ON ESTIMATORS OF VARIANCE OF ESTIMATE OF POPULATION TOTAL IN VARYING PROBABILITIES

BY G. R. SETH
Institute of Agricultural Research Statistics, New Delhi

It is well-known that by assigning varying probabilities of selection to different units in a population, it is possible to get estimates of the population parameters with smaller sampling error compared to those on sampling with varying probabilities. Horvitz and Thompson[2] have given a general sampling theory with unequal probabilities and without replacement. An unbiased estimate of the variance of their estimator of the population total is given by

$$\hat{V}(\hat{y}_{HT}) = \sum_{i=1}^{n} \frac{1-\pi_i}{\pi_i^2} y_i^2 + 2 \sum_{j>i=1}^{n} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j\pi_{ij}} y_i y_j$$

where $\hat{y}_{HT} = \Sigma \frac{y_i}{\pi_i}$, $V(\hat{y}_{HT}) = \sum_{i=1}^{N} \frac{y_i^2}{\pi_i} + \sum_{i \neq i}^{N} \frac{\pi_{ij}}{\pi_i\pi_j} y_i y_j - y^2$

and π_i is the probability that the ith unit in the population of N units enters in the sample of size n and π_{ij} is the probability that the ith and jth units in the population enter the sample.

One serious disadvantage with the proposed estimator $\hat{V}(\hat{y}_{HT})$ is that it may assume negative values. Yates and Grundy have proposed an alternative estimator of $V(\hat{y}_{HT})$ which is believed to be less often negative. Their estimator is given by

$$\hat{V}_{YG}(\hat{y}_{HT}) = \sum_{j>i=1}^{n} \frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2$$

Des Raj[1] has shown that when the first unit is selected with probabilities proportional to some measure of size and without replacement and the second unit with probabilities proportional to the same measure of sizes of the remaining units for sample size

$n=2$, Yates and Grundy estimator is always positive. In case of sample size $n>2$, he has also shown that when the first unit is selected with varying probabilities and without replacement and subsequent units with equal probability and without replacement, Yates and Grundy's estimate will still be positive. J.N.K. Rao[3] has generalised the result of Des Raj in the sense that for sample size $n>2$, Yates and Grundy's estimate has been shown to be positive where the probabilities of the first draw are proportional to size and without replacement, and at the second draw probabilities of the remaining units are proportional to the same measure of size as in the first draw and probabilities at the third and subsequent draws, being equal and without replacement.

In this note it is shown that Yates and Grundy's estimate is positive for a system of probabilities which is more general than discussed by J.N.K. Rao. Let $\pi_i(n)$, $\pi_j(n)$ and $\pi_{ij}(n)$ denote the probabilities of including the ith unit, jth unit and the pair of ith and jth units respectively in a sample of size n out of the total number of units N. In the extreme but not a practical case, when $n=N-1$, it can be easily seen that $\pi_i(n)\ \pi_j(n) > \pi_{ij}(n)$ is positive for any system of probabilities of selection at the various draws for any pair of units i and j $(i, =1, 2, \ldots\ldots N)$. Probability of exclusion of both ith and jth units in the sample is given by

$$1-\pi_i(n)-\pi_j(n)+\pi_{ij}(n)$$

and it is zero when $n=N-1$.

Thus in this case $\pi_{ij}(n)=\pi_i(n)+\pi_j(n)-1$

and

$$\pi_i(n)\ \pi_j(n)-\pi_i i(n)=[1-\pi_i(n)]\ [1-\pi_j(n)]$$

which is always positive for any pair (i, j). Thus Yates and Grundy's estimator of variance of YHT is always positive.

Next let us consider the system of varying probabilities and without replacement in which $\pi_i(2)\ \pi_j(2) \geqslant \pi_{ij}(2)$ and the probabilities at the third and subsequent draws are equal. This spstem is more general than the one discussed by J. N. K. Rao as the probabilities at the second draw need not be proportional to the same measure of size as at the first draw for the remaining units. Such a system is possible. For example, for $N=4$, $n=2$ and probabilities proportional to 1, 2, 3, 4 at first draw and probabilities proportional to 1, 2, 4, 5 for the remaining units at second draw gives $\pi_i(2)\,\pi_j(2) > \pi_{ij}(2)$ for all pairs i and j.

For such a probability system for making successive draws, we easily see that

$$\pi_i(n) = \pi_i(2)\frac{N-n}{N-2} + \frac{n-2}{N-2} \quad (i=1, 2, \ldots N)$$

$$\pi_{ij}(n) = \pi_{ij}(2) + [\pi_i(2)+\pi_j(2)-2\pi_{ij}(2)]\frac{n-2}{N-2}$$
$$+ [1-\pi_i(2)-\pi_j(2)+\pi_{ij}(2)]\frac{(n-2)\,(n-3)}{(N-2)(N-3)} \quad (i, j=1, 2, \ldots, N)$$

From the above

$$\pi_i(n)\,\pi_j(n)-\pi_{ij}(n)$$
$$= [\pi_i(2)\,\pi_j(2)-\pi_{ij}(2)]\left(\frac{N-n}{N-2}\right)^2 + \frac{(n-2)\,(N-n)}{(N-2)^2(N-3)}$$
$$[1-\pi_i(2)-\pi_j(2)+\pi_{ij}(2)]$$

Now

$$1-\pi_i(2)-\pi_j(2)+\pi_{ij}(2),$$

being the probability that first two draws neither include ith nor the jth unit, is non-negative. With $\pi_i(2)\,\pi_j(2)-\pi_{ij}(2)$ assumed to be non-negative for all pairs of i and j, it is clear that $\pi_i(n)\,\pi_j(n)-\pi_{ij}(n)$ is non-negative. This shows that Yates and Grundy's estimate is positive for such a system of varying probabilities of selection of units. This provides a simple proof of the result obtained by J. N. K. Rao[3] for probability system of drawing units discussed by him. Even when $\pi_{ij}(2)-\pi_i(2)\pi_j(2) > 0$ for some pairs i, j, the expression for $\pi_i(n)\,\pi_j(n)-\pi_{ij}(n)$ can be non-negative provided

$$[\pi_{ij}(2)-\pi_i(2)\,\pi_j(2)]\left[\frac{n-2}{(N-n)(N-3)}-1\right]$$
$$+\frac{(n-2)[1-\pi_i(2)][1-\pi_j(2)]}{(N-n)(N-3)}$$

is non-negative

or

$$\rho_{ij}\left[\frac{(N-n)(N-3)}{n-2}-1\right] \leq \sqrt{\left(\frac{1}{\pi_i(2)}-1\right)\left(\frac{1}{\pi_j(2)}-1\right)}$$

where ρ_{ij} is the correlation between the event of inclusion of the ith unit and the event of inclusion of jth unit in a sample of size 2. For the probability system in which $\pi_i(2)$ is constant,

$$\pi_i(2) = \frac{2}{N} ,$$

we get

$$\rho_{ij}\left[\frac{(N-n)(N-3)}{n-2}-1\right] \leqslant \left(\frac{N}{2}-1\right)$$

If

$$\frac{N}{2} \geqslant \frac{(N-n)(N-3)}{n-2} ,$$

Yates and Grundy's estimate will be non-negative. That is, when

$$Nn-2N \geqslant 2N^2-2nN-6N+6n$$

or $\quad 2N^2-3nN-4N+6n \leqslant 0$

or $\quad 2N(N-2)-3n(N-2) \leqslant 0$

or $\quad (2N-3n)(N-2) \leqslant 0$

i.e. $\quad n \geqslant \frac{2}{3} N$, Yates and Grundy's estimate is always positive.

For $n \leqslant \frac{2}{3}N$, the expression for $\pi_i(n)\,\pi_j(n)-\pi_{ij}(n)$ is positive

when

$$\rho_{ij} \leqslant \frac{\frac{N}{2}-1}{\frac{(N-n)(N-3)}{n-2}-1}$$

$$\leqslant \frac{(N-2)(n-2)}{2[(N-n)(N-3)-n+2]}$$

$$\leqslant \frac{(N-2)(n-2)}{2[(N-n-1)(N-2)]}$$

$$\leqslant \frac{n-2}{2(N-n-1)}$$

For $n=2$, ρ_{ij} has to be negative, the well known result. For $n>2$, Yates and Grundy's is non-negative provided

$$\rho_{ij} \leqslant \frac{1}{2(N-4)}$$

For example. Let us consider the case $n=4$ and $N=6$

$$\pi_{12}(2)=\pi_{34}(2)=\pi_{56}(2)=\frac{5}{27}\ ,\ \pi_{ij}=\frac{1}{27} \text{ for other } (i, j)$$

$$\pi_i\,(2)=\frac{1}{3} \qquad (i=1, 2, 3,..., 6)$$

$$\pi_{12}-\pi_1\pi_2=\pi_{34}-\pi_3\pi_4=\pi_{56}-\pi_5\pi_6=\frac{2}{27}$$

$$\pi_{ij}-\pi_i\pi_j=-\frac{2}{27}$$

for other pairs of i, j with these values

$$\rho_{ij}=\frac{1}{3}$$

for $i, j=(1, 2)\ (3, 4), (5, 6)$

$$=-\frac{1}{3} \text{ for other pairs } (i, j).$$

The expression for $\pi_i(n)\ \pi_j(n)-\pi_{ij}(n)$ takes the form

$$\pi_i(n)\pi_j(n)-\pi_{ij}(n)=\frac{2}{81} \text{ for } (i, j)=(1, 2), (3,4), (5, 6)$$

$$=\frac{4}{81} \text{ for other pairs } (i, j)$$

For $\pi_i(2)$ unequal, the expression $\pi_i(n)\pi_j(n)-\pi_{ij}(n)$ is positive, provided

$$\rho_{ij}\leq\left(\frac{1}{\lambda}-1\right)\frac{n-2}{(N-2)(N-n-1)}$$

where λ is the maximum value of

$\pi_i(2)$ for $i=1, 2,..., N$.

Summary

In this note, Yates and Grundy's estimate of the variance of Horvitz-Thompson estimate of the total is shown to be non-negative for a system of probabilities than hitherto discussed in the literature.

References

1. Des Raj — Some estimators in sampling with varying probabilities, without replacement—J. Amer. Statis. Ass. Vol. 51, 1956, pp. 269-284.

2. Horvitz, D. G. and Thompson, D. J. — A generalisation of sampling without replacement from finite universe, J. Amer. Math. Ass. Vol. 47 (1952), pp. 350-374.

3. Rao, J. N. K. — On the estimate of variance in unequal probability sampling, Annals of the Inst. of Stat. Maths., Vol. XII (1961) pp. 57-60.