Minimum Variance Unbiased Estimation based on Bootstrap Iterations

Stephen M.S. Lee*

Department of Statistics and Actuarial Science, The University of Hong Kong,

Pokfulam Road, Hong Kong

Email: smslee@hkusua.hku.hk

Abstract

Practical computation of the minimum variance unbiased estimator (MVUE) is often a difficult, if not impossible, task, even though general statistical theory assures its existence under regularity conditions. We propose a new approach, based on infinitely many iterations of bootstrap bias correction, to calculating the MVUE approximately. A numerical example is given to illustrate the effectiveness of our new approach.

Key words and phrases: bias; bootstrap iteration; Monte Carlo; MVUE.

1 Introduction

In cases where the maximum likelihood estimator (MLE) is a function of a complete sufficient statistic, an exact MVUE may be obtained by eliminating the bias of the MLE. For example, an MLE can be bias-corrected by subtracting an estimated bias term, but the result is usually not satisfactory. In this paper, we propose a new approach that makes use of infinitely many bootstrap iterations to completely eliminate the bias of the MLE, thus yielding the exact MVUE. Application of the bootstrap to estimate the bias was first conceived by Efron (1979). In general the bias of an estimator can be reduced by an order of $O(n^{-1})$, where n denotes the sample size, by subtracting the bootstrap bias estimate. Each iteration of the bootstrap reduces the bias successively by one order. In principle, we may iterate the bootstrap indefinitely so as to completely eliminate the bias, at least in an asymptotic sense. A unified account of the iterated bootstrap is given by Hall and Martin (1988). In practice, bootstrap iterations can be done

by nested levels of Monte Carlo simulations, which become computationally prohibitively expensive even for a moderate number of iterations. With the modern computer power, indefinitely iterating the bootstrap using simulation is infeasible. Moreover, the necessarily finite simulation size incurs simulation error which makes the algorithm increasingly unstable as the number of iterations increases. Chan and Lee (2001) propose an exact algorithm for nonparametric bootstrap bias elimination, which works for small sample sizes and does not involve any Monte Carlo simulation. Our new approach to computing the MVUE extends Chan and Lee's (2001) algorithm to a classical parametric context where bias correction is made to an MLE. It consists of Monte Carlo simulation of first-level parametric bootstrap samples, followed by implementation of Chan and Lee's (2001) nonparametric algorithm for bias elimination. It involves only one level of Monte Carlo simulation, which is considerably more stable and computationally more efficient than the conventional Monte Carlo approach. In practice, the numerical error of our approach is determined solely by the precision of the computer and the number of first-level Monte Carlo samples that we are prepared to simulate.

Section 2 discusses the theory of bias correction of MLE based on bootstrap iterations, and establishes its relevance to the computation of the exact MVUE. Our new approach is introduced in Section 3. Section 4 illustrates our method with gamma data in a numerical example. Concluding remarks are given in Section 5.

2 Theory

Let $\mathcal{X} = (X_1, \ldots, X_n)$ be a random sample drawn from a distribution F_θ in a parametric family indexed by $\theta \in \Theta \subset \mathbb{R}^d$. Under regularity conditions, the likelihood function of θ is maximized uniquely at $\hat{\theta}_n$, which is defined as the MLE of θ. Note that $\hat{\theta}_n$ depends on the observed sample $\mathcal{X}$. For a generic sample $\mathcal{Y} = (Y_1, \ldots, Y_n)$, we may define $\theta(\mathcal{Y})$ to be the MLE of θ based on $\mathcal{Y}$, so that $\theta(\mathcal{X}) = \hat{\theta}_n$. Let $\mathcal{X}^{(0)} = \mathcal{X}$, and $\mathcal{X}^{(j+1)}$ be a generic nonparametric bootstrap sample drawn from $\mathcal{X}^{(j)}$, for $j = 0, 1, 2, \ldots$, such that $\mathcal{X}^{(j+1)}$ constitutes a random sample of n observations drawn with replacement from $\mathcal{X}^{(j)}$.

Theorem 1.3 of Hall (1992) establishes a general formula for the bias-corrected estimate based on j bootstrap iterations, which is given by

$$\tilde{\theta}_j = \sum_{i=0}^{j} (-1)^i \binom{j+1}{i+1} \mathbb{E}[\theta(\mathcal{X}^{(i)}) \mid \mathcal{X}]. \tag{1}$$

It can be shown that the bias of $\tilde{\theta}_j$ is of the order $O(n^{-(j+1)})$, so that the bias vanishes as $j \uparrow \infty$ in an asymptotic sense.

We describe below two theorems which suggest two slightly different but related characterizations of the MVUE. These characterizations will be employed to develop our iterated bootstrap algorithm for calculating the MVUE.

Theorem 1 *(Lehmann-Scheffé) If there exists an unbiased estimator of θ that is a function of a complete sufficient statistic for θ, then it is the MVUE, which is unique almost surely under each θ.*

Theorem 2 *(Rao-Blackwell) Let U be an unbiased estimator of θ and T a sufficient statistic for θ. Then $\phi(T) \equiv \mathbb{E}_\theta[U|T]$ is unbiased for θ and $\mathrm{Var}_\theta \phi(T) \leq \mathrm{Var}_\theta U$.*

3 Computation of MVUE

Let T be a complete sufficient statistic. Abusing our notation slightly, $\tilde{\theta}_\infty = \lim_{j \to \infty} \tilde{\theta}_j$ is exactly unbiased and it follows from Theorems 1 and 2 that $\mathbb{E}[\tilde{\theta}_\infty | T]$ is the MVUE. This suggests an algorithm for computing MVUE, which requires preliminary simulation of Monte Carlo samples from the conditional distribution of $\mathcal{X}$ given T, followed by a separate iterated bootstrap bias correction for each conditional sample.

Define

$$\mathcal{L}(n) = \left\{ (l_1, \ldots, l_n) \in \mathbb{Z}^n : l_1 \geq l_2 \geq \cdots \geq l_n \geq 0, \sum_{i=1}^{n} l_i = n \right\}.$$

Identify a typical, possibly iterated, bootstrap sample with a unique $l \in \mathcal{L}(n)$, such that $l_1, \ldots, l_n$ give the "ordered" frequencies of appearances of the original observations in the bootstrap sample. In particular, the state $\mathbf{1} = (1, \ldots, 1)$ corresponds to the original sample $\mathcal{X}$. Bootstrap iterations may be viewed as a Markov chain process on $\mathcal{L}(n)$, since each bootstrap sample depends stochastically only on the bootstrap sample drawn at the preceding level. Denote the transition matrix for the Markov chain by $P_n = [p_n(s,t) : s, t \in \mathcal{L}(n)]$, where $p_n(s,t)$ is the transition probability from state s to state t and can be found by routine combinatoric calculations. The i-step transition probabilities can be obtained from the product matrix $P_n^i = [p_n^{(i)}(s,t) : s, t \in \mathcal{L}(n)]$. Denote by $\mathcal{M}(l)$ the set of all distinct permutations of $l \in \mathcal{L}(n)$. Let M_l be the size of $\mathcal{M}(l)$. For each $m = (m_1, \ldots, m_n) \in \mathcal{M}(l)$, define $\mathcal{X}(m)$ to be the sample which contains m_i replicates of X_i for $i = 1, \ldots, n$. Then we have

$$\tilde{\theta}_j = (j+1)\hat{\theta}_n + \sum_{i=1}^{j} \binom{j+1}{i+1}(-1)^i \sum_{l \in \mathcal{L}(n)} p_n^{(i)}(\mathbf{1}, l) M_l^{-1} \sum_{m \in \mathcal{M}(l)} \theta(\mathcal{X}(m)), \quad (2)$$

which is directly computable. Details of (2) and the properties of $p_n^{(i)}(s,t)$ can be found in Chan and Lee (2001).

We now describe an algorithm for approximating the MVUE. First generate a large number of independent Monte Carlo samples, $\mathcal{X}_1, \ldots, \mathcal{X}_B$ say, from the conditional distribution of $\mathcal{X}$ given a complete sufficient statistic T. For each j, approximate the estimate $\mathbb{E}[\tilde{\theta}_j | T]$ by

$$(j+1)\hat{\theta}_n + \sum_{i=1}^{j} \binom{j+1}{i+1} (-1)^i \sum_{l \in \mathcal{L}(n)} p_n^{(i)}(\mathbf{1}, l) M_l^{-1} \sum_{m \in \mathcal{M}(l)} B^{-1} \sum_{b=1}^{B} \theta(\mathcal{X}_b(m)), \quad (3)$$

where $\mathcal{X}_b(m)$ has the same definition as $\mathcal{X}(m)$ except that the observations $\mathcal{X}$ are now replaced by $\mathcal{X}_b$. Compute the expression (3) sequentially for $j = 1, 2, \ldots$ until numerical convergence is detected. The limiting value can then be taken as an approximation to the MVUE. Clearly a large B is required for a better approximation. Computational efficiency and accuracy of the above algorithm thus hinge upon the sample size n and the ease with which conditional samples $\mathcal{X}_1, \ldots, \mathcal{X}_B$ can be generated. The number of levels j has relatively little effects on its efficiency.

4   Examples

Consider a random sample $\mathcal{X} = (X_1, \ldots, X_n)$ from gamma $(\theta, 1)$, where θ denotes the shape parameter. Then $T = \prod_{i=1}^{n} X_i$ is complete sufficient for θ. Let $V = (X_1, \ldots, X_{n-1})$, so that $(X_1, \ldots, X_n) \mapsto (T, V)$ defines a one-one transformation φ. It is clear that the conditional density of V given T is proportional to

$$g_2(T, v) = g_2(T, x_1, \ldots, x_{n-1}) = \left(\prod_{i=1}^{n-1} x_i \right)^{-1} \exp \left\{ -\sum_{i=1}^{n-1} x_i - T / \prod_{i=1}^{n-1} x_i \right\}.$$

Choose $h(x_1, \ldots, x_{n-1})$ to be the joint density function of $n-1$ independent exponential random variables each of unit rate, so that $h(x_1, \ldots, x_{n-1}) = \exp\{-\sum_{i=1}^{n-1} x_i\}$. Note that $g_2(T, x_1, \ldots x_{n-1})/h(x_1, \ldots x_{n-1}) \leq T^{-1} e^{-1}$, which serves as a suitable upper bound for application of the acceptance-rejection algorithm to generate a random variate $W = (W_1, \ldots, W_{n-1})$ from the conditional density of V given T. Specifically, the algorithm generates $n - 1$ independent exponential random variables $V_1, \ldots, V_{n-1}$ of unit rate, calculates $R = T \exp(1 - T / \prod_i V_i) / \prod_i V_i$, and accepts $W = V$ only when a randomly generated uniform $(0, 1)$ variable $U \leq R$. The required conditional sample is then $\varphi^{-1}(T, W) = (W_1, \ldots, W_{n-1}, T / \prod_{i=1}^{n-1} W_i)$. The MVUE is approximated by evaluating (3) based on B independent conditional samples obtained by the above procedure. Note that the acceptance-rejection algorithm tends to yield samples with observations not too close to zero. For if some of the V_i's are close to zero, then R is also close to zero and we are likely to reject this set of V_i's.

In the simulation study we took $n = 5$ and drew $\mathcal{X}$ from gamma $(2, 1)$. The MLE was calculated to be $\hat{\theta}_n = 2.6262$. Table 1 shows the values of $\mathbb{E}[\tilde{\theta}_j | T]$,

Table 1: Example: gamma $(\theta, 1)$ random sample of size 5.

B	$\mathbb{E}[\theta_1|T]$	$\mathbb{E}[\theta_2|T]$	$\mathbb{E}[\theta_5|T]$	$\mathbb{E}[\theta_{10}|T]$	$\mathbb{E}[\theta_{20}|T]$	$\mathbb{E}[\theta_{25}|T]$
1000	2.5598	2.5478	2.5459	2.5457	2.5455	2.5455
2000	2.5610	2.5491	2.5471	2.5470	2.5468	2.5468
5000	2.5609	2.5489	2.5469	2.5468	2.5466	2.5466
10000	2.5614	2.5495	2.5474	2.5473	2.5472	2.5472
20000	2.5615	2.5497	2.5476	2.5476	2.5474	2.5474
50000	2.5620	2.5501	2.5481	2.5480	2.5479	2.5479
100000	2.5620	2.5501	2.5481	2.5480	2.5479	2.5479

Monte Carlo approximated $\theta_j^\dagger$:

B	$\theta_1^\dagger$	$\theta_2^\dagger$
5000	2.5407	2.4676
10000	2.5320	2.4354

approximated using different numbers of conditional samples, for up to $j = 25$ iterations. We see that increasing B has the effect of stabilizing our estimates at each iteration level. An equilibrium point seems to be reached when B and j increase beyond 50,000 and 20 respectively. It is evident that the MVUE should be around 2.5479. Table 1 also gives the Monte Carlo approximated $\theta_j^\dagger$ for $j = 1, 2$, based on B first- and B^2 second-level parametric bootstrap samples. Here each parametric bootstrap sample was drawn from a gamma distribution with shape parameter estimated from a sample drawn at the preceding level. In comparison with our estimates, the first-level Monte Carlo estimate is closer to the MVUE than the second-level one, which elucidates the undesirable instability of Monte Carlo iterations. One should be cautious about iterating Monte Carlo simulations, especially if other more accurate methods, such as our hybrid procedure in the present context, are available.

We remark that our procedure applies generally to problems in a parametric setup where a complete sufficient statistic is available and a Monte Carlo algorithm can be found to generate conditional samples. In particular, it adapts readily to multivariate settings to compute a multivariate MVUE. We have re-investigated the gamma example with unknown shape and scale parameters and computed the MVUE's of the two parameters by our procedure. Detailed findings will be reported elsewhere. Briefly, the required samples, conditioned on the sufficient statistic $(\sum_i X_i, \prod_i X_i)$, had to be simulated by the more computationally involved Metropolis-Hastings algorithm and the numerical results were not as stable as those from the gamma $(\theta, 1)$ example. The latter may be attributed to an intrinsic feature of the bootstrap, which has a positive probability to generate a sample of n identical observations and thus poses a numerical problem when

$\theta(\mathcal{Y})$ depends to some extent on the sample variance of $\mathcal{Y}$.

5   Conclusion

The classical problem of MVUE derivation has been extensively studied in the literature. However, in many situations the knowledge of the existence of the MVUE does not automatically lead to a simple analytic or numerical method for its computation. The contemporary bootstrap iteraton method has been shown to play a new and constructive role in this classical area. It provides a credible and efficient solution when there is no explicit formula available for the MVUE. The method involves very limited Monte Carlo effort, and has an efficiency largely independent of the number of iterations. It is particularly useful in cases where n is small. Our simulation findings are encouraging. The iterated estimates converge rapidly to the correct MVUE. On the contrary, conventional Monte Carlo simulation of bootstrap iterations fails to give an accurate and stable answer. We believe that our proposed algorithms will be applicable to large samples before long, with the rapid advance of computer power. Lastly, we remark that the hybrid procedure introduced in Section 3 is highly problem-specific and requires clever Monte Carlo techniques to uphold its efficiency.

References

[1] Chan, K.Y.F. and Lee, S.M.S. (2001). An exact iterated bootstrap algorithm for small-sample bias reduction. *Comput. Statist. Data Anal.*, **36**, 1–13.

[2] Efron, B. (1979). Bootstrap methods: another look at the jacknife. *Ann. Statist.*, **7**, 1-26.

[3] Hall, P. and Martin, M.A. (1988). On bootstrap resampling and iteration. *Biometrika*, **75**, 661-671.

[4] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer-Verlag: New York.