad (3\text{-}5a)$$

$$= \frac{\sigma^2}{n}[1+(n-1)\rho] \qquad (3\text{-}5b)$$

As shown previous paragraph $Var(\overline{X_c})$ is same as $Var(\overline{X_r})$, but $Var(\overline{X_i})$ is not same as $Var(\overline{X_r})$. We can use the variance $Var(\overline{X_i})$ as a criterion of what arrangement circular sampling is more efficient than SRSWOR. In order to show that, we can take equation (3-6) from $Var(\overline{X_r})$ and $Var(\overline{X_i})$,

$$\begin{aligned} Var(\overline{X_r}) &\geq Var(\overline{X_i}) \quad \text{if} \quad \rho \leq -1/(N-1) \\ Var(\overline{X_r}) &< Var(\overline{X_i}) \quad \text{if} \quad \rho > -1/(N-1) \end{aligned} \qquad (3\text{-}6)$$

Hence, when $\rho = -1/(N-1)$, circular sampling and SRSWOR give equal precision. When $\rho \leq -1/(N-1)$, circular sampling is more efficient than SRSWOR, and when $\rho > -1/(N-1)$, SRSWOR is more efficient than circular sampling. This result will give us a criterion when we should use circular sampling instead of SRSWOR.

4. Example of Sample Selection

Suppose sample size n=3 are selected from the hypothetical population S = {1,2,4,6,7} given in above example with circular sampling for the purpose of estimating the population mean μ. From this population we can get the population mean μ = 4, variance σ^2 = 5.2.

Considering the circular sampling for this example, the number of circular permutation of population size 5 is (5-1)! = 24. Let the reults be Po(i) separately, then they are followings :

Po(1) = {1,2,4,6,7}, Po(2) = {1,2,4,7,6}, Po(3) = {1,2,6,7,3},

Po(4) = {1,2,6,4,7}, Po(5) = {1,2,7,4,6}, Po(6) = {1,2,7,6,4},

Po(7) = {1,4,2,6,7}, Po(8) = {1,4,2,7,6}, Po(9) = {1,4,6,2,7},

Po(10) = {1,4,6,7,2}, Po(11) = {1,4,7,2,6}, Po(12) = {1,4,7,6,2},

Po(13) = {1,6,2,4,7}, Po(14) = {1,6,2,7,4}, Po(15) = {1,6,4,2,7},

Po(16) = {1,6,4,7,2}, Po(17) = {1,6,7,2,4}, Po(18) = {1,6,7,4,2},

Po(19) = {1,7,2,4,6}, Po(20) = {1,7,2,6,4}, Po(21) = {1,7,4,2,6},

Po(22) = {1,7,4,6,2}, Po(23) = {1,7,6,2,4}, Po(24) = {1,7,6,4,2}

When we select the sample size 3 with circular sampling, the total supports obtained from these 24 arrangements are 120 samples. For example if we select the sample size 3 from arrangement Po(1) = {1,2,4,6,7} with circular sampling, the samples $S(i)$ $i=1,2,\ldots,N$ are followings:

$$S(1)=\{1,2,3\},\ S(2)=\{2,4,6\},\ S(3)=\{4,6,7\},\ S(4)=\{6,7,1\},\ S(5)=\{7,1,2\}$$

5 samples can be selected from each arrangement, so the total number of support obtained from 24 arrangements are 120. The number of support obtained from circular sampling is always greater than that of SRSWOR. And the sample mean variance of SRSWOR $Var(\overline{X_r})$ $= \frac{(N-n)}{(N-1)}\frac{\sigma^2}{n}$ is equal to 13/15. We can find that the sample mean of $\overline{X_c}$ is the unbiased estimator of population mean μ and the estimator of sample mean variance $\widehat{Var}(\overline{X_c})$ is the unbiased estimator of the sample variance $Var(\overline{X_c})$, too. This results satisfy the theorem (3-1) and (3-2).

5. Conclusion and Remarks

In conculsion the expectation of sample mean and sample variance of the circular sampling are same as those of SRSWOR. But when we consider i-th arrangement, the expectation of sample mean of the circular sampling is same as that of SRSWOR, however the sample variance obtained from i-th arrangement is not same as that of SRSWOR because the sample is not balanced sample.

If the first order circular correlation coefficient of i-th arrangement ρ is less than or equal to $-1/(N-1)$, circular sampling is more efficient than SRSWOR, but if ρ is greater than $-1/(N-1)$, SRSWOR is more efficient than circular sampling. But it is impossible to know the value of ρ in advance to select samples. Therefore, one must be sure of that the sampling units of a population are arranged randomly to use of the circular sampling.

To explain the circular sampling theoratically, we labelled every N sampling units of a population. But, in reality, the labelling of the sample units are not necessary. The sample must be taken once without replacement. That is why the circular sampling is more economical and quicker than the SRSWOR in selecting samples. Also, as mensioned it above, the efficiency of the circular sampling is not less than that of SRSWOR, so circular sampling can be used intead of SRSWOR.

The drawback of the circular sampling is that it has the number of supports lot more than SRSWOR. Also, when the contiguous units of population are similar, the efficiency of the circular sampling is worse than SRSWOR. And especially we can appliy circular sampling to the nonprobability sampling methods to determine the reliability of the sample results in terms of probability like probability sampling process. It is possible to apply directly circular sampling to stratified sampling and cluster sampling in order to improve the efficiency.

References

[1] Cochran, W. G. (1977). *Sampling Techniques,* John Wiley and Sons, Inc. New York, 21-30.

[2] Godambe V. P. (1970). Foundations of Survey Sampling, *The American Statistician*, Vol. 24, 33-38.

[3] Hansen, M.H., Hurwitz W. N., and Madow W. G. (1953). Sample Survey, Methods and Theory. Vol. I, *Methods and Applications*, Vol. II, Theory. John Wiley & Sons, Inc. New York.

[4] Hartley H. O. and Rao J. N. K. (1971). Foundations of Survey Sampling(A Don Quixote), *The American Statistician*, Vol. 25, 21-27.

[5] Hedayat, A. S., Rao, C. R. and Stufkan, J. (1988). Sampling Plans Excluding Contiguous Units, *J. Statist. Plann. Infer.*, 19, 159-170.

[6] Hedayat, A. S., Rao, C. R. and Stufkan, J. (1988). Designs in Survey Sampling Avoiding Contiguous Units, *Handbook of Statistics*, Vol. 6, 575-583.

[7] Rao, C. R. (1975). Some Problems of Samples. *Supplement Advanced Applied Probability,* Vol. 7, 50-61.

[8] Royall, R. M.(1970). Finite Population Sampling-On Labels in Estimation, *The Annals of Mathematical Statistics,* Vol. 41, NO. 5, 1774-1779.