Computational Statistics and Data Analysis 52 (2008) 5136–5142

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Short communication

Bootstrap quantile estimation via importance resampling

Jiaqiao Hu[a], Zheng Su[b,*]

[a] Department of Applied Mathematics and Statistics, State University of New York, Stony Brook, NY 11794, United States
[b] Genentech Inc., South San Francisco, CA 94080, United States

ARTICLE INFO

Article history:
Received 3 August 2007
Received in revised form 22 May 2008
Accepted 23 May 2008
Available online 3 June 2008

ABSTRACT

We propose an adaptive importance resampling algorithm for estimating bootstrap quantiles of general statistics. The algorithm is especially useful in estimating extreme quantiles and can be easily used to construct bootstrap confidence intervals. Empirical results on real and simulated data sets show that the proposed algorithm is not only superior to the uniform resampling approach, but may also provide more than an order of magnitude of computational efficiency gains.

1. Introduction

Bootstrap confidence intervals are widely used to quantify uncertainties in the inferences that can be drawn from a data sample. To construct bootstrap confidence intervals, the bootstrap quantiles of some statistic need to be evaluated, for which the crude Monte Carlo uniform resampling method is frequently used. Importance resampling was first suggested by Johns (1988) and Davison (1988) as a way of improving the efficiency of Monte Carlo simulation methods for calculating bootstrap quantiles. Recently, Do et al. (2001) also proposed an empirical importance resampling approach for constructing bootstrap confidence intervals for proportional hazards regression. Under some regularity conditions, these approaches provide asymptotically approximately optimal importance resampling weights for normalized statistics. In parallel to the estimation of bootstrap quantiles, various importance sampling approaches have also been proposed for estimating bootstrap tail probabilities, including the two-stage variance minimization method of Do and Hall (1991) and different saddlepoint methods (cf. e.g., Chen and Do (1994), Hesterberg (1996) and Lee and Wong (2002)). The overall idea behind these approaches is to generate samples from a suitably chosen resampling measure so that the variance of the resultant estimator can be significantly reduced. More recently, the two-stage method of Do and Hall (1991) was generalized by Hu and Su (2008) to a multi-step variance minimization procedure for general statistics, where the idea was based on an iterative algorithm proposed in Rubinstein (1997) for estimating rare event probabilities in discrete event systems.

In this paper, we propose a new computational approach, which extends the sequential variance minimization approach of Hu and Su (2008) to the setting of estimating bootstrap quantiles. The motivation behind the approach is based on a simple interpretation of one important result established in Johns (1988). The key observation is that the problem of quantile estimation is essentially the dual problem of estimating bootstrap tail probabilities, so that the set of importance resampling weights derived for estimating tail probabilities can also be used as good resampling weights for estimating bootstrap quantiles. We remark that the approach outlined in Johns (1988) relies on asymptotic normality approximations, and is not directly applicable to general statistics, whereas the approach considered in this paper does not resort to normal asymptotics and can be easily used for efficient estimation of bootstrap quantiles for general statistics. Our simulation results on a real survival data set show that the proposed algorithm is not only superior to the uniform resampling approach, but may also provide more than an order of magnitude of computational efficiency gains.

* Corresponding author.
E-mail address: su.zheng@gene.com (Z. Su).

0167-9473/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2008.05.022

The rest of the paper is organized as follows. In Section 2, we provide a detailed description of the proposed algorithm for estimating bootstrap quantiles. In Section 3, we carry out simulation studies on a real survival data set to compare the performance of the proposed algorithm with that of the uniform resampling approach. Finally, we conclude this paper in Section 4.

2. Bootstrap quantile estimation for general statistics

Consider estimating the bootstrap tail probability $\rho = P(T(\mathbf{Z}) \le x|\mathbf{X})$ based on a random sample $\mathbf{X} = \{X_1, \ldots, X_n\}$, where T is a statistic of interest as a function of the random vector $\mathbf{Z}$, $\mathbf{Z}$ is a uniform resample from $\mathbf{X}$, and x is a prespecified threshold value. In the uniform resampling method, each sample value X_i is assigned a probability weight n^{-1} (i.e., all X_i's will have the same probability of being selected), and $P(T(\mathbf{Z}) \le x|\mathbf{X})$ is simply estimated by its sample average approximation $\rho_N^* = N^{-1}\sum_{j=1}^{N} I\{T(\mathbf{Z}_j) \le x\}$, where $\mathbf{Z}_j$ represents the jth uniform resample generated according to the discrete uniform distribution over $\{X_1, \ldots, X_n\}$, and N is the total number of uniform resamples generated.

The importance resampling approach generalizes the uniform resampling by allowing a different probability weight p_i for each sample value X_i, with $\sum p_i = 1$. In particular, for a given set of weights $p_1, \ldots, p_n$, an unbiased estimator of $P(T(\mathbf{Z}) \le x|\mathbf{X})$ can be obtained via a change of measure:

$$\rho_N = \frac{1}{N}\sum_{j=1}^{N} I\{T(\mathbf{Z}_j) \le x\}\frac{\prod_{i=1}^{n}(1/n)^{M_{j,i}}}{\prod_{i=1}^{n}(p_i)^{M_{j,i}}} = \frac{1}{N}\sum_{j=1}^{N} I\{T(\mathbf{Z}_j) \le x\}\prod_{i=1}^{n}(np_i)^{-M_{j,i}},$$

where $\mathbf{Z}_j$ denotes the jth resample drawn from the importance resampling distribution $p_1, \ldots, p_n$, and $M_{j,i}$ is the number of times X_i appears in the jth resample $\mathbf{Z}_j$. It is easy to see that the variance of the estimator is given by

$$\mathrm{Var}(\rho_N) = \frac{1}{N}E\left[I\{T(\mathbf{Z}) \le x\}\prod_{i=1}^{n}(np_i)^{-M_i}\,\middle|\,\mathbf{X}\right] - \frac{1}{N}\rho^2,$$

where M_i signifies the number of times X_i appears in a random resample from $\mathbf{X}$. The essence of importance sampling is to find the set of optimal probability weights $p_1, \ldots, p_n$ so that the variance of the estimator $\mathrm{Var}(\rho_N)$ is minimized. Note that minimizing $\mathrm{Var}(\rho_N)$ is equivalent to minimizing the quantity

$$E\left[I\{T(\mathbf{Z}) \le x\}\prod_{i=1}^{n}(np_i)^{-M_i}\,\middle|\,\mathbf{X}\right]. \tag{1}$$

Now consider the dual problem of estimating the ρth quantile x_ρ of the statistic T, where $P(T(\mathbf{Z}) \le x_\rho|\mathbf{X}) = \rho$. The main difficulty of applying importance sampling technique in quantile estimation is that a specified quantile usually cannot be easily represented as the expectation of a statistic with known distribution. The direct uniform resampling method for estimating x_ρ would be to generate i.i.d. resamples $\mathbf{Z}_1, \ldots, \mathbf{Z}_N$ from the discrete uniform distribution, order the statistics $T_{(1)} \le T_{(2)} \le \cdots \le T_{(N)}$ from the smallest to the largest, and then take the ρNth order statistic $T(\mathbf{Z}_{\lfloor \rho N \rfloor})$ as an estimate of the true quantile x_ρ, where $T_{(i)} = T(\mathbf{Z}_{(i)})$ for all $i = 1, \ldots, N$ and $\lfloor a \rfloor$ is the integer part of a. When importance sampling is used in estimating x_ρ, we may generate i.i.d. resamples $\mathbf{Z}_1, \ldots, \mathbf{Z}_N$ according to the resampling weights $p_1, \ldots, p_n$. Let $f_{\mathbf{Z}}(\cdot; \mathbf{p})$ be the distribution function of the resample $\mathbf{Z}$ generated from $\mathbf{X}$ according to the probability weights $\mathbf{p} = (p_1, \ldots, p_n)$, we have

$$\begin{aligned}\rho = P\left(T(\mathbf{Z}) \le x_\rho\right) &= E\left[I\{T(\mathbf{Z}) \le x_\rho\}\right] \\ &= E_{\mathbf{p}}\left[I\{T(\mathbf{Z}) \le x_\rho\}\prod_{i=1}^{n}(np_i)^{-M_i}\right] \\ &\approx \frac{1}{N}\sum_{j=1}^{N} I\{T_{(j)} \le x_\rho\}\prod_{i=1}^{n}(np_i)^{-M_{(j),i}}\end{aligned} \tag{2}$$

where $E_{\mathbf{p}}[\cdot]$ is the expectation taken with respect to $f_{\mathbf{Z}}(\cdot; \mathbf{p})$, $M_{(j),i}$ indicates the number of times X_i appears in the resample $\mathbf{Z}_{(j)}$ obtained according to the distribution $f_{\mathbf{Z}}(\cdot; \mathbf{p})$, and the approximation follows by replacing the expectation with its sample average. Thus, if we further define the quantity

$$S_r = \frac{1}{N}\sum_{j=1}^{r}\prod_{i=1}^{n}(np_i)^{-M_{(j),i}} \tag{3}$$

for all $r = 1, \ldots, N$, then Eq. (2) implies that the ρth quantile x_ρ of the statistic T under the original measure can be approximated by the Rth order statistic $T_{(R)}$ calculated based on the resamples from $\mathbf{X}$ under the new resampling distribution $f_{\mathbf{Z}}(\cdot; \mathbf{p})$ (cf. also Johns (1988)), where R is an integer such that

$$S_{(R)} \le \rho \quad \text{and} \quad S_{(R+1)} > \rho. \tag{4}$$

Johns (1988) showed that the estimator $\widetilde{x}_{\rho,N}$ characterized by Eqs. (3) and (4) is asymptotically normally distributed. Specifically, in the context of this paper, a simple interpretation of his result shows that as the number of resamples $N \to \infty$, we have

$$\sqrt{N}(\widetilde{x}_{\rho,N} - x_\rho) \xrightarrow{dist} N(0, \sigma^2),$$

where

$$\sigma^2 = C \left\{ E\left[I\{T(\mathbf{Z}) \le x_\rho\} \prod_{i=1}^{n} (np_i)^{-M_i} \middle| \mathbf{X} \right] - \rho^2 \right\},$$

and $C > 0$ is some constant. Thus, a good importance resampling weight vector $\mathbf{p} = (p_1, \ldots, p_n)$ would be the one that minimizes the variance σ^2. This result together with Eq. (1) suggests that the importance sampling estimation of quantiles is essentially equivalent to the estimation of tail probabilities, in that the set of optimal weights for estimating the probability $P(T(\mathbf{Z}) \le x_\rho|\mathbf{X})$ (assume that x_ρ is known) is also optimal for estimating the quantile x_ρ (assume that ρ is given) in an asymptotic sense.

However, as compared with the estimation of tail probabilities, the obvious difficulty in directly calculating the weights to minimize σ^2 is that the true quantile x_ρ is unknown. We now outline an iterative procedure based on the sequential variance minimization algorithm developed in Hu and Su (2008) to adaptively generate a sequence of weight vectors $\mathbf{p}_k,\ k = 0, 1, \ldots$ to approximate the optimal resampling weight vector for estimating x_ρ. The overall idea is to sequentially change the underlying resampling measure so that the ρth quantile under the original uniform resampling becomes less and less extreme under this sequence of measures.

We start by specifying a constant $\eta \in (0, 1)$ with $\eta > \rho$ and an initial estimation problem $\alpha_1 = E[I\{T(\mathbf{Z}) \le \gamma_0\}|\mathbf{X}]$, where γ_0 is a threshold value under which the probability α_1 is approximately equal to η, this is achieved primarily through a (crude) quantile estimate of the statistic T based on uniform resamples from $\mathbf{X}$. Given a set of importance resampling weights $\mathbf{p} = (p_1, \ldots, p_n)$, an unbiased estimator of α_1 is

$$\widetilde{\alpha}_1 = \frac{1}{N} \sum_{j=1}^{N} I\{T(\mathbf{Z}_j) \le \gamma_0\} \prod_{i=1}^{n} (np_i)^{-M_{j,i}},$$

and the variance of the estimator $\widetilde{\alpha}_1$ is given by

$$\mathrm{Var}(\widetilde{\alpha}_1) = \frac{1}{N} E\left[I\{T(\mathbf{Z}) \le \gamma_0\} \prod_{i=1}^{n} (np_i)^{-M_i} \right] - \frac{1}{N}\alpha_1^2. \tag{5}$$

Thus, if a weight vector $\mathbf{p}^* = (p_1^*, \ldots, p_n^*)$ that minimizes $\mathrm{Var}(\widetilde{\alpha}_1)$ can be found, then the associated parameterized distribution $f_{\mathbf{Z}}(\cdot; \mathbf{p}^*)$ will be an optimal importance sampling distribution (within the parameterized family $\{f_{\mathbf{Z}}(\cdot; \mathbf{p})\}$) for estimating α_1, and hence can also be used as a good resampling distribution for estimating the α_1th quantile γ_0 (assume that α_1 is given). By the definitions of α_1 and γ_0, if η is not too small, then for a reasonable sample size N, the optimal weight vector $\mathbf{p}^*$ can be closely approximated by the weight vector $\mathbf{p}^{(1)} = (p_1^{(1)}, \ldots, p_n^{(1)})$ that minimizes the stochastic counterpart of (5)

$$\min_{p_1,\ldots,p_n} \frac{1}{N} \sum_{j=1}^{N} I\{T(\mathbf{Z}_j) \le \gamma_0\} \prod_{i=1}^{n} (np_i)^{-M_{j,i}}. \tag{6}$$

Given the solution of the optimization problem (6), now proceed by defining another estimation problem $\alpha_2 = E[I\{T(\mathbf{Z}) \le \gamma_1\}|\mathbf{X}]$, where γ_1 is chosen so that the probability of the event $\{T(\mathbf{Z}) \le \gamma_1\}$, under the new distribution $f_{\mathbf{Z}}(\cdot; \mathbf{p}^{(1)})$, is approximately equal to η. The idea is that the event $\{T(\mathbf{Z}) \le \gamma_0\}$ will become less rare under the new distribution $f_{\mathbf{Z}}(\cdot; \mathbf{p}^{(1)})$, so that the new threshold γ_1 obtained will be less than γ_0. For an arbitrary importance sampling weight vector $\mathbf{p} = (p_1, \ldots, p_n)$ with $p_i > 0\ \forall i$, the variance of the importance sampling estimator $\widetilde{\alpha}_2$ for estimating α_2 can be written as

$$\mathrm{Var}(\widetilde{\alpha}_2) = \frac{1}{N} E_{\mathbf{p}^{(1)}} \left[I\{T(\mathbf{Z}) \le \gamma_1\} \prod_{i=1}^{n} (np_i)^{-M_i} \prod_{i=1}^{n} (np_i^{(1)})^{-M_i} \right] - \frac{1}{N}\alpha_2^2,$$

where $E_{\mathbf{p}^{(1)}}[\cdot]$ is the expectation with respect to $f_{\mathbf{Z}}(\cdot; \mathbf{p}^{(1)})$. Again, by the way γ_1 is constructed, $\{T(\mathbf{Z}) \le \gamma_1\}$ is not a rare event under $f_{\mathbf{Z}}(\cdot; \mathbf{p}^{(1)})$; thus the variance above can be reliably approximated by its stochastic counterpart with a moderate sample size N, the solution of which would lead to a reasonable approximation of the set of optimal weights minimizing $\mathrm{Var}(\widetilde{\alpha}_2)$.

By iterating in the same manner, we will obtain a sequence of (approximately) decreasing thresholds $\{\gamma_k,\ k = 0, 1, \ldots\}$, where each γ_k can be viewed as the (approximate) ηth quantile of T under the distribution $f_{\mathbf{Z}}(\cdot; \mathbf{p}^{(k)})$, and each distribution $f_{\mathbf{Z}}(\cdot; \mathbf{p}^{(k)})$ is an (approximately) optimal importance sampling distribution for estimating the previous quantile γ_{k-1}. The entire iteration procedure is stopped whenever it is detected that the current quantile γ_k is close to the (unknown) ρth quantile x_ρ for some k. This detection step can be conveniently carried out by using the relationship (4). Specifically, one can check at each step of the iteration procedure, and stop the iteration whenever the $\lfloor \eta N \rfloor$th order statistic $S_{(\lfloor \eta N \rfloor)}$ (cf. Eq. (3)) is less than or equal to ρ. Upon termination of the procedure, one can use the current distribution $f_{\mathbf{Z}}(\cdot; \mathbf{p}^{(k)})$ as a good importance sampling distribution to generate more resamples to obtain a more precise estimate of the quantile x_ρ.

A detailed description of the algorithm that we call sequential variance minimization is given below.

Sequential Variance Minimization (SVM)

1. Specify a total resample size M, the number of resamples $N > 1$ to be used at each iteration, and a constant $\eta \in (0, 1)$. Set the initial probability weights $p_i^{(0)} = \frac{1}{n}\ \forall i = 1, \ldots, n$, and the iteration counter $k = 0$.
2. Generate N resamples $\mathbf{Z}_1, \ldots, \mathbf{Z}_N$ from $\mathbf{X}$ according to the distribution $f_{\mathbf{Z}}(\cdot; \mathbf{p}^{(k)})$, and obtain the sample ηth quantile γ_k by taking the $\lfloor \eta N \rfloor$th order statistic of $T_{(1)} \le T_{(2)} \le \cdots \le T_{(N)}$, i.e., $\gamma_k := T_{(\lfloor \eta N \rfloor)}$.
3. If $S_{(\lfloor \eta N \rfloor)} > \rho$, update the probability weights by solving for all $i = 1, \ldots, n$,

$$p_i^{(k+1)} = \arg\min_{p_1,\ldots,p_n} \frac{1}{N} \sum_{j=1}^{N} I(T(\mathbf{Z}_j) \le \gamma_k) \prod_{i=1}^{n} (np_i)^{-M_{j,i}} \prod_{i=1}^{n} (np_i^{(k)})^{-M_{j,i}} \tag{7}$$

set $k = k + 1$ and reiterate from Step 2; **otherwise** generate the rest of the $M - (k + 1)N$ resamples from the current distribution $f_{\mathbf{Z}}(\cdot; \mathbf{p}^{(k)})$, which when combined with the previous N resamples generated from $f_{\mathbf{Z}}(\cdot; \mathbf{p}^{(k)})$ gives a total of $M - kN$ resamples $\mathbf{Z}_1, \ldots, \mathbf{Z}_{M-kN}$. Evaluate the statistic T at those resamples, order them from the smallest to the largest, i.e., $T_{(1)} \le \cdots \le T_{(M-kN)}$, and then use $T_{(R)}$ as an estimate of the target quantile x_ρ, where $S_R \le p$, $S_{R+1} > p$.

3. Simulation studies

In this section, we carry out simulation studies on two real data sets to illustrate the performance of SVM and compare its performance with that of uniform resampling. The Matlab code for our simulation studies can be found at http://www.ams.sunysb.edu/~jqhu/.

3.1. Estimating bootstrap quantiles in proportional hazards model

In a recent paper (Do et al., 2001), the practical value of importance resampling bootstrap methodologies has been explicitly emphasized in a survival framework. To illustrate the performance of the proposed algorithm, we consider a real survival data set, where the objective is to estimate the 0.0005, 0.005, 0.025 and 0.05 bootstrap quantiles of the regression parameter in Cox's proportional hazards model, which correspond to 99.9%, 99%, 95% and 90% two-sided confidence intervals.

A study was conducted on the effects of ploidy on the prognosis of patients with cancers of the mouth. 80 patients were selected who had a paraffin-embedded sample of the cancerous tissue taken at the time of surgery. Follow-up survival data was obtained on each patient. The tissue samples were examined using a flow cytometer to determine if the tumor had an aneuploid or diploid DNA profile, and Cox's model was applied to analyze the difference between survival times for these two groups of patients; see Klein and Moeschberger (2003, pp. 13–14). In this study, a sample consists of n ($n = 80$) three-tuples (T_i, δ_i, Z_i), $i = 1, \ldots, n$, where T_i is the time of study for the ith patient, δ_i is the death indicator for the ith patient ($\delta_i = 1$ if a death has occurred and $\delta_i = 0$ if the lifetime is censored), and Z_i is a profile indicator with $Z_i = 0\,(1)$ if the profile of the ith patient is diploid (aneuploid). Let $h(t|Z)$ be the hazard rate at time t for a patient with covariate Z. Cox's proportional hazards model assumes

$$h(t|Z) = h_0(t) \exp(\beta^{\mathrm{T}} Z),$$

where $h_0(t)$ is the baseline hazard rate. In practice, the parameter β can be estimated by $\hat{\beta}$, which maximizes the log partial likelihood function

$$\sum_{i=1}^{n} \delta_i \left\{ \beta^{\mathrm{T}} Z_i - \log \left(\sum_{j \in R_i} \exp(\beta^{\mathrm{T}} Z_j) \right) \right\},$$

where R_i is a risk set containing all patients under study just prior to T_i. The bootstrap distribution of $\hat{\beta}$ is crucial for constructing bootstrap confidence intervals and tests for β (c.f., e.g., Lai and Su (2006) and Bloch et al. (2006)). We are interested in estimating the 0.0005, 0.005, 0.025 and 0.05 quantiles of $\hat{\beta}^*$, where $\hat{\beta}^*$ is calculated based on a bootstrap resample. More specifically, bootstrap resamples are drawn from the distribution putting weight $1/n$ on each (T_i, δ_i, Z_i), $i = 1, \ldots, n$.

Table 1
Performance of SVM and UR on estimating bootstrap quantiles of $\hat{\beta}$, based on 100 independent runs

ρ (x_ρ)	UR	SVM	
	M.S.E.	M.S.E.	R.E.
0.05 (−0.9457)	1.78×10^{-4}	5.64×10^{-5}	3.2
0.025 (−1.0417)	4.83×10^{-4}	8.39×10^{-5}	5.8
0.005 (−1.2413)	0.0017	1.09×10^{-4}	15.6
0.0005 (−1.4964)	0.0280	5.47×10^{-4}	51.2

In our simulation study, the underlying true quantiles are estimated by a brute-force uniform resampling with 1 million uniform resamples. We have experimented with different sets of parameters, and found empirically that the performances of the algorithms are insensitive to the choices of N and η, as long as η is chosen not too small and that the value of ηN is at least in the order of a few tens. Our extensive simulation studies on various problems (not reported here) showed that the algorithm terminates fast; it took at most four iterations to terminate for a quantile as extreme as 0.0005 for all the problems considered. Since Efron and Tibshirani (1993) recommended to use at least 2000 bootstrap resamples to estimate bootstrap quantiles when bootstrap confidence intervals are constructed, the total resample size M is set to 2000, the resample size per iteration N is set to 500, and η is taken to be 0.2 throughout our simulation study.

Table 1 shows the performances of SVM for the respective cases $\rho = 0.0005, 0.005, 0.025$ and 0.05, where we have also included for comparison the performances of the uniform resampling method (UR). For each test case, we performed 100 independent simulation runs of both algorithms, and the mean squared errors (M.S.E.) and relative efficiencies (R.E.) are reported in Table 1. The performance comparison is based on the same amount of computational effort. Therefore, in our numerical experiments, the same amount of 2000 total resamples were assigned to UR to estimate the quantiles. From Table 1, we see that SVM provides superior performance over UR, and yields good results for all the four test cases considered. Note that performance of UR degrades quickly as ρ decreases.

Note that for each resample, the statistic $\hat{\beta}^*$ needs to be determined by a numerical optimization procedure, which makes the evaluation process computationally intensive. Our simulation results on an IBM PC with a 2.0 GHz CPU and 1 GB memory show that for the $\rho = 0.005$ case (corresponds to a 99% two-sided confidence interval), the average computational time (over 100 runs) required for SVM was 12 s. Just as an indication of the potential computational gains of SVM over the uniform resampling method, we also performed 100 independent replication runs of UR with 30 000 resamples, which yields a MSE of 1.13×10^{-4} and an averaged run time of 178 s. We see that in this case, to achieve the same amount of accuracy, SVM is able to reduce the computational efforts of UR by roughly a factor of 15, which is consistent with the relative efficiency result obtained in Table 1. Note that the computational savings can be even more significant if the statistic is more complex and/or a larger resample size is required.

To illustrate the performance of the proposed approach under different types of survival distributions and varying amounts of censoring, we consider a hypothetical clinical trial with 100 patients. The patient's failure time follows the proportional hazards model in which the baseline survival function is that of a Weibull distribution with scale parameter 1, shape parameter θ, and covariates $Z = (z_1, z_2)^{\mathrm{T}}$, where z_1 is the treatment indicator (taking the value 1 if the treatment is received and 0 otherwise) and z_2 is a patient characteristic, which is assumed to be uniformly distributed on the interval (0, 1). The censoring times c_i are i.i.d. exponential with intensity parameter λ being 0.5, 1 or 2, and $\beta = (\beta_1, \beta_2)^{\mathrm{T}} = (1, 0.5)^{\mathrm{T}}$. In Table 2, we consider nine different test cases by varying the values of the parameter pair (θ, λ), where in each respective case, a hypothetical clinical trial is simulated following the proportional hazards model, and 0.0005, 0.005, 0.025 and 0.05 bootstrap quantiles of $\hat{\beta}_1$ are estimated using both UR and SVM. For each test case, we performed 100 independent simulation runs of both algorithms, and the same amount of 2000 total resamples were assigned to both UR and SVM to estimate the quantiles. The relative efficiencies (R.E.) of SVM relative to UR are reported in Table 2, which clearly indicate the superior performance of SVM over UR.

3.2. Estimating bootstrap quantiles of a correlation coefficient

In this section, we further illustrate the performance of SVM on estimating bootstrap quantiles of a correlation coefficient. We consider a real law school data set given in Efron and Tibshirani (1993, p. 21), where LSAT (average score on a national law test) and GPA (average undergraduate grade-point average) were collected for a sample of 15 universities, and the goal is to estimate the bootstrap quantiles of the correlation coefficient between LSAT and GPA:

$$\mathrm{corr}(\mathrm{LSAT}, \mathrm{GPA}) = \frac{\sum_{j=1}^{15}(Y_j - \bar{Y})(Z_j - \bar{Z})}{[\sum_{j=1}^{15}(Y_j - \bar{Y})^2 \sum_{j=1}^{15}(Z_j - \bar{Z})^2]^{1/2}},$$

where (Y_j, Z_j) denotes the pair (LSAT, GPA) for the jth university, $j = 1, \ldots, 15$.

Table 2
Relative efficiencies (R.E.) on estimating bootstrap quantiles of $\hat{\beta}_1$, based on 100 independent runs

ρ	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$
(a) $\theta = .7$			
0.05	3.1	2.9	3.5
0.025	5.9	5.2	6.3
0.005	16.5	13.4	15.0
0.0005	51.6	50.6	47.2
(b) $\theta = 1$			
0.05	2.6	4.5	2.9
0.025	4.7	6.7	7.6
0.005	14.8	12.9	15.3
0.0005	43.5	49.8	47.0
(c) $\theta = 1.3$			
0.05	3.1	3.7	3.0
0.025	5.3	6.9	7.3
0.005	13.6	16.8	14.0
0.0005	51.8	51.5	55.1

Table 3
Performance of SVM and UR on estimating bootstrap quantiles of correlation coefficient, based on 100 independent runs

ρ (x_ρ)	UR	SVM	
	M.S.E.	M.S.E.	R.E.
0.05 (0.5230)	6.19×10^{-5}	2.02×10^{-5}	3.1
0.025 (0.4595)	1.08×10^{-4}	2.25×10^{-5}	4.8
0.005 (0.3227)	4.12×10^{-4}	3.72×10^{-5}	11.1
0.0005 (0.1503)	0.0120	1.17×10^{-4}	102.6

In Efron and Tibshirani (1993), the bootstrap quantiles for the correlation coefficient were calculated based on uniform resampling by drawing data with equal weight 1/15. In Table 3, the performance of their approach is compared with the proposed SVM algorithm on estimating the 0.0005, 0.005, 0.025 and 0.05 bootstrap quantiles of the correlation coefficient. Again, the performance comparison is based on the same amount of computational effort, and each test case is based on 100 independent simulation runs of both algorithms. The M.S.E. and R.E. are reported in Table 3. From the table, we see that SVM consistently outperforms UR in all four test cases. Moreover, for the two extreme cases $\rho = 0.005$ and $\rho = 0.0005$, SVM is able to reduce the computational effort of UR by at least an order of magnitude.

4. Conclusions

In this paper, we provide a general computational tool for bootstrap practitioners by introducing an importance resampling algorithm for estimating bootstrap quantiles of general statistics. Our work combines the adaptive importance resampling method in Hu and Su (2008) for estimating bootstrap tail probabilities with the asymptotic results obtained in Johns (1988); thus, in some sense this paper can also be viewed as a computational extension of Johns (1988). Simulation results indicate that the proposed algorithm not only outperforms the uniform resampling method, but may also provide significant computational efficiency gains.

Acknowledgments

The authors thank the associate editor and two anonymous referees for their insightful comments and suggestions, which have led to a substantially improved paper. The work of Jiaqiao Hu was supported by the Air Force Office of Scientific Research under grant FA95500710366.

References

Bloch, D.A., Lai, T.L., Su, Z., Tubert-Bitter, P., 2006. A combined superiority and non-inferiority approach to multiple endpoints in clinical trials. Stat. Med. 26 (6), 1193–1207.
Chen, Z., Do, K.A., 1994. Bootstrap method with saddlepoint approximations and importance resampling. Statist. Sinica 4, 407–421.
Davison, A.C., 1988. Discussion of paper by D.V. Hinkley. J. Roy. Statist. Soc. Ser. B 50, 356–357.
Do, K.A., Hall, P., 1991. On importance sampling for the bootstrap. Biometrika 78, 161–167.
Do, K.A., Wang, X., Broom, B.M., 2001. Importance bootstrap resampling for proportional hazards regression. Comm. Statist. Theory Methods 30 (10), 2173–2188.
Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman and Hall, New York.
Hesterberg, T., 1996. Control variates and importance sampling for efficient bootstrap simulations. Statist. Comput. 6 (2), 147–157.
Hu, J., Su, Z., 2008. Adaptive resampling algorithms for estimating bootstrap distributions. J. Statist. Plann. Inference 138, 1763–1777.

Johns, M.V., 1988. Importance sampling for bootstrap confidence intervals. J. Amer. Statist. Assoc. 83, 709–714.
Klein, J.P., Moeschberger, M.L., 2003. Survival Analysis: Techniques for Censored and Truncated Data. Springer, New York.
Lai, T.L, Su, Z., 2006. Confidence intervals for survival quantiles in the cox regression model. Lifetime Data Anal. 12 (4), 407–419.
Lee, S.M.S, Wong, I.O.L., 2002. A hybrid approach based on saddlepoint and importance sampling methods for bootstrap tail probability estimation. Statist. Comput. 12 (3), 209–217.
Rubinstein, R.Y., 1997. Optimization of computer simulation models with rare events. European J. Oper. Res. 99, 89–112.