Survey Methodology, December 1996
Vol. 22, No. 2, pp. 167-173
Statistics Canada

Asymptotic Variance for Sequential Sampling Without Replacement With Unequal Probabilities

YVES G. BERGER[1]

ABSTRACT

We propose a second-order inclusion probability approximation for the Chao plan (1982) to obtain an approximate variance estimator for the Horvitz and Thompson estimator. We will then compare this variance with other approximations provided for the randomized systematic sampling plan (Hartley and Rao 1962), the rejective sampling plan (Hájek 1964) and the Rao-Sampford sampling plan (Rao 1965 and Sampford 1967). Our conclusion will be that these approximations are equivalent if the first-order inclusion probabilities are small and if the size of the sample is large.

KEY WORDS: Sampling with replacement; Randomized systematic sampling plan; Rejective sampling plan; Rao-Sampford sampling plan; Inclusion probabilities; Horvitz-Thompson; Yates-Grundy.

1. INTRODUCTION

Consider a finite population U_N containing N units and a subset U_k of U_N comprising the first units k of U_N. Let $\pi_{(k;i)}$ denote the first-order inclusion probabilities for a population U_k. We assume that they are proportional to an auxiliary variable. These probabilities have two arguments: the size k of the population and the serial number i of the unit within the population. We assume that $\pi_{(k;i)} < 1$ for all i and that all $k > n$. This hypothesis has more chance of breaking down when k is small, *i.e.*, close to n. We can solve this problem by assuming that the values of the auxiliary variable show little dispersion for those units occurring at the beginning of the population.

Let $\pi_{(k;i,j)}$ denote the second-order inclusion probability of units i and j for a population U_k. These probabilities are dependent on the sampling plan used.

We will use the Horvitz-Thompson estimator (1951) to estimate the total $\sum_{i=1}^{N} Y_i$ of a variable Y. This estimator is given by

$$t_{HT} = \sum_{i \in S_N} \frac{Y_i}{\pi_{(N;i)}}; \tag{1}$$

where S_N is a sample of U_N. We assume that the size of S_N is constant and equal to n.

Given that the size of the sample is fixed, a variance estimator of (1) is given by the Yates-Grundy estimator (1953),

$$\hat{V} = \sum_{j \in S_N} \sum_{i \in S_N; i<j} \frac{-\Delta_{(N;i,j)}}{\pi_{(N;i,j)}} \left[\frac{Y_i}{\pi_{(N;i)}} - \frac{Y_j}{\pi_{(N;j)}} \right]^2, \tag{2}$$

where

$$\Delta_{(N;i,j)} = \pi_{(N;i,j)} - \pi_{(N;i)} \pi_{(N;i)}. \tag{3}$$

Let us consider the sample size sequence $\{n_1, n_2, ..., n_\nu, ...\}$ and the population size sequence $\{N_1, N_2, ..., N_\nu, ...\}$, where n_ν and N_ν increase whenever $\nu \to \infty$. To simplify the problem we eliminate the index ν.

The asymptotic approach used here is that of Hájek (1964):

$$d = \sum_{j=1}^{N} \pi_{(N;j)} [1 - \pi_{(N;j)}] \to \infty,$$

which means that $n \to \infty$ and $(N - n) \to \infty$, given that $d \le \sum_{j=1}^{N} [1 - \pi(N;j)] = N - n$ and that $d \le \sum_{j=1}^{N} \pi(N;j) = n$.

In section 2, we introduce the Chao sampling plan (1982) as well as three results linked to first and second-order inclusion probabilities. In section 3, we provide an approximation of $\pi_{(N;i,j)}$. In section 4, we propose an approximation of the Yates-Grundy variance. Section 5 compares this variance approximation with other approximations proposed for the randomized systematic plan, the rejective plan and the Rao-Sampford plan. Two numerical examples are provided in section 6.

2. CHAO SAMPLING PLAN

This is a sampling plan without replacement with unequal probabilities, of fixed size. This method is a generalization of the method used by McLeod and Bellhouse (1983) for a simple plan.

Let S_K denote a sample of size n of U_k with a set $\{\pi_{(k;i)}: i \in U_k\}$ of first-order inclusion probabilities. The Chao plan provides for a sample S_{k+1} of size n of U_{k+1} with a set $\{\pi_{(k+1;i)}: i \in U_{k+1}\}$ of first-order inclusion probabilities. The method entails selecting the $(k + 1)$-th unit with the probability $\pi_{(k+1;k+1)}$. If this unit is not selected, then we take $S_{k+1} = S_k$; otherwise we take $S_{k+1} = S_k \cup \{k + 1\} \backslash \{j\}$, where j is a unit selected at random within S_k. The procedure starts from an initial sample $S_n = U_n$ comprising the first units n of the population.

[1] Yves Berger, Université Libre de Bruxelles, Laboratoire de Méthodologie du Traitement des Données, C.P. 124, Avenue Jeanne, 44, Bruxelles, Belgique, E-Mail : yvberger@ulb.ac.be

The Chao plan provides the advantage of being sequential. In fact, it allows us to select a sample through a simple sequential run of the population. The systematic plan is another sequential plan that is often used. However, the latter is inconvenient in that it induces zero second-order inclusion probabilities. We can avoid this problem by randomizing the systematic plan. In such a case, the population is ordered at random before the sample is selected. This operation eliminates in part the problem of zero second-order inclusion probabilities. As will be seen at the end of this section, the Chao plan offers the advantage of not having any zero second-order inclusion probabilities. Randomization is therefore not needed for the latter.

The rejective plan and the Rao-Sampford plan are inconvenient in that they are not sequential. In fact, the units are selected at random with replacement within the population. If a unit is selected twice, we are forced to select a new sample. These two plans, although they are more easily understood, are more difficult to implement than the Chao plan.

The following theorem, which is a direct application of the theorem given by Chao (1982), provides a relation between the first-order inclusion probability $\pi_{(k;i)}$ of the i-th unit of U_k and the first-order inclusion probability $\pi_{(k+1;i)}$ of the i-th unit of U_{k+1}.

Theorem 1

$$\pi_{(k+1;i)} = \begin{cases} [1 - \pi_{(k+1;k+1)} R_{(k;i)}]\, \pi_{(k;i)}, & \text{for } i < k+1; \\ \pi_{(k+1;k+1)} & \text{for } i = k+1; \end{cases} \tag{4}$$

where

$$R_{(k;i)} = \begin{cases} \dfrac{1 - \pi_{(n+1;i)}}{\pi_{(n+1;n+1)}}, & \text{for } k = n, \\ \dfrac{1}{n}, & \text{for } k \geq n+1. \end{cases} \tag{5}$$

The second-order inclusion probabilities can be calculated iteratively using the following theorem:

Theorem 2 (Chao, 1982)

$$\pi_{(k;i,j)} =$$

$$\begin{cases} \{1 - \pi_{(k;k)}[R_{(k-1;i)} + R_{(k-1;j)}]\}\, \pi_{(k-1;i,j)}, & \text{for } i<j<k; \\ \pi_{(k;k)}\,[1 - R_{(k-1;i)}]\, \pi_{(k-1;i)} & \text{for } i<j=k. \end{cases}$$

Bethlehem and Schuerhoff (1984) give a sufficient and necessary condition for the second-order inclusion probabilities to be strictly positive for a population U_k:

$$\#\{i : i \leq \ell \text{ and } \pi_{(\ell;i)} = 1\} \neq n-1, \text{ for } \ell \text{ such that } n < \ell \leq k.$$

Since $\pi_{(\ell;i)} < 1$ for all i and ℓ such that $i \leq \ell \leq k$, this condition is always met. Therefore, within the framework of this article, we will never have zero second-order inclusion probabilities.

Moreover, the quantity $\Delta_{(N;i,j)}$ is always negative if we use the Chao plan (Chao 1982, p. 656). Then the Yates-Grundy variance offers the advantage of always being positive.

3. APPROXIMATION OF SECOND-ORDER INCLUSION PROBABILITIES

The following theorem provides us with an asymptotic expression for second-order inclusion probabilities for the Chao plan.

Theorem 3

$$\pi_{(N;i,j)} \approx \begin{cases} \pi_{(N;i)}\,\pi_{(N;j)}\,\dfrac{n-1}{n-p_{(j)}}, & \text{if } j > n+1; \\ \pi_{(N;i)}\,\pi_{(N;j)}\,\dfrac{\pi_{(n+1;i)} + \pi_{(n+1;j)} - 1}{\pi_{(n+1;i)}\,\pi_{(n+1;j)}}, & \text{if } j \leq n+1; \end{cases} \tag{6}$$

where $p_{(j)} = \pi_{(j;j)}$ and $i<j$.

The proof of this theorem can be found in Appendix I.

Note that this approximation has a different structure depending on whether $j > n+1$ or $j \leq n+1$. To avoid this problem, we will use a plausible condition for the auxiliary variable so that these two structures will be equivalent. Let us consider the hypothesis given in the introduction, that the values of the auxiliary variable show little dispersion for the first units $n+1$ of the population. More precisely, we assume that the auxiliary variable is constant for the first units $n+1$, *i.e.*:

$$\pi_{(n+1;i)} = \frac{n}{n+1} \quad \text{for } i \leq n+1.$$

In this case,

$$\frac{\pi_{(n+1;i)} + \pi_{(n+1;j)} - 1}{\pi_{(n+1;i)}\,\pi_{(n+1;j)}} = \frac{n-1}{n - \pi_{(n+1;j)}}.$$

By using (6), we have the following approximation for second-order inclusion probabilities

$$\pi_{(N;i,j)} \approx \pi_{(N;i)}\,\pi_{(N;j)}\,\frac{n-1}{n-p_{(j)}} \quad \text{if } i<j; \tag{7}$$

where

$$p_{(j)} = \begin{cases} \pi_{(j;j)}, & \text{if } j > n+1, \\ \pi_{(n+1;j)}, & \text{if } j \leq n+1. \end{cases} \tag{8}$$

4. VARIANCE ESTIMATOR

Relation (7) leads to the following approximation for $\Delta_{(N;i,j)}$:

$$\tilde{\Delta}_{(N;i,j)} = \pi_{(N;i)}\pi_{(N;j)}\frac{p_{(j)} - 1}{n - p_{(j)}}, \quad \text{if} \quad i < j. \tag{9}$$

(2), (7) and (9) provide an asymptotic expression for the Yates-Grundy estimator.

$$\tilde{V}_C = \frac{1}{[n-1]}\sum_{j \in S_N}[1 - p_{(j)}]\sum_{i \in S_N; i<j}\left[\frac{Y_i}{\pi_{(N;i)}} - \frac{Y_j}{\pi_{(N;j)}}\right]^2. \tag{10}$$

But this expression tends to underestimate the variance. In fact, to establish relation (6), we use approximation (19) from Appendix I. This approximation always implies that:

$$\pi_{(N;i,j)} < \pi_{(N;i)}\pi_{(N;j)}\frac{n-1}{n - p_{(j)}}. \tag{11}$$

This can easily be verified if we observe that (20) is obtained from (18) using approximation (19). Inequality (11) is therefore true for $j > n+1$. For $j \le n+1$, it is sufficient to observe that (21) is also obtained from (19). Inequality (11) implies that:

$$\frac{-\Delta_{(N;i,j)}}{\pi_{(N;i,j)}} > \frac{1 - p_{(j)}}{n-1}, \tag{12}$$

given that $\Delta_{(N;i,j)} < 0$. From (2), (10) and (12), we have effectively

$$\hat{V} > \tilde{V}_C.$$

To overcome this problem of variance underestimation, we plan to make an adjustment on (9). It is well known that:

$$\sum_{i=1; i \neq j}^{N} \pi_{(N;i,j)} = (n-1)\pi_{(N;j)}. \tag{13}$$

Approximation (7) does not abide by constraint (13). The adjustment involves assuming that the $p_{(j)}$ are unknown and selecting them so as to satisfy (13) for the second-order probability approximation, *i.e.*:

$$\sum_{i=1}^{j-1}\pi_{(N;i)}\pi_{(N;j)}\frac{n-1}{n-p_{(j)}} + \sum_{i=j+1}^{N}\pi_{(N;i)}\pi_{(N;j)}\frac{n-1}{n-p_{(i)}} = (n-1)\pi_{(N;j)}.$$

This constraint can be written as follows:

$$\sum_{i=1}^{j-1}\pi_{(N;i)} + \sum_{i=j+1}^{N}\pi_{(N;i)}\frac{n - p_{(j)}}{n - p_{(i)}} = n - p_{(j)}. \tag{14}$$

Given that $\sum_{j=1}^{N}\pi_{(N;j)} = n$, constraint (14) is practically verified if

$$p_{(i)} = \pi_{(N;i)} \tag{15}$$

$$\sum_{i=j+1}^{N}\pi_{(N;i)}\frac{n - \pi_{(N;j)}}{n - \pi_{(N;i)}} \approx \sum_{i=j+1}^{N}\pi_{(N;i)}. \tag{16}$$

Relation (16) is plausible given that the difference between the left and right sides of (16) has as its lower bound

$$\frac{1}{n}\sum_{i=j+1}^{N}\pi_{(N;i)}[\pi_{(N;i)} - \pi_{(N;j)}],$$

and as its upper bound

$$\frac{1}{n-1}\sum_{i=j+1}^{N}\pi_{(N;i)}[\pi_{(N;i)} - \pi_{(N;j)}].$$

These two bounds are close to zero when the $\pi_{(N;i)}$ show little dispersion. This means that solution (15) is appropriate when the $\pi_{(N;i)}$ are small. Furthermore, the greater the value of j, the closer the two bounds are to zero. Therefore, solution (15) verifies (13) all the more as j is large. This implies that our approximation (9) is very good for the duplicate pairs (i, j) $(i < j)$ such that the unit j is located at the end of the population. In fact, we want approximation (9) to be the best for the duplicate pairs (i, j) whose presence in the sample is highly probable (*i.e.*, for the pairs (i, j) $(i < j)$ for which $\pi_{(N;j)}$ is the largest). It is therefore preferable to place the units having high first-order inclusion probabilities at the end of the population.

If we choose to have $p_{(i)} = \pi_{(N;i)}$, we have $p_{(i)}$ smaller than (8). This leads to a larger variance approximation. This solution is all the more acceptable as it corresponds to the result of the simple plan without replacement. In fact, if we replace within (7)$\pi_{(N;i)}$, $\pi_{(N;j)}$ and $p_{(j)}$ by n/N, we obtain

$$\pi_{(N;i,j)} \approx \frac{n(n-1)}{N(N-1)}, \quad \text{if} \quad i > n+1.$$

This expression corresponds, quite clearly, to the result of the simple plan without replacement.

In conclusion, we approximate $\Delta_{(N;i,j)}$ through (9) with $p_{(i)} = \pi_{(N;i)}$. We assume that the population is ordered in such a way that the units having small $\pi_{(N;i)}$ are located at the beginning of the population and that the units having large $\pi_{(N;i)}$ are located at the end of the population. We also assume that the $\pi_{(N;i)}$ do not show too much dispersion for the first units $n+1$ of the population.

5. COMPARISON WITH OTHER PLANS

Instead of comparing the second-order inclusion probabilities, we will compare the quantities $-\Delta_{(N;i,j)}/\pi_{(N;i,j)}$ which are of some use in calculating the Yates-Grundy variance. We will examine what these quantities provide for the Chao plan, the randomized systematic plan (Hartley and Rao 1962), the rejective plan (Hájek 1964) and the Rao-Sampford plan (Rao 1965, and Sampford 1967).

Theorem 4

$$\frac{-\Delta_{(N;i,j)}}{\pi_{(N;i,j)}} \approx \begin{cases} \dfrac{1-\pi_{(N;j)}}{n-1}, & \text{For the Chao plan;} \\ \dfrac{1-\pi_{(N;i)}-\pi_{(N;j)}}{n-1}, & \text{for the randomized systematic plan;} \\ \dfrac{n[1-\pi_{(N;i)}][1-\pi_{(N;j)}]}{d(n-1)}, & \text{for the rejective plan and the Rao-Sampford plan.} \end{cases}$$

The proof of this theorem can be found in Appendix II.

It is important to note that the proposed approximation for the randomized systematic plan comes from Deville's approximation (p. 21) and not from the famous Hartley-Rao approximation (1962). We were not able to use the Hartley-Rao formula because the latter is based on the asymptotic hypothesis, n fixed and $N \to \infty$, which is different from that adopted in this paper.

We observe that if the $\pi_{(N;i)}$ are small, $-\Delta_{(N;i,j)}/\pi_{(N;i,j)}$ is equivalent for the Chao plan and for the systematic plan. However, we observe that $-\Delta_{(N;i,j)}/\pi_{(N;i,j)}$ is always smaller in the systematic case than it is in the Chao case. This is certainly due to the fact that the approximation for the systematic plan underestimates $-\Delta_{(N;i,j)}/\pi_{(N;i,j)}$. This can be confirmed by replacing $\pi_{(N;i)}$ and $\pi_{(N;j)}$ by n/N. We then have

$$\frac{-\Delta_{(N;i,j)}}{\pi_{(N;i,j)}} \approx \frac{N-2n}{N(n-1)},$$

for the randomized systematic plan. This is equivalent to a simple plan, thus

$$\frac{-\Delta_{(N;i,j)}}{\pi_{(N;i,j)}} = \frac{N-n}{N(n-1)}.$$

We intend to adjust the approximation of $-\Delta_{(N;i,j)}/\pi_{(N;i,j)}$ for the systematic plan by multiplying it by

$$\frac{N-n}{N-2n} = \frac{1-f}{1-2f},$$

where $f = n/N$ is the sampling rate.

The approximation of $-\Delta_{(N;i,j)}/\pi_{(N;i,j)}$ for the Chao plan is also of the same magnitude as that of the rejective plan. In fact, if the $\pi_{(N;i)}$ are small, we have the approximation

$$\frac{n[1-\pi_{(N;i)}]}{d} \approx \frac{n[1-\pi_{(N;i)}]}{[1-\pi_{(N;i)}]\sum_{j=1}^{N}\pi_{(N;j)}} \approx 1.$$

Therefore, the Yates-Grundy estimator is approximately the same whether we use the Chao plan, the randomized systematic plan, the rejective plan or the Rao-Sampford plan, for large n and small $\pi_{(N;i)}$.

6. NUMERICAL EXAMPLES

The two following examples correspond to two extreme cases. In the first example, the $\pi_{(N;i)}$ show little dispersion; in the second, they show much more dispersion. Let us consider a small sample of size 20. The population size is 50 so that the $\pi_{(N;i)}$ are not too small. We have willingly opted for a bad situation in order to show that even with a sample of size 20 and a small population, the asymptotic results nevertheless represent a good approximation.

Example 1

Let us consider the first-order inclusion probabilities represented in Figure 1.

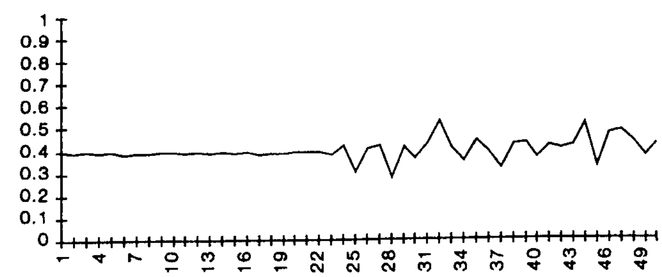

Figure 1. First-order inclusion probabilities in the case of Example 1

Figure 2 shows, on the Y axis, the true values of $-\Delta_{(N;i,j)}/\pi_{(N;i,j)}$ for the Chao plan and, on the X axis, the approximations. We have also represented the straight line where the approximations are equal to the true values. The approximations are all the better as the points are close to the straight line.

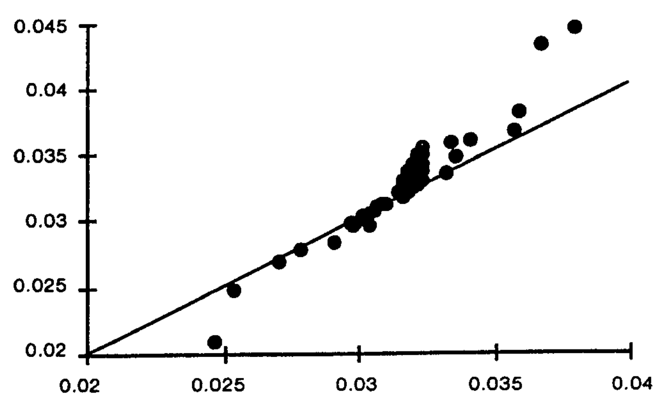

Figure 2. Approximations and true values of $-\Delta_{(N;i,j)}/\pi_{(N;i,j)}$, in the case of Example 1

We have a mean error of -0.000569 with a standard deviation of 0.0015996. This is very small in relation to the order of magnitude of the approximations. The centre of gravity of the scatter plot is located in (0.0313; 0.0318). It might seem surprising that there are less points at the left of the centre of gravity than at the right. This is simply due to the fact that most of the points at the left of the centre of gravity overlap.

We observe that the pairs (i, j) with $i < j$ such that $\pi_{(N;j)}$ is large correspond to points located on the left. They are the pairs showing the best approximation. Moreover, there is a high probability that these pairs are located within the sample given that $\pi_{(N;j)}$ is large. Therefore, our approximate variance (10) is definitely acceptable.

Example 2

The first-order inclusion probabilities are given in Figure 3. Here we notice that these probabilities are more dispersed than in Example 1. Figure 4 provides the true values as well as the approximations of $-\Delta_{(N;i,j)}/\pi_{(N;i,j)}$.

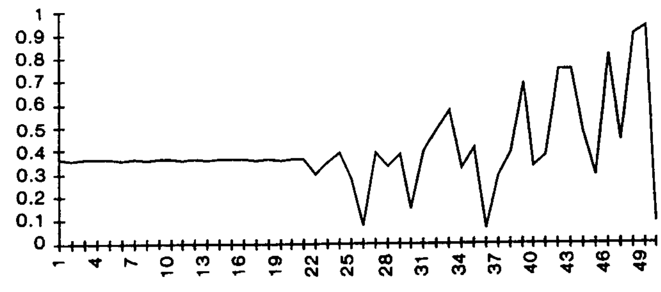

Figure 3. First-order inclusion probabilities in the case of Example 2

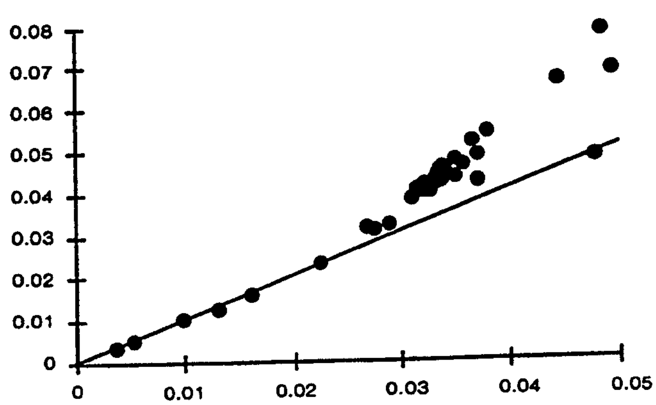

Figure 4. Approximations and true values of $-\Delta_{(N;i,j)}/\pi_{(N;i,j)}$, in the case of Example 2

We have a mean error of -0.006999 with a standard deviation of 0.006438. The centre of gravity of the scatter plot is located in (0.02957; 0.036606).

We reach the same conclusion as in Example 1. The second example leads to worse approximations. This is simply due to the high first-order inclusion probabilities.

7. CONCLUSION

The Chao plan provides a number of advantages: (i) it is sequential; (ii) the second-order inclusion probabilities are positive; and (iii) the Yates-Grundy variance is always positive. On the other hand, the second-order inclusion probabilities are difficult to calculate. That is why we propose to approximate them. We have observed that this approximation is better when the beginning of the population consists of units having small $\pi_{(N;i)}$ and the end of the population consists of units having large $\pi_{(N;i)}$. We have compared our approximation with other approximations provided for the randomized systematic plan, the rejective plan and the Rao-Sampford plan. We have concluded that these approximations are equivalent if the first-order inclusion probabilities are small and if the size of the sample is large. The two numerical examples which close this paper confirm the sound results of our approximation.

APPENDIX I

Proof of Theorem 3

Before proving this theorem, we will demonstrate the following two lemmas.

Lemma 1

$$\pi_{(k;i)} = p_{(i)}^* \prod_{\ell=a_i^*}^{k} \left[1 - \pi_{(\ell;\ell)} \frac{1}{n}\right];$$

where

$$p_{(i)}^* = \begin{cases} \pi_{(i;i)} & \text{if } i > n+1; \\ \pi_{(n+1;i)} & \text{if } i \leq n+1; \end{cases}$$

$$a_i^* = \begin{cases} i+1 & \text{if } i > n+1; \\ n+2 & \text{if } i \leq n+1. \end{cases} \tag{17}$$

Lemma 2

$$\pi_{(k;i,j)} = q_{(j)}^* \prod_{\ell=a_j^*}^{k} \left[1 - \pi_{(\ell;\ell)} \frac{2}{n}\right];$$

where $i < j$,

$$q_{(j)}^* = \begin{cases} \pi_{(j-1;i)} \pi_{(j;j)} \left(1 - \dfrac{1}{n}\right) & \text{if } j > n+1; \\ \pi_{(n+1;i)} + \pi_{(n+1;j)} - 1 & \text{if } j \leq n+1; \end{cases}$$

and a_i^* is defined by (17).

Now, with these two lemmas, we can demonstrate Theorem 3.

Proof of Theorem 3

Case 1: If $j > n+1$, using Lemma 2, we have

$$\pi_{(N;i,j)} = \pi_{(j-1;i)} \pi_{(j;j)} \left(1 - \frac{1}{n}\right) \prod_{\ell=j+1}^{N} \left[1 - \pi_{(\ell;\ell)} \frac{2}{n}\right].$$

On the basis of Lemma 1, this last expression becomes

$$\pi(N;i,j) = p^*_{(i)}\pi_{(j;j)}\left(1 - \frac{1}{n}\right)\prod_{\ell=a_i^\cdot}^{j-1}\left[1 - \pi_{(\ell;\ell)}\frac{1}{n}\right]\prod_{q=j+1}^{N}\left[1 - \pi_{(q;q)}\frac{2}{n}\right].$$

By multiplying this last expression by

$$\left[\frac{1 - \pi_{(j;j)}\frac{1}{n}}{1 - \pi_{(j;j)}\frac{1}{n}}\right]\prod_{\ell=j+1}^{N}\left[\frac{1 - \pi_{(\ell;\ell)}\frac{1}{n}}{1 - \pi_{(\ell;\ell)}\frac{1}{n}}\right]\left[\frac{1 - \pi_{(\ell;\ell)}\frac{2}{n}}{1 - \pi_{(\ell;\ell)}\frac{2}{n}}\right] = 1,$$

and by regrouping certain terms, we obtain

$$\pi_{(N;i,j)} = \pi_{(j;j)}p^*_{(i)}\left[\frac{n-1}{n - \pi_{(j;j)}}\right]\prod_{\ell=a_i^\cdot}^{N}\left[1 - \pi_{(\ell;\ell)}\frac{1}{n}\right]\prod_{q=j+1}^{N}\left[\frac{1 - \pi_{(q;q)}\frac{2}{n}}{1 - \pi_{(q;q)}\frac{1}{n}}\right].$$

On the basis of Lemma 1, this last expression becomes

$$\pi_{(N;i,j)} = \left[\frac{n-1}{n - \pi_{(j;j)}}\right]\pi_{(N;i)}\pi_{(j;j)}\prod_{\ell=j+1}^{N}\left[\frac{1 - \pi_{(\ell;\ell)}\frac{2}{n}}{1 - \pi_{(\ell;\ell)}\frac{1}{n}}\right]. \tag{18}$$

If n is sufficiently large

$$\frac{1 - \pi_{(\ell;\ell)}\frac{2}{n}}{1 - \pi_{(\ell;\ell)}\frac{1}{n}} \approx \left[1 - \pi_{(\ell;\ell)}\frac{2}{n}\right]\left[1 + \pi_{(\ell;\ell)}\frac{1}{n}\right];$$

$$\approx 1 + \frac{\pi_{(\ell;\ell)}}{n} - \frac{2\pi_{(\ell;\ell)}}{n} - \frac{2\pi^2_{(\ell;\ell)}}{n^2};$$

$$\approx 1 - \frac{\pi_{(\ell;\ell)}}{n}. \tag{19}$$

Then (18) becomes,

$$\pi_{(N;i,j)} \approx \left[\frac{n-1}{n - \pi_{(j;j)}}\right]\pi_{(N;i)}\pi_{(j;j)}\prod_{\ell=j+1}^{N}\left[1 - \pi_{(\ell;\ell)}\frac{1}{n}\right]. \tag{20}$$

Finally, on the basis of Lemma 1, this last expression can be written:

$$\pi_{(N;i,j)} \approx \pi_{(N;i)}\pi_{(N;j)}\frac{n-1}{n - \pi_{(j;j)}}.$$

Case 2: If $j \le n + 1$, Lemma 2 provides

$$\pi_{(N;i,j)} = [\pi_{(n+1;i)} + \pi_{(n+1;j)} - 1]\prod_{\ell=n+2}^{N}\left[1 - \pi_{(\ell;\ell)}\frac{2}{n}\right];$$

in other words

$$\pi_{(N;i,j)} = \prod_{\ell=n+2}^{N}\left[1 - \pi_{(\ell;\ell)}\frac{1}{n}\right]\prod_{q=n+2}^{N}\left[\frac{1 - \pi_{(q;q)}\frac{2}{n}}{1 - \pi_{(q;q)}\frac{1}{n}}\right][\pi_{(n+1;i)} + \pi_{(n+1;j)} - 1].$$

By using approximation (19), we obtain

$$\pi_{(N;i,j)} \approx \left\{\prod_{\ell=n+2}^{N}\left[1 - \pi_{(\ell;\ell)}\frac{1}{n}\right]\right\}^2 \pi_{(n+1;i)}\pi_{(n+1;j)}\frac{\pi_{(n+1;i)} + \pi_{(n+1;j)} - 1}{\pi_{(n+1;i)}\pi_{(n+1;j)}}.$$

On the basis of Lemma 1, we obtain finally

$$\pi_{(N;i,j)} \approx \pi_{(N;i)}\pi_{(N;j)}\frac{\pi_{(n+1;i)} + \pi_{(n+1;j)} - 1}{\pi_{(n+1;i)}\pi_{(n+1;j)}}. \tag{21}$$

Q.E.D.

APPENDIX II

Proof of Theorem 4

- For the Chao plan, it is sufficient to use (6), (9) and (15).
- For the randomized systematic plan, it is sufficient to use the approximation of the $\pi_{(N;i,j)}$ given by Deville (p. 21)

$$\pi_{(N;i,j)} \approx \pi_{(N;i)}\pi_{(N;j)}\frac{n-1}{n - \pi_{(N;i)} - \pi_{(N;j)}}. \tag{22}$$

This expression is obtained from the hypothesis

$$\mathrm{Max}_{1 \le i \le N}\left\{\frac{\pi_{(N;i)}}{n}\right\} \to 0.$$

This last hypothesis is verified since $n \to \infty$.

- For the rejective plan, using Hájek's result (1964, p. 1508), we have

$$\frac{-\Delta_{(N;i,j)}}{\pi_{(N;i,j)}} \approx \frac{[1 - \pi_{(N;i)}]\,[1 - \pi_{(N;j)}]}{d - [1 - \pi_{(N;i)}]\,[1 - \pi_{(N;j)}]}, \tag{23}$$

for $d \to \infty$. We note that (23) remains valid for the Rao-Sampford plan (see Hájek 1981, Theorem 8.2, p. 82). Using the approximation (Hájek 1964, p. 1521),

$$\{d - [1 - \pi_{(N;i)}]\,[1 - \pi_{(N;j)}]\}^{-1} \approx \frac{n}{d(n-1)},$$

we obtain the result of the theorem.

Q.E.D.

ACKNOWLEDGEMENTS

The author wishes to thank the referees who submitted a number of constructive comments that led to considerable improvements.

REFERENCES

BETHLEHEM, J.G., and SCHUERHOFF, H. (1984). Second-order inclusion probabilities in sequential sampling without replacement with unequal probabilities. *Biometrika*, 71, 642-644.

CHAO, M.T. (1982). A general purpose unequal probability sampling plan. *Biometrika*, 69, 653-656.

DEVILLE, J.-C. (No date). Cours de sondage, Chapitre III: les outils de bases. Lecture notes, ENSAE, Paris.

HÁJEK, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Annals of Mathematical Statistics*, 35, 1491-1523.

HÁJEK, J. (1981). *Sampling from a Finite Population*. New York and Bassel: Marcel Dekker, Inc.

HARTLEY, H.O., and RAO, J.N.K. (1962). Sampling with unequal probabilities without replacement. *Annals of Mathematical Statistics*, 33, 350-374.

HORVITZ, D.G., and THOMPSON, D.J. (1951). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47,663-685.

McLEOD, A.I., and BELLHOUSE, D.R. (1983). A convenient algorithm for drawing a simple random sample. *Applied Statistics*, 32, 2.

RAO, J.N.K. (1965). On two simple schemes of unequal probability sampling without replacement. *Journal of the Indian Statistical Association*, 3, 173-180.

SAMPFORD, M.R. (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika*, 54, 494-513.

YATES, F., and GRUNDY, P.M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society*, series B, 1, 253-261.