Statistical Inference for the Indicators of Inequality and Poverty

Y. Tillé[1]

[1] University of Neuchâtel, Institut de Statistique, Pierre à Mazel 7, 2000 Neuchâtel, Switzerland

Keywords: Gini Index, Indicators, Linearization, Quintile share ratio, Variance estimation, Conference on Indicators and Survey Methodology 2010.

1 Introduction

The Laeken indicators are a set of statistical indicators on poverty and social exclusion, established by the European Council in 2001. These indicators include the *At-risk-of-poverty rate* (ARPR), the *Income Quintile Share Ratio* (QSR) and the *Gini coefficient (G)*. These indicators are complex because they are non-linear functions of the income. Moreover, the ARPR and the QSR strongly depends on the queues of the income distribution. These indicators are easy to interpret but from the point of view of inference seem to be an unhappy choice. Indeed, the queues of income distributions are particularly difficult to estimate. An indicator that strongly depends on the queues has thus more chances to be unstable. The construction of confidence intervals for these indicators must thus take into account several technical problems: the non-linearity of the indicators, the use of a complex sampling designs, the asymmetry of the distributions of the indicators, the influence of the queues of the distributions.

2 Inference for survey sampling

Suppose that, in a population U of size N, we want to estimate a function θ of an interest variable y that takes value y_k on unit $k \in U$. In order to estimate θ, a sample S is selected by means of a sampling design $p(s) = Pr(S = s)$ that have first order inclusion probabilities π_k and joint inclusion probabilities $\pi_{k\ell}$. If θ is a function of totals, the interest function θ can be estimated by substitution i.e. by estimating the totals that are the arguments of θ. Two methods can be used to construct a confidence interval: linearization or resampling. Linearization is in general recommended because resampling methods are difficult to implement in complex designs. In fact, there does not exist a general method to resample from a sample selected by means of a complex sampling design.

3 Estimation of the indicators

The Gini (1912, 1914, 1921) index was propose almost one century ago, but the inference on this indicator is relatively recent (see Sandstrom et al., 1985, 1988; Dixon et al., 1987). Deville (1996, 1999) was probably the first to linearize the Gini Index. We will show that a simple result can allow deriving quickly the linearized variable.

Several authors (Ogwang, 2000, 2004, 2006; Giles, 2004, 2006) have pointed out that the Gini index is proportional to a parameter that can be written as a regression coefficient of the income y_k by the corresponding rank k. Since the inference on the regression coefficient is trivial, the inference on the regression coefficient directly follows. We will show that this reasoning is a deep methodological error and leads to an overestimation of the variance.

The quintile share ratio can be also be linearized. Nevertheless, its definition is far from being univocal because it depends on two quantiles that can be defined in according to at least 9 different ways (see Hyndman and Fan, 1996). The linerarization of a quantile depends of the density function

of the interest variable, which is also problematic when the sample comes from a finite population. Osier (2006) has linearized the QRS in order to estimate its variance. Unfortunately, this estimator of variance depends on estimated densities. Langel and Tillé (2009) have however showed that it is possible to lead statistical inference without estimating the density.

4 Recent development

The aim of our talk is to present the recent developments for leading appropriate statistical inference on indicators and to discuss the practical problems. The main questions are: How to estimate complex nonlinear parameters? How to derive their linearized variables? Why, for the Gini index, is the inference based on the regression coefficient false? What is the influence of the queues of the distributions? How to take into account the asymmetry of the distribution of the estimates? Which new indicators could be used to efficiently measure inequalities?

References

Deville, J.C. (1996). Estimation de la variance du coefficient de Gini estimé par sondage. *Actes des journées de Méthodologie Statistique, Insee*, 69-70-71:269–288.

Deville, J.C. (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology*, 25:193–204.

Dixon, P. M., Weiner, J., Mitchell-Olds, T., and Woodley, R. (1987). Bootstrapping the Gini coefficient of inequality. *Ecology*, 68:1548–1551.

Giles, D. E. A. (2004). Calculating a standard error for the Gini coefficient: some further results. *Oxford Bulletin of Economics and Statistics*, 66:425–433.

Giles, D. E. A. (2006). A cautionary note on estimating the standard error of the Gini index of inequality: comment. *Oxford Bulletin of Economics and Statistics,*, 68:395–396.

Gini, C. (1912). *Variabilit'a e Mutabilit'a*. Bologna: Tipografia di Paolo Cuppin.

Gini, C. (1914). Sulla misura della concertrazione e della variabilit' a dei caratteri. *Atti del R. Istituto Veneto di SS. LL. AA*, 73:1203–1248.

Gini, C. (1921). Measurement of inequality and incomes. *The Economic Journal*, 31:124–126.

Hyndman, R. J. and Fan, Y. (1996). Sample quantiles in statistical packages. *American Statistician*, 50:361–365.

Langel, M. and Tillé, Y. (2009). Statistical inference for the quintile share ratio. *Unpublished*.

Ogwang, T. (2000). A convenient method of computing the Gini index and its standard error. *Oxford Bulletin of Economics and Statistics*, 62:123–129.

Ogwang, T. (2004). Calculating a standard error for the Gini coefficient: some furtherresults: reply. *Oxford Bulletin of Economics and Statistics*, 66:435–437.

Ogwang, T. (2006). A cautionary note on estimating the standard error of the Gini index of inequality: comment. *Oxford Bulletin of Economics and Statistics*, 68:391–393.

Osier, G. (2006). Variance estimation: the linearization approach applied by Eurostat to the 2004 SILC operation. Technical report, Eurostat and Statistics Finland Methodological Workshop on EU-SILC, Helsinki, 7-8 November 2006.

Sandstrom, A., Wretman, J. H., and Walden, B. (1985). Variance estimators of the gini coefficient: Simple random sampling. *Metron*, 43:41–70.

Sandstrom, A., Wretman, J. H., and Walden, B. (1988). Variance estimators of the gini coefficient: Probability sampling. *Journal of Business and Economic Statistics*, 6:113–120.