Journal of Official Statistics, Vol. 28, No. 3, 2012, pp. 335–339

Discussion

Jean-François Beaumont[1]

I would first like to thank the Editors for offering me the opportunity to discuss this very interesting and thought-provoking paper on the calibrated Bayes (CB) approach for official statistics. I must congratulate Prof. Little for providing convincing arguments in favour of the CB approach and for writing an article that I hope will stimulate the debate and the research on the choice of an inferential approach for samples selected from finite populations. As explained in Prof. Little's article, both the design-based and model-based approaches, have strengths and weaknesses. The main idea underlying the CB approach for official statistics consists of making Bayesian (model-based) inferences that have good design properties.

Prof. Little argues that this capitalizes on the strengths of both approaches but I think it may also inherit their weaknesses. In practice, CB is implemented by incorporating design information in the model and by using weak prior distributions. Although the design-based approach is not a panacea and the CB idea seems conceptually interesting, I must admit that I will not instantly become a CB proponent except when specific problems are considered such as small area estimation. My point of view is mostly justified by practical concerns and some conceptual issues that I explain below. My point of view may change after reading Prof. Little's rejoinder.

1. Model-based Approaches are Dependent on the Validity of a Model

As pointed out above, inferences with the CB approach remain Bayesian and thus model-based. The main criticism of model-based inferences, be they frequentist or Bayesian, is that inferences rely on an appropriate specification of a model. The frequentist approach requires correct specification of the first two moments of a model. The Bayesian approach requires an appropriate specification of a full parametric model (not only the first two moments) as well as the specification of a prior distribution. This is restrictive, and one of the main reasons why national statistical offices continue to use the design-based approach modified to account for issues such as nonsampling errors and small samples. This criticism can possibly be attenuated by using nonparametric methods such as penalized splines, advocated by Prof. Little, or the Polya posterior (e.g., Lazar et al. 2008; Rao 2011), which is not discussed in this article. These techniques could be useful in problems in which the number of explanatory variables is small and the sampling design is simple. As far as I know, I think their efficiency in problems with a large number of explanatory variables and/or with complex sampling designs remains to be demonstrated. Nonparametric methods may also be more difficult to apply and explain to data users than design-based methods.

[1]Statistics Canada, RH Coats Bldg, 16-B,Ottawa, ON K1AOT6, Canada. Email: jean-francois.beaumont@statcan.gc.ca

2. Why is Calibration Necessary or Even Useful?

If I understand properly, CB for official statistics is implemented in practice by determining models that yield posterior inferences with good design properties such as design consistency. First, as Rao (2008) pointed out, it is not clear how one ensures that posterior inferences are well calibrated in complex situations without explicitly accounting for the design effect associated with the estimator. More importantly, I am not sure that I understand why such calibration is necessary or even useful. From a Bayesian perspective, inferences are model-based and I do not understand how achieving design consistency, or any other design property, helps in any way in protecting against model misspecifications. An estimator could be design-consistent but justified using a model that does not hold. In that case, Bayesian inferences are misleading even if the sample size is large. If good design properties are required, why not make design-based inferences and use models only to justify the form of estimators like in the model-assisted approach?

To better understand this issue, let us consider the example of simple random sampling without replacement combined with a simple linear regression model with constant variance, normal errors and including an intercept. Under this scenario, the standard model-assisted estimator of the population mean $\bar{Y}$ is the generalized regression estimator $\bar{Y}_{GREG}$. The design variance of $\bar{Y}_{GREG}$ can be estimated using some design-consistent variance estimator (e.g., see Särndal et al. 1992 p. 234– 238), and thus valid design-based inferences can be made without requiring the linear model to hold exactly. Using this linear model and assuming a noninformative prior distribution for its parameters, the posterior mean of $\bar{Y}$ is $\bar{Y}_{GREG}$ and its posterior variance is a design-consistent estimator of the design variance of $\bar{Y}_{GREG}$. Therefore, the posterior mean and variance can be used to make valid design-based inferences. At first glance, the design-based and Bayesian approaches may seem to be more or less equivalent in this example. However, the CB statistician is not interested in making design-based inferences. From a Bayesian viewpoint, inferences are valid only if the linear model holds, and ensuring design-based validity of inferences does not seem to protect in any way against model misspecifications. In this example, I would not mind using the posterior mean and variance, but I would rather prefer to make design-based (model-assisted) inferences because they are not dependent on the validity of the linear model. For instance, the design-based interpretation of confidence intervals continues to hold even when the linear model is not satisfactory, whereas the Bayesian interpretation of intervals is misleading if the linear model is misspecified regardless of the design-based validity of these intervals.

3. Inclusion of Design Variables in the Model

Prof. Little recommends the inclusion of the design variables into the model. This ensures that the sampling design is not informative and thus removes the selection bias. However, sampling designs are often complex and it is not always straightforward to incorporate design variables in the model. How to do that in multi-phase or even multi-stage sampling? For instance, suppose a stratified sample of clusters (households or enterprises) is selected in the first phase and then a stratified sample of elements (persons or establishments) is drawn in the second phase from all the elements in the selected clusters, with possibly a different stratification in each phase. It is not obvious now to properly incorporate all the design information in the model for that kind of sampling design, which is sometimes used in national statistical offices. A design-based approach in which weights are computed as the inverse of the selection probabilities at each phase is simpler by far. Calibration weighting can be used to improve the efficiency of design-based estimators without really sacrificing simplicity.

4. The Multipurpose Nature of Many Sample Surveys

Another issue with model-based approaches is that a different model may be needed for different variables and different parameters. In multipurpose surveys, this is inconvenient for the data producer as the modeling task may become tremendous. This is also inconvenient for data users because each model may lead to its own set of survey weights. Users are accustomed to using a rectangular data file with one set of survey weights attached to it and perhaps replicate weights for variance estimation. This is generally not possible if more than one model is considered. These are two practical reasons why model-based methods are restricted to specific problems in practice. These methods are used only where they are really needed.

5. The Analysis of Survey Data and Nonsampling Errors

Prof. Little argues in Section 4.1 that the analysis of survey data leads to confusion and conflict within the design-based approach. I do not share this view. The survey analyst will often postulate a model and will be interested in making inferences about model parameters. This seems to preclude using the design-based approach, since a strict design-based approach can only be used to make inferences about finite population parameters using the sampling design, which has nothing to do with any model. However, a design-based statistician will normally get rid of the apparent conflict by making inferences with respect to the joint distribution induced by the analyst's model and the sampling design (e.g., among many others, Binder and Roberts 2003; Demnati and Rao 2010; Beaumont and Charest 2012). The reason for involving the sampling design in the inferential framework is to protect against informative sampling.

In Section 4.7, Prof. Little points out that the design-based approach to inference cannot easily tackle both the sampling error and nonsampling errors. I agree that a strict design-based approach cannot handle nonsampling errors; models are needed. However, design-based statisticians would again handle nonsampling errors by making inferences with respect to the joint distribution induced by models for the nonsampling errors and the sampling design. This approach minimizes the reliance on models (e.g., see Rao 2011). It is frequently used, both in the literature and in practice, to handle the unit nonresponse error. If there are many sources of errors, the modeling task may become challenging. But it is not obvious to me why CB is simpler or better in this context.

6. Does Multiple Imputation Really Fit in the CB Framework?

Prof. Little describes multiple imputation as a method to handle uncertainty caused by missing values. Multiple imputation is certainly well justified under a Bayesian framework (see Chapter 3 of Rubin 1987). It is not obvious that multiple imputation is CB because it is known not to have good frequentist properties in general. If frequentist properties are evaluated with respect to the joint distribution induced by the sampling design and the missing data mechanism, then Rubin (1987) in his Chapter 4 gives conditions for proper imputation, which ensure frequentist validity. The problem is that these conditions are very difficult (or perhaps impossible) to achieve in practice under this inferential framework (see Binder and Sun 1996). If frequentist properties are instead evaluated with respect to the joint distribution induced by the imputation model (the model for the variable being imputed), the sampling design and the missing data mechanism, then Meng and Romero (2003) (see also Kott 1995; Kim et al. 2006) showed that the multiple imputation variance estimator is not consistent unless a complete response estimator that is self-efficient is considered. There are many practical scenarios where a self-inefficient complete response estimator is used (and useful). This lack of consistency of the multiple variance estimator has nothing to do with model misspecifications. Indeed, multiple imputation captures only two of the three components of the overall variance, namely, the sampling variance and the nonresponse/imputation variance. It completely ignores the mixed component (see Särndal 1992; or Beaumont and Bissonnette 2011 for the definition of these components) regardless of whether the imputation model holds or not. Brick et al. (2004) showed that the mixed component can be positive or negative and not always negligible. My experience using SEVANI, the System for the Estimation of Variance due to Nonresponse and Imputation (Beaumont and Bissonnette 2011) developed at Statistics Canada, is that the mixed component is often negative, but not always, and is sometimes quite large. Ignoring this component may in some cases lead to serious overestimation of the overall variance. Obviously things become worse if the imputation model is not properly specified. Therefore, I believe that multiple imputation does not fit in the CB framework because it can have bad frequentist properties even when the imputation model holds.

Let me once again congratulate Prof. Little for challenging the design-based approach to inference for official statistics. The design-based approach is not a panacea and, like Rao (2011), I believe that model-based approaches have a role to play in some contexts. For instance, in Beaumont (2008) I used a model to deal with the problem of highly variable design weights. The model was used to smooth the design weights so as to improve the efficiency of design-based estimators. Also, I think that nonparametric model-based approaches could eventually become interesting alternatives to the design-based approach, which, in passing, is itself a nonparametric approach. However, I doubt national statistical offices will adopt any strictly model-based approach to inference, including CB, as a general purpose approach. The main reasons are that: i) model-based approaches are dependent on the validity of a model; ii) the inclusion of design variables in the model is not always straightforward and iii) the modeling task may become tedious in multipurpose surveys.

7. References

Beaumont, J.-F. (2008). A New Approach to Weighting and Inference in Sample Surveys. Biometrika, 95, 539–553.

Beaumont, J.-F. and Bissonnette, J. (2011). Variance Estimation Under Composite Imputation: The Methodology Behind SEVANI. Survey Methodology, 37, 171–179.

Beaumont, J.-F. and Charest, A.-S. (2012). Bootstrap Variance Estimation with Survey Data when Estimating Model Parameters. Computational Statistics and Data Analysis, 56, 4450–4461.

Binder, D.A. and Roberts, G.R. (2003). Design-Based and Model-Based Methods for Estimating Model Parameters. Analysis of Survey Data, R.L. Chambers and C.J. Skinner (eds). New-York: Wiley.

Binder, D.A. and Sun, W. (1996). Frequency Valid Multiple Imputation for Surveys with a Complex Design. Proceedings of the Section on Survey Research Methods, American Statistical Association, 281–286.

Brick, J.M., Kalton, G., and Kim, J.K. (2004). Variance Estimation with Hot Deck Imputation Using a Model. Survey Methodology, 30, 57–66.

Demnati, A. and Rao, J.N.K. (2010). Linearization Variance Estimators for Model Parameters from Complex Survey Data. Survey Methodology, 36, 193–201.

Kim, J.K., Brick, J.M., Fuller, W.A., and Kalton, G. (2006). On the Bias of the Multiple-Imputation Variance Estimator in Survey Sampling. Journal of the Royal Statistical Society, Series B, 68, 509–521.

Kott, P.S. (1995). A Paradox of Multiple Imputation. ASA Proceedings of the Section on Survey Research Methods, 380–383.

Lazar, R., Meeden, G., and Nelson, D. (2008). A Non-Informative Bayesian Approach to Finite Population Sampling Using Auxiliary Variables. Survey Methodology, 34, 51–64.

Meng, X.-L. and Romero, M. (2003). Discussion: Efficiency and Self-Efficiency with Multiple Imputation Inference. International Statistical Review, 71, 607–618.

Rao, J.N.K. (2008). Discussion of "Weighting and Prediction in Sample Surveys by R. Little". Calcutta Statistical Association Bulletin, 60, 29–39.

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

Rubin, D.B. (1987). Multiple Imputation for Nonresponse in Surveys. New York: Wiley.

Särndal, C.-E. (1992). Methods for Estimating the Precision of Survey Estimates when Imputation has been Used. Survey Methodology, 18, 241–252.

Särndal, C.-E., Swensson, B., and Wretman, J. (1992). Model Assisted Survey Sampling. New York: Springer.

Received July 2012