Better Bootstrap Confidence Intervals

by Bradley Efron

Gregory Imholte

University of Washington, Department of Statistics

May 29, 2012

We wish to make a confidence interval for some parameter $\theta \equiv T(F)$ (e.g. $\theta = E_F\, X$), based on data

$$X_i \overset{\text{i.i.d.}}{\sim} F \in \mathcal{F}.$$

- Exact intervals
- Normal approximation ($\hat{\theta} \pm 1.96 \times se(\hat{\theta})$)
- Approximations based on series expansions
- Bootstrap confidence intervals

Exact intervals and series expansions are hard, solutions differ from problem to problem. Normal approximation can be poor. Even certain bootstrap confidence intervals can be sub-par!

For simplicity, we start by assuming that $\hat{\theta} \sim f_\theta$ estimates $\theta \in \Theta$. We make the following assumptions regarding our estimator. For our family $\{f_\theta : \theta \in \Theta\}$, there exists a monotone increasing transformation h and constants z_0 and a such that

$$\hat{\phi} = h(\hat{\theta}) \qquad\qquad \phi = h(\theta)$$

satisfy

$$\hat{\phi} = \phi + \sigma_\phi(Z - z_0) \qquad\qquad Z \sim N(0,1)$$

with

$$\sigma_\phi = 1 + a\phi$$

- The constant z_0 is the bias correction constant
- The constant a is the acceleration constant

Let $\hat{G}(s) = P_{\hat{\theta}}\{\hat{\theta}^* < s\}$ denote the bootstrap distribution function. We can either estimate this via Monte Carlo (i.e., resample data x^* with replacement and compute $\hat{\theta}^*$ over and over), or use $\hat{\theta}^* \sim f_{\hat{\theta}}$.

Lemma

Under the conditions in the previous slide, if $1 - \Phi\left(\frac{1}{|a|} - |z_0|\right) < \alpha < .5$ *the correct central confidence interval of level 1 - 2α for θ is*

$$\left[\hat{G}^{-1}(\Phi(z[\alpha])), \hat{G}^{-1}(\Phi(z[1-\alpha]))\right]$$

where

$$z[\alpha] = z_0 + \frac{(z_0 + z^{(\alpha)})}{1 - a(z_0 + z^{(\alpha)})}$$

The interval in Lemma 1 defines the BC_a method, and we compute it with suitable estimates of z_0 and a.
We can consider the BC_a interval as a generalization of two methods. The *percentile method*:

$$\left[\hat{G}^{-1}(\alpha), \hat{G}^{-1}(1-\alpha)\right]$$

The *BC method*:

$$\left[\hat{G}^{-1}(\Phi(2z_0 + z_\alpha), \hat{G}^{-1}(\Phi(2z_0 + z_{(1-\alpha)}))\right]$$

Essentially, the percentile method works if we assume existence of a normalizing transformation with $a = 0$ and $z_0 = 0$. The BC method is a predecessor to the BC_a method, and arises if we assume existence of a normalizing transformation with $a = 0$ and $z_0 \neq 0$.

The BC and BC_a methods provide increasingly powerful corrections to the percentile interval, by adjusting the endpoint for error in our estimation of the sampling distribution by the bootstrap distribution.

How much in error is the ordinary percentile interval? How do the BC and BC_a intervals compensate? To answer the question, the Edgeworth and Cornish-Fisher expansions will come in handy. Consider the quantity

$$S_n = \frac{\sqrt{n}(\hat{\theta} - \theta)}{\sigma}.$$

Under a certain class of asymptotically standard normal estimators, we can express the c.d.f. and quantiles of S_n in terms of polynomials and standard normal related quantities.

The Edgeworth expansion gives a c.d.f. approximation

$$P(S_n \leq x) = \Phi(x) + n^{-1/2} p_1(x)\phi(x) + O(n^{-1}).$$

The quantity $p_1(x)$ is an even polynomial of x with coefficients related to the cumulants of S_n. The Cornish-Fisher expansion gives an expansion of the quantiles of S_n. Let u_α denote the α quantile of S_n. We have

$$u_\alpha = z_\alpha - n^{-1/2} p_1(z_\alpha) + O(n^{-1}).$$

We can also produce valid expansions for a bootstrap distribution:

$$\begin{aligned} P\{n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma} \leq x|\mathcal{X}\} &= \Phi(x) + n^{-1/2}\hat{p}_1(x) + O_p(n^{-1}) \\ \hat{u}_\alpha &= z_\alpha - n^{-1/2}\hat{p}_1(z_\alpha) + O_p(n^{-1}). \end{aligned}$$

The polynomial $\hat{p}_1$ is the same as p, except bootstrap estimates replace population valued coefficients. Generally,

$$\hat{p}_1(x) - p_1(x) = O_p(n^{-1/2}).$$

Our percentile endpoints are defined via $P(\hat{\theta}^* \le \hat{y}_\alpha | \mathcal{X}) = \alpha$. We can write $\hat{y}_\alpha$ in terms of expansion quantities:

$$\begin{aligned} P\{n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma} \le \hat{u}_\alpha | \mathcal{X}\} &= \alpha \\ P\{\hat{\theta}^* \le \hat{\theta} + n^{-1/2}\hat{\sigma}\hat{u}_\alpha | \mathcal{X}\} &= \alpha \end{aligned}$$

This implies

$$\begin{aligned} \hat{y}_\alpha &= \hat{\theta} + n^{-1/2}\hat{\sigma}\hat{u}_\alpha \\ &= \hat{\theta} + n^{-1/2}\hat{\sigma}[z_\alpha - n^{-1/2}\hat{p}_1(z_\alpha)] + O_p(n^{-3/2}) \\ &= \hat{\theta} + n^{-1/2}\hat{\sigma}z_\alpha - n^{-1}\hat{\sigma}p_1(z_\alpha) + O_p(n^{-3/2}). \end{aligned}$$

So we have our percentile endpoint in terms of Edgeworth quantities. How does this compare to an "ideal" interval?

We introduce yet another Edgeworth/Cornish Fisher pair for the quantity

$$T_n = \frac{\sqrt{n}(\hat{\theta} - \theta)}{\hat{\sigma}}:$$

$$\begin{aligned} P(T_n \leq x) &= \Phi(x) + n^{-1/2} q_1(x)\phi(x) + O(n^{-1}) \\ \nu_\alpha &= z_\alpha - n^{-1/2} q_1(z_\alpha) + O(n^{-1}). \end{aligned}$$

If we knew our sampling distribution for T_n, we would know ν_α, and we could make the ideal interval endpoint

$$\hat{\theta} - n^{-1/2}\hat{\sigma}\nu_{(1-\alpha)} = \hat{\theta} + z_\alpha n^{-1/2}\hat{\sigma} + n^{-1} q_1(z_\alpha)\hat{\sigma} + O_p(n^{-3/2})$$

Relating these two quantities, if we shift the percentile interval $\hat{y}_\alpha$ by

$$n^{-1}\hat{\sigma}\left[p_1(z_\alpha) + q_1(z_\alpha)\right] = n^{-1}\hat{\sigma}\left[2p(0) + \frac{1}{6}C\sigma^{-3}z_\alpha^2\right],$$

we would have a *second order correct* interval. We'll discuss C soon. Plugging in $\hat{p}$ and $\hat{q}$ retains this property.

We'll show that the BC interval gets us part of the way, and that the BC_a interval achieves the full correction.

Recall from before the BC interval

$$\left[\hat{G}^{-1}(\Phi(2z_0 + z_\alpha), \hat{G}^{-1}(\Phi(2z_0 + z_{(1-\alpha)}))\right]$$

We estimate

$$\begin{aligned}\hat{z}_0 &= \Phi^{-1}\left(\hat{G}(\hat{\theta})\right) \\ &= n^{-1/2}\hat{p}_1(0) + O_p(n^{-1}).\end{aligned}$$

Using a Taylor series expansion about z_α,

$$\begin{aligned}\Phi(2\hat{z}_0 + z_\alpha) &= \Phi(z_\alpha) + 2\hat{z}_0\phi(z_\alpha) + O_p(n^{-1}) \\ &= \alpha + n^{-1/2}2\hat{p}_1(0)\phi(z_\alpha) + O_p(n^{-1}) \\ &\equiv \beta\end{aligned}$$

For $\hat{u}_\beta$ defined by $P\left\{\frac{\sqrt{n}(\hat{\theta}^* - \hat{\theta})}{\hat{\sigma}} \leq \hat{u}_\beta | \mathcal{X}\right\} = \beta$, via a Cornish-Fisher expansion and Taylor series arguments for Φ^{-1} and p_1,

$$\begin{aligned} \hat{u}_\beta &= z_\alpha - n^{-1/2}\hat{p}_1(z_\alpha) + n^{-1/2}2\hat{p}_1(0) + O_p(n^{-1}) \\ \hat{G}^{-1}(\Phi(2z_0 + z_\alpha)) &= \color{red}{\hat{\theta} + \hat{\sigma} n^{-1/2} z_\alpha - n^{-1}\hat{p}_1(z_\alpha)\hat{\sigma}} \\ &\quad + \color{blue}{n^{-1}\hat{\sigma}2p_1(0)} + O_p(n^{-3/2}) \\ &= \color{red}{\hat{y}_\alpha} + \color{blue}{n^{-1}\hat{\sigma}2p_1(0)} + O_p(n^{-3/2}) \end{aligned}$$

The *BC* interval only makes part of the required correction, in blue.

We can play a similar game with the BC_a endpoint. Define $a = n^{-1/2}\frac{1}{6}C\sigma^{-3}$. This a is the same acceleration constant from earlier. Then, as from Lemma 1

$$\begin{aligned} z[\alpha] &= \hat{z}_0 + \frac{(\hat{z}_0 + z_\alpha)}{1 - \hat{a}(\hat{z}_0 + z_\alpha)} \\ &= z_\alpha + 2\hat{z}_0 + \hat{a}z_\alpha^2 + O_p(n^{-1}) \end{aligned}$$

We end up with

$$\hat{G}^{-1}(\Phi(z[\alpha])) = \hat{y}_\alpha + n^{-1}\hat{\sigma}\left[2p(0) + \frac{1}{6}C\sigma^{-3}z_\alpha^2\right] + O_p(n^{-3/2})$$

This is exactly the correction we require! The BC_a intervals are second order correct.

Meanwhile, suppose we bootstrap the quantities

$$t_b^* = \frac{\sqrt{n}(\hat{\theta}^* - \hat{\theta})}{SE(\hat{\theta}^*)}$$

and let $\hat{\nu}_\alpha$ be the α quantile of the t^*s. The bootstrap-t interval is

$$\left[\hat{\theta} - n^{-1/2}\hat{\sigma}\hat{\nu}_{(1-\alpha)}, \hat{\theta} - n^{-1/2}\hat{\sigma}\hat{\nu}_\alpha\right]$$

and this interval is also second order correct. This interval directly estimates the "ideal interval".

To understand C and a we need to go back and look at when these expansions are valid. Suppose that $X_1, X_2, \ldots$ are i.i.d. column vectors of fixed dimension k with mean μ. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ denote the vector of component-wise means. Let $A : \mathcal{R}^d \to \mathcal{R}$ be a smooth function satisfying $A(\mu) = 0$. We are interested in functions such as

$$A(\bar{X}) = \frac{g(\bar{X}) - g(\mu)}{h(\mu)} \quad \text{or} \quad A(\bar{X}) = \frac{g(\bar{X}) - g(\mu)}{h(\bar{X})}$$

where our parameter of interest is $\theta = g(\mu)$ and $h(\mu)^2$ is the asymptotic variance of $\sqrt{n}\hat{\theta}$. Note that this model covers maximum likelihood estimation in exponential families as well as any number of nonparametric quantities.

Calculating $\hat{a}$ can be a chore. We let

$$\begin{aligned}
\hat{a}_i &= \frac{\partial}{\partial x^{(i)}} A(x)\bigg|_{x=\bar{X}} \\
\hat{\mu}_{ijk} &= \frac{1}{n}\sum_{l=1}^{n}(X_l - \bar{X})^{(i)}(X_l - \bar{X})^{(j)}(X_l - \bar{X})^{(k)}.
\end{aligned}$$

Then

$$\hat{a} = n^{-1/2}\frac{1}{6}\hat{\sigma}^{-3}\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{d}\hat{a}_i\hat{a}_j\hat{a}_k\hat{\mu}_{ijk}.$$

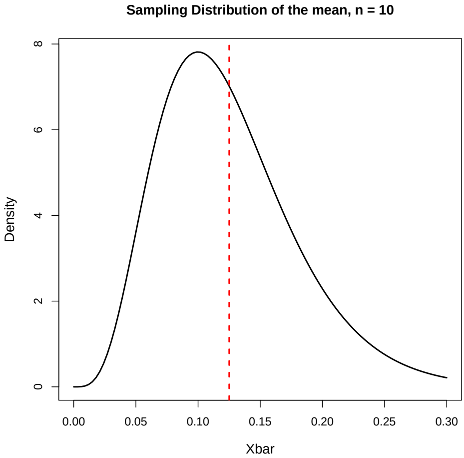
To estimate a in a non-parametric framework, consider parameters θ defined as functions of arbitrary vectors of populations moments μ, $\theta = t(\mu)$. We define the *empirical influence function*

$$U_i = \lim_{\Delta \to 0} \frac{t\left[(1-\Delta)\bar{X} + X_i\Delta\right] - t(\bar{X})}{\Delta}.$$

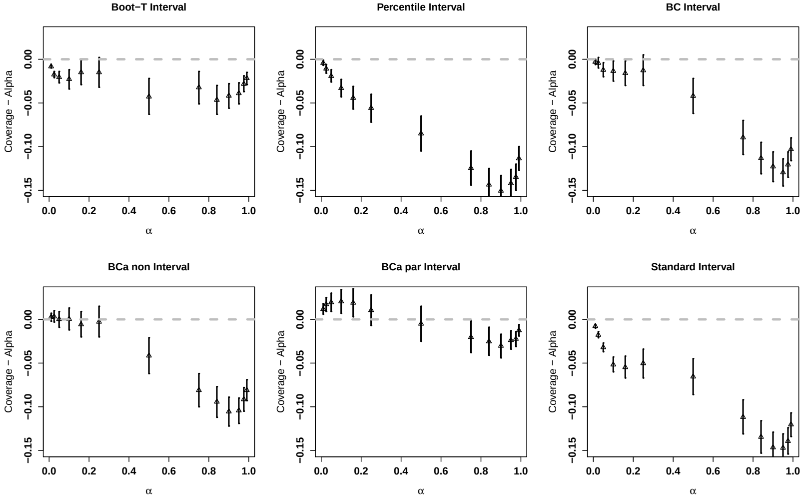
We then have

$$\hat{a} = \frac{1}{6} \frac{\sum_{i=1}^{n} U_i^3}{\left(\sum_{i=1}^{n} U_i^2\right)^{(3/2)}}.$$

As a quick example, we let $x_1, \ldots, x_n \sim i.i.d.$ Gamma$(.5, \text{rate} = 4)$ for $n = 10$. The sampling distribution looks like



We can also examine one-sided interval coverage: $P(\theta \leq \hat{\theta}[\alpha])$.



At the 95% level, these intervals have width and nominal coverage as below:

	Actual Coverage	Mean Interval Length
Boot-T Interval	0.94	0.45
Percentile Interval	0.83	0.18
BC Interval	0.83	0.19
BCa non Interval	0.86	0.20
BCa par Interval	0.91	0.31
Standard Interval	0.83	0.19

Table: 4

- The percentile, BC, and BC_a methods can be seen as extending each other
- We can understand their performance via the Edgeworth expansion
- The BC_a interval is not as "automatic" as Efron might want you to believe
- The bootstrap-t interval may be preferable in the presence of a stable estimate of variance