

Some reflections on Bayesian inference in Official Statistics

Survey sampling and, more generally, Official Statistics are experiencing an important renovation time. On one hand, there is the need to exploit the huge information potentiality that the digital revolution made available in terms of data. On the other hand, this process occurred simultaneously with a progressive deterioration of the quality of classical sample surveys, due to a decreasing willingness to participate and an increasing rate of missing responses.

The switch from survey-based inference to a hybrid system involving register-based information has made more stringent the debate and the possible resolution of the design-based versus model-based approaches controversy. In this new framework, the use of statistical models seems unavoidable and it is today a relevant part of the official statistician toolkit. Models are important in several different contexts, from Small area estimation to non-sampling error adjustment, but they are also crucial for correcting bias due to over and under-coverage of administrative data, in order to prevent potential selection bias, and to deal with different definitions and/or errors in the measurement process of the administrative sources.

The progressive shift from a design-based to a model-based approach in terms of super-population is a matter of fact in the practice of the National Statistical Institutes. However, the introduction of Bayesian ideas in official statistics still encounters difficulties and resistance. In this work, we attempt a non-systematic review of the Bayesian development in this area and try to highlight the extra benefit that a Bayesian approach might provide. Our general conclusion is that, while the general picture is today clear and most of the basic topics of survey sampling can be easily rephrased and tackled from a Bayesian perspective, much work is still necessary for the availability of a ready-to-use platform of Bayesian survey sampling in the presence of complex sampling design, non-ignorable missing data patterns, and large datasets.