An Exact Iterated Bootstrap Algorithm For Small-Sample Bias Reduction

Kenny Yuen Fai Chan
Department of Statistics and Actuarial Science, The University of Hong Kong
Pokfulam Road
Hong Kong
E-mail: h9407526@hkusua.hku.hk

Stephen Man Sing Lee
Department of Statistics and Actuarial Science, The University of Hong Kong
Pokfulam Road
Hong Kong
E-mail: smslee@hkuxa.hku.hk

1. Introduction

By considering bootstrap iteration as a Markov process, we propose an exact algorithm for its implementation in the context of small-sample bias reduction. The algorithm caters for any number of bootstrap iterations without the need for extensive Monte Carlo resampling , in the context of bootstrap bias reduction for small samples. The performance of high-level bootstrap iteration is then investigated.

2. Exact computation of bias-corrected estimator based on j bootstrap iterations

Let $\mathcal{X} = \{X_1, \ldots, X_n\}$ be a random sample drawn from an unknown distribution F_0, and F_1 be the empirical distribution of $\mathcal{X}$. Let $\theta = t(F_0)$ be a functional of F_0. Then a natural estimator of θ is $\hat{\theta} = t(F_1)$, provided t is defined on the space of all discrete distributions. Hall (1992) gives a general formula for the bias-corrected estimator of θ based on j bootstrap iterations, namely

$$\hat{\theta}_j = \sum_{i=1}^{j+1} \binom{j+1}{i} (-1)^{i+1} \mathbb{E}\{t(F_i) \mid F_1\},$$

for $j = 1, 2, \ldots$, where F_i denotes the empirical distribution of a random sample of size n drawn from F_{i-1}.

We see that bias reduction by j bootstrap iterations involves calculation of the conditional expectations $\mathbb{E}\{t(F_i)|F_1\}$. Approximation to the latter is usually obtained by Monte Carlo simulation of nested bootstrap resamples, which becomes impractical if j is large. Following the notation of Fisher and Hall (1991), define $\mathcal{L}(n)$ to be the set of all distinct n-tuples $(l_1, \ldots, l_n)$ such that the l_i's are nonnegative integers satisfying $l_1 \geq l_2 \geq \cdots \geq l_n$ and $\sum_{i=1}^n l_i = n$. For each $l \in \mathcal{L}(n)$, define $\mathcal{M}(l)$ to be the set of all distinct permutations of l. Note that each, possibly iterated, bootstrap resample arising from $\mathcal{X}$ can be uniquely identified with an n-tuple $(k_1, \ldots, k_n)$ in $\mathcal{M}(l)$ for a unique $l \in \mathcal{L}(n)$, where k_i gives the number of appearances of X_i in the bootstrap resample. A jth level bootstrap resample relates stochastically to $\mathcal{X}$ by the j-step transition probabilities, which are obtainable as components of the product matrix $P_n^j = [p_n^{(j)}(s,t) : s, t \in \mathcal{L}(n)]$. For each $k = (k_1, \ldots, k_n) \in \mathcal{K}(n)$, define $F_{(k)}$ to be the discrete distribution that places a mass of k_i/n on the observation X_i, $i = 1, \ldots, n$. Then we have

$$\hat{\theta}_j = \sum_{i=1}^{j+1} \binom{j+1}{i} (-1)^{i+1} \sum_{l \in \mathcal{L}(n)} p_n^{(i-1)}(\mathbf{1}, l) |\mathcal{M}(l)|^{-1} \sum_{k \in \mathcal{M}(l)} t(F_{(k)}),$$

which follows from the general formula and is directly computable. The factor $|\mathcal{M}(l)|^{-1}$ reflects the equal weights shared by bootstrap resamples from the same set $\mathcal{M}(l)$. Our algorithm computes $\hat{\theta}_j$ direct from our equation.

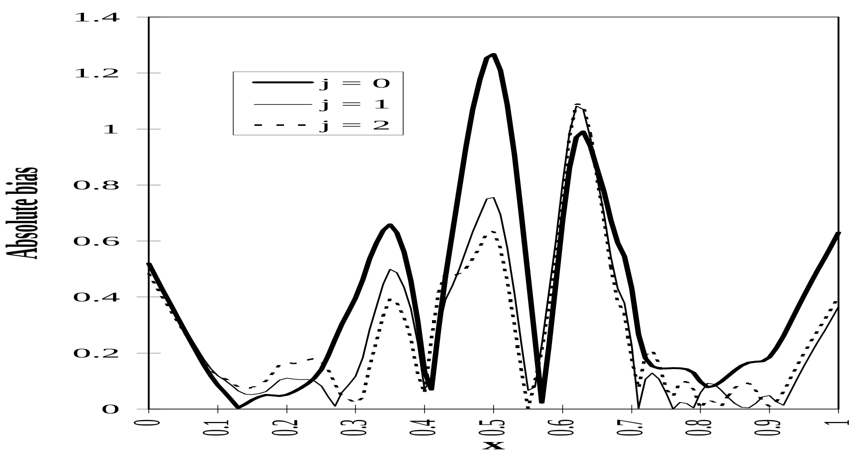
3. Simulation studies

Simulation studies are conducted to examine the effectiveness of our algorithm in reducing the bias of the Nadaraya-Watson estimate. Hall and Presnell (1999) apply their b-bootstrap algorithm to reduce the bias of the Nadaraya-Watson estimator. Our example is taken from their simulation study but focuses on a smaller sample size without fixed design points.

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent data pairs such that X_i is uniformly distributed over $[0,1]$ and Y_i is normally distributed with mean $m(X_i)$ and variance $1/2$ conditional on X_i, where $m(x) = 5\{1 - x + \exp[-100(x-1/2)^2]\}$. Consider the Nadaraya-Watson estimator of the mean curve $m(x)$, defined as

$$\hat{m}(x) = \sum_{i=1}^{n} K_h(x - X_i)Y_i / \sum_{i=1}^{n} K_h(x - X_i),$$

where $K_h(u) = h^{-1}K(u/h)$, $K(u) = (2\pi)^{-1/2}\exp(-u^2/2)$ and the bandwidth h is set to 0.04. Denote by $\hat{m}_j$ the bias-corrected version of $\hat{m}$ based on j bootstrap iterations. The result is based on 50 random samples and sample size n is 10 throughout the study. The curves for $j \geq 2$ are visually indistinguishable. We see that bootstrap iteration is effective in reducing the bias of $\hat{m}(x)$, especially for moderate values of x. Convergence of the process is reached at approximately the second level and any further iterations do not improve the estimator significantly.



REFERENCES

Fisher, N.I. and Hall, P. (1991). Bootstrap algorithms for small samples. *J. Statist. Planning and Inference*, **27**, 157-169.

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

Hall, P. and Presnell, B. (1999). Intentionally biased bootstrap methods. *J. Roy. Statist. Soc. Series B*, **61**, 143-158.