Bayesian Ideas in Survey Sampling: The Legacy of Basu

Marco Di Zio
Istituto Nazionale di Statistica, Roma, Italy
Brunero Liseo
MEMOTEF, Sapienza Università di Roma, Roma, Italy
Maria Giovanna Ranalli
Dept. of Political Science, University of Perugia, Perugia, Italy

Abstract

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 undercoverage 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.

AMS (2000) Primary 62F15; Secondary 62D05.
Keywords and phrases. Calibrated Bayes, Model-based survey sampling, Polya posterior.

1   Introduction

The role of Debabrata Basu in the critical development of survey sampling could be hardly overstated. His paper on the foundations of the subject (Basu, 1971) is a landmark which, for the first time - at least with such clarity - unveiled the irreconcilability between design-based inference and the likelihood principle. His criticism against the use of the Horvitz-Thompson estimator in order to guarantee unbiasedness of the estimates, extremely and colorfully expressed with the elephant example (see e.g. Welsh, 2010), has caused an incredibly vivid and interesting debate among statisticians; as a consequence, survey sampling has experienced, in the last decades, several attempts of radical restructuring of the foundations.

The aim of this paper is to explore and discuss the potential role of Bayesian ideas and techniques in modern survey sampling. The paper is structured as follows: § 2 discusses the theoretical conflict between design-based methods and the likelihood principle and highlights the role that a Bayesian approach could have. § 3 goes beyond the basic framework discussed in § 2 and discusses the ineluctability of a shift towards model-based techniques in modern survey statistics. § 4 reviews the most prominent and promising ideas for a Bayesian theory of inference for finite populations, namely

- the Polya posterior approach, proposed in a series of papers by Glen Meeden and collaborators (see for example Ghosh and Meeden, 1997; Strief and Meeden, 2013).

- the Calibrated Bayesian approach, popularized in several papers by Roderick Little (see for example Little 2006; Little 2011; Little 2022).

Then, § 5 considers a real case study, which we consider paradigmatic of the issues and the open problems discussed above. Finally, § 6 provides some concluding remarks.

2   The conflict

2.1   Basu's criticism of design-based methods   To fix notation and ideas, we consider the simplest situation where a random sampling without replacement is drawn from a population P with N identified units. Here N is assumed to be known and units are identified through their labels, say $\{1, 2, \ldots, N\}$. We draw a sample s of size n and assume that the randomization scheme assigns a probability $p(s)$ to this specific sample. The quantity

of interest is the vector of values of a variable Y observed on the entire population, say $Y_P = (y_1, y_2, \ldots, y_N)$, or a specific function of it, say $\tau = f(Y_P)$. However, Y_P is observed only on the units belonging to the sample s. Let Y_s be the set of observed values: then the goal is to make inferential statements on the values $Y_{P \setminus s}$.

A design-based technique for producing an unbiased estimator of τ is based on the Horvitz-Thompson strategy which suggests using an empirical version of τ, where the n observations are weighted with the inverse of their corresponding probabilities to be included in the sample. Unbiasedness is calculated with respect to the randomization scheme (the sampling design); this is usually inspired by, but not necessarily related to, Y_P, which is considered fixed but unknown. Then Basu formally proved that the likelihood function for Y_P is flat, i.e. it is equal to a positive constant, for all values compatible with the observed Y_s, and it is zero otherwise.

Many scientists have interpreted this result as a proof of a general inadequacy of the likelihood function - and the likelihood principle - as the main tool of the inferential process in this framework. On the other hand, Basu and other Bayesian statisticians believe that this specific context offers an example where the likelihood function provides obvious but correct results, and this clarifies the insufficient level of modeling of the design-based methods.

From a historical perspective, the first attempt to overcome the difficulties of using a likelihood function in a finite population sampling framework can be considered the *scale load* approach described in Hartley and Rao (1968), where the support of the quantity of interest Y is discretized into T different values, having frequencies $N_1, N_2, \ldots, N_T$ at the population level and $n_1, n_2, \ldots n_T$ at a sample level. Here the choice of T is not crucial and the Authors consider, as a likelihood function for the unknown vector $(N_1, \ldots, N_T)$, the hypergeometric distribution associated with the observed sample:

$$L(N_1, \ldots, N_T) = \prod_{t=1}^{T} \binom{N_t}{n_t} / \binom{N}{n}. \tag{1}$$

The Authors described how to make consistent inferences on functions of $(N_1, \ldots, N_T)$, through maximization of the likelihood or combining it with a suitable prior distribution. The scale-load approach can be considered prodromic to the more general notion of *empirical likelihood*, developed after Owen (1988) and reconsidered, in the context of finite population sampling in several papers: see for example Zhong and Rao (2000) and Berger (2018).

2.2 Design-based, likelihood and Bayes In the last decades, the scientific debate around survey sampling has been quite vivid and it has dealt mainly with the contrast between design-based and model-based approaches. This is not the place to recall the details of the conflict, and we will only highlight the main points. A vivid and deep comparative analysis of different inferential approaches in finite population sampling can be found in Beaumont and Haziza (2022).

In survey sampling, many different issues must be considered in the optimization of the sampling plan. In spite of that, in a design-based philosophy, the construction of point estimators and confidence intervals is based on the randomization process and it is not always directly related to the quantity of interest Y_P. In other terms, the design-based analysis leads to conclusions about the finite population quantity totally free of assumptions about the structure of the variation in the population (Cox 2006).

A pure likelihood analysis of such a problem is bound to provide trivial conclusions (Basu 1971, Godambe 1966): all the configurations of the parameter Y_P compatible with Y_s receive the same support from the likelihood; the likelihood function itself is not able to introduce any sort of similarities/dissimilarities among the units in the sample and those not observed. This is done surreptitiously in a design-based approach by silently assuming a sort of exchangeability among the units. The Bayesian road seems at least clearer. In order to provide inference on the vector Y_P, which is now treated as a random vector, one needs to introduce a prior distribution in the game, and the prior is precisely the instrument that formalizes potential external information about the mutual similarities among units.

More in detail, and following (Little 2022), let us denote by $S_P = (S_1, \ldots, S_N)$ the vector of selection indexes for the N units of the population, that is

$$S_i = \begin{cases} 1 & i\text{-th individual is in the sample} \\ 0 & \text{otherwise.} \end{cases}$$

If Z_P denotes a vector including all other design-related variables, the goal is to make inference on some quantity $Q(Y_P)$, using possible covariate information Z_P. A model-based inference approach will be based on the joint distribution

$$p_{s,y;z}(s, y; z, \theta, \psi) = p_{y;z}(y; z, \theta) p_{s|y;z}(s|y; z, \psi), \tag{2}$$

where θ is a vector of parameters directly related to the variable of interest y and ψ only refers to the mechanism of inclusion.

Bayesian inference in this context requires the introduction of a prior distribution $p_{y;z}(y;z)$ for the population values. Inferences are then based on the posterior predictive distribution of the non-sampled values $Y_{P\setminus s}$ of Y_P, given the sampled values. The prior distribution is often specified in a hierarchical way, where a parametric model $p_{y;z}(y;z,\theta)$ indexed by parameters θ, in practice the one appearing in Eq. 2, is combined with a prior distribution $p(\theta;z)$. Then,[1] if we assume that - for the sake of simplicity - the sampling mechanism is ignorable,

$$p(y;z) = \int p(y|\theta,z)p(\theta;z)d\theta. \tag{3}$$

The posterior predictive distribution of the non-sampled values $Y_{P\setminus s}$ is then

$$P(Y_{P\setminus s}|Y_s,z) = \int p(Y_{P\setminus s}|Y_s;z,\theta)p(\theta|Y_s;z)d\theta \tag{4}$$

where $p(\theta|Y_s,z)$, is the posterior distribution of the hyper-parameters θ.

The issues of deep irreconcilability among different inferential paradigms, in general, do not cause huge differences in practice. In simple situations, if the units are approximately exchangeable, the use of the Horvitz-Thompson estimator would provide the same numerical answer that one could obtain with a weakly informative prior, or even using a Bayesian method *without* a prior, as in Ghosh and Meeden (1997) and Strief and Meeden (2013).

The debate between design-based and model-based approaches is basically internal to the non-Bayesian world and it appeared - and has increasingly become - relevant because of the more and more complex problems faced by modern survey sampling.

Much of the theory of model-based methodology has been developed in a non-Bayesian fashion, starting from the seminal and thought-provoking papers of Royall (1970, 1976) where the role of the likelihood function, when properly defined and interpreted, is deemed central also in a finite population framework and particularly for predictive purposes, provided that a superpopulation perspective is considered. The prediction approach is extensively discussed and supported in Valliant et al. (2000), and some Bayesian ver-

---

[1] A word on notation: we use different notation $p(u|v)$ and $p(u;v)$ according to whether at least some component of v must be considered a random quantity or not, respectively.

sions of super-population modeling can also be found in Zacks (2002) and Bolfarine and Zacks (1992).

The most prominent cases where a design-based approach has provided unsatisfactory results can be listed as follows:

- *Small Area Estimation*, where the sample size in a sub-population/domain of interest may be so small (small area) to jeopardize the reliability of the estimates;

- the presence of *non-sampling errors* in the selection of the sample, that can hardly be introduced in the randomization process;

- the presence of non-random patterns of non-response and/or missingness among units; in these cases, some units might have a negligible or zero probability of being included in the sample and this occurrence practically destroys any possible assumption of exchangeability among the units, so requiring extra modeling;

- inference based on the integration among survey data and other data sources, for instance, register-based data (Lohr and Raghunathan 2017).

These issues will be discussed in the next § 3.

3   Modern survey sampling

Basu's classical Elephant example highlighted a few settings in which Horvitz-Thompson estimation could be inefficient, particularly when the sampling plan cannot be designed to proxy the distribution of the values of the variable(s) of interest. The use of auxiliary information has been the first device to improve the efficiency of design-based estimates. Indeed, the statistician in the Elephants' example could have well saved his job at the circus by exploiting the knowledge of the population size and simply using the Hajek estimator! The model-assisted framework has kept researchers busy for more than 50 years and allowed them to employ auxiliary information at the estimation stage using countless assisting models to describe the relationship with the response variable(s). See Breidt and Opsomer (2017) for a recent review including modern regression techniques.

The model-assisted approach has brought models out of the shadow in design-based inference. Nonetheless, there are instances in which the design-based – even if model-assisted – framework is not enough to allow for consistent and/or efficient estimation strategies. The first and more notable is

small area estimation (SAE), in which the sample size available in a sub-population of interest is so small that Horvitz-Thompson *direct* estimates, albeit unbiased, have unduly large variances.

SAE methods are *indirect* as they make use of observations coming from other domains/areas to obtain estimates for a particular area and are essentially model-based. SAE methods have seen tremendous development in the past 25 years: a complete review updated to the year of publication can be found in Rao and Molina (2015). In a presentation of the first edition of this book from 2003, Jon Rao admitted his genuine surprise when he finished writing it and realized that the longest chapters were those dedicated to methods using a hierarchical Bayesian approach. Then, if SAE can undoubtedly be considered the Trojan horse that brought model-based estimates into National Statistical Offices to address the need for timeliness and granularity of estimates, we may hint at a similar role for the Bayesian approach. See as a noticeable example the SAIPE program in the US United States Census Bureau (2021).

Another challenge that has reduced the suitability of design-based methods in survey sampling in the past years is the increased impact of **non-sampling errors**. When the outcome of under and/or over-coverage, of (item or unit) non-response, measurement error (including mode effects) depends on unknown processes, then it is impossible to draw a purely design-based inference, and the recourse to modeling is unavoidable. Modeling is required to reduce (a possibly non-negligible) bias in this context, rather than variance. The assumptions under which such bias reduction is achieved can seldom be verified and the inference is therefore model-based. This is true also when calibration or other reweighting methods such as raking are used to address under-coverage and/or non-response because, in this framework, modeling choices are implicit in the (possibly generalized) calibration/reweighting procedure (Haziza and Lesage 2016, Lesage et al. 2019).

In a Bayesian framework, non-response and under-coverage are often adjusted using inverse propensity weighting; see e.g. Little (1986). Indeed, propensity score adjustment was originally developed by Rosenbaum and Rubin (1984) to address selection bias in experimental designs but has been used extensively to control for selection bias in **non-probability samples**. Elliott and Valliant (2017) and the recent discussion paper by Wu (2022) provide a review of methods to draw inferences from such data. These can be seen as situations in which non-response and/or under-coverage get extreme consequences. Voluntary (typically online) surveys provide a large amount of (usually cheap) information that can provide misleading conclusions if bias

is not properly mitigated. The large amount of information available with non-probability samples, and big data more in general, can lead to less trustworthy conclusions because of the apparent large sample size available (Meng 2018). In this context, (Lee 2006) provides evidence that a reference probability sample must be available to obtain reliable inference. **Data integration** is a very active field of research to develop techniques for combining a probability sample with a non-probability data source (see, for a review, Yang and Kim 2020). These are essentially model-based and can be grouped into two main approaches. The first approach is weighting and can be based on propensity score adjustments: propensity scores are pseudo-inclusion probabilities estimated based on covariates available for sampled and non-sampled units. Calibration is another weighting approach that estimates the weights directly by calibrating auxiliary information in the non-probability sample with that in the probability sample. In the second approach, superpopulation modeling for the variable(s) of interest collected on sample units is used to predict values for non-sampled units. This approach is closely related to mass imputation and multiple imputation (Rubin 2004) can be used in this framework. Doubly robust estimation methods combine the weighting and imputation approaches to improve the robustness against model misspecification (Kim and Haziza 2014).

Data integration can occur at the micro-level when the information coming from a small survey may be enriched by the extra information coming from administrative non-probabilistic lists. The link step is not generally flawless, due to measurement errors and changing status of the statistical units involved in the process. This kind of problem calls for **record linkage** techniques. Record linkage is a class of statistical and algorithmic methods that aim at identifying whether two or more observed records refer to the same statistical entity or not. Duplications of the same entity within one single source or across different files may be interpreted as "clusters of records", showing strong similarities across their fields. Then the record linkage process may be also viewed as a formal Bayesian or non-Bayesian micro-clustering model; see Johndrow et al. (2018) and Tancredi et al. (2020)

All the above-mentioned issues become particularly relevant in the production of official statistics, where the problem of harmonizing and merging information coming from different sources becomes central as the general framework is moving towards an integrated system of statistical production and dissemination (D'Orazio et al. 2006). In fact, the National Statistical Institutes of well-developed countries, are progressively shifting from a survey-based system, where the sampling design played a decisive role, to

an integrated system where administrative lists may help to build a more complex data structure that represents different populations of interest. This data structure may be then combined with specific surveys and/or the use of other types of non-probability/big data to produce statistical information at (a possibly very) granular level of domains.

4 New ideas from a Bayesian perspective

The logical conflict between design-based methods and Bayesian philosophy has generated a sort of practical separation between Official Statistics and Bayesian methodology with the unpleasant result that survey sampling is not a typical theme of research among Bayesian-oriented Ph.D. students, despite its relevance from an applied perspective. Currently, much of the academic research is devoted to creating or developing the Bayesian versions of the model-based procedures, already become a relevant part of the applied survey statistician toolkit. Nevertheless, there have been more systematic attempts to reformulate the entire survey sampling methodology from a Bayesian perspective.

From a historical perspective, the first instance of the practical relevance of Bayesian methodology in survey sampling other than the above-mentioned small area estimation problem can be dated back to the introduction of multiple imputation techniques (Rubin 2004) for dealing with non-response and, more generally, missing data issues.

A nonparametric Bayesian approach has been initially proposed by Ericson (1969) in the context of a simple random sampling. To overcome the problem of the already mentioned flatness of the likelihood function (Godambe 1966), an exchangeable prior on the N-dimensional parameter Y_P is assumed. Using a weakly informative prior, one re-obtains design-based results from a completely different perspective. Similar results were obtained by Lo (1986), where a Dirichlet-Multinomial process is introduced, which converges, as $N \to \infty$, to a standard Dirichlet process. In the case of stratified sampling, priors are assumed exchangeable only within strata, in the spirit of hierarchical modeling (Rao 2011). Lo (1988) introduced the finite population Bayesian bootstrap (FPBB), which is defined in terms of a Polya's urn scheme and it is implemented by simulating a posterior distribution starting from a flat Dirichlet-Multinomial prior, as described in Lo (1986).

4.1 The Polya Posterior   Building on the seminal work of Lo (1988), an alternative approach to inference for finite populations is described in a series of papers by Glen Meeden and his collaborators: see for example

(Ghosh and Meeden 1997, Lazar et al. 2008, Strief and Meeden 2013); it is known as the **Polya Posterior** approach and it can be considered the finite population adaptation of the Bayesian Bootstrap proposed by Rubin (1981). Consider the following simple scenario. Suppose we have a population of N units and we draw a simple random sample of size n, say Y_s: assume that the goal is to estimate the mean θ of some function $h(Y_P)$. We put the n observed units in another urn U_2 and let the other $N-n$ units in the original urn U_1. Then we proceed as follows:

1. we draw a unit from U_2 and observe its value y;

2. we draw a unit from U_1, attach to it the y value and replace *both* units in U_2;

3. repeat steps 1-2 until U_1 is empty.

This way we have simulated a realization of the entire population. We repeat this simulation a huge number M of times, in order to get a *posterior* distribution for the quantity of interest Y_P, which can be summarized using descriptive statistics. Of course, this should be only interpreted as a pseudo-Bayesian posterior, since no prior has been introduced. Nonetheless, (Lo 1988) provides important theoretical results for this procedure. Assume that in Y_s there are k distinct values and, for $j = 1; \ldots, k$, let n_j be the corresponding frequencies in the sample. In performing an FPBB, let m_j^* be the random frequencies of the k distinct values.

Theorem 1 *(Lo, 1988) The following statements hold:*

- *The random vector $(m_1^*, \ldots, m_k^*)|Y_s$ has a Dirichlet-Multinomial distribution (Mosimann 1962) with parameters $(N - n; n_1, n_2, \ldots, n_k)$*

- *As $N \to \infty$,*

$$\left(\frac{m_1^*}{N - n}, \ldots, \frac{m_k^*}{N - n} \right) |Y_s \xrightarrow{d} Dirichlet(n_1, n_2, \ldots, n_k),$$

where $\xrightarrow{d}$ denotes convergence in distribution.

The Dirichlet-Multinomial distribution, cited in Theorem 1 can be interpreted as the multivariate extension of the Beta-Binomial distribution, that is a mixture of Binomial(n, p) distribution, with fixed n and p following a Beta distribution. Mosimann (1962) provides a general account of the prop-

erties of the Dirichlet-Multinomial distribution. Theorem 1, part b) can be used to say that, when the sampling fraction $f = n/N$ is negligible, one can avoid actually performing simulations and approximate the posterior distribution with the Dirichlet distribution. This idea is crucial in the development of Polya's Posterior methodology, especially when extra information on the population is available, and it can be translated into linear constraints on the Dirichlet random vector.

Although the Polya's posterior does not stem from any specific prior distribution, (Lo 1988) also proved that it can be derived as the posterior distribution on θ when the prior on the values of Y_s is a "flat" Dirichlet-Multinomial.

Polya's Posterior approach then simulates the entire population and allows simple inferences on specific parameters of the population; it is particularly useful when many parameters need to be estimated at the same time. There have been several attempts to extend this approach to more general contexts. Strief and Meeden (2013) proposed an alternative step-wise Bayesian justification of the use of the sampling weights which is not directly related to the sampling design, and that makes use of the standard kind of information present in auxiliary variables: however, it does not assume a model relating the auxiliary variables to the characteristic of interest. Dong et al. (2014) made an attempt to extend the finite population Bayesian bootstrap of Lo (1988) to account for complex sample designs. The paper takes the same goal of the inverse sampling technique and it can be treated as the Bayesian finite population version of inverse sampling. Lazar et al. (2008) considers the problem of implementing a Polya's Posterior approach in the presence of genuine partial information about auxiliary variables.

A limitation of the Polya posterior approach is that it requires an exchangeability assumption, not always tenable. In addition, (Rao 2011) noticed that *"Also, it is not clear how this method can handle complex designs, such as stratified multistage sampling designs, or even single-stage unequal probability sampling without replacement with non-negligible sampling fractions, and provide design-calibrated Bayesian inferences."*

The use of the Bayesian Bootstrap in a finite population setting is also discussed in Aitkin (2008), Carota (2009), and Cocchi et al. (2022) where a procedure for estimating the variance in a multiple frame context is proposed.

4.2   Calibrated Bayes   In a series of papers, during the last 15 years, Roderick Little has strongly advocated the use of Bayesian methods in survey sampling and, more generally, in official statistics. To summarize in a few words, Little advocates a compromise between various approaches. While

inference procedures should follow a Bayesian road, design features like clustering and stratification should be explicitly incorporated into the model to avoid the sensitivity of inference to model misspecification. In other terms, a purely design-based approach to finite population inference is no longer able to *"adequately address many of the problems of modern sample survey"* (Little 2022) and a model-based approach is deemed necessary: however, the model-based approach should be dressed in a Bayesian suit in order to easily incorporate survey sample design features. This compromise would guarantee good frequentist properties and would also benefit from the richness of information that the predictive posterior distribution allows obtaining.

Consider again the model-based framework expressed by the Eq. 2. If we ignore, for the sake of simplicity here, the issue of non-response, the distribution of $S|Z, Y$ does not actually depend on Y and the likelihood function contribution to inference is restricted to the term $p_{y|z}(y; z, \theta)$, which is combined with a suitable prior on θ in order to produce the posterior predictive distribution Eq. 4 for the non-observable quantity $Y_{P \setminus s}$.

This obvious consideration simply rules out any chance that the Bayesian answers could be efficient from a frequentist perspective if the word "frequentist" is meant in terms of the sampling mechanism. It is then clear that the frequentist properties should be considered either with respect to the conditional model induced by the family of distributions $p_{y;z}(y; z, \theta)$, or to the joint distribution $p_{y,s;z}(y, s; z, \psi, \theta)$.

It is well known (see Berger et al. 2009; Consonni et al. 2018) that a correct frequentist coverage of Bayesian procedure can be obtained only through the use of formal "noninformative" priors, whose exact expression depends on the specific statistical model. The derivation of a sensible noninformative prior is then not always easy. For example, usual improper priors which are routinely used in standard statistical models are not adequate for small area estimation and more generally for hierarchical models. See, as a general reference, (Berger et al. 2020) where the Authors derive a proper prior on the boundary of admissibility, which results as diffuse as possible without resulting in inadmissible procedures. A more specific analysis for small area models is described in Burris and Hoff (2019), where an alternative confidence interval procedure for the area means and totals is proposed under normally distributed sampling errors.

In general, the calibration of Bayesian procedures under complex sampling design is problematic and some approximations are often unavoidable. Things are even more complicated in the presence of non-ignorable non-

response patterns, which must be taken into account in the sampling model. The next section is devoted to the description of such a real case study.

An alternative route that tries to combine design and Bayesian properties is proposed in Wang et al. (2017). Here the likelihood is replaced by the sampling distribution of some summary statistics with "design-based" properties: this "pseudo-likelihood" is then combined with a prior reflecting genuine or vague prior information. This approach, although approximated in principle, provided "calibrated Bayes" procedures when combined with noninformative priors.

5   A real case study

In this section, we discuss a real case study where we consider the potential benefits and the inherent difficulties of a fully Bayesian treatment of the problem. During the last years, the Italian National Statistical Institute (Istat, hereafter) has begun a long and complex process of reorganization of data production and dissemination, called *modernization*, which can basically be described in three steps.

1. A main global infrastructure consisting of an integrated system of statistical registers.

2. The introduction of repeated sample surveys with the goal of constructing, updating, and enriching the statistical registers by observing new variables.

3. An integrated use of non-probability data coming from different kinds of sources (e.g., big data) for producing new information such as Trusted Smart Statistics.

An important case study, illustrative of the new data production system, is the *Italian Permanent Census (IPC)*, which replaces the general Population census, previously carried out every 10 years (the last one dates back to 2011): the new IPC system is a prototypical example of the new data production process and we now briefly describe it.

The starting point is the construction of the BRI (Base Register of Individuals). BRI is a "list" of N_R people who are residents in Italy, collected from all Italian municipalities; the BRI contains some core information such as gender, citizenship, and age, based on administrative data, which are considered highly reliable. BRI is then enriched with the reconstruction, through the implementation of suitable statistical models, of additional vari-

ables, namely educational level and employment status. The former is reconstructed via a log-linear model based on administrative data, while the latter is predicted through a suitably tailored hidden Markov model (Boeschoten et al. 2021).

Istat conducts two surveys to obtain an area sample s^A and a list sample s^L in order to evaluate the probabilities of *under*-coverage and of *over*-coverage, respectively, of BRI at the municipality level; these estimates are then used to correct administrative counts and to obtain estimates of the *resident* population. Population counts corrected for coverage errors are obtained through weighted counts of the BRI, where the weights are calculated as the ratio between the above probabilities.

The area and the list surveys are carried out using a sampling design that is quite common in National Statistical Institutes. In fact, they both follow a two-stage complex design where municipalities are Primary Sampling Units (PSUs) and households (for the list sample) or administrative geographical areas (for the area sample) are Secondary Sampling Units (SSUs). In particular, for the area sample s^A, the SSUs are addresses and enumeration areas. For each year of the census cycle, both the area and the list surveys share the same sample PSUs. Nevertheless, the samples of households in the two surveys are negatively coordinated from one to the other and for different survey occasions. It is worthwhile noticing that an allocation step of the sample size of SSUs (frequently combined with balancing procedures) actually determines their inclusion probabilities.

With the goal of discussing the potential use of Bayesian models in a real NSI case, it is important here to report some details of the sampling design adopted for the permanent census. For the sake of brevity, we discuss in detail the sampling design of s^L for estimating over-coverage probabilities. Then, we discuss the modeling of under-coverage probabilities.

As detailed in Righi et al. (2021), at the first stage all municipalities with a population size larger than 18,000 inhabitants and all municipalities selected in the Labor Force Survey (LFS, hereafter) are classified as self-representative (SR), while the others are considered non-self-representative (NSR). All SR municipalities are included in the sample, and for the NSR municipalities, a sample is drawn according to a probabilistic sampling design as follows. Within each province (LAU1), NSR municipalities are stratified in order to obtain homogeneous strata in terms of population size. Each stratum consists of four PSUs, and one single PSU is drawn from each stratum each year according to simple random sampling without replacement; this way,

in a four-year period, all the Italian municipalities can be observed. Then households are selected from each municipality adopting a simple random sampling without replacement design.

The allocation of the sample is then performed via a first sample size allocation among provinces, which is based on a trade-off between an equal sampling fraction and a sampling fraction inversely proportional to the population size of the provinces; indeed, a larger survey fraction is planned for smaller than for larger provinces. Afterward, in each province, the household sample was allocated within the municipalities as follows:

- for SR municipalities, a trade-off between an equal sampling rate and a proportional allocation is considered, in order to limit the number of households in the larger municipalities;

- for SR municipalities coming from the LFS, a proportional allocation is planned;

- for SNR municipalities, the sampling fraction assigned to each stratum is proportional to the population size of the stratum, so that each municipality is assigned a sample of households that is also representative, at least in terms of size, of the other three municipalities included in the stratum but not included in the sample for that specific year.

In addition, a minimum number of 100 households has to be included in each municipality. As a consequence, municipalities with a smaller number of households are completely enumerated. Finally, all members of selected households are interviewed.

The complex, although quite standard, structure of this sampling plan is difficult to render from a Bayesian perspective in order to make the CB approach operative. Let N_D be the population count of interest for subpopulation (domain) P_D and let $D_{\sum_{hailj}}$ be an indicator variable that takes the value 1 if unit j of household l in municipality i of enumeration area a of stratum h of the population belongs to P_D and 0 otherwise. Then $N_{DR} = \sum_h \sum_a \sum_i \sum_l \sum_j D_{\sum_{hailj}}$ is the number of people in BRI for domain P_D. The count estimates of the living population N_D can be obtained as

$$\hat{N}_D = \sum_h \sum_a \sum_i \sum_l \sum_j D_{\sum_{hailj}} \frac{1 - \hat{p}^o_{hilj}}{1 - \hat{p}^u_{hailj}},$$

where $\hat{p}^o_{hilj}$ and $\hat{p}^u_{hilj}$ are the estimated *over* and *under*-coverage probabilities for unit $hilj$ computed from s^L and s^A, respectively (see, for a similar approach, Pfeffermann 2015). A Bayesian treatment of the quantities $\hat{p}^o_{hilj}$ and $\hat{p}^u_{hilj}$ would easily allow to produce a posterior distribution for the overall quantity N_D and then produce a suitable measure of uncertainty. As noted before, samples in the two surveys are in general negatively coordinated, although in practice we consider them independent. Indeed, $\hat{p}^o_{hilj}$ and $\hat{p}^u_{hilj}$ are estimated for socio-demographic profiles for which the probabilities of over-coverage can be considered homogeneous and where the assumptions of the capture/recapture model hold (see, for more details, Righi et al. 2021).

More in detail, let us first focus on the over-coverage probabilities estimated using a Bayesian logistic model on data from s^L. The latter is a two-stage sampling design and, following following Little (2006, 2022), the covariates that determine the sampling design must be included in the model to render the inclusion mechanism ignorable. In addition, a Bayesian hierarchical model should be used to deal with the within-cluster correlation.

Let y_{hilj} be the dichotomous random variable that is 1 when unit j of household l in municipality i of stratum h selected in s^L from BRI is *not* found for the interview and 0 otherwise. When $y_{hilj} = 1$, the unit in the register should not be counted in the population (over-coverage). Then, a possible hierarchical model for over-coverage can be written as follows

$$y_{hilj}|p^o_{hilj} \overset{ind}{\sim} \textbf{Bernoulli}(p^o_{hilj})$$
$$\textbf{logit}(p^o_{hilj}) = \mathbf{x}^T_{hilj}\boldsymbol{\beta} + u_{hil} + v_{hi} + \gamma_h \tag{5}$$
$$u_{hil} \overset{iid}{\sim} N(0; \sigma^2_u) \tag{6}$$
$$v_{hi} \overset{iid}{\sim} N(0; \sigma^2_v) \tag{7}$$
$$p(\boldsymbol{\beta}, \gamma_h, \sigma^2_u, \sigma^2_v) \propto 1$$

where

- $\mathbf{x}_{hilj}$ collects individual-level covariates such as gender, age-class, citizenship, household-level covariates such as type of household or number of components, municipality-level covariates such as population size, type (urban/non-urban), and stratum level covariates such as macro-region;

- u_{hil} is a household-level random effect;

- v_{hi} is a municipality-level random effect;

- γ_h is a stratum fixed effect.

Only variables included in the BRI can be in $\boldsymbol{x}$ since the model will be used to make a prediction of the variable y on the units in BRI not observed in the sample. In addition, the vector of first-order inclusion probabilities could also be included in $\boldsymbol{x}$ to account for extra-variability introduced by the complex survey design not explained by the design variables already introduced in the model.

For the regression parameters $\boldsymbol{\beta}$ and γ_h, diffuse normal priors can also be considered that are sufficiently non-informative and computationally more convenient than flat priors over the real line. The normality of the random effects is a standard assumption in hierarchical models, while the choice of the prior for the variance components has been vastly debated, as in Bayesian mixed models the posterior distributions of these parameters are known to be sensitive to prior specification (Gelman 2006). Alternative choices can be the inverse Gamma for the variance or the half-Cauchy for the standard deviation.

The choice of the distribution for the household level random effect in Eq. 6 can be made more flexible by considering a different variance component for each possible household type, i.e.

$$u_{hil} \overset{iid}{\sim} N(0; \sigma^2_{uk(l)}), \text{for } k = 1, \ldots, K. \tag{8}$$

Here, $k(l)$ denotes the group to which household l belongs to. In fact, there might be different household types determined by their size, the relationship among members, and other characteristics, for instance, households with one single person, of couples, of couples and children, and so on. The variance σ^2_{uk} represents the similarity of the outcome variable on people in the same household typology. The use of different variance parameters σ^2_{uk} on the u_{hil} would also allow removing some of the random effects when their posterior distributions pile up in a neighborhood of zero.

The municipality-level random effect in Eq. 7 can be further generalized by allowing for an interaction with a subset of the covariates in $\boldsymbol{x}$, say $\boldsymbol{z}$ of dimension q. Then, the equation for the linear predictor in Eq. 5 can be enhanced to be

$$\mathbf{logit}(p^o_{hilj}) = \mathbf{x}^T_{hilj}\boldsymbol{\beta} + u_{hil} + \mathbf{z}^T_{hilj}\boldsymbol{v}_{hi} + \gamma_h$$

where $\boldsymbol{v}_{hi} \stackrel{iid}{\sim} N_q(\boldsymbol{0}; \boldsymbol{\Sigma}_v)$ and $p(\boldsymbol{\Sigma}_v) \propto 1$. An alternative choice for the prior could be a Wishart distribution for $\boldsymbol{\Sigma}_v^{-1}$. Variables in $\boldsymbol{z}$ could reflect the information used in the allocation of the sample of SSUs and/or directly the sample size. Similarly, the stratum-level fixed effect γ_h could be further generalized by including the interaction with a subset of the vector $\boldsymbol{x}$.

A similar modeling exercise can be developed for under-coverage that makes use of the data from the area survey s^A. In this case, the administrative geographical area characterizing the sampling design may introduce an intra-cluster correlation that should be taken into account. Let y_{hialj} be the dichotomous random variable that is 1 when unit j of household l in administrative area a of municipality i of stratum h selected in s^A is found for the interview and is not in BRI and 0 when the unit is found for the interview and is in BRI. When $y_{hialj} = 1$, the unit should be counted in the population (under-coverage). Then, a possible hierarchical model for under-coverage can be written as follows:

$$
\begin{aligned}
y_{hialj}|p_{hialj}^u &\stackrel{ind}{\sim} \textbf{Bernoulli}(p_{hialj}^u) \\
\textbf{logit}(p_{hialj}^u) &= \mathbf{x}_{hialj}^T\boldsymbol{\beta} + w_{hial} + u_{hia} + v_{hi} + \gamma_h \\
w_{hial} &\stackrel{iid}{\sim} N(0; \sigma_w^2) \\
u_{hia} &\stackrel{iid}{\sim} N(0; \sigma_u^2) \\
v_{hi} &\stackrel{iid}{\sim} N(0; \sigma_v^2) \\
p(\boldsymbol{\beta}, \gamma_h, \sigma_w^2, \sigma_u^2, \sigma_v^2) &\propto 1
\end{aligned}
\tag{9}
$$

where w_{hial} is a household-level random effect, u_{hia} is a random effect related to the administrative geographical area, and v_{hi} and γ_h have an interpretation similar to that of Eq. 5. Also in this case, it can be useful to consider different characteristics of the administrative geographical areas, such as rural/urban, type of dwelling, and use an approach similar to that used for households in Eq. 8 to model u_{hia}. Alternatively, these random effects can be assumed to be spatially correlated according to the distance $d_{aa'}$ between areas a and a'. For example, Eq. 9 can be replaced by

$$
\boldsymbol{u} \sim N_A(\boldsymbol{0}; \boldsymbol{\Sigma}_u), \quad \Sigma_{aa'} = \sigma_u^2 \exp\left(-\phi d_{aa'}\right), \quad p(\phi) \propto 1
$$

where A is the number of areas, σ_u^2 is the variance at any given point, and ϕ is a smoothing parameter that controls the scale of the correlation between areas. A Conditional Autoregressive specification can also be considered in

which the conditional distribution of u_{hia} given values in all the remaining areas only involves the neighboring areas.

6   Conclusions

Finite population sampling is an important chapter of statistical theory that deserves particular attention and a specific methodology. Bayesian inference is based on a solid prescriptive and coherent mathematical theory, sometimes difficult to combine with the practical difficulties of survey sampling. Basu himself noticed, as reported in Zacks (2002):

> The Bayesian as a surveyor must make all kinds of compromises... He may even agree to introduce an element of randomization into his plan... I can not put this enormous speculative process into a jacket of a theory. I happen to believe that data analysis is more than a scientific method...

The same concept is reiterated in Basu (1978)

> I do not think that it is realistic to ask for a well-defined theory of survey sampling. The problem is too complex and too varied from case to case. I have no clear-cut prescription for the planning of a survey. Apart from saying that we ought to hold the data as fixed and speculate about the parameters I have indeed very little else to offer.

However, we believe that the Bayesian contribution to the development of a more efficient quantification of uncertainty in survey sampling can be valuable. In particular, the role of the prior distribution is crucial.

In the absence of genuine prior information, or when some sort of "objectivity" of the estimation process in the field of official statistics is required, the use of formal noninformative priors must be recommended in order to provide "calibrated answers" with good frequentist properties (Berger et al. 2022). In complex design, the derivation of the formal noninformative prior is really too difficult to obtain and approximations are necessary, as for example in Berger et al. (2020). However, approximations should not be confused with weakly informative priors, which could provide silly - and, even worse, prior-dependent - answers (Berger 2006), and this should be absolutely avoided.

In a completely different scenario, the use of available genuine prior information can be crucial and sometimes necessary. There are many cases where population parameters smoothly vary in time and space, like in Demography, and it is relatively easy to guess a priori the reasonable range of such quantities. The introduction of such information in the model would ease the

calibration of the simulation algorithm on one hand. Of course, a sensitivity analysis to the prior inputs would be unavoidable in these cases; a significant dependence on the final answer to the prior inputs, however, should not be interpreted as a failure of the Bayesian approach but, rather, an indication that there might be too many parameters in the model and the data information is simply not enough to *update* all of them.

Finally, the last decades have experienced a real explosion, both in theoretical and applied terms, of Bayesian nonparametric methods of inference. Survey sampling has not yet been hit by this wave although the seminal papers by Lo (1986, 1988) seem to have paved the way. Some recent exceptions are Mendoza et al. (2021) and Savitsky and Toth (2016), and in the context of multiple imputation, (Paddock 2002).

To better reiterate our consideration of Basu's work, we would like to conclude with another quotation, taken from Casella and Gopal (2011)

(Re-)Reading Basu's papers, which combine an inimitable style of writing with impactful examples, is an educating, enlightening and entertaining experience. At best, we question our assumptions and beliefs, which leads us to gain new insights into classical statistical concepts. At "worst", we embark on a journey to becoming Bayesian.

Declarations

References

Aitkin, M. (2008). Applications of the Bayesian Bootstrap in finite population inference. *Journal of Official Statistics 24*, 21–51.

Basu, D. (1971). An essay on the logical foundations of survey sampling. I. In *Foundations of statistical inference (Proc. Sympos., Univ. Waterloo, Waterloo, Ont., 1970)*, pp. 203–242. Holt, Rinehart and Winston of Canada, Toronto, Ont.

Basu, D. (1978). On the Relevance of Randomization in Data Analysis. In *Survey Sampling and Measurement*, N. K. Namboodiri, ed, pp. 267–292. Academic Press, New Tork.

Beaumont, J.-F. and D. Haziza (2022). Statistical inference from finite population samples: A critical review of frequentist and bayesian approaches. *Canadian Journal of Statistics 50*(4), 1186–1212.

Berger, J. (2006). The case for objective Bayesian analysis. *Bayesian Analysis 1*(3), 385 – 402.

Berger, J., J. Bernardo, and D. Sun (2009). The formal definition of reference priors. *Annals of Statistics 37*, 905–938.

Berger, J., J. Bernardo, and D. Sun (2022). Objective Bayesian inference and its relationship to frequentism. In *Handbook of Bayesian Fiducial and Frequentist Inference (J.O. Berger, X.-L. Meng, N. Reid and M. Xie eds.).*, pp. (in press). Blackwell, Hoboken, NJ.

Berger, J., D. Sun, and C. Song (2020). An objective prior for hyperparameters in normal hierarchical models. *Journal of Multivariate Analysis 178*(104606).

Berger, Y. (2018). Empirical likelihood approaches under complex sampling designs. *The Survey Statistician 78*, 22–31.

Boeschoten, L., D. Filipponi, and R. Varriale (2021). Combining multiple imputation and hidden markov modeling to obtain consistent estimates of employment status. *Journal of Survey Statistics and Methodology 9*(3), 549–573.

Bolfarine, H. and S. Zacks (1992). *Prediction Theory for Finite Populations*. Springer Series in Statistics, Springer-Verlag.

Breidt, F. and J. Opsomer (2017). Model-assisted survey estimation with modern prediction techniques. *Statistical Science 32*(2), 190–205.

Burris, K. and P. Hoff (2019, 05). Exact Adaptive Confidence Intervals for Small Areas. *Journal of Survey Statistics and Methodology 8*(2), 206–230.

Carota, C. (2009). Beyond Objective Priors for the Bayesian Bootstrap Analysis of Survey Data. *Journal of Official Statistics 25*(3), 405–413.

Casella, G. and V. Gopal (2011). Basu's Work on Randomization and Data Analysis. In *Selected Works of Debabrata Basu, Selected Works in Probability and Statistics, A. DasGupta (ed.)*, pp. 1–4. Springer Science.

Cocchi, D., L. Marchi, and R. Ievoli (2022). Bayesian bootstrap in multiple frames. *Stats 5*(2), 561–571.

Consonni, G., D. Fouskakis, B. Liseo, and I. Ntzoufras (2018). Prior Distributions for Objective Bayesian Analysis. *Bayesian Analysis 13*(2), 627 – 679.

Cox, D. (2006). *Principles of Statistical Inference*. Cambridge University Press.

Dong, Q., M. Elliott, and T. Raghunathan (2014). A nonparametric method to generate synthetic populations to adjust for complex sampling design features. *Survey Methodology 40*(1), 29.

D'Orazio, M., M. Di Zio, and M. Scanu (2006). *Statistical Matching: Theory and Practice*. John Wiley & Sons.

Elliott, M. and R. Valliant (2017). Inference for nonprobability samples. *Statistical Science 32*(2), 249–264.

Ericson, W. (1969). Subjective Bayesian models in sampling finite populations. *J. Roy. Statist. Soc. Ser. B 31*, 195–233.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis 1*(3), 515–533.

Ghosh, M. and G. Meeden (1997). *Bayesian methods for finite population sampling*. Chapman & Hall, London.

Godambe, V. P. (1966). A new approach to sampling from finite populations. I. Sufficiency and linear estimation. *J. Roy. Statist. Soc. Ser. B 28*, 310-319.

Hartley, H. and J. N. K. Rao (1968, 11). A new estimation theory for sample surveys. *Biometrika 55*(3), 547–557.

Haziza, D. and É. Lesage (2016). A discussion of weighting procedures for unit nonresponse. *Journal of Official Statistics 32*(1), 129–145.

Johndrow, J., K. Lum, and D. Dunson (2018). Theoretical limits of record linkage and microclustering. *Biometrika 105*, 431–446.

Kim, J. K. and D. Haziza (2014). Doubly robust inference with missing data in survey sampling. *Statistica Sinica 24*(1), 375–394.

Lazar, R., G. Meeden, and D. Nelson (2008). A noninformative Bayesian approach to finite population sampling using auxiliary variables. *Survey Methodology 34*, 51–64.

Lee, S. (2006). Propensity score adjustment as a weighting scheme for volunteer panel web surveys. *Journal of official statistics 22*(2), 329.

Lesage, É., D. Haziza, and X. D'Haultfœuille (2019). A cautionary tale on instrumental calibration for the treatment of nonignorable unit nonresponse in surveys. *Journal of the American Statistical Association 114*(526), 906–915.

Little, R. (1986). Survey nonresponse adjustments for estimates of means. *International Statistical Review/Revue Internationale de Statistique 54*(2), 139–157.

Little, R. (2006). Calibrated Bayes: a Bayes-frequentist roadmap. *Amer. Statist. 60*(3), 213–223.

Little, R. (2011). Calibrated Bayes, an alternative inferential paradigm for Official Statistics. *Journal of Official Statistics 28*(3), 309–320.

Little, R. (2022). Bayes, buttressed by design-based ideas, is the best overarching paradigm for sample survey inference. *Survey Methodology 48*, 257–281.

Lo, A. (1986). Bayesian Statistical Inference for Sampling a Finite Population. *Annals of Statistics 14*(3), 1226-1233.

Lo, A. (1988). A Bayesian bootstrap for a finite population. *Annals of Statistics 16*, 1684-1695.

Lohr, S. and T. Raghunathan (2017). Combining survey data with other data sources. *Statistical Science 32*(2), 293–312.

Mendoza, M., A. Contreras-Cristán, and Gutièrrez-Pena E (2021). Bayesian Analysis of Finite Populations under Simple Random Sampling. *Entropy 23*, 318.

Meng, X.-L. (2018). Statistical paradises and paradoxes in big data (i) law of large populations, big data paradox, and the 2016 us presidential election. *The Annals of Applied Statistics 12*(2), 685–726.

Mosimann, J. (1962). On the compound multinomail distribution, the multivariate β-distribution and correlations among proportions. *Biometrika 49*, 65–77.

Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika 75*, 237–249.

Paddock, S. (2002). Bayesian nonparametric multiple imputation of partially observed data with ignorable nonresponse. *Biometrika 89*(3), 529–538.

Pfeffermann, D. (2015). Methodological issues and challenges in the production of official statistics: 24th annual morris hansen lecture. *Journal of Survey Statistics and Methodology 3*(4), 425–483.

Rao, J. (2011). Impact of Frequentist and Bayesian Methods on Survey Sampling Practice: A Selective Appraisal. *Statistical Science 26*(2), 240–256.

Rao, J. and I. Molina (2015). *Small area estimation*. John Wiley & Sons.

Righi, P., P. Falorsi, S. Daddi, E. Fiorello, P. Massoli, and M. Terribili (2021). Optimal sampling for the population coverage survey of the new italian register based census. *Journal of Official Statistics 37*(3), 655–671.

Rosenbaum, P. R. and D. B. Rubin (1984). Reducing bias in observational studies using subclassification on the propensity score. *J. Amer. Stat. Assoc. 79*(387), 516–524.

Royall, R. (1970). Finite population sampling—On labels in estimation. *Ann. Math. Statist. 41*, 1774–1779.

Royall, R. (1976). Likelihood Functions in Finite Population Sampling. *Biometrika 63*, 605–614.

Rubin, D. B. (1981). The Bayesian bootstrap. *Annals of Statistics 9*, 130–134.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*, Volume 81. John Wiley & Sons.

Savitsky, T. and D. Toth (2016). Bayesian estimation under informative sampling. *Electronic Journal of Statistics 10*, 1677–1708.

Strief, J. and G. Meeden (2013). Objective Stepwise Bayes Weights in Survey Sampling. *Survey Methodology 39*(1), 1-28.

Tancredi, A., R. Steorts, and B. Liseo (2020). A unified framework for de-duplication and population size estimation (with discussion). *Bayesian Anal. 15*(2), 633–682.

United States Census Bureau (2021). Small Area Income and Poverty Estimates (SAIPE) Program. https://www.census.gov/programs-surveys/saipe.html. Accessed: 2023-04-06.

Valliant, R., A. Dorfman, and R. Royall (2000). *Finite population sampling and inference*. Wiley Series in Probability and Statistics. Wiley-Interscience, New York.

Wang, Z., J. K. Kim, and S. Yang (2017, 12). Approximate Bayesian inference under informative sampling. *Biometrika 105*(1), 91–102.

Welsh, A. (2010). Basu on survey sampling. In *Selected Works of Debabrata Basu*, Volume 6 of *Selected Works in Probability and Statistics*, pp. 45–49. Springer, New York.

Wu, C. (2022). Statistical inference with non-probability survey samples. *Surv. Methodol 48*, 283–311.

Yang, S. and J. K. Kim (2020). Statistical data integration in survey sampling: A review. *Japanese Journal of Statistics and Data Science 3*, 625–650.

Zacks, S. (2002). In the footsteps of Basu: The Predictive Modelling Approach to Sampling from Finite Population. *Sankhya, A 64*, 532–544.

Zhong, C. and J. Rao (2000). Empirical likelihood inference under stratified sampling using auxiliary population information. *Biometrika 87*, 929–938.

Marco Di Zio
Istituto Nazionale di Statistica,
Roma, Italy
E-mail: dizio@istat.it

Brunero Liseo
MEMOTEF, Sapienza Università di
Roma, Roma, Italy

E-mail: brunero.liseo@uniroma1.it

Maria Giovanna Ranalli
Dept. of Political Science,
University of Perugia, Perugia,
Italy
E-mail: maria.ranalli@unipg.it