A prediction approach to sampling design

Li-Chun Zhang and Ib Thomsen
Statistics Norway

In standard approach to sampling, whether model-based or not, the main aim is to estimate one or several finite population totals, or some predefined sub-totals. Within the model-based prediction approach the implications on sampling design can be extreme as e.g. in the case of purposive selection for ratio regression populations. This is mainly because a purposive sample is unlikely to be suitable for other uses than the inference of population totals (or means). For instance, we may want to use the data for micro simulations in an econometric model. Or we may need the data for small area (or domain) estimation.

An alternative point of departure is individual prediction. Formally, consider the class of functions $\sum_{i \in U} \lambda_i y_i$, where $U = \{1,...,N\}$ denotes the finite population, and the λ_i's are fixed constants such that $\sum_{i \in U} |\lambda_i| = 1$, and y_i is the variable of interest for the *i*th unit. The population mean is given by setting $\lambda_i = 1/N$. Whereas in the prediction of any individual we set $\lambda_i = 1$ for that chosen unit and let $\lambda_i = 0$ otherwise, which is in some sense the linear function within the above class that differs most from the one implied by the prediction of population total. Moreover, for general database-like uses of the survey data, a natural criterion for sampling design is to make the unconditional individual prediction MSE equal for all the units in the population.

We derive the equal prediction designs for a number of populations. The models we consider here are primarily suitable for continuous variables. It turns out that balancing between equal prediction of the individuals and optimal prediction of the population totals provides a useful model-based approach to sampling design. In this talk we will concentrate on some implications on small area estimation by this prediction approach to sampling design.