2003 Proceedings of the Autumn Conference, Korean Statistical Society

Probability Sampling Using Nonlinear Programming : a Feasibility Study

Sun-Woong Kim[1)]

Abstract

We show how some probability nonreplacement sampling designs can be implemented using nonlinear programming. The efficiency of the proposed approach is compared with selected probability sampling schemes in the literature. The approach is simple to use and appears to have reasonable variance.

Key Words: Joint Probability, Nonlinear Programming, Efficiency

1. Introduction

A number of schemes for selecting probability nonreplacement (PNR) samples have been proposed. But it is not clear yet which method perform consistently well with respect to statistical efficiency. It may be due, in part, to the fact that in the case of small populations, the variances of the estimates of interest are very sensitive to population characteristics as well as the particular method of selection.

Jessen (1969) suggested four sampling schemes and examined their properties. The four differ primarily in the degree of control that is exercised over the joint probability that two sampling units are both in the sample. One of them, called method 4, would be an interesting scheme because it shows comparatively high statistical efficiency, as presented in his paper. Since the method requires an iterative solution to achieve the desired control over the joint probability, it may be reluctant to use it in practical problems.

In this paper, we show how the Jessen's scheme is improved using nonlinear programming (NLP) and suggest a different method on the basis of the approach. We compare those methods by NLP with some selected methods in the literature. The NLP approach is simple to employ and appears to have high efficiency.

2. Applying Nonlinear Programming to Sampling Designs

2.1 Some Methods of Probability Sampling

Hansen and Hurwitz (1943) first introduced the general idea of sampling where sampling units are selected with unequal probabilities. Narain (1951) and Horvitz and Thompson (1952) suggested some theory and methods of sampling finite populations with unequal probabilities and without replacement. These papers give the same unbiased estimator of the

1)Full-time lecturer, Department of Statistics, Dongguk University, e-mail:sunwk@dongguk.edu

population total, Y , and its variance, that is,

$$\hat{Y}=\sum_{i=1}^{n}\frac{y_i}{\pi_i}\ , \tag{2.1}$$

$$Var(\hat{Y})=\sum_{i=1}^{N}\frac{y_i^2}{\pi_i}+2\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{\pi_{ij}}{\pi_i\pi_j}y_iy_j-Y^2\ , \tag{2.2}$$

where y_i is the value of the i th sampling unit where $i=1,2,\cdots,N$, π_i is the probability that the i th sampling unit is selected in a sample of n and π_{ij} is the joint probability with which two sampling units i and j are both included in a sample of n .

Jessen (1969) presented four selection methods. In methods 1 and 2 the sampler has the considerable control over the combination of sampling units that consist of each possible sample. Method 3, which is completely objective, generates the samples by a weighted system of randomizations. The three methods have the same characteristic, namely, that $\pi_i=nA_i$ where A_i is the relative size of each sampling unit i . This is a sufficient condition for minimizing variance of Y .

He also examined the influence of π_{ij} on $Var(\hat{Y})$ expressed in another form

$$Var(\hat{Y})=\sum_{i=1}^{N}\sum_{j>i}^{N}(\pi_i\pi_j-\pi_{ij})\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2\ , \tag{2.3}$$

which was derived by Yates and Grundy (1953).

Jessen begins with the following idea to explore the effect of π_{ij} . Let $W_{ij}=\pi_i\pi_j-\pi_{ij}$ and obtain

$$W=\sum_{i=1}^{N}\sum_{j>i}^{N}W_{ij}=\frac{n-\sum_{i=1}^{N}\pi_i^2}{2}\ . \tag{2.4}$$

Let consider the situation where the weight W_{ij} is a constant $\overline{W}$, denoted by

$$\overline{W}=\frac{W}{N(N-1)/2}=\frac{n-\sum_{i=1}^{N}\pi_i^2}{N(N-1)}\ . \tag{2.5}$$

Then (2.3) can be written

$$Var(\hat{Y})=\overline{W}\sum_{i=1}^{N}\sum_{j>i}^{N}\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2\ . \tag{2.6}$$

But it is impossible for all $W_{ij}=\overline{W}$ because the A_i s are unequal. Therefore he proposed a method for selecting samples of $n=2$ with $W_{ij} \doteq \overline{W}$ and called this as method 4. This method first needs the computations of the joint probability π_{ij}^* , the approximation to the desired π_{ij} . However, since π_{ij}^* s do not meet the relation that $\pi_i = nA_i$, he finds a set of adjusted joint probabilities by using the marginal constraints for the tableau. The problem is that although this approach is quite flexible, there exist many different solutions that we may choose. Thus we should use the average variance of the results of several trials if we compare statistical efficiency of this approach with that of other methods.

2.2 Nonlinear Programming Approaches

Method 4, suggested by Jessen (1969), would be useful because it could be more efficient than others, although his method has some disadvantages mentioned above. Note that he referred to the fact that it is not the primary purpose of his paper to examine the general problem such as statistical efficiency but rather to suggest several methods.

Those problems of method 4 can be apparently improved using NLP. We first show how his method is implemented by NLP and an optimal set of π_{ij} is provided.

Consider the following simple approach. Suppose that there are many sets of joint probabilities satisfying several constraints. Our concern is to determine an appropriate set of joint probabilities so that variance of $\widehat{Y}$ is minimized. In fact, his method tries to find a set of π_{ij} such that (2.3) is minimized, as described above. In this case we can apply method 4 to a NLP problem subject to certain constraints. Constructing for $W_{ij} = \pi_i \pi_j - \pi_{ij}$ and $\overline{W}$, we get the following NLP problem:

$$\text{Minimize} \quad \sum_{i=1}^{N} \sum_{j>i}^{N} \{ (\pi_i \pi_j - \pi_{ij}) - \overline{W} \}^2 \tag{2.7}$$

$$\text{subject to} \quad \sum_{j \neq i}^{N} \pi_{ij} = \pi_i \tag{2.8}$$

$$\pi_{ij} \geq 0 \tag{2.9}$$

Note that $\overline{W}$ is a constant, and hence (2.7) is equivalent to $\sum_{i=1}^{N} \sum_{j>i}^{N} (\pi_i \pi_j - \pi_{ij})^2$. The NLP approach can avoid the iterative trials and yield an unique solution of π_{ij} , which is optimized by some nonlinear programming algorithms. We simply name this approach as *NLPA1*.

On the other hand, the *NLPA1* would be helpful to consider a different approach. Recall that (2.2) is available instead of (2.3). More specifically, the second term of (2.2) is the similar form to (2.3). That is,

$$\sum_{i=1}^{N} \sum_{j>i}^{N} \frac{\pi_{ij}}{\pi_i \pi_j} y_i y_j \tag{2.10}$$

Since (2.10) is a function of $\pi_{ij}/(\pi_i\pi_j)$, the problem of deciding π_{ij} can be stated as follows:

$$\text{Minimize} \quad \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{\pi_{ij}}{\pi_i\pi_j} \tag{2.11}$$

$$\text{subject to} \quad \sum_{j\neq i}^{N}\pi_{ij}=\pi_i \tag{2.12}$$

$$\pi_{ij}\geq 0 \tag{2.13}$$

We label the second approach as *NLPA2*, which is compared to *NLPA1*.

For our approaches, it would be interesting to examine the influence of π_{ij} on $Var(\hat{Y})$ through application to an example from the literature.

3. An Example

Jessen (1969) compared methods 3 and 4 of his paper with some selected schemes. Table 3.1 presents the data for the $N=5$ case he used. This data is originally available from Cochran (1963). The comparison is limited to samples of size $n=2$.

Table 3.1 Three Populations of $N=5$

Element	i :	1	2	3	4	5
Relative Size	A_i :	0.1	0.1	0.2	0.3	0.3
Population *A*	y_i :	0.3	0.5	0.8	0.9	1.5
Population *B*	y_i :	0.3	0.3	0.8	1.5	1.5
Population *C*	y_i :	0.5	0.5	0.8	0.9	0.9

Using the data, Table 3.2 presents the exact variances of estimates of Y , resulting from *NLPA1* and *NLPA2*, compared with the two methods of Jessen and several schemes in the literature.

Table 3.2 Comparison of Variances of *NLPA1* and *NLPA2* with Other Schemes

Population	$Var(\hat{Y})$									
	PR	*N*	*M3*	*M4*	*NLPA1*	*NLPA2*	*HR*	*CO*	*RHC*	*HT*
A	0.400	0.244	0.200	0.243	0.247	0.300	0.233	0.220	0.320	0.279
B	0.320	0.252	0.340	0.257	0.247	0.140	0.273	0.300	0.256	0.434
C	0.320	0.252	0.340	0.257	0.247	0.140	0.273	0.300	0.256	0.120
Average	0.347	0.249	0.293	0.252	0.247	0.193	0.260	0.273	0.277	0.278
Rel. Eff.	100	140	115	138	141	180	134	127	136	125

Note: PR - Sampling with probability proportional to size with replacement
N - Narain (1951)

M3 - Method 3
M4 - Method 4, used the average of variances in three trials
H-R - Hartley and Rao (1962)
CO - Cochran (1963), equal sizes of groups
RHC - Rao, Hartley and Cochran (1962)
HT - Horvitz and Thompson (1952)

The relative efficiencies of Narain method, *NLPA1* and *NLPA2* are smaller, *NLPA2* being clearly better, due to populations B and C . *NLPA2*'s efficiency to population is very sensitive, whereas *NLPA1*'s is less sensitive. It is noted that *NLPA1* is slightly better than method 4.

4. Conclusion

We have showed how Jessen's method 4, PNR sampling design, is implemented using NLP. The approach (*NLPA1*) by NLP would be preferable to method 4. We also proposed a different nonlinear programming approach (*NLPA2*). Both of them are simple to use and seem to be more efficient than other schemes including method 4. But it would be helpful to have more studies about a variety of populations.

References

[1] Cochran, W. G. (1963). *Sampling Techniques*, Second edition, John Wiley and Sons.

[2] Hansen, M. H. and Hurwitz, W. N. (1943). On the theory of sampling from finite populations, *Annals of Mathematical Statistics*, Vol. 14, No. 4, 333-362.

[3] Hartley, H. O. and Rao, J. N. K. (1962). Sampling with unequal probabilities and without replacement, *Annals of Mathematical Statistics*, Vol. 32, No. 2, 350-374.

[4] Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe, *Journal of the American Statistical Association*, Vol. 47, 663-685.

[5] Jessen R. J. (1969). Some methods of probability non-replacement sampling, *The Journal of the American Statistical Association*, Vol. 64, 175-193.

[6] Narain, R. D. (1951). On sampling without replacement with varying probabilities, *Journal of the Indian Society of Agricultural Statistics*, Vol. 3, No. 2, 169-175.

[7] 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 Statistical Society*, Series B, Vol. 24, 482-491.

[8] Yates, F. and Grundy, P. M. (1953). Selection without replacement from within strata and with probability proportional to size, *Journal of the Royal Statistical Society*, Series B, Vol. 15, 253-261.