An Edgeworth expansion for the m out of n bootstrapped median

Anat Sakov
University of California, Berkeley

Peter J. Bickel
University of California, Berkeley

July 16, 1998

Abstract

It is well known (Singh 1981) that the bootstrap distribution of the median has the correct limiting distribution. In this note we prove the existence of the next term in the Edgeworth expansion if the bootstrap sample size is $m = o(n)$.

Keywords: bootstrap, m out of n bootstrap, median, Edgeworth expansion.

1 Introduction

It is well known (Singh 1981) that the bootstrap distribution of the normalized median for a bootstrap sample of size n has the same limiting distribution as the normalized median in a non-bootstrap setting. However, the error is of order $O\left(\frac{\sqrt{\log\log n}}{n^{1/4}}\right)$.

Recent literature (Bickel, Götze & van Zwet 1997, Politis & Romano 1994) suggests using bootstrap samples of size m where $m \to \infty$ and $\frac{m}{n} \to 0$ in cases of bootstrap failure. Here, by bootstrap failure we mean that the bootstrap does not have the correct limiting distribution (section 6 of Bickel & Freedman 1981, Bickel & Ren 1995, Bickel et al. 1997).

For the bootstrapped median, although bootstrap samples of size n have the correct limiting distribution, the next term in an Edgeworth expansion does not exist. Using bootstrap samples of size m as above rectifies the problem. We consider bootstrap samples taken with replacement. For simplicity we refer to n-bootstrap or m-bootstrap as bootstrap samples of sizes n and m, respectively.

The problem with the n-bootstrap is the following: the leading term in the expansion for the non-bootstrapped median depends on the population density evaluated at the population median, and therefore higher-order terms in the expansion depend on the derivatives of the density. When sampling from $\hat{F}_n$, the empirical distribution function, neither the density nor its derivatives exist.

When taking bootstrap samples of size m as above the next term in the expansion does exist, in spite resampling from a discrete distribution. The reason is that relative to bootstrap samples of size m, the 'jumps' in $\hat{F}_n$ are very small. This has the same effect as sampling from a smooth distribution which results in one more term in the Edgeworth expansion. The m-bootstrap expansion agrees with the expansion of the normalized median in the non-bootstrap setting (Reiss 1976).

1.1 Expansion in the non-bootstrap setting

Assume $X_1, \ldots, X_n$ are iid from a distribution F, and let $\theta = F^{-1}\left(\frac{1}{2}\right)$ denote the median of the distribution. Assume $f = F'$ exists and is positive and continuous in a neighborhood of θ. Let $\hat{\theta}_n$ be the sample median. It is well known (for example Serfling 1980) that

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \overset{\mathcal{L}}{\Rightarrow} \mathrm{N}\left(0, \sigma_\theta^2\right), \qquad \text{where} \qquad \sigma_\theta = \frac{1}{2f(\theta)}. \tag{1}$$

Assume n is odd so that the median is a sample quantile. If we assume further that f has two bounded derivatives then Reiss (1976) shows that

$$\mathrm{P}\left(\frac{\sqrt{n}(\hat{\theta}_n-\theta)}{\sigma_\theta}\le t\right)=\Phi(t)+\frac{1}{\sqrt{n}}\frac{f'(\theta)}{4f^2(\theta)}\phi(t)t^2+O\left(\frac{1}{n}\right),$$

or

$$\mathrm{P}\left(\sqrt{n}(\hat{\theta}_n-\theta)\le t\right)=\Phi\left(2tf(\theta)\right)+\frac{1}{\sqrt{n}}f'(\theta)t^2\phi\left(2tf(\theta)\right)+O\left(\frac{1}{n}\right). \qquad (2)$$

In the equations above Φ and ϕ denote the cdf and the density of the standard normal distribution, respectively. Assuming more than two derivatives on f results in more terms in the expansion (Reiss 1976).

A few papers discuss the expansion for the studentized median, i.e. an expansion for $\mathrm{P}\left(\frac{\sqrt{n}(\hat{\theta}_n-\theta)}{\hat{\sigma}_\theta}\le t\right)$, where $\hat{\sigma}_\theta$ is an estimator of σ_θ. Hall & Martin (1991) and Hall (1992) on p. 321 consider the bootstrap estimate of σ_θ. Here the error term is $O\left(\frac{1}{n^{3/4}}\right)$. Hall & Sheather (1988) consider the Siddiqui-Bloch-Gastwirth estimator of the variance. This estimator is based on the width of a window which includes $2h+1$ ordered observations: h observations smaller than the median and h larger. If $h=o(n)$, then the expansion takes the form

$$\begin{aligned}\mathrm{P}\left(\frac{\sqrt{n}(\hat{\theta}_n-\theta)}{\hat{\sigma}_\theta}\le t\right) &= \Phi(t)+\frac{1}{\sqrt{n}}p_1(t)\phi(t)+\frac{1}{h}p_2(t)\phi(t)\\ &\quad+\left(\frac{h}{n}\right)^2p_3(t)\phi(t)+o\left(\frac{1}{h}+\frac{h^2}{n^2}\right).\end{aligned}$$

Falk & Janas (1992) establish an expansion when the kernel density estimator at θ is used as the variance estimator.

2 Bootstrapping the median

Let $X_1^*,\ldots,X_m^*$ be a bootstrap sample from $\hat{F}_n$. Let $\hat{\theta}_m^*$ be the median of the bootstrap sample. Singh (1981) showed that if F has a bounded second derivative in a neighborhood of θ and $f(\theta)>0$ then

$$\mathrm{P}^*\left(2f(\theta)\sqrt{n}(\hat{\theta}_n^*-\hat{\theta}_n)\le t\right)=\Phi(t)+O\left(\frac{\sqrt{\log\log n}}{n^{1/4}}\right).$$

Singh considered bootstrap samples of size n.

The Edgeworth expansion for the median in the non-bootstrap setting depends on the smoothness of F at θ, and therefore there is a problem with such an expansion in the bootstrap setting. A comparison between the bootstrap distribution of the median and the smoothed bootstrap distribution of the median (Falk & Reiss 1989, Babu & Rao 1993, DeAngelis, Hall & Young 1993) indicates that the smoothed bootstrap gives a better error rate than Singh's (1981) error rate for the ordinary bootstrap.

Nevertheless, for bootstrap samples of size $m = o(n)$ from the empirical distribution function, the bootstrap distribution of the median has the correct expansion as the following theorem shows.

Theorem 1 *: Assume f has two bounded derivatives and $f(\theta) > 0$. Then for bootstrap samples of size m where $m \to \infty$ and $\frac{m}{n} \to 0$:*

$$\begin{aligned} P^*\left(\sqrt{m}(\hat{\theta}^*_m - \hat{\theta}_n) \le t\right) &= \Phi\left(2tf(\theta)\right) + \frac{1}{\sqrt{m}}\phi\left(2tf(\theta)\right) f'(\theta) t^2 \\ &\quad + O_p\left(\sqrt{\frac{m}{n}} + \frac{1}{m}\right). \end{aligned}$$

Proof: For simplicity assume that m is odd and therefore the median of a bootstrap sample is the kth ordered statistic, where $k = \frac{m+1}{2}$. Conditional on $\hat{F}_n$,

$$m\bar{X}^*_t \equiv \sum_{i=1}^{m} 1\left(X^*_i \le t\right)$$

has the Binomial distribution with parameters m and π^*_t, where

$$\pi^*_t = \mathrm{P}^*\left(X^*_i \le t\right) = \frac{1}{n}\sum_{i=1}^{n} 1\left(X_i \le t\right). \tag{3}$$

The event $\left\{\hat{\theta}^*_m \le t\right\}$ is equivalent to the event $\left\{m\bar{X}^*_t > \frac{m-1}{2}\right\}$. Note that $\frac{m-1}{2}$ is a lattice point of $m\bar{X}^*_t$. Define $t_m = \hat{\theta}_n + \frac{t}{\sqrt{m}}$. Then

$$\begin{aligned} \mathrm{P}^*\left(\sqrt{m}(\hat{\theta}^*_m - \hat{\theta}_n) \le t\right) &= \mathrm{P}^*\left(m\bar{X}^*_{t_m} > \frac{m-1}{2}\right) \\ &= 1 - \mathrm{P}^*\left(\frac{\sqrt{m}(\bar{X}^*_{t_m} - \pi^*_{t_m})}{\sqrt{\pi^*_{t_m}(1-\pi^*_{t_m})}} \le \frac{\frac{m-1}{2} - m\pi^*_{t_m}}{\sqrt{m\pi^*_{t_m}(1-\pi^*_{t_m})}}\right). \end{aligned}$$

The point at the right-hand-side of the last probability statement is a lattice point. Let Y be distributed according to the binomial distribution with parameters m and π, and let $F(t) = \mathrm{P}\left(\frac{\sqrt{m}(\bar{Y}-\pi)}{\sqrt{\pi(1-\pi)}} \le t\right)$. Denote by $F_{\#}$ the polygonal approximation of F (p. 540 of Feller 1966). Then for t which is a midpoint between two lattice points of $\frac{\sqrt{m}(\bar{Y}-\pi)}{\sqrt{\pi(1-\pi)}}$,

$$F_{\#}(t) = \Phi(t) - \frac{1-2\pi}{6\sqrt{m\pi(1-\pi)}} \left(\mathrm{H}_2\phi\right)(t) + o\left(\frac{1}{\sqrt{m}}\right),$$

where $\mathrm{H}_2(t) = t^2 - 1$ is the second Hermite polynomial.

Let $\mathrm{P}^*_{\#}$ be the polygonal approximation of P^*. It follows from the definition of a polygonal approximation (Feller 1966) that

$$\begin{aligned}
\mathrm{P}^*\left(\sqrt{m}(\hat{\theta}^*_m - \hat{\theta}_n) \leq t\right) &= 1 - \mathrm{P}^*_{\#}\left(\frac{\sqrt{m}(\bar{X}^*_{t_m} - \pi^*_{t_m})}{\sqrt{\pi^*_{t_m}(1-\pi^*_{t_m})}} \leq \frac{\frac{m-1}{2} - m\pi^*_{t_m} + \frac{1}{2}}{\sqrt{m\pi^*_{t_m}(1-\pi^*_{t_m})}}\right) \\
&= 1 - \Phi\left(\frac{\frac{m}{2} - m\pi^*_{t_m}}{\sqrt{m\pi^*_{t_m}(1-\pi^*_{t_m})}}\right) \qquad (4) \\
&\quad + \frac{1-2\pi^*_{t_m}}{6\sqrt{m\pi^*_{t_m}(1-\pi^*_{t_m})}}(\mathrm{H}_2\phi)\left(\frac{\frac{m}{2} - m\pi^*_{t_m}}{\sqrt{m\pi^*_{t_m}(1-\pi^*_{t_m})}}\right) \\
&\quad + o\left(\frac{1}{\sqrt{m}}\right).
\end{aligned}$$

Note that Feller's (1966) result is for fixed π while here π_{t_m} is random. However, his result holds uniformly for all π in a neighborhood of $\frac{1}{2}$ which is our case (see equation 5).

Recall that $t_m = \hat{\theta}_n + \frac{t}{\sqrt{m}}$ and define $W_n(t) = \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^n 1\left(X_i \leq t\right) - F(t)\right]$. Using (3) and a Taylor expansion of F around θ:

$$\begin{aligned}
\pi^*_{t_m} &= \frac{1}{n}\sum_{i=1}^n 1\left(X_i \leq t_m\right) = \frac{1}{n}\sum_{i=1}^n 1\left(X_i \leq \theta\right) + \left(F(t_m) - \frac{1}{2}\right) \\
&\quad + \frac{1}{\sqrt{n}}\left[W_n(t_m) - W_n(\theta)\right] \\
&= \frac{1}{2} + \left(\frac{1}{n}\sum_{i=1}^n 1\left(X_i \leq \theta\right) - \frac{1}{2}\right) + \left(\hat{\theta}_n - \theta + \frac{t}{\sqrt{m}}\right) f(\theta) \\
&\quad + \frac{1}{2}\left(\hat{\theta}_n - \theta + \frac{t}{\sqrt{m}}\right)^2 f'(\theta) + \frac{1}{\sqrt{n}}\left[W_n(t_m) - W_n(\theta)\right] \\
&\quad + O_p\left(\frac{1}{m^{3/2}}\right).
\end{aligned}$$

From Theorem 1 (p. 542) of Shorack & Wellner (1986) it follows that

$$|W_n(t_m) - W_n(\theta)| \leq \sup_s |W_n(s) - W_n(\theta)| = O\left(\sqrt{\frac{\log\sqrt{m}}{\sqrt{m}}}\right),$$

where the sup is taken over all s such that

$$|s-\theta| \leq \left|\hat{\theta}_n - \theta + \frac{t}{\sqrt{m}}\right|,$$

and therefore,

$$\pi^*_{t_m} = \frac{1}{2} + \Delta_m, \tag{5}$$

where

$$\begin{aligned}\Delta_m &= \frac{t}{\sqrt{m}} f(\theta) + \frac{1}{2}\frac{t^2}{m} f'(\theta) + \left(\frac{1}{n}\sum_{i=1}^n 1\left(X_i \leq \theta\right) - \frac{1}{2}\right) + \left(\hat{\theta}_n - \theta\right) f(\theta) \\ &\quad + O_p\left(\frac{1}{n} + \frac{1}{m^{3/2}} + \sqrt{\frac{\log\sqrt{m}}{n\sqrt{m}}}\right).\end{aligned}$$

Let $g(x) = \frac{1-2x}{\sqrt{x(1-x)}}$. The Taylor expansion of $g(x)$ around $\frac{1}{2}$ is

$$g\left(\frac{1}{2} + \Delta\right) = -4\Delta + o\left(\Delta^3\right). \tag{6}$$

The argument inside Φ and $\mathrm{H}_2\phi$ in (4) has the form

$$\begin{aligned}\frac{\sqrt{m}}{2} g\left(\frac{1}{2} + \Delta_m\right) &= -2\sqrt{m}\Delta_m + o\left(\sqrt{m}\Delta_m^3\right) \\ &= -2tf(\theta) - \frac{1}{\sqrt{m}} t^2 f'(\theta) - 2\sqrt{m}\left(\frac{1}{n}\sum_{i=1}^n 1\left(X_i \leq \theta\right) - \frac{1}{2}\right) \\ &\quad - 2\sqrt{m}\left(\hat{\theta}_n - \theta\right) f(\theta) \qquad (7) \\ &\quad + O_p\left(\frac{\sqrt{m}}{n} + \frac{1}{m} + \sqrt{\frac{\sqrt{m}\log\sqrt{m}}{n}}\right) \\ &= -2tf(\theta) - \frac{1}{\sqrt{m}} t^2 f'(\theta) + O_p\left(\sqrt{\frac{m}{n}} + \frac{1}{m}\right). \qquad (8)\end{aligned}$$

Combining (4), (5) and (8) gives

$$\begin{aligned}\mathrm{P}^*\left(\sqrt{m}(\hat{\theta}^*_m - \hat{\theta}_n) \leq t\right) &= \Phi\left(2tf(\theta) + \frac{1}{\sqrt{m}} t^2 f'(\theta) + O_p\left(\sqrt{\frac{m}{n}} + \frac{1}{m}\right)\right) \\ &\quad + O_p\left(\frac{1}{m}\right) \\ &= \Phi(2tf(\theta)) + \frac{t^2}{\sqrt{m}} f'(\theta)\phi(2tf(\theta)) \\ &\quad + O_p\left(\sqrt{\frac{m}{n}} + \frac{1}{m}\right).\end{aligned}$$

This completes the proof.

Aside from the error term, which is worse, the expansion is the same as Reiss's expansion (equation 2).

Comments:

1. Note the error term of order $\sqrt{\frac{m}{n}}$ which illustrates the problem of using a bootstrap sample size $m = n$.

2. Using (7) instead of (8) in the last step of the proof gives

$$\begin{aligned}
\mathrm{P}^*\left(\sqrt{m}(\hat{\theta}^*_m - \hat{\theta}_n) \le t\right) &= \Phi(2tf(\theta)) + \frac{t^2}{\sqrt{m}} f'(\theta)\phi(2tf(\theta)) \\
&\quad + 2\sqrt{m}\phi(2tf(\theta)) \left(\frac{1}{n}\sum_{i=1}^{n} 1\left(X_i \le \theta\right) - \frac{1}{2}\right) \\
&\quad + 2\sqrt{m}\phi(2tf(\theta)) \left(\hat{\theta}_n - \theta\right) f(\theta) \\
&\quad + O_p\left(\frac{1}{m} + \sqrt{\frac{\sqrt{m}\log\sqrt{m}}{n}}\right).
\end{aligned}$$

   The two error terms are working in different directions: one gets smaller as m increases while the other gets smaller as m decreases. The optimal m is the one which makes the two error terms be of the same order and is almost $n^{2/5}$. For such an order of m the error term is $\frac{\sqrt{\log n}}{n^{2/5}}$ which is larger than $\frac{1}{\sqrt{n}}$, but is smaller than $\frac{1}{n^{1/3}}$. Note that the expansion depends on the unknown median and the unknown density. However, by using extrapolation (Bickel & Yahav 1988, Sakov 1998) it is possible to get the bootstrap distribution correct to an order of (almost) $\frac{1}{n^{2/5}}$. This rate is very satisfactory if we notice that, indirectly, we need to estimate the density at the median.

3 Summary

Using the m-bootstrap with $m \to \infty$ and $\frac{m}{n} \to 0$ is suggested in situations where the bootstrap fails to estimate the limiting distribution correctly. Here we give an example of a statistic for which the bootstrap does obtain the correct limiting distribution for a bootstrap sample of size n, but not the next term in an Edgeworth expansion. However, using bootstrap samples of smaller size gives the correct expansion.

A point which will be addressed in more details in a forthcoming paper by Bickel and Sakov is that using the m-bootstrap has an advantage in terms of execution time. Naive calculation of a median requires sorting the data. Sorting a sample of size $m << n$ is much faster than sorting a sample of size n. For more complicated statistics the saving in time using the m-bootstrap can be more impressive and essential.

References

Babu, G. & Rao, C. (1993), Bootstrap methodology, *in* C. Rao, ed., 'Handbook of Statistics', Vol. 9, Elsevier Science Publishers, pp. 627–659.

Bickel, P. & Freedman, D. (1981), 'Some asymptotic theory for the bootstrap', *The Annals of Statistics* **9**(6), 1196–1217.

Bickel, P., Götze, F. & van Zwet, W. (1997), 'Resampling fewer than n observations: gains, losses and remedies for losses.', *Statistica Sinica* **7**(1), 1–31.

Bickel, P. & Ren, J. (1995), The m out of n bootstrap and goodness of fit tests with doubly censored data, *in* 'Robust Statistics: data, analysis and computer intensive methods', Springer-Verlag, pp. 35–47.

Bickel, P. & Yahav, J. (1988), 'Richardson extrapolation and the bootstrap', *Journal of the American Statistical Association* **83**, 387–393.

DeAngelis, D., Hall, P. & Young, G. (1993), 'A note on coverage error of bootstrap confidence intervals for quantiles', *Math. Proc. Camb. Phil. Soc.* **114**, 517–531.

Falk, M. & Janas, D. (1992), 'Edgeworth expansions for studentized and prepivoted sample quantiles', *Statistics and Probability letters* **14**, 13–24.

Falk, M. & Reiss, R. (1989), 'Weak convergence of smoothed and nonsmoothed bootstrap quantile estimates', *The Annals of Probability* **17**, 362–371.

Feller, W. (1966), *An introduction to probability theory and its applications.*, Vol. II, 2nd edn, John Wiley & Sons.

Hall, P. (1992), *The bootstrap and Edgeworth expansion*, Springer-Verlag.

Hall, P. & Martin, M. (1991), 'On the error incurred using the bootstrap variance estimate when constructing confidence intervals for quantiles', *Journal of multivariate analysis* **38**, 70–81.

Hall, P. & Sheather, S. (1988), 'On the distribution of a studentized quantile', *Journal of the Royal Statistical Soc. B* **50**(3), 381–391.

Politis, D. & Romano, J. (1994), 'Large sample confidence regions on subsamples under minimal assumptions.', *The Annals of Statistics* **22**(4), 2031–2050.

Reiss, R. (1976), 'Asymptotic expansions for sample quantiles', *The Annals of Probability* **4**(2), 249–258.

Sakov, A. (1998), Using the m out of n bootstrap in hypothesis testing, PhD thesis, University of California, Berkeley.

Serfling, R. (1980), *Approximation theorems to Mathematical Statistics*, Wiley.

Shorack, G. & Wellner, J. (1986), *Empirical processes with applications to statistics*, Wiley.

Singh, K. (1981), 'On the asymptotic accuracy of Efron's bootstrap', *The Annals of Statistics* **9**(6), 1187–1195.