Better Bootstrap Confidence Intervals

by Bradley Efron

Gregory Imholte

University of Washington, Department of Statistics

April 12, 2012

Suppose we wish to make inference on some parameter $\theta \equiv T(F)$ (e.g. $\theta = E_F\, X$), based on data

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

We might suppose that

$$\mathcal{F} = \{F : E_F\, X^2 < \infty\},$$

or we might let

$$\mathcal{F} = \{F_\theta : \theta \in \Theta \subset \mathbb{R}^p\}.$$

We often concern ourselves with two tasks:

- Point estimates $\hat{\theta}$
- Interval estimates $(\hat{\theta}[\alpha], \hat{\theta}[1-\alpha])$

where $\hat{\theta}[\alpha]$ denotes an α level endpoint of a confidence interval.

In our frequentist hats, the two tasks usually start (and end) with maximum likelihood estimation. We turn our frequentist crank, finding

- The MLE $\hat{\theta}_{MLE}$
- The observed Fisher information $\hat{I}_n$.

We then use the relationship

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta) \approx_d N\left(0, \hat{I}_n^{-1}\right)$$

to pivot around θ and generate an interval estimate

$$\hat{\theta}_{MLE} \pm z^{(\alpha)} \frac{\hat{I}_n^{-1}}{\sqrt{n}}$$

But what can go wrong?

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta) \approx_d N\left(0, \hat{I}_n^{-1}\right)$$

is just an approximation. For small sample sizes, we have no guarantees. Strong skewness or heavy tails in data distribution may distort the sampling distribution of $\hat{\theta}$. Result:

Badly calibrated confidence intervals!

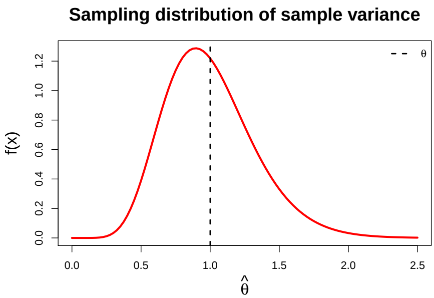
So what can we do about that? Ad-hoc solutions:

- Find some monotone transformation g that makes $g(\hat{\theta})$ *really* approximately normal
- Get lucky: identify the exact distribution of $\hat{\theta}$ or some transformation thereof
- This can be hard!

For example, consider

$$X_i \overset{i.i.d.}{\sim} N(\mu, \theta) \qquad \hat{\theta} = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2.$$

Suppose $\theta = 1$. Normal theory yields that $\hat{\theta} \sim \frac{\chi^2_{(n-1)}}{(n-1)}$.



When the data are normal we also have that $Var(\hat{\theta}) = \frac{2\sigma^4}{n-1}$. Applying asymptotic approximations, an approximate 95% confidence interval would be

$$\hat{\theta} \pm 1.96 \times \sqrt{\frac{2\hat{\theta}^2}{n-1}}$$

and an exact 95% confidence interval for θ is

$$(n-1)\hat{\theta}\left(\frac{1}{\chi^2_{n-1,.975}}, \frac{1}{\chi^2_{n-1,.025}}\right)$$

For $n = 20$, the approximate intervals have 88% coverage.

Suppose that last two ad-hoc solutions are *hard*, and our interval estimates perform badly. What else can we do?

Could we adjust for the skewness or kurtosis that mucks up our intervals? This is the road taken by Abramovitch & Singh (1985), Bartlett (1953), Hall (1983), to name a few. The general idea is that the true confidence limits often exist in the form of

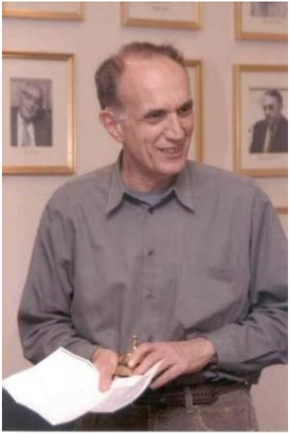
$$\hat{\theta}[\alpha] = \hat{\theta} + \hat{\sigma}\left(z^{(\alpha)} + \frac{A_n^{(\alpha)}}{\sqrt{n}} + \frac{B_n^{(\alpha)}}{n} + \cdots\right).$$

These methods generally boil down to estimating some function of higher order moments to correct the interval.

Are we in the clear yet? No. These methods can become exceedingly cumbersome to apply in multiparameter families, especially for more exotic transformations of parameters (e.g. ratios of means, ratios of regression parameters, products of parameters...).

What are we to do?

"You can pull yourself up by your bootstraps and you don't need anything else" -Efron



The bootstrap is a data resampling technique that allows us to use the empirical distribution function $\mathbb{F}_n$ to estimate the sampling distribution of our estimator $\hat{\theta}$. Recall $\mathbb{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{[t \leq x_i]}$.

For $b = 1, \ldots, B$:

- Sample $X_i^{*(b)} \overset{i.i.d.}{\sim} \mathbb{F}_n \quad i = 1, \ldots, n$.
- Compute $\hat{\theta}^{*(b)}$ from data $(X_1^{*(b)}, \ldots, X_n^{*(b)})$.

Use the bootstrap distribution of $\hat{\theta}^*$ to compute quantities of interest. E.g.

$$(\hat{\theta}^*_{(B\alpha)}, \hat{\theta}^*_{(B(1-\alpha))})$$

would give an approximate 100(1-2α)% confidence interval for θ.

The bootstrap technique applies to a huge variety of parameters in both parametric and nonparametric families. Since its inception the technique has seen various refinements.

The primary contribution of this talk's namesake is the accelerated bootstrap confidence interval (BC_a).

The method's goal:

Automatically create confidence intervals that adjust for underlying higher order effects.

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

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

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 $\hat{G}(s) = P\{\hat{\theta}^* < s\}$ denote the bootstrap distribution function. We can either estimate this via Monte Carlo or use $\hat{\theta}^* \sim f_{\hat{\theta}}$. The BC_a interval is defined as below.

Lemma

Under the conditions in the previous slide 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

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

Note that the form of the transformation g never comes into play. In a sense, the method automatically selects a transformation that brings $\hat{\theta}$ to normality, computes an exact 95% interval, and then transforms backwards to reach the θ scale again.

We require estimates of a and z_0 to make this work:

$$z_0 \approx \Phi^{-1}(\hat{G}(\hat{\theta})) \quad a \approx \frac{1}{6} SKEW_{\theta=\hat{\theta}}(\dot{l}_\theta)$$

I.e. the skew of the score transformation of $\hat{\theta}$

A main theorem of the paper is that this interval is *second-order correct* in the sense that the endpoints of the BC_a confidence intervals are very close to the true exact endpoints,

$$\hat{\theta}_{BC}[\alpha] - \hat{\theta}_{EX}[\alpha] = O_p(n^{-\frac{3}{2}}).$$

That is to say, the BC_a method successfully captures $\frac{A_n^{(\alpha)}}{\sqrt{n}}$ below:

$$\hat{\theta}[\alpha] = \hat{\theta} + \hat{\sigma}(z^{(\alpha)} + \frac{A_n^{(\alpha)}}{\sqrt{n}} + \frac{B_n^{(\alpha)}}{n} + \cdots).$$

In the previous sample variance example, BC_a intervals have estimated coverage essentially identical with the exact intervals.

Looking ahead:

- This method is also second order correct in multiparameter families, with different estimates of a and z_0
- The method also works in nonparametric settings, via an extension from multiparameter families
- Lots of math