

Fast Bootstrapping by Combining Importance Sampling and Concomitants

Tim Hesterberg
Research Department, MathSoft Inc.
Seattle, WA 98109
timh@statsci.com

Abstract

Importance sampling is the old standby method for obtaining accurate tail quantiles of a bootstrap distribution more quickly. A newer method, a variation of control variates called concomitants, is especially attractive in larger problems because its efficiency relative to simple Monte Carlo sampling increases at the rate of \sqrt{n} , where n is the sample size.

We show how to combine these complementary methods. Doing so successfully requires two modifications to classical importance sampling — a weighted average estimate and a mixture design distribution — and the use of saddlepoint estimates for the concomitants. These methods can be programmed to run automatically, and offer improved moment estimation simultaneous with quantile estimation. The efficiency gains in can be large, e.g. by a factor of 30, even with small n .

We also obtain promising results by smoothing the distribution estimates produced by concomitants, with and without importance sampling.

Keywords

Variance reduction, importance sampling, concomitants, bootstrap, saddlepoint.

1 Introduction

The bootstrap is a general statistical technique, usually implemented using computer-intensive Monte Carlo simulation. A variety of methods have been used to reduce the computational effort, to reduce the number of Monte Carlo samples required to obtain acceptable accuracy. Our focus in this article is on two methods which are effective for estimating tail quantiles of the bootstrap distribution — concomitants of order statistics and importance sampling — and how they may be combined. The combination (“CC.IS”) is effective for quantiles and moments, and may be further improved by smoothing.

We concentrate on the nonparametric bootstrap; see e.g. Efron and Tibshirani (1993) for an introduction. The original data is $\mathcal{X} = (x_1, x_2, \dots, x_n)$, a sample from an unknown distribution, which may be multivariate. Let $\mathcal{X}^* = (X_1^*, X_2^*, \dots, X_n^*)$ be a “resample” (a bootstrap sample) of size n chosen with replacement from \mathcal{X} . We wish to estimate something about the distribution of a random variable $T^* = T(\mathcal{X}^*)$. T^* may be a parameter estimate or a pivotal statistic used for inference.

Let $G(a) = P(T^* \leq a)$. The simple Monte Carlo implementation of the nonparametric bootstrap begins by generating a large number B of samples \mathcal{X}_b^* , $b = 1, \dots, B$, of size n with replacement from the original data. Compute $T_b^* = T(\mathcal{X}_b^*)$ for each such resample. Then the bootstrap distribution estimate is $\hat{G}(a) = (1/B) \sum_{b=1}^B I(T_b^* \leq a)$, where I is the usual indicator function.

In some applications we need to estimate moments of the distribution of T^* for bootstrap estimates of bias or standard error. In other applications we need to estimate quantiles of the distribution, particularly in the tails, for bootstrap confidence intervals. Estimating tail quantiles accurately is harder than estimating moments — Efron (1987) finds that reasonable standard error estimates are obtained with only $B = 100$, or even 25 resamples, but that 1000 resamples are needed for accurately estimating tail quantiles for nonparametric confidence intervals — so we focus here on estimating quantiles.

The Monte Carlo simulation can be expensive, especially if the statistic T is hard to compute. A number of techniques have been used for reducing the computational effort of bootstrapping, including importance sampling (Johns, 1988; Davison, 1988), antithetic variates (Therneau 1983, Hall 1989), control variates (Therneau 1983; Davison, Hinkley, and Schechtman 1986; Efron 1990), balanced sampling (Davison, Hinkley, and Schechtman 1986, Gleason 1988, Graham et al. 1990), concomitants of order statistics (Efron 1990, Do and Hall 1992, Hesterberg 1995b), quasi-random resampling

(Do and Hall 1991), and post-stratification (Hesterberg 1995b). Various combinations of methods have been investigated, including concomitants with balanced or antithetic sampling (Do 1992), importance sampling with balanced sampling (Booth et al. 1993), and importance sampling with control variates (“CV.IS”) (Hesterberg 1996). The combination CC.IS we discuss here is even more effective.

The first element of CC.IS is importance sampling, usually the most effective single method for estimating bootstrap quantiles, the second is concomitants, which is particularly useful in large sample problems. We begin in Section 2 with a discussion of linear approximations for T^* , which are needed by both importance sampling and control variates. We discuss concomitants in Section 3 and importance sampling in Section 4. Certain variations of both methods are necessary to make the combination effective, which we discuss in Section 5. We discuss smoothing the distribution estimates produced by concomitants, with and without importance sampling, in Section 6.

2 Linear Approximations

Both concomitants and importance sampling methods depend in one way or another on an accurate “generalized linear” approximation to T^* , which is determined by a vector \mathcal{L} of length n , with elements L_j corresponding to each of the original observations x_j , such that

$$\psi(T(\mathcal{X}^*)) \doteq \sum_{j=1}^n L_j P_j^* = L^* \quad (1)$$

where ψ is a smooth monotone increasing function, $P_j^* = M_j/n$, and M_j is the number of times x_j is included in \mathcal{X}^* . The special case where $\psi(T^*) = T^* - T(\mathcal{X})$ is a standard linear approximation.

Efron (1982) chooses \mathcal{L} based on an empirical influence function. We assume that T is invariant to permutations of its arguments, so that a resample can be described by the number of times each original observation is included in the resample. Then we may express T as a function of weights on the original observations, $T^* = T(\mathbf{P}^*)$, where $\mathbf{P}^* = (P_1^*, \dots, P_n^*)$. Let $\mathbf{P}_0 = (1/n, \dots, 1/n)$ be the vector of weights that corresponds to the original sample \mathcal{X} . The components of $\mathcal{L}^{\text{influence}}$ are

$$L_i^{\text{influence}} = \frac{d}{d\epsilon} T(\mathbf{P}_0 + \epsilon(\mathbf{P}_{(i)} - \mathbf{P}_0)) \quad (2)$$

where $\mathbf{P}_{(i)}$ is the vector with 1 in position i and zeroes elsewhere. $\mathcal{L}^{\text{influence}}$ can sometimes be calculated ana-

lytically, or may be approximated using a small value of ϵ . The positive jackknife corresponds to $\epsilon = 1/(n+1)$.

Figure 1 shows scatterplots of T^* vs. two versions of L^* , for the studentized mean (one-sample t -statistic) for data (9.6, 10.4, 13.0, 15.0, 16.6, 17.2, 17.3, 21.8, 24.0, 26.9, 33.8) from (Graham et al. 1990). In both cases the relationship is nonlinear. A scatterplot smooth of L^* vs T^* may be used to estimate ψ , or vice versa to estimate ψ^{-1} . The right panel uses a “tail-specific” linear approximation (Hesterberg 1995b), which gives more accurate results in the corresponding tail.

Influence function approximations should only be used if $T(\mathbf{P})$ is a smooth function of \mathbf{P} , which is true for most common statistics T . Approximations obtained by linear regression (Efron 1990, Hesterberg 1995b) may be used for other statistics.

Let F denote the distribution function for L^* . We estimate quantiles of L^* by reversing the saddlepoint formula of Lugannani and Rice (Daniels 1987, Davison and Hinkley 1988, Hesterberg 1994). Let

$$p(\tau) = \Phi(\zeta) + \phi(\zeta)(1/\zeta - 1/\hat{z}),$$

where ϕ and Φ are the standard normal density and distribution functions,

$$\kappa(\tau) = n \log(n^{-1} \sum_{j=1}^n \exp(\tau L_j/n)) \quad (3)$$

is the cumulant generating function of L^* , $\zeta = \text{sgn}(\tau) \sqrt{2(\tau \kappa'(\tau) - \kappa(\tau))}$, and $z = \tau \sqrt{\kappa''(\tau)}$. Then $\hat{F}^{-1}(\alpha) = \kappa'(\tau)$, where τ is the (numerical) solution of $p(\tau) = \alpha$.

3 Concomitants

Davison and Hinkley (1988) use the saddlepoint for linear bootstrap problems; here we essentially use the saddlepoint for the linear part of a statistic, and Monte Carlo simulation for the nonlinear part.

For simplicity of notation, we sort the resamples by the values of L^* , so that L_b^* is the b 'th order statistic of the linear approximation. The concomitants estimate of the bootstrap distribution is

$$\hat{G}(a) = (1/B) \sum_{b=1}^B I(T_b^\dagger \leq a) \quad (4)$$

where

$$T_b^\dagger = \hat{\psi}^{-1}(L_b^\dagger) + T_b^* - \hat{\psi}^{-1}(L_b^*) \quad (5)$$

and where L_b^\dagger is an estimate of the $(b-0.5)/B$ quantile of the distribution of L^* . A simple variation uses $\psi(x) = x$.

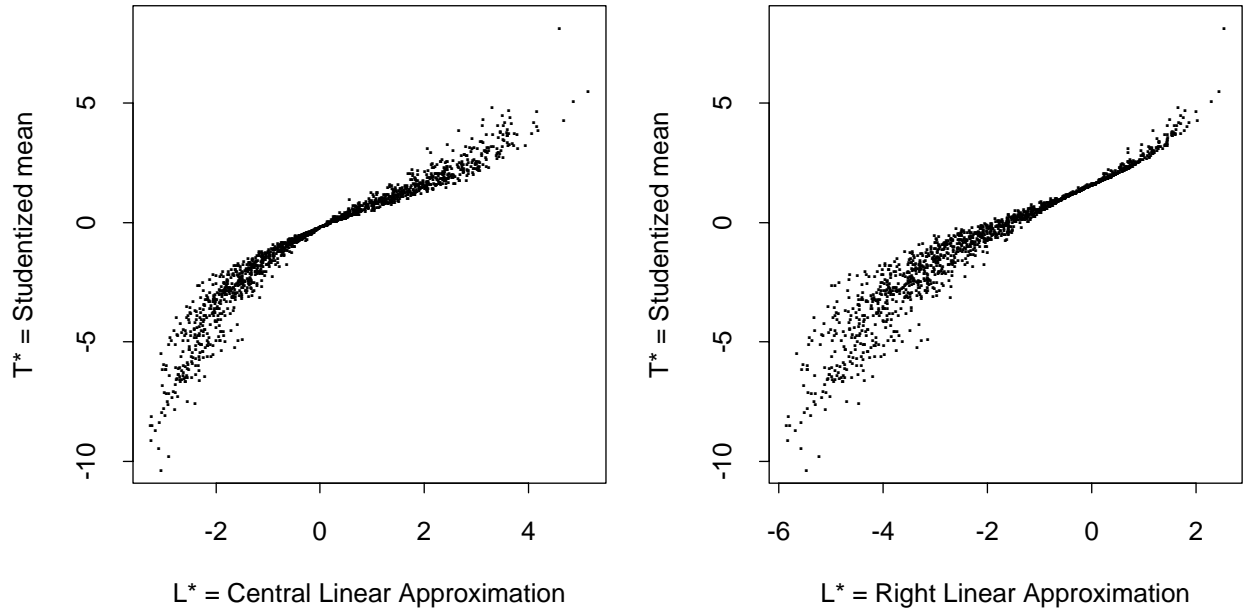


Figure 1: Central and Right Linear Approximations. Studentized mean T^* vs. linear approximation L^* . The same 1500 resamples are shown in both panels, with the same values of T^* but two different linear approximations — the approximation in the right panel is customized for the right tail. For display purposes the randomly generated points have heavier tails than under simple bootstrap sampling.

Efron (1990) estimates ψ^{-1} using a cubic regression of T^* against L^* , and Hesterberg (1995b) uses smoothing and other variations.

We interpret concomitants as decomposing T_b^* into its estimated expected value given L_b^* ,

$$\hat{\psi}^{-1}(L_b^*), \quad (6)$$

and the residual

$$R_b^* = T_b^* - \hat{\psi}^{-1}(L_b^*), \quad (7)$$

then replacing the random observed order statistic L_b^* in (6) with a value near the center of its distribution, L_b^\dagger .

Efron (1990) lets L_b^\dagger be the b 'th normal score $\Phi^{-1}((b-0.5)/B)$, iteratively transformed using cubic Cornish-Fisher transformations so that the first four sample moments match the theoretical moments of L^* , but suggests that using the saddlepoint would be more accurate. We use the saddlepoint, with $L_b^\dagger = \hat{F}^{-1}((b-0.5)/B)$. Rather than evaluate each quantile estimate individually, we evaluate the parametric equations $(p(\tau), \kappa'(\tau))$ for a number of values of τ to obtain points on the curve $(\hat{F}(x), x)$, and smoothly interpolate, adding additional values of τ if needed.

Using the saddlepoint is especially important in CC.IS because the weighted sample moments needed by the Cornish-Fisher method can be extremely variable when obtained using importance sampling.

4 Importance Sampling

The usual derivation of importance sampling (Hammerley and Hanscomb 1964) is designed for “Monte Carlo integration”—estimating an integral, or equivalently the expected value of a random variable. The integral may be rewritten

$$\begin{aligned} \mu &= E_f[Q(\mathcal{X}^*)] = \int Q(\mathcal{X}^*)f(\mathcal{X}^*) \\ &= \int \frac{Q(\mathcal{X}^*)f(\mathcal{X}^*)}{g(\mathcal{X}^*)}g(\mathcal{X}^*) = E_g[Y(\mathcal{X}^*)] \end{aligned} \quad (8)$$

where $Y(\mathcal{X}^*) = Q(\mathcal{X}^*)f(\mathcal{X}^*)/g(\mathcal{X}^*)$. In the bootstrap context f is the discrete distribution corresponding to a simple random sample with replacement from the original data, and g is a “design distribution” which also samples in some manner from the original data. The classical “integration” estimate based on B resamples

from g is

$$\hat{\mu}_{\text{int}} = \bar{Y} = \frac{1}{B} \sum_{b=1}^B \frac{Q(\mathcal{X}_b^*)f(\mathcal{X}_b^*)}{g(\mathcal{X}_b^*)}. \quad (9)$$

This corresponds to a weighted sum, with weight

$$(1/B)W_b$$

on Q_b , where $W_b = f(\mathcal{X}_b^*)/g(\mathcal{X}_b^*)$. The weights do not add to 1, which can cause a number of problems. Two alternate estimates (Hesterberg 1988, 1995a) for which the weights do add to 1 are the “ratio estimate” which uses weights $W_b/\sum_{k=1}^B W_k$, and the “regression estimate” which uses weights

$$V_b = (1/B)W_b(1 + c(W_b - \bar{W})) \quad (10)$$

where $c = (1 - \bar{W})/((1/B)\sum_{b=1}^B(W_b - \bar{W})^2)$. In the bootstrap context the weights are used to create a weighted empirical distribution for T^* , e.g. $\hat{G}(a) = \sum_{b=1}^B V_b I(T_b^* \leq a)$.

4.1 Design Distributions

We use importance sampling design distributions of the form

$$g(\mathcal{X}^*) = 0.2f(\mathcal{X}^*) + 0.4g_1(\mathcal{X}^*) + 0.4g_2(\mathcal{X}^*),$$

where g_1 and g_2 are concentrated on the left and right tails of the bootstrap distribution. In particular, g_k indicates sampling with unequal probabilities $P(X_i = x_j) = c_k \exp(\tau_k/nL_{(k)j})$ for $k = 1, 2$, where $\tau_1 < 0 < \tau_2$, the c_k are normalizing constants, and $\mathcal{L}_{(k)}$ are vectors that define generalized linear approximations. This is a combination of exponential tilting (Johns 1988, Davison 1988) with defensive mixture distributions (Hesterberg 1988, 1995a). This design has 20% of the bootstrap samples are chosen by simple random bootstrap sampling and 40% each from distributions biased toward the left and right tails. The weight

$$W_b = (0.2 + 0.4 \exp(\tau_1 L_{(1)b}^* - \kappa_{(1)}(\tau_1)) + 0.4 \exp(\tau_2 L_{(2)b}^* - \kappa_{(2)}(\tau_2)))^{-1}$$

is independent of which distribution (f , g_1 , or g_2) was used to generate resample b , and is bounded above by 5.

We use tail-specific $\mathcal{L}_{(k)}$ here, but a single \mathcal{L} can be used. We choose the tilting parameters τ_k so that the expected value of $L_{(k)}^*$ under g_k is at approximately α quantile under f , where α is 0.025 and 0.975 for the left ($k = 1$) and right ($k = 2$) tails, respectively; we find τ by solving the saddlepoint equation $p(\tau) = \alpha$; the solution need not be very accurate.

5 Combining Concomitants and Importance Sampling

In concomitants without importance sampling the concomitants use quantiles for L^* which are evenly spaced on the probability scale, $L_b^\dagger = \hat{F}^{-1}((b - 0.5)/B)$ for $b = 1, \dots, B$. With importance sampling they must be unequally spaced to match the weights, say V_b (10). After sorting the bootstrap samples by L^* , we let

$$L_b^\dagger = \hat{F}^{-1}(\sum_{i=1}^{b-1} V_i + V_b/2)$$

for $b = 1, \dots, B$. These are used in (5) to obtain a weighted distribution estimate

$$\hat{G}(a) = \sum_{b=1}^B V_b I(T_b^\dagger \leq a) \quad (11)$$

which can be used both for quantile and moment estimation.

It is very important that the saddlepoint variation of concomitants be used instead of the cumulant transformation variation, or that robust importance sampling methods (a defensive mixture design and the ratio or regression estimate) be used; we use both the saddlepoint variation and robust methods. The reason is that if the importance sampling design uses only simple exponential tilting biased toward one tail, then values from the other tail are infrequently observed but receive large weights when observed, the sum of the weights can differ greatly from 1, and weighted moment estimates used by the cumulant transformation variation are highly variable.

The use of tail-specific linear approximations introduces a complication. The importance sampling weights are determined by the design distribution actually used, which is based on one or more vectors $\mathcal{L}_{(k)}$. Now concomitants is also based on an \mathcal{L} (via L^*) which can but need not match any used in importance sampling; we may even use different values of \mathcal{L} to obtain different concomitants distribution estimates. Our practice when estimating quantiles is use the $\mathcal{L}_{(k)}$ which gives most accurate results for each quantile, i.e. to do concomitants based on a tail-specific linear approximations when estimating a quantile in the tail. However for estimating moments we use a single central \mathcal{L} .

Results for importance sampling, concomitants, CC.IS, and other variance reduction methods are shown in in Table 1, for the Studentized mean example in Figure 1. Numbers in the table are the estimated efficiency for each method, relative to simple Monte Carlo bootstrap sampling. CC.IS more accurate for estimating tail

Table 1: Variance Reduction

Method	Moments		Quantiles		
	mean	std. dev.	.025	.5	.975
Antithetic	6.8	0.72	1.0	3.5	1.1
Balanced	7.8	0.74	1.1	2.9	1.2
Importance	2.1	22.3	14.2	0.50	9.1
Control Var.	15.1	2.3	1.4	80.2	1.4
Concomitants	20.2	4.2	5.8	63.0	9.7
CV.IS	33.1	38.5	30.0	8.6	25.0
CC.IS	23.6	10.9	30.0	12.1	34.0
CC+Smooth	20.2	4.3	11.2	97.0	23.4
CC.IS+Smooth	19.8	3.9	41.4	38.7	57.1

The values in the table are the estimated variance using simple Monte Carlo bootstrap sampling divided by the estimated variance using each variance reduction techniques, for the bootstrap distribution of the Studentized mean for the same data used in Figure 1. Numbers are based on 2000 bootstrap experiments with $B = 200$ bootstrap samples in each. Most standard errors are between 4% and 6% of the entry. A value of 30 in the table indicates that simple Monte Carlo requires roughly 30 times as many bootstrap samples for comparable accuracy, or the method needs roughly 1/30 as many samples for comparable accuracy (except that some methods suffer small-sample effects if B is small).

Estimates involving importance sampling use the regression method and the 20%/40%/40% mixture design described in the text. The balanced method used here is not the usual biased method, but a variation in which observations in a single bootstrap sample are independent. The control variate estimates use different sets of control variates for estimating moments and quantiles.

quantiles than any other method, with variance reduction factors of 30 or better, though CV.IS is close.

The combination CC.IS has several advantages over CV.IS. First, it provides both quantile and moment estimates simultaneously. In contrast, Hesterberg (1996) found that the best results for CV.IS were obtained using different sets of control variates for moments than for quantiles. Second, it appears to be more accurate for quantiles. The asymptotic variance of concomitants quantile estimates (Hesterberg 1995b) is the same as that of control variates if both are optimally tuned, i.e. the optimal (nonlinear) control variate is used and the correct ψ is used; in practice discrete control variates are used, resulting in a loss of efficiency by about a factor of $\sqrt{2}$ if the conditional distribution of T^* given L^* is approximately normal with small variance. Hesterberg and Nelson (1996) discuss optimal control variates and discrete approximations. Estimating ψ is also easier than estimating the optimal control variate, and the optimal control variate is different for every quantile.

There are also disadvantages for CC.IS. First, it requires computation of n saddlepoint quantile estimates of L^* , while CV.IS requires only a small number like 1, 2, or 3. Second, the method appears to be less accurate for moments than is CV.IS — and in this context CV.IS requires no saddlepoint estimates. Third, the best results for concomitants are obtained only if the nonlinear transformation ψ is estimated. Fourth, the use of

tail-specific L^* complicates matters. Fifth, the values T_b^\dagger differ from the corresponding T_b^* , so that this procedure is not suitable for discrete distributions (but Hall (1986) shows that bootstrap distributions are practically continuous under fairly general conditions).

Either combination does substantially better than using the component methods in isolation, and both are substantially better than methods such as balanced bootstrap sampling or antithetic variates.

Results are not as good for all statistics and datasets as shown in Table 1. The extremely small conditional variance of T^* given L^* apparent in Figure 1 does not occur in all problems and is particularly favorable to control variates and (to a lesser extent) to concomitants. This is the reason for the exceptional variance reductions for estimating the median of the bootstrap distribution using either control variates or concomitants, without importance sampling. On the other hand, the sample size here is small, and both control variates and concomitants performance becomes better as n increases, with asymptotic variances for quantiles of order $O(n^{-1/2}B^{-1})$; we have obtained some excellent results in large sample problems. And the gains offered by smoothing, discussed next, should be even more valuable for other datasets where heteroskedasticity is less extreme than in Figure 1.

6 Smoothing Concomitants

Finally, we consider the effect of smoothing concomitants distribution estimates. The empirical results in Table 1 indicate that the method of smoothing we used is effective for estimating quantiles. We conjecture that smoothing distributions with concomitants is much more effective than without; we return to this point below, but first motivate and describe our smoothing method.

We may rewrite (4) as

$$\hat{G}(a) = (1/B) \sum_{b=1}^B \hat{P}(\hat{\psi}^{-1}(L_b^*) + R_b^* \leq a),$$

where (5) corresponds to estimating each individual probability using an indicator function:

$$\hat{P}(\hat{\psi}^{-1}(L_b^*) + R_b^* \leq a) = I(\hat{\psi}^{-1}(L_b^\dagger) + R_b^* \leq a). \quad (12)$$

It may be possible to improve on this estimate by either vertical or horizontal smoothing of (12); these correspond to smoothing the residuals (7) and predicted values (6), respectively.

The idea in vertical smoothing is to continue to fix L_b^* at L_b^\dagger as in (12), but to replace the indicator function with an estimate of the probability based on the distribution of random R_b^* , the residual for the b 'th order statistic L_b^* . Results in Table 1 are obtained by a simple procedure in which the distribution of R^* is estimated by the local nearest-neighbor empirical distribution,

$$\hat{P}_b = 7^{-1} \sum_{j=-3}^3 I(\hat{\psi}^{-1}(L_b^\dagger) + R_{b+j}^* \leq a),$$

with adjustments for extreme values of b . We use a local rather than global estimate because the distribution of R^* may depend on L^* , e.g. Figure 1 shows definite heteroskedasticity.

In horizontal smoothing, we keep the residual R_b^* fixed, and replace the observed order statistic L_b^* not by a single value L_b^\dagger , but rather by an estimate of the distribution for the b 'th random order statistic, $\hat{P}_b = \hat{P}(\hat{\psi}^{-1}(L_{(b)}^*) + R_b^* \leq a)$, where now $L_{(b)}^*$ is considered random and R_b^* fixed.

Results (Table 1) are promising, with smoothing improving quantil estimates. This is still work in progress. We have not implemented horizontal smoothing, or combined it with vertical smoothing. Other ways to estimate the local distribution of R^* may be more effective, or we might achieve essentially the same result more simply by kernel smoothing the final distribution estimate. For smoothing the importance sampling concomitants combination we ignored the importance sampling weights

associated with the local neighbors, which may explain the disappointing performance for estimating moments.

We have not calculated the bias caused by smoothing, but have reason to believe it is small; this relates to our conjecture that smoothing with concomitants can be particularly effective. For the sake of argument we focus on kernel smoothing. Smoothing quantile estimates involves a tradeoff between bias and variance; the greater the amount of smoothing the smaller the variance, but the greater the bias. Both bias and variance depend on the ratio between the standard deviation (s.d.) of the kernel and the s.d. of the distribution being smoothed. We conjecture that here the bias is governed by the ratio of the kernel s.d. to the s.d. of R^* , and the variance by the ratio of the kernel s.d. to the s.d. of T^* ; this is favorable because the s.d. of the residuals R^* is much smaller than than of T^* . This would allow a greater degree of variance reduction by smoothing before the increase in bias negates the gains.

Simulation Details

Simulations are run in S-Plus (Becker et al. 1988; Statistical Sciences, 1991) and C, using the Super Duper random number generator of Marsaglia. All methods are evaluated on 2000 bootstrap experiments of $B = 200$ samples each. Common random numbers are used, with the original observations sorted according to the values of $\mathcal{L}_{\text{central}}$.

Summary

We obtain accurate estimates of both quantiles and moments using a combination of importance sampling and concomitants of order statistics. This combination is more accurate than the combination of importance sampling and control variates, and substantially more accurate than other available variance reduction procedures. Promising work in progress involves smoothing distribution estimates from concomitants, with and without importance sampling.

Acknowledgments

The author is grateful to Brad Efron, Rob Tibshirani, Kim Do, Jim Booth and Jim Schimert for helpful discussion and comments.

References

- Becker, R. A., J. M. Chambers, and A. R. Wilks (1988), *The New S Language*, Pacific Grove, CA: Wadsworth and Brooks/Cole.
- Booth, J. G., P. Hall, and A. T. A. Wood (1993), "Balanced Importance Resampling for the Bootstrap," *Annals of Statistics*, **21**, 286-298.
- Daniels, H. E. (1987), "Tail Probability Approximations," *International Statistical Review*, **55**, 1, 37-48.
- Davison A. C. (1988), Discussion of paper by D. V. Hinkley, *Journal of the Royal Statistical Society, Series B*, **50**, 356-57.
- Davison A. C. & Hinkley D. V. (1988), "Saddlepoint Approximations in Resampling Methods," *Biometrika*, **75**, 417-31.
- Davison A. C., D. V. Hinkley, and E. Schechtman (1986), "Efficient Bootstrap Simulation," *Biometrika*, **74**, 555-66.
- Do, K. and P. Hall (1991), "Quasi-random resampling for the bootstrap," *Statistics and Computing*, **1**, 13-22.
- Do, K. and P. Hall (1992), "Distribution Estimation using Concomitants of Order Statistics, with Application to Monte Carlo Simulation for the Bootstrap," *Journal of the Royal Statistical Society, Series B*, **54**, 595-607.
- Efron, B. (1982), *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics, Philadelphia.
- Efron, B. (1987), "Better Bootstrap Confidence Intervals (with discussion)," *Journal of the American Statistical Society*, **82**, 171-200.
- Efron, B. (1990), "More Efficient Bootstrap Computations," *Journal of the American Statistical Society*, **85**, 79-89.
- Efron, B. and R. J. Tibshirani (1993), *An Introduction to the Bootstrap*, Chapman and Hall.
- Gleason, J. R. (1988), "Algorithms for Balanced Bootstrap Simulations," *American Statistician*, **42**, 263-6.
- Graham, R. L., D. V. Hinkley, P. W. M. John, and S. Shi (1990), "Balanced Design of Bootstrap Simulations," *Journal of the Royal Statistical Society, Series B*, **52**, 185-202.
- Hall, P. (1986), "On the number of bootstrap simulations required to construct a Confidence Interval," *Annals of Statistics*, **14**, 1453-1462.
- Hall, P. (1989), "On Efficient Bootstrap Simulation," *Biometrika*, **76**, 613-617.
- Hammersley, J. M., and D. C. Hanscomb (1964), *Monte Carlo Methods*, London: Methuen.
- Hesterberg, T. C. (1988), "Advances in Importance Sampling," Ph.D. dissertation, Statistics Department, Stanford University.
- Hesterberg, T. C. (1994) "Saddlepoint Quantiles and Distribution Curves, with Bootstrap Applications," *Computational Statistics*, **9**, 207-212.
- Hesterberg, T. C. (1995a) "Weighted Average Importance Sampling and Defensive Mixture Distributions," *Technometrics*, **37**, 185-194.
- Hesterberg, T. C. (1995b), "Tail-specific Linear Approximations for Efficient Bootstrap Simulations," *Journal of Computational and Graphical Statistics*, **4**, 113-133.
- Hesterberg, T. C. (1996), "Control Variates and Importance Sampling for Efficient Bootstrap Simulations," *Statistics and Computing*, **6**, 147-157.
- Hesterberg, T. C. and B. L. Nelson (1996), "Control Variates for Probability and Quantile Estimation," *submitted to Management Science*.
- Johns, M. V. (1988), "Importance Sampling for Bootstrap Confidence Intervals," *Journal of the American Statistical Association*, **83**, 709-714.
- Statistical Sciences, Inc. (1991), *S-PLUS Reference Manual, Version 3.0*, Seattle: Statistical Sciences, Inc., 1991.
- Therneau, T. M. (1983), "Variance Reduction Techniques for the Bootstrap," *Technical Report No. 200 (Ph.D. Thesis)*, Department of Statistics, Stanford University.