

Resampling Methods for Sample Surveys

James G. Booth and Brett Presnell

*Department of Statistics, University of Florida
Gainesville, FL 32611, USA*

Application of resampling methods in sample survey settings presents considerable practical and conceptual difficulties. Various potential solutions have recently been proffered in the statistical literature. This paper provides a brief critical review of these methods. Our main conclusion is that, while resampling methods may be useful in some problems, there is little evidence of their usefulness as general purpose techniques for the analysis of complex surveys.

Introduction

In this paper we will review some of the resampling methods for the analysis of sample surveys that have been proposed in the literature and discuss the extent to which they are worthwhile. Reasons for considering the use of resampling methods, which justify the additional computational burden they impose, include (a) improved accuracy over standard but less computationally intensive procedures in simple settings; and (b) ease of generalization and automation in complex problems to which standard methods either do not apply or break down.

Throughout most of this paper we will emphasize stratified random sampling and estimation using the stratified sample mean because this setting, despite its simplicity, is broad enough to illustrate our main points. In fact, the simple random sampling without replacement setting is sufficient for describing the resampling schemes we shall consider. We omit most technical details and all proofs. These can be found in a technical report written by the authors or in other references cited.

Simple and Stratified Random Sampling

We follow the notation of Cochran (1977, Chapters 2 & 5). Thus, $\mathcal{P} = \{Y_1, \dots, Y_N\}$ denotes a population of measurements and $\mathcal{S} = \{y_1, \dots, y_n\}$ denotes a random sample selected without replacement from \mathcal{P} . The sample mean \bar{y} then estimates the population mean \bar{Y} with variance $V(\bar{y}) = \frac{1}{n}(1-f)S^2$, where S^2 is the population variance (with divisor $N-1$) and $f = n/N$ is the sampling fraction. An unbiased estimate of $V(\bar{y})$, denoted by $v(\bar{y})$, is obtained by replacing S^2 by its unbiased estimator s^2 based on the sample.

In stratified random sampling from L strata an additional index, $h = 1, \dots, L$, is required. The estimate of the population mean based on a stratified random sample is $\bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h$, where $W_h = N_h/N$ is the stratum weight. The variance of \bar{y}_{st} is $V(\bar{y}_{st}) = \sum_{h=1}^L W_h^2 \frac{1}{n_h}(1-f_h)S_h^2$.

Bootstrap Methods

The basic principle underlying bootstrap estimation based on with replacement sampling schemes is the plug-in rule, whereby a population characteristic or functional $\tau = \tau(F)$ is estimated by $\hat{\tau} = \tau(\hat{F})$; i.e. by plugging in an empirical version of the population distribution function F (Efron and Tibshirani, 1993, Chapter 4). When sampling without replacement, the finite population \mathcal{P} plays the role of the unknown distribution function. The key step in extending

the bootstrap to this context is appropriately defining an empirical population $\hat{\mathcal{P}}$.

When the population size is an integer multiple of the sample size, $N = kn$, then an empirical population can be obtained by replicating the sample values k times (Gross, 1980). Booth, Butler and Hall (1994) propose a modification to this definition for the situation where $N = kn + l$, for some integer $1 \leq l < n$, whereby the additional l values in the empirical population are obtained by sampling without replacement from \mathcal{S} . In this case the empirical population is a random quantity, $\hat{\mathcal{P}}^*$ say, conditional on the sample. This requires a modification of the plug-in principle in order that the bootstrap estimate be well defined; namely $\hat{\tau} = E\{\tau(\hat{\mathcal{P}}^*)|\mathcal{S}\}$. That is, the bootstrap estimate is defined as the average value of the characteristic over all empirical populations.

An alternative approach is motivated by an assumption that the population \mathcal{P} is a realization of a with replacement sample from a superpopulation with unknown distribution function F . This suggests empirical populations be obtained as with replacement samples of size N from the sample \mathcal{S} . As with the previous approach, bootstrap estimates are defined as averages over all possible empirical populations.

Both the replication and superpopulation bootstrap methods extend in an obvious way to stratified samples by defining empirical populations as the union of empirical strata. Also, note that sampling with replacement is probabilistically equivalent to sampling from an infinite population. If N is allowed to increase fast enough that f converges to zero, both approaches described in this section converge to the familiar, with replacement, bootstrap plug-in rule.

Variance Estimation

An obvious characteristic of interest associated with any estimate, $\hat{\theta}$ say, is its variance in repeated sampling, $\tau = \text{Var}(\hat{\theta}|\mathcal{P})$. In the case of the mean the variance is known and an exactly unbiased estimate is available, and so it is instructive to compare how well various resampling methods perform in this case.

Both replication and superpopulation bootstrap estimates of $V(\bar{y})$ turn out to be proportional to the unbiased estimate; $\hat{\tau} = cv(\bar{y})$. The constant c depends on the value of l for the replication method and equals $(n - 1)/n$ using the superpopulation approach. In both cases the bias is negligible if n is moderately large and the sampling fraction is small. However, bootstrap variance estimation is potentially problematic in stratified samples if the sample sizes are small in some strata. While the bias of the bootstrap estimate of $V(\bar{y})$ might be considered a serious deficiency, it is worth noting that the same bias issue occurs in the infinite population setting (see Efron 1982, Chapter 5).

Distribution Estimation

The variance (or standard error) of an estimator provides a crude assessment of its accuracy. Often, however, the performance of an estimator can be more measured more precisely by examining its sampling distribution or that of its Studentized form.

In the stratified sampling setting the Studentized estimate is given by $T = (\bar{y}_{st} - \bar{Y})/\sqrt{v(\bar{y}_{st})}$. Let t_α denote the α -quantile of the distribution of T , so that $P(T \leq t_\alpha|\mathcal{P}) = \alpha$. Then, the

percentile- t interval $(\bar{y}_{st} - t_{1-\alpha}\sqrt{\bar{v}}, \bar{y}_{st} - t_{\alpha}\sqrt{\bar{v}})$ contains \bar{Y} with probability $1 - 2\alpha$ exactly in repeated sampling. Unfortunately, use of this formula in practice requires full knowledge of the sampling distribution of \bar{y}_{st} and hence of \mathcal{P} . A solution is to replace t_{α} by its asymptotic approximation z_{α} , the corresponding standard normal quantile, or by an estimate \hat{t}_{α} , obtained by inverting the bootstrap distribution function $E\{P(T^* \leq t | \hat{\mathcal{P}}^*) | \mathcal{S}\}$.

The arguments of Booth, Butler and Hall (1994) show that, if the responses are continuous and the number strata is fixed, the error in the bootstrap estimate of the distribution function is asymptotically smaller than that of the normal approximation. This translates into more accurate approximation of the ideal quantile function t_{α} and closer-to-nominal, one-sided coverage to the left and right of the corresponding confidence interval. Specifically, if $n = n_1 + \dots + n_L$ is the total sample size and the sampling fractions remain fixed between zero and one, then the endpoints of the standard normal interval are first-order correct in the sense that $z_{\alpha} = t_{\alpha} + O(n^{-1/2})$, whereas $\hat{t}_{\alpha} = t_{\alpha} + O_p(n^{-1})$ making the bootstrap endpoints second-order correct. These results are analogous to well known results for the infinite population bootstrap (see e.g. Hall, 1993, Chapter 3).

Unfortunately, the improvements offered by the bootstrap in the traditional n -asymptotic setting with a fixed number of strata do not carry over to the L -asymptotic setting, applicable in situations in which small samples are taken from a large number of strata. It is also worth emphasizing that the theoretical arguments, involving Edgeworth expansions, that underly the second-order properties of bootstrap confidence intervals do not apply if the responses are binary, as is often the case in sample surveys.

Other Resampling Methods

Numerous other bootstrap-like methods have been proposed recently for the finite population sampling setting, including the bootstrap with replacement (McCarthy and Snowden, 1985), variations on the replication method (Bickel and Freedman, 1984, Chao and Lo, 1965), the rescaling method (Rao and Wu, 1988) and the mirror-match method (Sitter, 1993b). These methods are all described and analyzed in detail by Presnell and Booth (1994). For the most part these resampling schemes are ad hoc methods motivated by a desire to reproduce the standard unbiased variance estimate, $v(\bar{y}_{st})$, for the stratified sample mean and not by a general principle such as the bootstrap plug-in rule. For this reason, they are cumbersome to apply to distribution estimation even though they were advocated for this purpose in several cases. For example, it is not always obvious how to create a resample version of a Studentized estimate even in the simplest case of the sample mean. Moreover, Presnell and Booth (1994) show that, however the resample version of T is calculated, none of these methods produce second-order correct confidence intervals either in the n -asymptotic or L -asymptotic framework and they may not even be first-order correct if not applied with extreme care.

Conclusions

A brief summary of our findings with respect to the two objectives stated in the Introduction is as follows:

In the stratified sampling setting with a fixed number of strata, bootstrap procedures are avail-

able that provide improvements over classical approaches for constructing confidence intervals based on the normal approximation. However, the improvements are of second order and are generally only noticeable when the sample sizes are small. Moreover, in the stratified sampling setting with an increasing number of strata, none of the methods yet proposed provide any asymptotic improvement over the standard normal approximation.

It could be argued that some of the resampling methods mentioned in last section are useful if only for variance estimation, as is the case with more traditional resampling methods such as the jackknife and balanced repeated replication. However, none of the methods discussed provide any asymptotic improvement over much less computationally intensive, traditional variance estimators which are available for most standard sampling designs and hence they appear to be of little use in these settings. Resampling-based variance estimates have been shown to be useful in certain specialized problems (see e.g. Canty and Davison, 1999). However, since it is generally unclear how to extend resampling methods beyond even stratified random sampling, they should be applied with extreme care and, in general, we are skeptical about their usefulness for the analysis of complex surveys.

References

- Booth, J.G., Butler, R.W. and Hall, P. (1994). "Bootstrap methods for finite populations," *J. Amer. Statist. Assoc.* **89**: 1282-9.
- Canty, A.J. and Davison, A.C. (1999). "Resampling-based variance estimation for labour force surveys," *The Statistician*. To appear.
- Cochran, W.G. (1977). *Sampling Techniques*, 3rd Ed., Wiley: New York.
- Chao, M.-T. and Lo, S.-H. (1985). "A bootstrap method for finite populations," *Sankhya B* **47**: 399-405.
- Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*, CBMS-NSF Regional Conference Series, Vol. 38: SIAM.
- Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*, Chapman and Hall: New York.
- Gross, S.T. (1980). "Median estimation in sample surveys," *ASA Proc. of Survey Rsch. Methods Sect.*: 181-4.
- Hall, P. (1993). *The Bootstrap and Edgeworth Expansion*, Springer: Berlin, New York.
- McCarthy, P.J. and Snowden, C.B. (1985). "The bootstrap and finite population sampling," *Vital and Health Statistics* **2**, Public Health Service Publication 85-1369.
- Presnell, P. and Booth, J.G. (1994). "Resampling methods for sample surveys," Technical Report No. 470, Department of Statistics, University of Florida.
- Rao, J.N.K. and Wu, C.F.J. (1988). "Resampling inference with complex survey data," *J. Amer. Statist. Assoc.* **83**: 231-41.
- Sitter, R.R. (1992), "A resampling procedure for complex surveys data," *J. Amer. Statist. Assoc.* **87**: 755-65.