COMMENT

Stephen M. S. Lee and G. Alastair Young

The University of Hong Kong and University of Cambridge

1. Introduction

This article by Chuang and Lai provides a very nice summary of hybrid resampling methods and their properties. We believe that it contributes significantly to the establishment of an effective and reliable resampling methodology for the construction of accurate confidence intervals. While congratulating the authors on the clarity of their discussion, which in particular provides a useful presentation of conventional bootstrap methods as a special case of hybrid resampling, we should like to remark on some specific aspects of the methodology.

2. Bootstrap Inconsistency

Of major focus in recent times has been the establishment of resampling methods of inference which are valid, in the sense of consistency, even when the conventional bootstrap fails, and especially for circumstances where it fails for particular values of the model parameter, as in the first-order autoregressive example of Section 5 of the paper. A key tool for this purpose has been the "m out of n" bootstrap, as examined by Bickel, Götze and van Zwet (1997). Of interest would be a detailed comparison of the properties of hybrid resampling methods with those of the m out of n bootstrap. A potential disadvantage of the m out of n bootstrap is that, while it may provide a consistent estimate, the accompanying efficiency losses noted by Bickel, Götze and van Zwet (1997) might, in examples such as those considered by Chuang and Lai, produce an order of coverage error inferior to that given by hybrid resampling. Whether hybrid resampling is to be generally preferred, in terms of efficiency loss or its remedy, to the m out of n bootstrap remains an open question.

3. Choice of Root

Historically, much focus within the bootstrap literature has involved the issues and benefits of studentization and/or prepivoting, the latter taken to include ideas of bootstrap calibration and "double bootstrapping". The paper of Chuang and Lai presents an interesting idea on the choice of the root $R(X, \theta)$ used in construction of the confidence interval, which we believe is practically important, and worthy of further development. They suggest that stability of the hybrid resampling approach in small to moderate sample sizes can be enhanced by use of a hybrid pivot $R(X, \theta)$, which depends on the value of θ, as exemplified by (3.8) and (3.15) of the paper. As the discussion following (3.15) of the paper makes clear, use of a modified form of pivot can be made to automatically incorporate both studentization and prepivoting ideas. Implementation depends, however, on interpretation, for the problem at hand, of what constitutes '$\widehat{\theta}$ is not too far from θ'. In the examples given in the paper the authors give no specific guidance on how this question should be met. Some adaptive procedure, based on empirical assessment of the stability of the hybrid root $R(X, \theta)$ seems natural.

4. Non-parametric Inference

We were particularly interested to read the authors' recommendations, in Section 6 of the paper, on the choice of resampling family recommended for the hybrid resampling methodology in nonparametric problems. Their discussion advocates a particular one-parameter tilting family of distributions, as given by (6.1) of the paper. We have argued in Lee and Young (1999a) the advantages of such a tilting family in the construction of nonparametric likelihood ratio confidence intervals. The simplicity of the tilting family allows us to propose and analyze various asymptotic and bootstrap correction techniques as a means of producing, via the nonparametric likelihood, confidence intervals of low coverage error, comparable to those obtained by more computationally-intensive methods such as the iterated bootstrap. Direct comparison of these methods with hybrid resampling methods would also be of practical interest.

5. Iterated Hybrid Resampling

Chuang and Lai discuss the possibility of applying the hybrid resampling method to both non-pivotal and approximately pivotal $R(X, \theta)$, to achieve both second order accuracy and correctness of the hybrid confidence region; the correlation coefficient example is used to illustrate the latter. In fact, it is possible to prove rather stronger results about the effect of using hybrid resampling, rather than conventional bootstrap resampling. We provide here a brief description of these results; full details will be given elsewhere.

Suppose we assume the smooth function model, where $\widehat{\theta}$ is a smooth function of sample means, and consider construction of a (one-sided) nonparametric confidence set for θ from the non-pivotal root $R(X, \theta) = \sqrt{n}(\widehat{\theta} - \theta)$. Then the conventional bootstrap, which resamples from the empirical distribution function of the observed sample, yields a coverage error of order $O(n^{-1/2})$. Hybrid resampling improves this error to one of order $O(n^{-1})$. On the other hand, if we proceed from the approximately pivotal root $R(X, \theta) = (\widehat{\theta} - \theta)/\widehat{\sigma}$, the conventional bootstrap yields coverage error of order $O(n^{-1})$. Hybrid resampling improves this to $O(n^{-3/2})$, which is what is achieved by a conventional bootstrap calibration or double bootstrap method; see Martin (1990). The latter method provides, in our view, a satisfactory pragmatic solution to the problem of producing nonparametric confidence intervals of low coverage error, but with appropriate stability, which may not be enjoyed by other more sophisticated bootstrap procedures. It will be interesting to undertake a more extensive empirical analysis of how hybrid resampling intervals and the double bootstrap compare in practice. Evidence presented by Chuang and Lai for the correlation coefficient example suggests that hybrid resampling ought to be capable of challenging the double bootstrap gold standard.

In terms of computational expense, hybrid resampling is clearly preferable to the double bootstrap, as it only requires one level of resampling. But if we are willing to undertake a second level of resampling, might it not be advantageous to iterate the hybrid resampling, rather than use the conventional double bootstrap?

We sketch here the theoretical effects of iterated hybrid resampling. For simplicity of presentation, consider a nonpivotal root $R(X, \theta)$, and denote by $G(\cdot, \theta)$ its sampling distribution, as estimated by the hybrid resampling scheme using the tilting family (6.1) of Chuang and Lai's paper. As we have noted, the confidence limit based on the appropriate quantile of $G(\cdot\ , \theta)$ typically has coverage error of order $O(n^{-1})$. The concept of iteration is that an improved confidence limit can be obtained from the sampling distribution of the root $R_1(X, \theta) = G(R(X, \theta), \theta)$. There are two natural ways of estimating this sampling distribution: (a) by conventional bootstrapping, or (b) by hybrid resampling again.

It turns out that (a) yields a confidence limit with coverage error of order $O(n^{-3/2})$, an improvement in order terms over the $O(n^{-1})$ coverage error obtained by the conventional double bootstrap, and of the same order as the coverage error obtained if the sampling distribution $G(\cdot\ , \theta)$ is estimated by the conventional bootstrap, but hybrid sampling used to estimate the sampling distribution of $R_1(X, \theta)$. However, the benefits of hybrid resampling over conventional bootstrapping ensure that possibility (b) yields a confidence limit with coverage error of order $O(n^{-2})$. This means that a two-level resampling analysis which uses hybrid resampling at both levels, rather than conventional uniform resampling, produces an interval whose error is reduced by two orders of magnitude. Stated simply, single level hybrid resampling has the theoretically beneficial effect of conventional double bootstrapping, while a double level hybrid resampling has an effect similar to a conventional "quadruple" bootstrap.

Of course, as Chuang and Lai discuss, hybrid resampling requires rather more sophisticated computation than ordinary bootstrapping. In their notation, the sampling distribution of the root $R(X, \theta)$ must be simulated under $\widehat{F}_\theta$, for a set of different θ values, which amounts to weighted bootstrapping if the tilting family (6.1) is employed. Ordinary boostrapping just requires simulation of the sampling distribution of $R(X^*, \widehat{\theta})$ under the empirical distribution $\widehat{F}$. Iterated hybrid resampling would presumably involve weighted bootstrapping at a number of selected values of θ at each resampling level, but will still be substantially less expensive computationally than the quadruple bootstrap.

6. Monte Carlo Implementation

We should finally like to make some brief remarks on the conventional approach, as adopted in this paper, to the need for Monte Carlo simulation in the implementation of resampling methods of inference. Traditionally, the prevailing attitude has automatically been to seek an implementation which uses the maximum number of Monte Carlo samples possible, within the limitations imposed by the need to control the overall computational burden. We have recently challenged this attitude in showing that there may sometimes be advantage, in terms of coverage accuracy, in more careful control of the Monte Carlo simulation. In Lee and Young (1999b) we provide an analysis of the coverage accuracy of the calibrated percentile method confidence set, which takes into account both the inherent bootstrap and Monte Carlo errors. We demonstrate that by suitable control of the size of the Monte Carlo simulation we may actually reduce the order of coverage error below that of the 'infinite simulation' interval. In times of readily available computational power, it seems appropriate to think more deeply about implementation, not just in terms of overall computational expense.

Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong.
E-mail:

Statistical Laboratory, University of Cambridge, 16 Mill Lane, Cambridge CB2 1SB, U.K.
E-mail: g.a.young@statslab.cam.ac.uk

REJOINDER

Chin-Shan Chuang and Tze Leung Lai

Although it has been two decades since Efron's seminal paper on bootstrap methods, there are still many unresolved problems in resampling methodology. As noted by Bickel, Götze and van Zwet (1997), "Practical anecdotal experience seems to support theory in the sense that the bootstrap generally gives reasonable answers but can bomb". Indeed, our motivation for developing hybrid resampling