IS THERE A BEST FIXED SIZE πPS SAMPLING DESIGN?

Lennart Bondesson
Umeå university, Sweden
e-mail: Lennart.Bondesson@math.umu.se

Abstract

Fixed size πps sampling with prescribed inclusion probabilities is considered. It is discussed whether there is a best πps design. Several candidates are presented, as the Sampford design, the adjusted conditional Poisson design, the adjusted Pareto design, and some other designs. No definite conclusion is presented.

1. Introduction

A population of units $1, 2, \ldots, N$ is considered. We want to take a sample without replacement of size n according to given inclusion probabilities $\pi_1, \pi_2, \ldots, \pi_N$ with sum $\sum_{i=1}^{N} \pi_i = n$. The inclusion probabilities are assumed to be roughly proportional to the y_i-values of an interesting y-variable. We intend to estimate the population total by the Horvitz-Thompson estimator

$$\hat{Y}_{HT} = \sum_{i=1}^{N} \frac{y_i}{\pi_i} I_i,$$

where I_i is 1 if unit i is sampled and otherwise 0. We set $a_i = \breve{y}_i = y_i/\pi_i$ and then $\hat{Y}_{HT} = \sum_{i=1}^{N} a_i I_i$. The variance of the HT-estimator can be written as

$$Var(\hat{Y}_{HT}) = \mathbf{a}^T \mathbf{\Sigma} \mathbf{a},$$

where $\mathbf{\Sigma} = (c_{ij})$ is the matrix of covariances $c_{ij} = Cov(I_i, I_j)$ with $c_{ii} = d_i = \pi_i(1 - \pi_i)$. We also have, the Sen-Yates-Grundy form of the variance:

$$Var(\hat{Y}_{HT}) = \frac{1}{2} \sum_{i,j} \tilde{c}_{ij}(a_i - a_j)^2, \quad \text{where} \quad \tilde{c}_{ij} = -c_{ij}.$$

Dividing here $\tilde{c}_{ij}$ by $\pi_{ij} = E(I_i I_j)$ and then summing instead over i and j in the sample, we get the SYG variance estimator.

There is no sampling design with smallest variance uniformly in $\mathbf{a}$. In fact, if such a design with covariance matrix $\mathbf{\Sigma}_0$ existed, we would have $\mathbf{a}^T \mathbf{\Sigma}_0 \mathbf{a} \leq \mathbf{a}^T \mathbf{\Sigma} \mathbf{a}$ for all $\mathbf{a}$ and all other $\mathbf{\Sigma}$ with diagonal elements $d_i = \pi_i(1 - \pi_i)$, $i = 1, \ldots, N$. Then $\mathbf{D} = \mathbf{\Sigma} - \mathbf{\Sigma}_0 \geq 0$ (in matrix sense) and hence the eigenvalues of $\mathbf{D}$ are nonnegative. But they sum to 0 because the diagonal elements of $\mathbf{D}$ are 0 and hence trace($\mathbf{D}$) = 0. Thus all the eigenvalues are 0 and hence $\mathbf{\Sigma} = \mathbf{\Sigma}_0$, which is a contradiction.

We have not any superpopulation in mind except possibly the simplest one: $a_i = \alpha + \epsilon_i$, where the ϵ_is are uncorrelated with mean zero and variance σ^2. For that model, with E here denoting expected value w.r.t. the superpopulation, we have, since $\sum_{j;\, j\neq i} c_{ij} \equiv -\pi_i(1-\pi_i)$,

$$\mathrm{E}(Var(\mathbf{a}^T\mathbf{I})) = \frac{1}{2}\sum_{i,j;\, i\neq j} \tilde{c}_{ij}\mathrm{E}((a_i-a_j)^2) = \frac{1}{2}\sum_{i,j;\, i\neq j} \tilde{c}_{ij} 2\sigma^2 = \sigma^2 \sum_{i=1}^{N} \pi_i(1-\pi_i).$$

Hence all designs are equally efficient with respect to this superpopulation.

So to single out a 'best' design further considerations are needed. In the literature, e.g. Brewer *et al.* (1983), many πps designs are discussed. Some of them are approximate in the sense that $E(I_i) \approx \pi_i$ only. Such designs are not considered here. The following three designs deserve a lot of attention as candidates for being at least very good πps designs.

1. The Sampford design
2. The adjusted conditional Poisson design
3. The adjusted Pareto design.

These are discussed in sections 2 and 3. We give motivations for them and present advantages and drawbacks of them. In section 4 we derive some slightly more theoretical designs related to the conditional Poisson design. In section 5 and in an appendix we derive and discuss some further designs, which are of 2nd order type, i.e. are only given by their 2nd order inclusion probabilities. They are also more theoretical than practical. We illustrate and compare the methods in section 6 by looking at a small but not trivial population for which $N = 6, n = 3$, and $\pi_1 = \pi_2 = \pi_3 = 1/3$ and $\pi_4 = \pi_5 = \pi_6 = 2/3$. This population, the TBM-population, has earlier been considered in Traat *et al.* (2004). The paper ends with a brief discussion in section 7.

2. The Sampford, the adjusted conditional Poisson, and the adjusted Pareto designs

Here we look at the three designs mentioned in the introduction. We present the designs mainly by their probability functions (pf) $p(\mathbf{x}) = \Pr(\mathbf{I} = \mathbf{x})$, where $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ with $x_i = 0$ or 1.

The **Sampford design** was introduced by Sampford (1967). Its pf is given by

$$p_S(\mathbf{x}) = C_S \prod_{i=1}^{N} \pi_i^{x_i}(1-\pi_i)^{1-x_i} \times \sum_{k=1}^{N}(1-\pi_k)x_k, \quad |\mathbf{x}| = \sum_{i=1}^{N} x_i = n.$$

It is a profound result that the true inclusion probabilities really equal π_i. The constant C_S is inexplicit but otherwise the pf is very explicit. It is possible to sample from this pf by first sampling one unit with replacement according to the probabilities π_i/n, $i = 1, \dots, N$, and then with replacement $n-1$ units according to the probabilities $p'_i \propto \pi_i/(1-\pi_i)$. If all these n units are distinct, the sample is accepted, otherwise the whole procedure is repeated. This is often a slow procedure but there are also other methods to sample from the pf (e.g. Grafström, 2005).

The **adjusted conditional Poisson design** was introduced by Hajek (1964, 1981). Tillé (2005) gives it a careful treatment. The pf is

$$p_{CP}(\mathbf{x}) = C_{CP} \times \prod_{i=1}^{N} p_i^{x_i}(1-p_i)^{1-x_i}, \quad |\mathbf{x}| = n,$$

where p_i with $\sum_{i=1}^{N} p_i = n$ must be chosen so that the desired inclusion probabilities π_i are obtained. Hajek presented various approximations but nowadays it is also possible to calculate the desired p_is numerically by a computer program (e.g. Tillé, 2005). A simple recent good approximation is, with $d = \sum_{k=1}^{N} \pi_k(1-\pi_k)$,

$$\frac{p_i}{1-p_i} \propto \frac{\pi_i}{1-\pi_i} \exp(\frac{1-\pi_i}{d}).$$

This approximation was derived in Bondesson *et al.* (2006) via the assumption that $p_{CP}(\cdot)$ is close to the Sampford pf. It turns out to yield a very good approximation. To sample from the conditional Poisson design is easy, one samples from the Poisson design (independent I_is, with $I_i \sim Bin(1, p_i)$) but only samples of the desired size are accepted.

The **Pareto design** was introduced by Rosén (1997a,b). The main idea dates back to Ohlsson (1990) and Saavedra (1995). Target probabilities λ_i such that $\sum_{i=1}^{N} \lambda_i = n$ are used. Let $U_1, U_2, \dots, U_N$ be random numbers from $U(0,1)$ and let

$$Q_i = \frac{U_i/(1-U_i)}{\lambda_i/(1-\lambda_i)}, \quad i = 1, \dots, N,$$

be *ranking variables.* Now select the n units with the smallest Q_is. If we put $\lambda_i = \pi_i$, $i = 1, \dots, N$, the true inclusion probabilities will approximately equal the π_is but not exactly. It is possible to make an adjustment so that the true inclusion probabilities will be π_i (Aires, 2000). A very good approximation in this direction is provided by, with $d = \sum_{k=1}^{N} \pi_k(1-\pi_k)$,

$$\frac{\lambda_i}{1-\lambda_i} \propto \frac{\pi_i}{1-\pi_i} \exp(-\frac{\pi_i(1-\pi_i)(\pi_i - \frac{1}{2})}{d^2}).$$

It is derived in Bondesson *et al.* (2006) from the assumption that the adjusted Pareto pf is close to the Sampford pf. Hence the Q_is above, with $\lambda_i = \pi_i$, only have to be multiplied by the factor $\exp(\pi_i(1-\pi_i)(\pi_i - \frac{1}{2})/d^2)$ to yield a sample with inclusion probabilities π_i.

The pf for the Pareto design is given by

$$p_{Par}(\mathbf{x}) = \prod_{i=1}^{N} \lambda_i^{x_i}(1-\lambda_i)^{1-x_i} \times \sum_{k=1}^{N} c_k x_k,$$

where the constants c_k are given by certain integrals (Traat *et al.*, 2004, Bondesson *et al.*, 2006). Approximately, $c_k \propto 1 - \lambda_k$, which shows that the Pareto and the adjusted Pareto pfs are close to the Sampford pf.

3. Advantages and drawbacks of the designs

The sampling designs in section 2 have pfs that are very close to each other. Should one of these designs be preferred? We look here at advantages and drawbacks of each of them, in the order: Sampford sampling, Pareto sampling and conditional Poisson sampling.

Sampford sampling. The main advantage of this design is that the pf is very explicit. A main drawback has been that the original methods to get a Sampford sample are slow. However, since the pf is explicit except for the normalizing constant, one can easily use MCMC methods as Gibbs sampling to sample from it. There are now also very rapid methods that use Pareto sampling in a first step and then acceptance/rejection technique (Bondesson *et al.*, 2006, Grafström, 2005). Another small drawback is that there is no known optimality property of Sampford sampling.

Pareto sampling. The big advantage of this method is that it is very easy to get a sample. Without adjustment the method gives a slight bias of the estimators. Although the bias is small, it is slightly disturbing and therefore one may advocate at least simple adjustment; cf. section 2. Another advantage of the method is that it permits the use of permanent random numbers. A drawback of the method is that there is no simple pf. It is also complicated to calculate the true inclusion probabilities. The method has no known optimum property except that it is asymptotically the best method among all order sampling procedures.

Adjusted conditional Poisson sampling. A main advantage of this design is that the entropy $-\sum p(\mathbf{x}) \log(p(\mathbf{x}))$ is maximized under the given restrictions (Hajek, 1981). Thus the probabilities are spread over the possible samples as much as possible in some sense. The probability function belongs to an exponential family. A drawback is that the pf is not very explicit since the p_is must be calculated. Another drawback is that the standard rejective procedure for sampling takes some time for large populations and samples but there are list-sequential methods also (Chen & Liu 1997, Öhlund, 1999, Traat *et al.*, 2004, Tillé, 2005). There are also rapid methods based on preliminary Pareto samples which are accepted or rejected. (Bondesson *et al.*, 2006, Grafström, 2005).

4. Other designs related to the adjusted conditional Poisson design

Using two different starting points, we derive some further designs related to the conditional Poisson design.

The maximum entropy property for the adjusted conditional Poisson design can be expressed in another way too. Let $p_I(\mathbf{x})$ denote the pf for Poisson sampling with probabilities $\pi_i, i = 1, 2, \ldots, N$. For a fixed size design with pf $p(\mathbf{x})$ and inclusion probabilities π_i, let us minimize the Kullback-Leibler divergence $KL = \sum_{\mathbf{x};\,|\mathbf{x}|=n} p(\mathbf{x}) \log(p(\mathbf{x})/p_I(\mathbf{x}))$. We have

$$KL = \sum_{\mathbf{x};\,|\mathbf{x}|=n} p(\mathbf{x}) \log p(\mathbf{x}) \; - \sum_{\mathbf{x};\,|\mathbf{x}|=n} \left[p(\mathbf{x}) \sum_{i=1}^{N} (x_i \log \pi_i + (1-x_i)\log(1-\pi_i)) \right]$$

$$= -\text{Entropy} \, + \sum_{i=1}^{N} \left[\log \pi_i \sum_{\mathbf{x};\,|\mathbf{x}|=n} x_i p(\mathbf{x}) \, + \, \log(1-\pi_i) \sum_{\mathbf{x};\,|\mathbf{x}|=n} (1-x_i) p(\mathbf{x}) \right]$$

$$= -\text{Entropy} \, + \sum_{i=1}^{N} (\pi_i \log \pi_i + (1-\pi_i)\log(1-\pi_i)).$$

Now since the entropy is maximized for the adjusted conditional Poisson pf, which can be proved by a use of Lagrange multipliers, it follows that KL is minimized for that pf. Of course, one could then also try to minimize another distance measure, the squared Hellinger metric

$$d_H^2 = \sum_{\mathbf{x}} (\sqrt{p(\mathbf{x})} - \sqrt{p_I(\mathbf{x})})^2 = 2 \, - \, 2 \times E_I(\sqrt{\frac{p(\mathbf{x})}{p_I(\mathbf{x})}})$$

given that $\sum_{\mathbf{x}} x_i p(\mathbf{x}) = \pi_i$, $i = 1, 2, \ldots, N$, and $p(\mathbf{x}) = 0$ for $|\mathbf{x}| \neq n$. It is more difficult to minimize d_H^2 but it is possible for small populations for which the number of different samples is limited. It would have been some extra support for the adjusted conditional Poisson design if the 'Hellinger design' had been equal to that design. As will be seen in section 6 it is not the case.

The maximum entropy for the adjusted conditional Poisson design ought to guarantee that the variance of the HT-estimator is small though not in a very direct way. Since $p \log p$ is a limit of $p(p^\epsilon - 1)/\epsilon$ as $\epsilon \downarrow 0$, we see that maximum entropy corresponds to minimization of $\sum_{\mathbf{x};\,|\mathbf{x}|=n} (p(\mathbf{x}))^{1+\epsilon}$ for an ϵ close to 0. In this connection, Hölder's inequality may give some additional insight. We have, with $x_i = I_i$ and $\hat{Y}(\mathbf{x}) = \hat{Y}_{HT}$,

$$Var(\hat{Y}) = \sum_{\mathbf{x};|\mathbf{x}|=n} (\hat{Y}(\mathbf{x}) - Y)^2 p(\mathbf{x}) \leq \left(\sum_{\mathbf{x};\,|\mathbf{x}|=n} (p(\mathbf{x}))^{1+\epsilon} \right)^{\frac{1}{1+\epsilon}} \left(\sum_{\mathbf{x};\,|\mathbf{x}|=n} (\hat{Y}(\mathbf{x}) - Y)^{2\frac{1+\epsilon}{\epsilon}} \right)^{\frac{\epsilon}{1+\epsilon}}.$$

But of course here we could also use $\epsilon = 1$ (Cauchy's inequality) or $\epsilon = \infty$. This would lead to designs where we minimize

$$\sum_{\mathbf{x};\,|\mathbf{x}|=n} (p(\mathbf{x}))^2 \quad \text{or} \quad \max_{\mathbf{x};\,|\mathbf{x}|=n} p(\mathbf{x})$$

given the restrictions. At least for symmetry reasons, it may seem more natural to use $\epsilon = 1$ than ϵ very close to 0. We call these designs the minsum$-p^2$ and the minmax$-p$ designs, respectively. They are more difficult to manage than the conditional Poisson design but at least for small populations and samples they can be handled.

5. Some second order designs

In this section we look at designs defined by 2nd order inclusion probabilities only. Although it is not completely true that in sampling higher order inclusion probabilities are irrelevant, we focus on the 2nd order ones here.

Hajek (1981) thought that it would be desirable to have a design with $c_{ij} = Cov(I_i, I_j)$ of the simple product form $c_{ij} = -c_i c_j,\ i \neq j$. Then there is a simple expression for the variance of the HT-estimator. Moreover a good approximation of the covariances of the adjusted conditional Poisson design is obtained. It is possible to solve these equations by iterative methods but the solution is inexplicit. The solution is of the form

$$c_{ij}^H = -\frac{\pi_i(1-\lambda_i)\pi_j(1-\lambda_j)}{\sum \pi_k(1-\lambda_k)},$$

where $\lambda_i \approx \pi_i$. Hajek was not able to show that there really is a sampling design with the derived covariances and the 2nd order inclusion probabilities $\pi_{ij} = c_{ij}^H + \pi_i\pi_j$. Nowadays at least for small populations one can use linear programming to find such designs: we should solve the linear equations $\sum_{\mathbf{x};\ |\mathbf{x}|=n} x_i x_j p(\mathbf{x}) = \pi_{ij}$ for $p(\mathbf{x})$ under nonnegativity restrictions. It is also possible to use a pf of the form $p(\mathbf{x}) = \prod_{i=1}^N \pi_i^{x_i}(1-\pi_i)^{1-x_i}\,Q(\mathbf{x})$, where Q is a quadratic form that has to be calculated (Lundqvist & Bondesson, 2005).

Hajek also derived his product form by maximizing the 'entropy' $\sum_{i,j;\,i\neq j} c_{ij}\log(-c_{ij})$. Bondesson *et al.* (2006) instead minimized the measure

$$\text{SSCorr} = \sum_{i,j;\,i\neq j} \rho_{ij}^2,$$

where $\rho_{ij} = Corr(I_i, I_j)$. The restrictions $\sum_{j=1}^N c_{ij} = 0$ together with Lagrange multiplier technique show that there is an explicit solution:

$$c_{ij}^{BTL} = -\pi_i(1-\pi_i)\pi_j(1-\pi_j)(\gamma_i+\gamma_j), \quad i \neq j,$$

where

$$\gamma_i = \frac{\frac{1}{d-2\pi_i(1-\pi_i)}}{1+\sum \frac{\pi_k(1-\pi_k)}{d-2\pi_k(1-\pi_k)}} \quad \text{with} \quad d = \sum \pi_k(1-\pi_k).$$

Most often these covariances are very close to c_{ij}^H. Of course, to minimize SSCorr gives in some sense as much pairwise independence as possible to the inclusion variables I_i. We also have, with $a_i = y_i/\pi_i$, $\tilde{\rho}_{ij} = -\rho_{ij}$, and $d_i = \pi_i(1-\pi_i)$, by Cauchy's inequality,

$$Var(\hat{Y}_{HT}) = Var(\sum a_i I_i) = \frac{1}{2}\sum_{i,j;\, i\neq j} \tilde{c}_{ij}(a_i - a_j)^2$$

$$= \frac{1}{2}\sum_{i,j;\, i\neq j} \tilde{\rho}_{ij}\sqrt{d_i d_j}\,(a_i - a_j)^2 \;\leq\; \frac{1}{2}\sqrt{\text{SSCorr}} \times \sqrt{\sum_{i,j;\, i\neq j} d_i d_j (a_i - a_j)^4}.$$

For a fixed population and fixed inclusion probabilities, the last factor above is constant. However, we can affect SSCorr and by minimizing it we get in a direct way some guarantee that $Var(\hat{Y}_{HT})$ becomes small.

There are several simple variants of the approach above. Instead of focusing on the correlation, we may focus on the covariance. Minimizing the sum of squared covariances under the appropriate restrictions, we get the solution:

$$c_{ij} = \frac{1}{N-2}\left(\frac{d}{N-1} - \pi_i(1-\pi_i) - \pi_j(1-\pi_j)\right), \quad i \neq j.$$

Often this is not a useful solution since the signs of the covariances may vary. They should preferably be nonpositive to give a stable Sen-Yates-Grundy variance estimator. Instead of Cauchy's inequality, we may use Hölder's inequality. In particular, we may use Hölder's inequality with the exponents $p = \infty$ and $q = 1$. This leads to second order minimax-designs.

We now turn to such designs and use first the covariance. Let $a_i = y_i/\pi_i$ and $\mathbf{a} = (a_1, a_2, \ldots, a_N)^T$. Set $\mathbf{1} = (1, 1, \ldots, 1)^T$. Then $\mathbf{a} = \bar{a}\,\mathbf{1} + \mathbf{b}$, where $\mathbf{b} = (a_1 - \bar{a}, \ldots, a_N - \bar{a})$ is orthogonal to $\mathbf{1}$. Now, since $\sum I_i = n$,

$$Var(\hat{Y}_{HT}) = Var(\mathbf{a}^T\mathbf{I}) = Var(\mathbf{b}^T\mathbf{I}) = \mathbf{b}^T\mathbf{\Sigma}\mathbf{b} \;\leq\; \lambda_{\max}||\mathbf{b}||^2 = \lambda_{\max}\sum_{i=1}^{N}(a_i - \bar{a})^2,$$

where $\lambda_{\max}$ is the maximal eigenvalue of $\mathbf{\Sigma}$. Now we could try to choose a covariance matrix $\mathbf{\Sigma}$ with given diagonal elements $d_i = \pi_i(1-\pi_i)$ and eigenvector $\mathbf{1}$ with eigenvalue $\lambda_1 = 0$ and such that its maximal eigenvalue is minimal. We should add the condition that $\tilde{c}_{ij} \geq 0$, $i \neq j$, where $\tilde{c}_{ij} = -c_{ij}$. This is a problem that in some cases can be solved.

We can alternatively describe the problem as follows. We have

$$Var(\mathbf{a}^T\mathbf{I}) = \frac{1}{2}\sum_{i,j;i\neq j}\tilde{c}_{ij}(a_i-a_j)^2 \leq \max_{i,j;\,i\neq j}\tilde{c}_{ij}\,\frac{1}{2}\sum_{i,j;\,i\neq j}(a_i-a_j)^2 = \max_{i,j;\,i\neq j}\tilde{c}_{ij}\times N\sum_{i=1}^{N}(a_i-\bar{a})^2.$$

Now we should try to find a covariance matrix with $\mathbf{1}$ as eigenvector with eigenvalue 0 and such that $\max_{i,j;\,i\neq j}\tilde{c}_{ij}$ is minimal. Additionally we should require that $\tilde{c}_{ij}\geq 0$, $i\neq j$, to get a stable variance estimator. This is a problem that in small cases can be solved by linear programming for determination of the appropriate $\tilde{c}_{ij}$. Often, but not always the solution is of the form that the matrix has all its elements in the row with the largest $d_i = \pi_i(1-\pi_i)$ equal and if that row is the first row then equal to $-d_1/(N-1)$ (since all rows sums are 0). We then have

$$Var(\mathbf{a}^T\mathbf{I}) \leq \frac{Nd_1}{N-1}\sum_{i=1}^{N}(a_i-\bar{a})^2.$$

In fact, in this case the inequality becomes an equality for $\mathbf{a}\propto(N-1,\,-1,\,-1,\ldots,\,-1)$ which is an eigenvector with eigenvalue $Nd_1/(N-1)$. Hence the inequality is sharp in some sense.

We can also use minimax designs w.r.t. the correlation. We have, with $d_i=\pi_i(1-\pi_i)$,

$$Var(\mathbf{a}^T\mathbf{I}) \leq \max_{i,j;\,i\neq j}\tilde{\rho}_{ij}\,\frac{1}{2}\sum_{i,j;\,i\neq j}(a_i-a_j)^2\sqrt{d_id_j} = \max_{i,j;\,i\neq j}\tilde{\rho}_{ij}\,(\sum_{i=1}^{N}\sqrt{d_i})^2\sum_{i=1}^{N}(a_i-\bar{\bar{a}})^2p_i',$$

where $p_i'\propto\sqrt{d_i}$ with $\sum p_i'=1$ and $\bar{\bar{a}}$ is a weighted mean. We may therefore try to minimize $\max\tilde{\rho}_{ij}$ under simple restrictions (see below). An alternative, but not equivalent approach, is to set $b_i=a_i\sqrt{d_i}$ and then use the inequality

$$Var(\mathbf{a}^T\mathbf{I}) = \mathbf{b}^T\mathbf{R}\mathbf{b}\ \leq\ \lambda_{\max}||\mathbf{b}||^2$$

for the correlation matrix $\mathbf{R}$. The maximal eigenvalue of $\mathbf{R}$ should then be minimized under the restrictions that $(\sqrt{d_1},\sqrt{d_2},\ldots,\sqrt{d_N})^T$ is an eigenvector of $\mathbf{R}$ with eigenvalue 0 and $\tilde{\rho}_{ij}\geq 0$, $i\neq j$.

6. Example: The TBM population

Here we return to the TBM-population in section 1 with $N=6, n=3$, and $\pi_1=\pi_2=\pi_3=\frac{1}{3}, \pi_4=\pi_5=\pi_6=\frac{2}{3}$. The population is simple but it illustrates many things in a good way. There are 20 possible samples of size $n=3$ but only 4 with distinct probabilities: $\{1,2,3\},\{1,2,4\},\{1,4,5\},\{4,5,6\}$. Each of the samples 2 and 3 has 8 variants. We set

$p_1 = \pi_{123}$, $p_2 = \pi_{124}$, $p_3 = \pi_{145}$, $p_4 = \pi_{456}$. Then $p_1 + 6p_2 + 3p_3 = 1/3$ and $3p_2 + 6p_3 + p_4 = 2/3$ implying that $p_1 + 9p_2 + 9p_3 + p_4 = 1$. It is easy to experiment with this population since there are only a few parameters to vary, e.g. p_1 and p_2. In Table 1 different characteristics have been calculated for all the designs considered.

Table 1: The TBM-population; 2nd order inclusion probabilities, sample probabilities, and some other characteristics for seven different sampling designs.

	MinSSCorr	Sampford	Pareto(adj)	CP(adj)	Hellinger	minsum-p^2	minmax-p
π_{12}	0.06667	0.06918	0.06973	0.07081	0.07170	0.04762	0.03030
π_{14}	0.17778	0.17610	0.17574	0.17501	0.17442	0.19048	0.20202
π_{45}	0.400	0.40252	0.40306	0.40415	0.40503	0.38095	0.36364
p_1	0	0.00629	0.00647	0.00686	0.00730	0	0
p_2	0.02222	0.02096	0.02109	0.02132	0.02147	0.01587	0.01010
p_3	0.06667	0.06709	0.06678	0.06619	0.06574	0.07937	0.09091
p_4	0.20	0.20126	0.20272	0.20556	0.20780	0.14286	0.09091
SSCorr	**1.2**	1.2025	1.2037	1.2071	1.2103	1.3469	1.7355
$\max \tilde{\rho}_{ij}$	**0.20**	0.2075	0.2092	0.2125	0.2151	0.2857	0.3636
Entropy	2.7080	2.7150	2.7151	**2.7152**	2.7151	2.6796	2.5976

There are several solutions, $\mathbf{p} = (p_1, p_2, p_3, p_4)$ in the MinSSCorr case. Above an extreme solution is given. All the different designs considered in section 5 lead for this simple population to the MinSSCorr solution in the second column. Since $d_i = \pi_i(1 - \pi_i) \equiv 2/9$, it does not even matter whether we consider the correlation or the covariance. Because of the very symmetric character of this design, $\rho_{ij} \equiv -0.2$ for $i \neq j$, one may think that this is a very good solution in this simple case. Its entropy can be increased to 2.7142 by the choice of a less extreme MinSSCorr design among the possible variants. The Sampford, the Pareto-adjusted, the CP-adjusted, and the Hellinger designs do not agree with the MinSSCorr design although they are very close to it. These latter four designs are pairwise very equal in this example with {Sampford, Pareto(adj)} and {CP(adj), Hellinger} as the pairs. This is a relation that is true in general as found by Lundqvist (2006). The minsum-p^2 and the minmax-p designs seem a bit extreme compared to the other designs. They have much higher SSCorr and lower entropy than the other five designs. They have also very low values of π_{12} which leads to less stable variance estimators.

It was mentioned in the introduction that for the simple superpopulation model $a_i = \alpha + \epsilon_i$ with i.i.d. ϵ_is, the expected (design) variance of the HT-estimator is the same for all πps designs with the given inclusion probabilities. We may then also look at the expected design variance of the Sen-Yates-Grundy estimator of the variance of the HT-estimator. Assuming that the fourth central moment of ϵ_i equals $3Var(\epsilon_i) = 3\sigma^4$, we got the following expected values.

Table 2: Expected values of the design variance of the SYG variance estimator under a simple superpopulation model.

MinSSCorr	Sampford	Pareto(adj)	CP(adj)	'Hellinger'
$1.844\sigma^4$	$1.816\sigma^4$	$1.806\sigma^4$	$1.789\sigma^4$	$1.778\sigma^4$

Thus the Hellinger design gives a slightly more stable variance estimator than the other included designs. On the other hand, by setting $p_1 = 0.0813$, $p_2 = 0$, $p_3 = 0.0840$, and $p_4 = 0.1626$, we get the smallest possible expected value: $1.650\sigma^4$. Of course, it is well known that the requirement of a small variance for an estimator is in conflict with the requirement of a stable variance estimator.

7. Discussion

It is a bit annoying that it is not really possible to single out a best fixed size πps sampling design. There is no doubt about that at present the (adjusted) Pareto design is the best one to select a sample easily. On the other hand the adjusted conditional Poisson design has a very attractive maximum entropy property. The Sampford design has a simple and nice pf and is very attractive from that point of view. Fortunately these three designs are close to each other. Brewer (2002) classifies the Pareto design as a high entropy design. The minimax designs considered in section 5 are at present not very practical but focus more directly on making the variance small than the other designs. Gabler (1990) presents many results on strongly related minimax designs.

Finally, it is appropriate to add that if there is relevant auxiliary information, many other designs are possible, as e.g. systematic πps sampling designs, and may be better than the ones considered here.

References

Aires, N. (2000). Comparisons between conditional Poisson sampling and Pareto πps sampling designs. *J. Statist. Plann. Inference* **88**, 133-147.

Bondesson, L., Traat, I., and Lundqvist, A. (2006). Pareto sampling versus Sampford and conditional Poisson sampling. *Scand. J. Statist.* **33**, to appear.

Brewer, K.R.W. & Hanif, M. (1983). *Sampling with unequal probabilities.* Lecture Notes in Statistics, No. 15. Springer-Verlag, New York.

Brewer, K.R.W (2002). Combined survey sampling inference: weighing of Basu's elephant. Hodder Arnold, London.

Chen, S.X. & Liu, J.S. (1997). Statistical applications of the Poisson-Binomial and conditional Bernoulli distributions. *Statistica Sinica* **7**, 875-892.

Gabler, S (1990). *Minimax solutions in sampling from finite populations.* Lecture Notes in Statistics **64**. Springer-Verlag, Berlin, Heidelberg.

Grafström, A. (2005). Comparisons of methods to generate conditional Poisson samples and Sampford samples. Master's thesis, Department of mathematics and mathematical statistics, Umeå university.

Hajek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Ann. Math. Statist.* **35**, 1491-1523.

Hajek, J. (1981). *Sampling from a finite population.* Marcel Dekker, New York.

Lundqvist, A. (2006). Comparing distances between some distributions that occur in sampling. Manuscript.

Lundqvist, A. & Bondesson, L. (2005). On sampling with desired inclusion probabilities of first and second order. Research report 3 in mathematical statistics, Umeå university, Sweden.

Ohlsson, E. (1990). Sequential Poisson sampling from a business register and its application to the Swedish consumer price index. Statistics Sweden R&D Reports 1990:6.

Öhlund, A. (1999). Comparisons of different methods to generate Bernoulli distributed random numbers given their sum. Master's thesis, Department of mathematical statistics, Umeå university, Sweden (in Swedish).

Rosén, B. (1997a). Asymptotic theory for order sampling. *J. Statist. Plann. Inference* **62**, 135-158.

Rosén, B. (1997b). On sampling with probability proportional to size. *J. Statist. Plann. Inference* **62**, 159-191.

Saavedra, P. (1995). Fixed sample size PPS approximations with a permanent random number. 1995 Joint Statistical Meetings American Statistical Association, Orlando, Florida.

Sampford, M.R. (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika* **54**, 499-513.

Tillé, Y. (2005). *Sampling algorithms.* Technical Report, Neuchâtel, Switzerland.

Traat, I., Bondesson, L. & Meister, K. (2004). Sampling design and sample selection through distribution theory. *J. Statist. Plann. Inference* **123**, 395-413.