Poisson Mixture (Pomix) Sampling for Business Sample Coordination

Ismo Teikari
Statistics Finland, statistical methodology R&D
FIN-00022 Statistics Finland
Finland
ismo.teikari@stat.fi

1. Controlling the response burden

Users of business survey data have an insatiable demand for detail, which is only limited by cost and the response burden. The response burden can be relieved by using administrative data as much as possible, by profiling and by good survey programme designing. The information is still needed, however, so that some response burden must exist. Three basic features of a useful definition of this response burden are listed by Sunter (1977):

1. Each survey questionnaire (j=1,2,..,J) is assessed for its *response 'load'*.
2. Each business (k=1,2, ..,N) is assessed for its "*response obligation*",
3. The response burden allocation system seeks to ensure that the assessed response obligation of each business is exceeded only rarely by its actual response burden

Let β_j denote the response load imposed by the j^{th} survey in a survey program, and π_{kj} the probability of inclusion of the k^{th} business in the j^{th} survey. It is now possible to derive an equation for the expected response burden for the business k

$$RB_k = \sum_{j=1}^{J} \pi_{kj} \beta_j \quad , \qquad (1)$$

which should not exceed the response obligation of this business.

The formula (1) shows that there are two factors that create a response burden: the response load β_j, which could be minimized by the use of administrative data, profiling and a proper survey programme, and π_{kj} , which makes the actual response burden a random variable. This means that the response burden may be distributed unevenly among businesses, many of which suffer in this system which is fair only in that the burden occurs by chance. Thus minimizing the response burden is not enough; steps should also be taken to see that it is distributed as evenly as possible. It is possible to do this by coordinating samples. Coordination can have one of two contrasting aims. Negative coordination aims to achieve the smallest overlap between samples, and so to avoid the situation in which a respondent is burdened with two successive questionnaires, while positive coordination aims to improve the quality of the survey by including successive responses from as many units as possible. Fortunately there is a compromise: the rotation of samples

2. Poisson Mixture sampling

Poisson Mixture (PoMix) sampling was introduced by Kröger , Särndal and Teikari in 1999 for coordinating business samples. PoMix sampling is based on the use of permanent random numbers (PRN), r_k, which are assigned to each population units. PRN is a realisation of the uniform distribution *Unif(0,1)* and units keeps their PRN's until they quit the frame. A size measure Q_k is calculated for every unit. The largest units are assigned to a take-all stratum U^{TA} so that $0 \le Q_k < 1$. Q is a linear transformation of well known size measure A_k for Poisson πps sampling

$$Q_k = \frac{1 - B / f^R}{1 - B} A_k$$

where f^R is the sample fraction n^R/N^R in U^R and B ($0 \le B \le f^R$) is so-called Bernoulli width. The sample consist of two parts (Bernoulli sampling and Poisson sampling) so that the inclusion probabilities get the form

$$\pi_k = \begin{cases} B + Q_k(1-B), & if\ k \in U^R \\ 1 & if\ k \in U^{TA} \end{cases}$$

The second order inclusion probabilities are easy to compute because r_k and r_l are independent random numbers. We have

$$\pi_{kl} = \pi_k \pi_l = [B + Q_k(1-B)][B + Q_l(1-B)],\ \text{for } k \neq l \in U^R .$$

If we put $B=0$ we get Poisson πps sampling and if we put $B=f^R = n/N$ we get Bernoulli sampling.

In MC simulation studies it was interestingly found that the minimum variance is not obtained for Poisson πps with $B = 0$, as one might expect, but rather for a value of B apparently somewhere between 0.02 and 0.03. The improvements achieved in the case $B = 0.02$ relative to $B = 0$ are substantial close to 50% for estimators using auxiliary information in the estimation stage. One possible explanation for this surprising result is that when B is close to zero, the units with the lowest x values, when selected, will have unduly large weights, which implies high variability. This can be avoided by choosing a B that is a long way from zero.

3. Order Poisson Mixture sampling

To ensure a fixed sample size, order PoMix sampling was introduced (Kröger et al. 2000) on the basis of the method described by Ohlsson (1990, 1996, 1998). When the value of the Bernoulli part was set at zero, we obtained Ohlsson's sequential Poisson sampling. when the size of the Bernoulli part was set at exactly f^R, we did not get Bernoulli sampling but Sequential Simple random sampling because the sample size is fixed.

It was found that Order PoMix gave the same improvement in variances as in the case of random size PoMix sampling. Using some part of the Bernoulli routine the variance was reduced by about 50 percent. It was also found that Sequential Poisson sampling improved the variances compared with Poisson πps sampling only when an ordinary HT estimator without auxiliary information was used.

REFERENCES

Kröger, H., Särndal, C-E., and Teikari, I. (1999). Poisson Mixture Sampling: A Family of designs for Coordinated selection Using permanent Random Numbers. *Survey Methodology*, 25, pp. 3-11

Kröger, H., Särndal, C-E., and Teikari, I. (2000). Poisson Mixture Sampling Combined with Order sampling: a Novel Use of the Permanent Random Number technique. Manuscript submitted in December 2000 for publication in *Journal of Official Statistics*.

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

Ohlsson, E. (1995). Sequential Poisson Sampling. *Institute of Actuarial Mathematics and Mathematical Statistics,* Stockholm University, Report No. 182, June 1995.

Ohlsson, E. (1995). Coordination of Samples Using Permanent Random Numbers. *Business survey Methods.* Ed. B.G. Cox, D.A. Binder, B.N. Chinnappa, A. Christianson, M.J. Colledge, and P.S. Kott. New York 1995.

RESUME

L'accroissement de la demande d'enquêtes dans les entreprises augmente la charge de réponse des entreprises. Les autorités de la statistique ont relevé le défi de diminuer cette charge. Cette diminution ne suffira toutefois pas, car les enquêtes ne pourront pas être complètement éliminées. La charge qui reste devra être répartie équitablement entre les entreprises interrogées. Poisson Mixture Sampling est une méthode d'échantillonnage (design) destinée à la coordination des échantillons des entreprises.