Statistical Science
2003, Vol. 18, No. 2, 185–190

Impact of Bootstrap on the Estimating Functions

Subhash R. Lele

Abstract. Estimating functions form an attractive statistical methodology because of their dependence on only a few features of the underlying probabilistic structure. They also put a premium on developing methods that obtain model-robust confidence intervals. Bootstrap and jackknife ideas can be fruitfully used toward this purpose. Another important area in which bootstrap has proved its use is in the context of detecting the problem of multiple roots and searching for the consistent root of an estimating function. In this article, I review, compare and contrast various approaches for bootstrapping estimating functions.

Key words and phrases: Model-robust confidence intervals, multiple roots, stochastic processes, Wu's wild bootstrap.

1. INTRODUCTION

Estimating functions (Godambe, 1960; Godambe and Kale, 1991) have proved to be an extremely useful methodology in applied as well as theoretical statistics. Two of the main attractions of this methodology are (1) its dependence on only a few features (e.g., the mean and the variance) of the underlying probability model (Godambe and Thompson, 1984) and (2) ease of handling nuisance parameters. Estimating functions yield estimators that are not strongly model dependent. A natural issue that arises is that of obtaining the standard errors and the confidence intervals that are not strongly model dependent as well. Bootstrap and jackknife techniques (Efron and Tibshirani, 1993) are the natural candidates for obtaining such standard errors and confidence intervals.

To facilitate further discussion, I introduce the following notation. Let $Y_1, Y_2, \ldots, Y_n$ denote the sequence of random variables following some probability distribution. This may be a sequence of i.i.d. random variables or a sequence of independent but not identically distributed random variables or it might even conceivably be a sequence of stationary or nonstationary dependent variables. Let θ denote the feature of this probabilistic structure that we are interested in estimating using the observed data. For example, we might be interested in the mean, the variance, the regression parameters or the dependence parameters. In the estimating function context, we construct a function of the random variables and the parameter of interest such that its expected value is zero. Let us denote this function by $g(\underline{Y}, \theta)$. In principle, there may be several different probabilistic structures under which the estimating function has expected value zero. This yields the celebrated robustness against the model specification as pointed out by Godambe and Thompson (1984). I do not specify all the regularity conditions here; the reader may refer to Heyde (1997), for example, for the theoretical underpinnings of estimating function theory.

In many practical situations, an estimating function can be expressed as a sum of component estimating functions. Let us call such estimating functions *linear estimating functions*. Thus, $g(\underline{Y}, \theta) = \sum_{i=1}^{n} g_i(Y_{(i)}, \theta)$, where $Y_{(i)}$ denotes a subset of the observations. For example, in the regression context $Y_{(i)} = (Y_i, X_i)$, whereas in the first order autoregression case, $Y_{(i)} = (Y_i, Y_{i-1})$. Because this is a sum of random variables, we usually can apply the weak law of large numbers as well as the central limit theorem, after proper standardization, to these sums. Inverting the Normal distribution thus obtained provides approximate confidence intervals for the parameter estimates (Godambe and Thompson, 1999; Rajarshi and

Subhash R. Lele is Professor of Statistics, Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1 (e-mail: slele@ualberta.ca).

Godambe, 2001). Estimation of the variance of this asymptotic distribution without full specification of the underlying parametric model is, then, the crux of the problem.

The issue of model-robust confidence intervals was first addressed by Royall (1986), who provided an analogue of the observed Fisher information in the estimating function context when the data are independent. The jackknife technique was extended to estimating functions, particularly in the context of dependent data, in Lele (1991a); see also Künsch (1989) for similar developments. It was obvious that an extension of the bootstrap technique for estimating functions would come along. Lele (1991b) suggested one such extension. The main idea in Lele (1991a, b) was to use the sequence of estimating functions, g_i in the above notation, rather than the original observations for jackknifing and bootstrapping. Later Hu and Zidek (1995) and Hu and Kalbfleisch (2000) developed this idea theoretically and showed its applicability in various important situations. In the following section, I briefly describe various approaches to bootstrapping estimating functions and also compare, contrast and extend them.

2. BOOTSTRAPPING ESTIMATING FUNCTIONS: CONTRASTS AND COMPARISONS

Perhaps the most straightforward and natural approach to bootstrapping estimating functions is to treat the estimating functions g_i evaluated at the estimated parameter as independent, identically distributed random variables and use them as the basis for resampling. This approach was followed by Hu and Zidek (1995) and Hu and Kalbfleisch (2000), who used the bootstrap only in the independent data context. Their algorithm, in a single parameter context, can be described as follows:

2.1 Multinomial Sampling-Based Algorithm

STEP 1. Solve the equation $\sum_{i=1}^{n} g_i(Y_{(i)}, \theta) = 0$ to obtain the estimate $\hat{\theta}$. Let $Z_i = g_i(Y_i, \hat{\theta})$, $i = 1, 2, \ldots, n$.

STEP 2. Obtain a sample of size n, with replacement, from $Z_1, Z_2, \ldots, Z_n$. Let us denote it by $Z_1^*, Z_2^*, \ldots, Z_n^*$.

STEP 3. Obtain a Studentized bootstrap estimating function, namely,

$$S_t^* = V^{*-1/2} \sum_{i=1}^{n} Z_i^*, \quad \text{where } V^* = \frac{1}{n} \sum_{i=1}^{n} (Z_i^* - \bar{Z}^*)^2.$$

STEP 4. Repeat Steps 2 and 3 B number of times to obtain the bootstrap distribution of S_t^*.

This distribution is inverted to obtain the confidence interval for the parameter. The lower limit is given by the solution of the equation $S_t^*(1 - \alpha/2) = S_t(y, \theta)$ and the upper confidence limit is given by the solution of the equation $S_t^*(\alpha/2) = S_t(y, \theta)$.

The main difference between bootstrapping estimating functions and the classical bootstrap is that in the classical bootstrap, we are interested in estimation of the distribution of the estimator, whereas in the estimating function bootstrap, we are interested in estimating the distribution of the normalized sum of the estimating functions. This distribution is then inverted to find the corresponding distribution of the parameter. One important advantage of this procedure is that of parameterization invariance as pointed out by Hu and Kalbfleisch (2000). A second advantage is that we need to solve only two equations and need not obtain the solution for every bootstrap sample.

Hu and Kalbfleisch (2000) strongly argued in favor of bootstrapping estimating functions rather than the estimator. They provided theoretical as well as practical justification for such a claim. The algorithm provided above mimics the classical bootstrapping algorithm, which I term the *multinomial bootstrap algorithm*. It is well known in the classical bootstrap literature (Wu, 1986) that extending this idea to either independent but not identically distributed random variables (e.g., in the regression setup) or the dependent data situation (both stationary and nonstationary situations such as time series or spatial models) is quite difficult. For stationary stochastic processes, one possibility is to generalize Künsch's blockwise bootstrap idea to estimating functions (Rajarshi and Godambe, 2001). In this case, we maintain the order of the estimating functions and construct blocks of consecutive estimating functions. We can then pretend that these blocks are independent and identically distributed random variables and resample them using the above algorithm. However, this is not necessarily a practical idea. A large amount of data is needed to have enough blocks, each block being large enough by itself. This idea also does not generalize to clustered data such as those obtained in longitudinal studies or small area estimation problems in survey sampling. The blockwise bootstrap idea also does not apply when the underlying process is nonstationary. The second problem with this idea is that, in the regression situation, the coverage is obtained unconditionally on the covariates rather than conditionally on the covariates.

Wu (1986) specifically addressed the second problem with the multinomial bootstrap algorithm in the regression context and suggested what is sometimes called Wu's wild bootstrap. Liu (1988) provided some extensions and theoretical justification for Wu's bootstrap. Lele (1991b) suggested an extension of Wu's bootstrap to the dependent data situation using estimating functions. In the following section, I briefly describe this approach when the data are independent but not identically distributed.

2.2 Wu's Bootstrap Algorithm Extended to Estimating Functions

STEP 1. Solve the equation $\sum_{i=1}^{n} g_i(Y_{(i)}, \theta) = 0$ to obtain the estimate $\hat{\theta}$. Let $Z_i = g_i(Y_i, \hat{\theta})$, $i = 1, 2, \ldots, n$.

STEP 2. Generate independent, identically distributed random variables $t_1, t_2, \ldots, t_n$ from a distribution such that $E(t) = 0$, $\mathrm{var}(t) = 1$ and $E(t^3) = 1$. Obtain $Z_i^* = Z_i t_i$.

STEP 3. Obtain a Studentized bootstrap estimating function, namely,

$$S_t^* = V^{*-1/2} \sum_{i=1}^{n} Z_i^*, \quad \text{where } V^* = \frac{1}{n} \sum (Z_i^* - \bar{Z}^*)^2.$$

STEP 4. Repeat Steps 2 and 3 B number of times to obtain the bootstrap distribution of S_t^*.

This distribution is inverted to obtain the confidence interval for the parameter as in the previous algorithm. The main difference between the two algorithms is in Step 2. The proof given in Liu (1988) can be mimicked to prove that this bootstrap corrects the same order terms as in the multinomial bootstrap. I applied Wu's bootstrap and the multinomial bootstrap to the example used by Hu and Kalbfleisch (2000) to compare the cutoff points obtained by these two algorithms to those obtained by the parametric bootstrap. Overall, it appears that Wu's bootstrap is somewhat closer to the parametric bootstrap than the multinomial bootstrap.

Godambe (1985) provided the foundations of inference for stochastic processes using the estimating function approach. Godambe and Heyde (1987) considered the class of estimating functions where although the data are dependent, the estimating functions are uncorrelated with each other. In such a situation, Lele (1991a) provided a jackknife approach for obtaining asymptotic standard errors and Lele (1991b) provided an extension of Wu's bootstrap. Essentially the algorithm described above remains the same but the sequence of t_i's is generated in such a fashion that the third mixed moments are nonzero. However, this extension provides only a partial correction.

For stationary dependent data, we can combine Wu's bootstrap with blockwise bootstrap in such a fashion that, at least theoretically, we can obtain the full second order correction as in the usual bootstrap. The algorithm described below can be used in the clustered data situation (after all, clusters are blocks of observations, albeit of different length) as well as in the case of nonstationary dependent data.

2.3 Wu's Blockwise Bootstrap for Estimating Functions Algorithm

STEP 1. Solve the equation $\sum_{i=1}^{n} g_i(Y_{(i)}, \theta) = 0$ to obtain the estimate $\hat{\theta}$. Let $Z_i = g_i(Y_i, \hat{\theta})$, $i = 1, 2, \ldots, n$.

STEP 2. Split the sequence of estimating functions in M_n blocks of size m_n, each block consisting of consecutive estimating functions. Let us denote these blocks by $G_1, G_2, \ldots, G_{M_n}$.

STEP 3. Generate independent, identically distributed random variables $t_1, t_2, \ldots, t_{M_n}$ from a distribution such that $E(t) = 0$, $\mathrm{var}(t) = 1$ and $E(t^3) = 1$. Obtain $G_i^* = G_i t_i$. That is, each estimating function within a block is multiplied by the same t. Put these perturbed blocks together in the original order.

STEP 4. Obtain a Studentized bootstrap estimating function, namely,

$$S_t^* = V^{*-1/2} \sum_{i=1}^{n} Z_i^*, \quad \text{where } V^* = \frac{1}{n} \sum (Z_i^* - \bar{Z}^*)^2.$$

STEP 5. Repeat Steps 3 and 4 B number of times to obtain bootstrap distribution of S_t^*.

This algorithm can be used successfully in various other situations as well. For example, Prasad (2002) has provided an application in small area estimation and linear mixed models.

2.4 Multiple Parameters

If the bootstrap distribution is obtained based on estimators rather than estimating functions, the equations typically have to be solved repeatedly. This may pose a substantial computational problem, especially in the multiparameter case. However, the bootstrap based on estimating functions involves solving the equations only once. The confidence contours correspond to inverting a score test based on estimating functions (Hu and Kalbfleisch, 2000). Consider the approximate pivotal quantity $Q(y, \theta) = S_t^T(y, \theta) S_t(y, \theta)$. We can then obtain the distribution of this quadratic form using the bootstrap and hence the estimate of its $100(1-\alpha)\%$ percentile q_α. The bootstrap confidence region is then obtained by $C_{1-\alpha}(y) = \{\theta : Q(y,\theta) < q_\alpha\}$. He and Hu (2002) suggested using a one parameter at a time algorithm in such a situation. However, solving one equation at a time does not guarantee the correct solution unless the estimating function has a unique root. He and Hu also resorted to simply computing the estimator and using its empirical distribution rather than the inversion described earlier.

3. MULTIPLE ROOTS AND OTHER COMPUTATIONAL ISSUES

The above algorithms work well when there exists a unique root to the estimating function. However, in many situations there are multiple roots to the estimating function. Some attempts have been made using the bootstrap to throw light on this issue.

3.1 Multiple Roots Problem

One of the major obstacles in the application of estimating function theory is the existence of multiple roots to the estimating function. See Small, Wang and Yang (2000) for an extensive discussion of this problem. If the estimating function is obtained from the likelihood, Wald (1949) showed that the root corresponding to the global maximum is the correct one. However, it is well known that a likelihood-type function corresponding to a given estimating function may not exist (Li and McCullagh, 1994). In those situations where such a function exists, Li (1996) showed that the root corresponding to the minimax point of this function is the correct root. However, to show that a root is a minimax point, it seems we have to find all the roots of the equation. This can be quite tricky, if not impossible, in the multiparameter situation. Markatou, Basu and Lindsay (1998) used the bootstrap to investigate the space of the roots of estimating functions. Having obtained a good understanding of this space, the problem of choosing the correct root still remains. The main idea behind choosing the "correct" root is to see which root is most consistent with the data. The consistency of the root with the data can be tested in many ways. A goodness of fit test is one possibility. Another possibility, suggested by Small, Wang and Yang (2000), is in terms of the approximate pivotal quantity $Q(y,\theta) = S^T(y,\theta)V^{-1}S(y,\theta)$. Suppose the underlying estimating functions are information unbiased, that is, $E(gg^T) = -E((d/d\theta)g)$. Then, a nonbootstrap-based estimator of V^{-1} is given by

$$I(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\frac{d}{d\theta}g_i(\hat{\theta}).$$

Consider the bootstrap distribution of $Q^*(y,\theta) = S^{*T}(y,\theta)I(\theta)S^*(y,\theta)$ evaluated for each root. Since the distribution of this quadratic form is χ^2_p only at the consistent root, we can test for the correctness of the root based on how close the bootstrap distribution is to the χ^2_p distribution. Generalization of this idea to noninformation unbiased estimation is, as yet, not available.

3.2 One-Step Estimation and Bootstrap

If we insist on using the estimator rather than estimating function to obtain the bootstrap distribution, we can simplify computation substantially by applying the one-step estimation procedure. We can use the original estimator as the initial value in the Newton–Raphson algorithm to solve the bootstrapped estimating functions. Lipsitz, Dear and Zhao (1994) suggested this improvement in the context of jackknifing estimating functions.

4. AN APPLICATION OF BOOTSTRAPPING TO THE MANTEL–HAENSZEL ESTIMATOR

Wu (1986) pointed out that the bootstrap for regression based on the pairs (X_i, Y_i) provides unconditional coverage that integrates over the variability of the covariates. However, there are situations where one may want to obtain confidence intervals that are correct "conditionally" on the observed covariate values. I will not go into the philosophical argument about which one is the right answer. The main issue that I want to point out is that the multinomial bootstrap adapted to estimating functions provides "unconditional" answers, whereas the adaptation of Wu's wild bootstrap provides conditionally correct answers. As an example, I will describe an application of bootstrapping estimating functions to obtain inferences for the Mantel–Haenszel estimator of common odds ratio.

I will paraphrase, in very rough and lay terms, the biological situation in the following discussion. More detailed discussion and background information was provided by Boyce, Lele and Johns (2003). Because the original article is still unpublished and for the sake of propriety, I report the analysis of the data set that is not corrected for several known confounding factors. This does not affect the illustrative aspect of the example.

The whooping crane (*Grus americana*) is an endangered bird species. These birds typically lay two eggs and usually both eggs hatch, but at the end of the season, in most years, only one chick survives. To help increase the stock, the park service decided to remove one egg from the nests and let some other species of crane incubate that egg. In this fashion, at the end of the season there are two chicks instead of only one. One of the plausible evolutionary reasons for laying two eggs is that it is an insurance policy against bad environmental conditions. The question naturally arises: would the strategy of removing an egg have a detrimental effect on the probability of a chick surviving from the one-egg nest?

The available data have the following general structure. Whooping cranes nest in several areas in the northern parts of Alberta. Within each area, an egg was removed from some of the nests and others were left unperturbed. At the end of the season, the presence or absence of a chick in each nest was noted. Thus, corresponding to each area we have a 2×2 table: number of control nests where the egg was not removed and number of treatment nests where egg was removed. The response is the presence or absence of a chick in each of these nests. This experiment was conducted over several years. We modelled the data using logistic regression, namely,

$$\log \frac{\mathrm{P}(\text{presence}|X = x)}{\mathrm{P}(\text{absence}|X = x)} = \alpha + \beta x,$$

where $X = 1$ if an egg was removed and $X = 0$ if the nest was unperturbed. We are interested in estimating the common odds ratio β. Clearly the intercept parameter α depends on the spatial location as well as the year of collection. The Mantel–Haenszel estimator of the common odds ratio is obtained by solving the estimating equation

$$\sum_{k=1}^{K} \frac{1}{n_{\cdot\cdot k}} (n_{11k} n_{22k} - \beta n_{12k} n_{21k}) = 0,$$

where K is the total number of 2×2 tables. Based on 56 tables, the estimate of the common odds ratio was 2.86. The 90% confidence interval using bootstrapping of the estimating functions (Wu's wild bootstrap) is (1.85, 4.44). The confidence interval based on Royall's robust variance estimator (Royall, 1986) and normal approximation is given by (1.72, 4.01). It seems that bootstrapping incorporates the skewness of the distribution effectively. The difference between Wu's wild bootstrap and the multinomial bootstrap is insignificant and hence the issue of conditioning does not seem critical in this particular data set. The estimate and the associated confidence interval indicate that removing an egg might, in fact, be beneficial to the survival of the chick. This result, although counterintuitive at first glance, in fact, can also be supported using biological and evolutionary arguments.

5. SUMMARY

Estimating functions provide an effective way to conduct statistical inference without specifying the full probability structure. Bootstrap methods are an effective way to obtain model independent standard errors and confidence intervals for parameters of interest. A combination of the two methodologies has begun in earnest only recently. There are many possible ways in which this methodology can be extended. Application of saddlepoint approximations in conjunction with the bootstrap is an interesting approach that has not been explored in the estimating function context. Adaptation of various corrections to bootstrap intervals in the context of estimating functions is also an interesting open problem. Bootstrap methods might also have a role to play in solving the thorny issue of detection of multiple roots and selection of the correct root.

ACKNOWLEDGMENTS

This work was supported in part by a grant from the National Science and Engineering Research Council of Canada. I acknowledge useful discussions with Professor N. G. N. Prasad.

REFERENCES

BOYCE, M. S., LELE, S. R. and JOHNS, B. W. (2003). Whooping crane recruitment enhanced by egg removal. Unpublished manuscript.

EFRON, B. and TIBSHIRANI, R. J. (1993). *An Introduction to Bootstrap*. Chapman and Hall, New York.

GODAMBE, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.* **31** 1208–1212.

GODAMBE, V. P. (1985). The foundations of finite sample estimation in stochastic processes. *Biometrika* **72** 419–428.

GODAMBE, V. P. and HEYDE, C. C. (1987). Quasi-likelihood and optimal estimation. *Internat. Statist. Rev.* **55** 231–244.

GODAMBE, V. P. and KALE, B. K. (1991). Estimating functions: An overview. In *Estimating Functions* (V. P. Godambe, ed.) 3–20. Oxford Univ. Press.

GODAMBE, V. P. and THOMPSON, M. E. (1984). Robust estimation through estimating equations. *Biometrika* **71** 115–125.

GODAMBE, V. P. and THOMPSON, M. E. (1999). A new look at confidence intervals in survey sampling. *Survey Methodology* **25** 161–173.

HE, X. and HU, F. (2002). Markov chain marginal bootstrap. *J. Amer. Statist. Assoc.* **97** 783–795.

HEYDE, C. C. (1997). *Quasi-likelihood and Its Application*. Springer, New York.

HU, F. and KALBFLEISCH, J. D. (2000). The estimating function bootstrap. *Canad. J. Statist.* **28** 449–499.

HU, F. and ZIDEK, J. V. (1995). A bootstrap based on the estimating equations of the linear model. *Biometrika* **82** 263–275.

KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241.

LELE, S. R. (1991a). Jackknifing linear estimating equations: Asymptotic theory and applications in stochastic processes. *J. Roy. Statist. Soc. Ser. B* **53** 253–267.

LELE, S. R. (1991b). Resampling using estimating equations. In *Estimating Functions* (V. P. Godambe, ed.) 295–304. Oxford Univ. Press.

LI, B. (1996). A minimax approach to consistency and efficiency for estimating equations. *Ann. Statist.* **24** 1283–1297.

LI, B. and MCCULLAGH, P. (1994). Potential functions and conservative estimating functions. *Ann. Statist.* **22** 340–356.

LIPSITZ, S. R., DEAR, K. B. G. and ZHAO, L. (1994). Jackknife estimators of variance for parameter estimates from estimating equations with applications to clustered survival data. *Biometrics* **50** 842–846.

LIU, R. Y. (1988). Bootstrap procedures under some non-iid models. *Ann. Statist.* **16** 1696–1708.

LIU, R. Y. and SINGH, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billard, eds.) 225–248. Wiley, New York.

MARKATOU, M., BASU, A. and LINDSAY, B. G. (1998). Weighted likelihood equations with bootstrap root search. *J. Amer. Statist. Assoc.* **93** 740–750.

PRASAD, N. G. N. (2002). Personal communication.

RAJARSHI, M. B. and GODAMBE, V. P. (2001). Confidence intervals based on estimating functions from stationary observations. Working Paper 2001-18, Dept. Statistics and Actuarial Science, Univ. Waterloo, Ontario, Canada.

ROYALL, R. M. (1986). Model robust confidence intervals using maximum likelihood estimators. *Internat. Statist. Rev.* **54** 221–226.

SMALL, C. G., WANG, J. and YANG, Z. (2000). Eliminating multiple root problems in estimation (with discussion). *Statist. Sci.* **15** 313–341.

WALD, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20** 595–601.

WU, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.* **14** 1261–1350.