Standard Error Estimation for Distributional Measure from Complex Surveys

Yves G. BERGER
University of Southampton,
Department of Social Statistics,
Southampton, SO17 1BJ
United Kingdom
ygb@soton.ac.uk

As far as variance estimation of complex non-linear statistics is concerned, two methods are usually used: the jackknife method (Tukey 1958; Durbin 1959) and the linearisation method (Kovaèeviæ and Binder, 1997; Deville, 1999).

Although most finite population parameters of interest can be defined in terms of a smooth function of population totals, or as the solution of a population estimating equations, there remains situations where the definition of the parameter of interest is so complex that application of linearisation methods is difficult and not widely used in practice (Kokic 1998). In such cases, we can use Jackknife method that is simple to implement, but is typically numerically intensive. In this contributed paper, we consider the Jackknife variance estimation and we show that the Jackknife and the linearisation method may be viewed as an estimation of an influence function.

Suppose that the sampling design is a balanced on a series of design variables. This often the case in single stage sampling. For example, stratified sampling design is balanced on stratification variables. Consider $F(y)$ being the distribution function of the survey variable of interest. Consider the sample estimator $F_n(y)$ of $F(y)$. Consider a functional $T(F)$ estimated by $T(F_n)$. We assume this functional to be continuously Gâteaux differentiable. We are concerned with the estimation of variance of $T(F_n)$.

A variance estimator of $T(F_n)$ is given by the variance of the Horvitz-Thompson total of the influence function (Deville, 1999). In this paper, we propose Jackknife method to estimate the influence function. This method involves the computation of pseudo-values. We show that variance estimation needs the pseudo-values and the joint inclusion probabilities when we consider a design-based approach. Therefore, we should use these probabilities, when we compute the variance among the jackknife pseudo-values. By ignoring these probabilities, we introduce a bias in the variance estimation. These probabilities should remove the bias of the jackknife variance estimator. We show that if we assume simple random sampling, the variance estimator proposed is equivalent to the usual jackknife estimator.

In order to approximate the variance by a quantity free of second-order inclusion probabilities, we propose a method based on residuals (Deville 1999, Berger et al. 2001; Berger 2001) that can be computed with any usual statistical packages. These residuals involved regression between the pseudo-values and the design variables. Systematic sampling will be also considered (Berger 2001).

The Economic and Social Research Council (United Kingdom) supports this research (grant R000.22.3397).

REFERENCE

Berger, Y.G., Tirari, M.E.H, and Tillé, Y. (2001), Optimal Regression Estimation with Balanced Sampling. Southampton working paper (Submitted for publication).

Berger, Y.G. (2001). Sampling Variance for ps Systematic Sampling", Southampton working paper (Submitted for publication).

Deville, J.C. (1999). Variance Estimation for Complex Statistics and Estimators: Linearization and Residual Techniques. *Survey Methodology*. **25**, 193-203.

Durbin, J. (1959). A Note of the Application on Quenouille's Method of Bias Reduction to the Estimation of Ratios. *Biometrika.* **46**, 477-480.

Kokic,P.N. (1998). Estimating the Sampling Variance of the UK Index of Production. *Journal of Official Statisitics*. **14**, 163-179.

Kovaèeviæ, M.S. and Binder, D.A. (1997). Variance Estimation for Measure of Income Inequality and Polarization - The Estimating Equations Approach. Journal of Official Statistics, **13**, 41-58.

Tukey, J.W. (1958). Bias and Confidence in Not-quite large Sample (Abstract). Annals of Mathematical Statisitics, **29**, 614.

RÉSUMÉ EN FRANÇAIS

Dans cette communication, nous considérons l'estimation de variance pour des statistiques complexes par la méthode d'Eustache. Nous montrons que le problème d'estimation peut se ramener à une estimation d'une fonction d'inflence. L'estimateur considéré nécessite les pseudovaleurs de la méthode d'Eustache et les probabilités d'inclusion d'ordre deux. Cet estimateur se ramène à l'estimateur d'Eustache classique lorsqu'un échantillon aléatoire simple est sélectionné. Nous proposons une méthode basée sur des résidus pour approximer notre estimateur par une quantité qui ne nécessite pas de probabilités d'inclusion d'ordre deux.