Comparison of PRN Techniques for Small Size PPS Sample Coordination

Esbjörn Ohlsson*

March 1999

Abstract

Consider multi-stage sampling from a stratified finite population, with a few primary sampling units selected in each stratum using probabilities proportional to size (pps). In a repeated survey, it is at times desired to redesign the sample with new size measures and new strata, while retaining as many units as possible from the old sample. In a former paper (Ohlsson, 1996) we considered the case with sample size $n = 1$, for which we gave an overview of existing methods and proposed a new method based on the use of permanent random numbers (PRN). In the present paper we focus on the case with small stratum sample sizes ($2 \leq n \leq 4$). We discuss the properties of different PRN methods and present a simulation study of their achieved sample overlap.

KEY WORDS: overlap maximization; overlap control; sample redesign; permanent random numbers; probabilities proportional to size.

---

*Address: Mathematical Statistics, Stockholm University, S-106 91 Stockholm, Sweden. e-mail: esbj@matematik.su.se. Financial support from the Bank of Sweden Tercentenary Foundation is gratefully acknowledged.

Contents

1 INTRODUCTION

In stratified multi-stage sampling designs it is common to draw primary sampling units with probability proportional to some measure of the units size (pps). Often, just a few units are selected in each stratum, and this is the case we consider here.

In a repeated survey, it is often desirable to retain the sample over time, both for cost and efficiency reasons. On the other hand, at times there is a need to update the sample to account for new size measures and/or to adjust the stratification in accordance with changes in the population (or in the budget of the survey). In this paper we consider the problem of properly updating a pps sample while retaining as many units as possible, i.e., maximizing the overlap of the old and new sample.

In the literature there are various procedures for sample coordination using *permanent random numbers* (PRN), covering the equal probability case for arbitrary n, and the pps case with moderate or large n, see Ohlsson (1995) for an overview. Such techniques are used by statistical agencies in several countries both for positive coordination (retaining units) and for negative coordination (reducing the overlap between different surveys).

In an earlier report (Ohlsson, 1996) we presented PRN *exponential sampling*, which is an efficient technique for the case $n = 1$, and compared it to other (non-PRN) solutions. In the present paper we discuss pps sample coordination with general n, focusing on cases with small sample sizes, which appears to be the most important in stratified multi-stage situations. Together with the mentioned results for moderate or large n, all possible situations are then covered by PRN techniques.

The starting point for this work was the need of the Swedish National Road Administration to update a master sample of roads, with stratum sample sizes $n = 1$. Other surveys at the road administration are now using one of the PRN techniques considered here, sequential Poisson sampling, with sample sizes in the range $2 \leq n \leq 10$.

We present a few candidate pps PRN procedures and discuss their properties. A major part of the paper presents a simulation study of the expected overlap and the inclusion probabilities of the procedures. We also give a brief overview of non-PRN solutions to the sample overlap problem.

2 CRITERIA FOR PPS SAMPLE COORDINATION

Even though we consider a stratified design, it is sufficient for our purposes to introduce notation for a single stratum only. The (stratum) population is $U = \{1, 2, \ldots, N\}$. The population is assumed to be recorded in a list frame, where there is a non-negative auxiliary variable $\mathbf{p} = (p_1, p_2, \ldots, p_N)$. In applications, p_i is usually a measure of the size of unit i. We assume that the p_i's have been normed so that $\sum p_i = 1$, within each stratum.

We search for a sampling procedure that enables the selection of pps samples at a first occasion, plus selection of pps samples at a second occasion using updated stratification and size measures, while retaining as many units as possible from the first sample.

We now specify our requirements on such a procedure in more detail. All quantities relating to the second sample s' will be equipped with a prime, as in p'_i.

Following Särndal et al. (1992, p. 90) we list the following desirable properties of a pps procedure, with π_i denoting actual inclusion probabilities, $\pi_i = \Pr(i \epsilon s)$:

(i) Relative simplicity in application.

(ii) For the first sample

$$\pi_i = np_i \quad ; \quad i \epsilon U \tag{1}$$

(iii) For the second sample

$$\pi'_i = n'p'_i \quad ; \quad i \epsilon U' \tag{2}$$

Särndal et al., add three conditions that enable variance estimation with the Sen-Yates-Grundy estimator. In a later article Särndal (1996) argues for the use of procedures that allow simple, single-sum variance estimation. This is our fourth property. Finally, we add two conditions that are particular for the problem of overlap control.

(iv) Availability of a variance estimator, preferably expressed as a single sum.

(v) Maximization of the expected sample overlap, i.e., of the expected number of units in common to the two samples, which is

$$\sum_i \Pr(i\epsilon s, i\epsilon s') \tag{3}$$

(vi) Independent sampling within strata.

Item (vi) is usually taken for granted in stratified sampling. It is included here since there are procedures proposed in the literature for which the second sample does not yield independent stratum samples.

As noted by Särndal et al. (1992, p. 90) it is not easy to devise a procedure having the properties desirable for pps sampling even at a single occasion. In Ohlsson (1996) we saw that there is no single procedure fulfilling (i)-(iii) plus (v)-(vi) in the case $n = 1$. An obvious conclusion is that we can not expect to find an optimal procedure in terms of (i)-(vi) in the case with general n. Instead we discuss procedures by Rosén (1997) and Ohlsson (1990, 1998) that fulfill (i), (iv) and (vi) and are not too far from the other. Closeness to (ii), (iii) and (v) will then be used to discriminate between the proposed procedures.

3 PRN PROCEDURES

In PRN sampling, each unit in the frame is given an independent uniform random number between 0 and 1, denoted by X_i for unit i. These numbers are permanently associated with the unit, i.e. they are *permanent random* *numbers.* Brewer, Early and Joyce (1972) suggested the use of PRN in connection with Poisson sampling, where unit i is included in the sample s if $X_i/p_i \leq n$. At the second occasion, the same rule is applied to the updated population, using the same X_i, so that unit i is included in s' if $X_i/p'_i \leq n'$. Note that the desired inclusion probability np_i may change due to change in allocation, i.e. n, or a up-dated p_i either because of a change in size measure or because of stratum changes.

It is readily seen that this procedure is optimal in terms of our criteria (i)-(vi) above. In particular, the probability of including unit i in both samples is

$$\Pr(i\epsilon s, i\epsilon s') = \min(np_i, n'p'_i) \tag{4}$$

which is obviously the maximum, so that after summation over i a maximum in (v) is reached.

The problem is that Poisson sampling is unfeasible for small n, since it yields a random sample size. The actual sample size is approximately Poisson distributed with parameter n. Hence, for small n there is a substantial probability of getting an empty sample. There can also be a substantial loss in (conditional) efficiency of the design from having a randomly varying number of psu's in the sample.

We conclude that Poisson sampling can not be used for pps sampling with small n, say $n \leq 10$, which is our main issue here. Equation (4) is useful as an upper limit for sample overlap, though.

For the case $n = 1$ we saw in Ohlsson (1996) that Exponential sampling is an exact pps procedure with fixed sample size that yields a large expected sample overlap. Here we compute the *transformed random numbers*

$$\xi_i = -\frac{\log(1 - X_i)}{p_i}; \quad i = 1, 2, ..., N \tag{5}$$

In each stratum we then select the unit that has the smallest transformed random number ξ_i. The natural extension of this procedure to the case with general n is to select the units with the n smallest ξ_i. This yields so called successive sampling, which is not strictly pps. In fact, the larger n is, the closer we come to equal probability sampling.

Cochran (1977, Section 9A.8-9) presents several attempts to cope with this problem. For the case $n = 2$, we consider Brewer's method of adjusting the draw probabilities p_i so that successive sampling becomes strictly pps. In our PRN setting Brewer's method amounts to using the following transformed random numbers for the first draw:

$$\xi_i = -\frac{\log(1 - X_i)}{p_i(1 - p_i)/(1 - 2p_i)}; \quad i = 1, 2, ..., N \tag{6}$$

so that the selection probability in the first draw is proportional to $p_i(1 - p_i)/(1 - 2p_i)$. For the second draw, we use (5), after having removed the unit selected in the first draw, so that the probability of selecting unit i in the second draw is $p_i/(1 - p_j)$, where j is the unit selected in the first draw, $j \neq i$.

We will call this procedure 'Brewer's successive sampling'. See Cochran (1977, Section 9A8) for a proof that the probability of inclusion after two draws is $2p_i$, i.e. the procedure is strictly pps. We conclude that properties (i)-(iii) of the previous section are met. As for property (iv), the Sen-Yates-Grundy estimator can be used, but no single-sum variance estimator is available. Expected overlap (v) will be investigated in Section 5 and, finally, property (vi) is trivially fulfilled.

Note that the validity of this procedure relies on the fact that the transformed numbers in (6), like those in (5), are exponentially distributed, and that the exponential distribution is memoryless so that the relation among the non-selected units remains the same after conditioning on the outcome of the transformed random number for the first drawn unit, j.

Sampford (1967) extended Brewer's method to general $n > 2$. This approach is quite involved and computationally heavy and will therefore not be considered further.

Instead we will turn to PRN methods that use other transformations of the PRN X_i. Ohlsson (1990 and 1998) give a fixed size alteration of Poisson sampling, called *sequential Poisson sampling*, which selects the n smallest of

$$\xi_i = \frac{X_i}{p_i}; \quad i = 1, 2, ..., N \tag{7}$$

In Ohlsson (1998) it is shown, by asymptotics and simulation, that this procedure approximately fulfills (ii) and (iii). This theoretic result is supported by a simulation study for $n \geq 5$.

Rosén (1997), presents an improvement of sequential Poisson sampling, called *Pareto sampling*, for which π_i is somewhat closer to the target probability in (ii). Here, the transformed random numbers are

$$\xi_i = \frac{X_i/(1 - X_i)}{np_i/(1 - np_i)}; \quad i = 1, 2, ..., N \tag{8}$$

Again, asymptotic and numerical results in the papers by Rosén (1997 and 1998) show that this procedure approximately fulfills (ii) and (iii). Note that Rosén does not consider sampling on two occasions.

From asymptotic considerations, Rosén (1997) also suggests an alteration of successive sampling, which can be considered as a simple alternative to the Brewer-Sampford approach. It consists of changing the denominator in (5) to

$$\xi_i = \frac{\log(1 - X_i)}{\log(1 - np_i)}; \quad i = 1, 2, ..., N \tag{9}$$

We will call the procedure that selects the n smallest of these transformed random numbers 'Rosén's successive sampling'.

Sequential Poisson sampling, Pareto sampling and Rosén's successive sampling share the following properties: They have the desired properties (i), (iv) and (vi) listed in Section 2. They are asymptotically pps; the closeness to the ideal in (ii) and (iii) for small n will be discussed in Section 5. Their expected overlap has not been investigated in the literature. A major part of the present paper is the simulation study in Section 5 on the expected overlap and the inclusion probabilities for sequential Poisson sampling, Pareto sampling and Rosén's successive sampling in case $n = 1, 2, 3, 4$.

4 OTHER APPROACHES

Here we briefly describe some non-PRN methods that have been suggested in the literature.

Causey, Cox and Ernst (1985) suggested a procedure that maximizes the expected overlap by solving a special type of linear programming problem, a so called transportation problem. The procedure is recapitulated in detail in Ohlsson (1996).

Ernst and Ikeda (1995) note two difficulties with the procedure "which can make it unusable in practice". One is that the strata become dependent so that (vi) is not fulfilled, which in particular means that it can be difficult or impossible to apply the procedure repeatedly for the same survey, and to compute a variance estimator. The second difficulty is that the "the transportation problem may be too large to solve in practice". Even when the problem is solvable, the procedure is quite complicated and will not be considered further.

Sunter (1989) presents an interesting method which in principle is applicable for any sample size and any sampling procedure with completely known sample distribution. The latter requirement is not fulfilled for most pps procedures, but is true for the particular procedure presented in Sunter (1986). This procedure is a bit more complicated than the PRN methods, and has been shown by Rosén (1997) to be less efficient.

A restriction with Sunter's (1989) method is that the old and new sample must be the same size, i.e. $n = n'$. Sunter's method appears to be designed primarily for updating size measures, but it can also handle stratum changes by treating units that are new in a stratum as if they had $p_i' = 0$. This can be expected to substantially reduce the amount of overlap when there are many stratum changes. In case $n = 1$, Sunter's method specializes to the well-known procedure by Keyfitz (1951), which was shown to give a substantially less expected overlap than PRN Exponential sampling in the numerical study in Ohlsson (1996).

Our conclusion is not to consider Sunter's procedure further here even though we believe that it may be useful under special circumstances.

Finally, we mention the procedure by Fellegi (1966), which is restricted to the case $n = 2$ and does not handle stratum changes.

5 A SIMULATION STUDY

Here we present a simulation study of expected overlap and inclusion probabilities for Rosén's successive sampling (SU), sequential Poisson sampling (SE), and Pareto sampling (PA), for $n = 1, 2, 3, 4$. In the important special cases $n = 1$ and $n = 2$, we will also consider exponential sampling (EX) and Brewer's successive sampling (BR), respectively, which are specially designed for these cases.

The simulation study uses data from a master frame of the Swedish National Road Administration. The statistical units are 'stretches of road' and the size measure is derived from traffic mileage per year. The units are stratified according to region and type of road, altogether 28 strata. The data are available at the address `http://www.matematik.su.se/~esbj/roads.dat`

Table 0 below gives some elementary statistics for the normed size measures

p_i within the 28 strata. Originally, there were 2524 units, but one unit in stratum 32 with $p_i = 0.50$ was removed to ensure $np_i < 1$, leaving us with 2523 units. Note that the strata represent quite a variety of different sizes and spread in the p_i's.

The population at the second occasion was generated by letting 10% of the units move to another stratum as follows: 5% of the units move to the next stratum, 2.5% move to the second next and 2.5% to the third, in a circular fashion.

The number of iterations, 14000, was chosen so that when estimating a probability of 0.10, the 95% confidence limits for the true probability are ± 0.005.

The simulations where run in SAS 6.12, using the built-in, congruential modulo $(2^{31} - 1)$ pseudo-random number generator. A single stream of 14000×2523 pseudo-random numbers was used, with a haphazard starting point. For each new simulation ($n = 1, 2, 3, 4$), a new starting point was used. In each iteration, we used the same 2523 PRN for all procedures. This makes the results for different procedures positively correlated and, if anything, increases the precision in comparisons between them.

5.1 Inclusion probabilities

Here we investigate the deviation of the actual inclusion probabilities (AIP) $\pi_i = \Pr(i \epsilon s)$, from the target probabilities np_i for SU, SE and PA. In this section, results are given for the first sample only, since the second sample

Table 0. Normed size measures. Percent.
Note: '0.0' means a value less than 0.05.

Stratum	N	Min	Median	Max	CV(%)
11	44	0.1	2.5	5.0	51
12	41	0.4	2.0	8.4	73
13	58	0.2	1.4	7.0	80
14	292	0.0	0.2	3.5	123
21	35	0.3	2.5	6.2	47
22	80	0.2	1.2	6.1	69
23	82	0.1	0.8	3.9	84
24	304	0.0	0.2	2.2	96
31	12	1.6	6.5	18.5	67
32	6	12.6	16.5	19.9	15
33	34	0.4	1.7	10.3	85
34	61	0.4	1.2	5.4	78
41	39	0.7	1.7	9.7	78
42	79	0.1	1.0	5.8	68
43	73	0.1	1.2	6.6	72
44	318	0.0	0.2	1.8	92
51	30	1.7	2.8	7.3	45
52	56	0.6	1.6	3.7	39
53	44	0.5	2.1	5.2	55
54	170	0.1	0.5	2.5	77
61	37	1.0	2.1	10.1	74
62	68	0.2	1.4	5.1	46
63	53	0.2	1.7	5.0	55
64	301	0.0	0.2	2.1	84
71	21	1.4	3.4	12.5	64
72	34	1.4	2.7	7.1	46
73	29	1.1	3.4	7.2	49
74	122	0.1	0.7	3.6	68

gave similar results. The estimated AIP from the simulations will be called 'Monte Carlo AIP'.

The bias of the 'pseudo-Horvitz-Thompson' estimator of the total Y, denoted $\hat{Y}_{HT}$, that is used with these procedures is

$$E(\hat{Y}_{HT}) - Y = \sum \frac{y_i}{np_i}(\pi_i - np_i)$$

We apply pps sampling in the belief that the target variable y is more or less proportional to the size measure p_i. Then it is natural to measure the difference in absolute terms $|\pi_i - np_i|$. On the other hand, the most serious bias may occur when proportionality does not hold, in which case it may be more relevant to measure the difference in relative terms $|(\pi_i - np_i)/(np_i)|$. In the present study, we present both absolute and relative differences.

For the case $n = 1$, Table 1a gives the median, the 75% and 90% quantiles, plus the maximum of $|\pi_i - np_i|$. Note that for brevity, these quantiles are given across all strata. All values have been multiplied by 100.

Table 1b gives the median, the 75% and 90% quantiles and the maximum of $|(\pi_i - np_i)/(np_i)|$. Tables 2a-4b are the $n = 2, 3, 4$ counterparts of Tables 1a and 1b.

In the tables we have included simulation results for Poisson sampling (PO), even though we already know that PO is unbiased. The PO values are included because they serve as a benchmark for the other procedures.

We have also investigated the mean of $|(\pi_i - np_i)/(np_i)|$ stratum by stratum. For the case $n = 2$ the results are given in Table 5. The investigation for other n gave similar conclusions and is omitted for brevity.

5.1.1 Conclusions on inclusion probabilities

The stratum by stratum comparison in Table 5 of *relative* deviations shows that in most strata the (potentially) biased procedures SU, SE and PA are not far from the unbiased PO, and there is little to choose between the procedures. In a few strata we have a notably larger deviation for SU and especially for SE. The strata where we have the largest deviations for SE are 31, 32 and 71. These strata are also the ones in which we have the largest maximum p_i, cf. table 0. The conclusions for other n are similar.

As for *absolute* deviation, again most units give no indication of bias when compared to Poisson sampling. For SU, and in particular for SE, there are a few units with a notable deviation. The largest absolute deviation for SE consistently occurred for the largest unit in stratum 31, with $p_i = 19\%$. As an example, for $n = 3$ this unit has Monte Carlo AIP 59.1% where we wanted $np_i = 55.6\%$. Even though PA shows a slight bias when compared to PO, at least in the case $n = 3$, this bias is very small.

Rosén (1998) studies exact inclusion probabilities for SE, PA and SU in an artificial, but nevertheless interesting, situation, viz. when all units have the same inclusion probability except for 'one odd unit'. Rosén concludes that PA 'differs only negligibly' from strict pps if $\min(n, N-n) \geq 5$, and that SE and SU 'also converge rapidly to target values, but not as fast as for' PA.

In our opinion, Roséns conclusions are a bit too conservative. We believe that

Table 1a. $n = 1$. Absolute deviation from target inclusion probabilities. Percent.

Procedure	Median	75%	90%	Max
SU	0.040	0.078	0.137	0.967
SE	0.040	0.078	0.142	1.677
PA	0.040	0.077	0.137	0.689
PO	0.040	0.078	0.140	0.685

Table 1b. $n = 1$. Relative deviation from target inclusion probabilities. Percent.

Procedure	Median	75%	90%	Max
SU	7.25	14.61	24.62	140.12
SE	7.36	14.68	24.70	140.12
PA	7.17	14.60	24.66	140.12
PO	7.24	14.37	25.23	134.67

Table 2a. $n = 2$. Absolute deviation from target inclusion probabilities. Percent.

Procedure	Median	75%	90%	Max
SU	0.055	0.112	0.201	1.682
SE	0.055	0.116	0.210	3.404
PA	0.055	0.110	0.198	0.855
PO	0.053	0.106	0.197	0.867

Table 2b. $n = 2$. Relative deviation from target inclusion probabilities. Percent.

Procedure	Median	75%	90%	Max
SU	5.22	10.14	17.15	97.74
SE	5.40	10.25	17.30	97.74
PA	5.13	10.14	17.45	97.74
PO	4.96	9.92	17.13	84.86

Table 3a. $n = 3$. Absolute deviation from target inclusion probabilities. Percent.

Procedure	Median	75%	90%	Max
SU	0.067	0.134	0.238	1.545
SE	0.069	0.140	0.266	3.459
PA	0.066	0.132	0.229	1.427
PO	0.064	0.132	0.228	1.048

Table 3b. $n = 3$. Relative deviation from target inclusion probabilities. Percent.

Procedure	Median	75%	90%	Max
SU	3.98	8.00	14.38	89.94
SE	4.10	8.16	14.33	89.94
PA	3.94	7.99	14.10	89.94
PO	3.95	8.07	13.97	79.81

Table 4a. $n = 4$. Absolute deviation from target inclusion probabilities. Percent.

Procedure	Median	75%	90%	Max
SU	0.079	0.163	0.282	1.729
SE	0.080	0.170	0.304	3.708
PA	0.078	0.159	0.276	1.065
PO	0.080	0.163	0.275	0.968

Table 4b. $n = 4$. Relative deviation from target inclusion probabilities. Percent.

Procedure	Median	75%	90%	Max
SU	3.58	7.36	13.26	59.65
SE	3.75	7.42	13.31	59.65
PA	3.56	7.40	13.23	59.65
PO	3.59	7.27	12.73	66.91

Table 5. $n = 2$. Mean (absolute) relative deviation by stratum, per cent.

Stratum	NH	SU	SE	PA	PO
11	44	4.4	4.3	4.7	3.7
12	41	4.0	4.7	3.6	3.9
13	58	4.3	4.4	4.1	4.3
14	292	10.8	10.8	10.9	11.1
21	35	2.7	2.8	2.7	2.6
22	80	4.3	4.5	4.3	4.0
23	82	6.2	6.3	6.1	5.9
24	304	11.0	11.0	10.9	11.0
31	12	4.0	7.2	2.1	2.8
32	6	2.0	3.2	1.4	0.9
33	34	4.0	4.6	3.7	4.4
34	61	5.2	5.4	5.0	4.5
41	39	2.9	3.6	2.8	3.5
42	79	5.2	5.4	5.0	5.2
43	73	5.4	5.5	5.3	4.4
44	318	10.5	10.4	10.5	10.1
51	30	2.5	2.6	2.6	2.3
52	56	4.2	4.2	4.1	3.8
53	44	3.5	3.6	3.4	3.7
54	170	7.6	7.6	7.7	7.9
61	37	3.8	4.3	3.5	3.8
62	68	4.9	4.9	4.9	4.0
63	53	4.4	4.4	4.5	4.1
64	301	10.3	10.3	10.4	9.9
71	21	2.2	3.3	1.7	2.0
72	34	3.3	3.7	3.1	2.8
73	29	2.6	3.0	2.4	2.8
74	122	5.3	5.4	5.3	6.2

Roséns study is consistent with the idea of using PA even with sample sizes as small as $n = 2$. When $2 \leq n \leq N - 3$ the relative error in the AIP for PA is never larger than 2% and in most cases it is much smaller. SU and SE are a bit further from the target probability.

Our conclusion so far is that it seems safe to use PA even with very small sample sizes, $n \geq 2$. PA is almost consistently closer to the target than SU and SE, and SU is somewhat better than SE.

5.2 Expected overlap

The major result of the simulation study is a comparison of the expected overlap, i.e. the expected number of units that are retained from the old sample to the new. In Table 6 we present results stratum by stratum for the case $n = n' = 2$. Since the results are similar in all strata we confine to giving only the totals for the other values of n. The totals are given in Table 7.

For comparison we have included the expected overlap for the situation when s are s' are independent, $\sum(np_i \times n'p'_i)$, denoted by IND. Further, we give the upper limit in (4), plus the total sample size n. Besides SU, SE and PA, we include figures for exponential sampling (EX) and Brewer's procedure (BR), as applicable.

Note that the upper limit is typically not achievable for a fixed size procedure, cf. Ohlsson (1996). The precision in Table 7 is roughly ± 0.1, as measured by a crude approximation to a 95% c.i.

Table 6. $n = 2$. Expected overlap by stratum.

Stratum	NH	IND	BR	SU	SE	PA	Limit
11	44	0.12	1.77	1.77	1.77	1.77	1.97
12	41	0.13	1.66	1.66	1.66	1.66	1.78
13	58	0.12	1.80	1.79	1.79	1.79	1.93
14	292	0.03	1.75	1.75	1.74	1.74	1.93
21	35	0.14	1.78	1.78	1.78	1.78	1.94
22	80	0.07	1.73	1.72	1.72	1.72	1.91
23	82	0.08	1.72	1.72	1.72	1.72	1.91
24	304	0.03	1.73	1.73	1.73	1.73	1.91
31	12	0.49	1.77	1.77	1.77	1.78	1.93
32	6	0.63	1.61	1.61	1.60	1.61	1.77
33	34	0.19	1.56	1.54	1.54	1.55	1.79
34	61	0.11	1.77	1.76	1.76	1.77	1.96
41	39	0.16	1.72	1.73	1.72	1.73	1.95
42	79	0.07	1.70	1.70	1.70	1.70	1.88
43	73	0.08	1.69	1.69	1.69	1.69	1.91
44	318	0.02	1.73	1.73	1.73	1.73	1.94
51	30	0.16	1.70	1.70	1.69	1.70	1.95
52	56	0.08	1.70	1.70	1.70	1.70	1.89
53	44	0.12	1.66	1.66	1.66	1.66	1.91
54	170	0.04	1.67	1.66	1.66	1.66	1.92
61	37	0.16	1.73	1.73	1.73	1.73	1.90
62	68	0.08	1.75	1.75	1.75	1.75	1.96
63	53	0.10	1.66	1.66	1.65	1.66	1.92
64	301	0.02	1.65	1.64	1.64	1.64	1.86
71	21	0.25	1.66	1.66	1.66	1.66	1.86
72	34	0.15	1.65	1.64	1.64	1.65	1.94
73	29	0.17	1.63	1.62	1.62	1.63	1.93
74	122	0.05	1.71	1.71	1.71	1.71	1.92

Table 7. Expected overlap, aggregated over strata.

Sample size	IND	EX/BR	SU	SE	PA	Limit	n
$n = 1$	1.0	23.0	23.0	23.0	23.0	26.7	28
$n = 2$	3.9	47.6	47.6	47.5	47.6	53.4	56
$n = 3$	8.7	.	72.9	72.8	72.9	80.1	84
$n = 4$	15.5	.	98.7	98.6	98.7	106.7	112

We conclude that all PRN methods perform equally well as regards expected overlap and that they give a result that, compared to the (non-achievable) upper limit and to the independent case, is very good. PRN sampling heavily reduces the cost for sample renewal in the studied situation.

6 CONCLUSIONS

The various PRN techniques investigated here are simple and efficient for maximizing the overlap of two successive samples. Except for Brewer's method, they allow for simple, single-sum variance estimation when $n \geq 2$. The choice between them should then primarily be based on how close the first order inclusion probabilities are to the target np_i.

For $n = 1$, our choice is exponential sampling, which is strictly pps, while the others have potential problems (manifested in the study by Rosén, 1998).

Pareto sampling is almost consistently closer to having the right inclusion probabilities, than sequential Poisson and successive sampling. For $n \geq 2$, our choice is thus Pareto sampling, which is not far from unbiased in Roséns and our own study. If exact inclusion probabilities are deemed important, and if all strata sample sizes are $n = 2$, one might consider using Brewer's method, at the cost of a bit more complicated variance estimation.

It should also be noted that sequential Poisson sampling has the virtue of giving separate sequential Poisson samples within post-strata. If this property is deemed important, as it has been for the Swedish Purchase Power Parity survey, the choice may fall on this procedure, but then one should avoid small sample sizes and/or move all extremely large units to a take-all stratum.

In most cases, Pareto sampling is the best choice. In our opinion, it gives simple and efficient sample coordination.

REFERENCES

Brewer, Early and Joyce (1972). Selecting several samples from a single population. "Australian Journal of Statistics," 14, 231-239.

Causey, B. D., Cox, L. H. and Ernst, L. R. (1985), "Application of Transportation Theory to Statistical Problems" *Journal of the American Statistical Association,* 80, 903-909.

Cochran, W. G. (1977). *Sampling Techniques 3d ed.,* New York: John Wiley.

Ernst, L. R. and Ikeda, M. M. (1995), "A Reduced-Size Transportation Algorithm for Maximizing the Overlap Between Surveys," *Survey Methodology,* 21, 147-157.

Fellegi, I. P. (1966), "Changing the Probabilities of Selection when Two Units are Selected with PPS Without Replacement," *Proceedings of the Social Statistics Section, American Statistical Association,* 434-442.

Keyfitz, N. (1951), "Sampling with Probabilities Proportional to Size," *Journal of the American Statistical Association,* 46, 105-109.

Ohlsson, E. (1990), "Sequential Poisson Sampling from a Business Register and its Application to the Swedish Consumer Price Index," *R & D Report 1990:6, Statistics Sweden.*

Ohlsson, E. (1995), "Coordination of Samples Using Permanent Random Numbers," In *Business Survey Methods,* New York: Wiley, 153-169.

Ohlsson, E. (1996), "Methods for PPS Size One Sample Coordination," Research Report No. 194, Institute of Actuarial Mathematics and Mathematical Statistics, Stockholm University.

Ohlsson, E. (1998), "Sequential Poisson Sampling," *Journal of Official Statistics*, 14, 149-162.

Rosén, B. (1997), "On Sampling with Probability Proportional to Size," *Journal of Statistical Planning and Inference*, 62, 159-191.

Rosén, B. (1998), "On Inclusion Probabilities for Order Sampling," *R & D Report 1998:2, Statistics Sweden.*

Sampford, M. R., (1967), "On Sampling Without Replacement with Unequal Probabilities of Selection," *Biometrika,* 54, 499-513.

Särndal, C. E., Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling,* New York: Springer-Verlag.

Särndal, C. E., (1996), "Efficient Estimators With Simple Variance in Unequal Probability Sampling," *Journal of the American Statistical Association,* 91, 1289-1300.

Sunter, A. B (1986), "Solutions to the Problem of Unequal Probability Sampling Without Replacement," *International Statistical Review,* 54, 33-50.

Sunter, A. B (1989), "Updating Size Measures in a PPSWOR Design," *Survey Methodology,* 15, 253-260.