# Bootstrap Confidence Intervals

Patrick Breheny

September 18

# Introduction

- So far, we have discussed the idea behind the bootstrap and how it can be used to estimate standard errors
- Standard errors are often used to construct confidence intervals based on the estimate having a normal sampling distribution:

$$\hat{\theta} \pm z_{1-\alpha/2} SE;$$

alternatively, the interval could be based on the $t$ distribution
- The bootstrap SE can be used in this way as well

# Introduction

- However, recall that the bootstrap can also be used to estimate the CDF $G$ of $\hat{\theta}$
- Thus, with the bootstrap, we do not need to make assumptions/approximations concerning the sampling distribution of $\hat{\theta}$ – we can estimate it as part of the confidence interval procedure
- This has been an active area of theoretical research into the bootstrap

# The bootstrap-$t$ interval

- For example, suppose that the standard error of an estimate varies with the size of the estimate
- If this is so, then our confidence interval should be wider on the right side than it is on the left
- One way to implement this idea is to estimate the SE separately for each bootstrap replication — this is the idea behind the *bootstrap-$t$ interval*

# The bootstrap-$t$ interval: Procedure

The procedure of the *bootstrap-$t$* interval is as follows:

(1) For each bootstrap sample, compute

$$z_b^* = \frac{\hat{\theta}_b^* - \hat{\theta}}{\widehat{SE}_b^*}$$

where $\widehat{SE}_b^*$ is an estimate of the standard error of $\hat{\theta}^*$ based on the data in the $b$th bootstrap sample

(2) Estimate the $\alpha$th percentile of $z^*$ by the value $\hat{t}_\alpha$ such that

$$B^{-1} \sum_b I(z_b^* \leq \hat{t}_\alpha) = \alpha$$

(3) A $1 - \alpha$ confidence interval for $\theta$ is then

$$(\hat{\theta} - \hat{t}_{1-\alpha/2}\widehat{SE}, \hat{\theta} - \hat{t}_{\alpha/2}\widehat{SE})$$

# The bootstrap-$t$ interval: Example

- As a small example, the survival times of 9 rats were 10, 27, 30, 40, 46, 51, 52, 104, and 146 days
- Consider estimating the mean; the point estimates are $\hat{\theta} = 56.2$ and $\widehat{SE} = 14.1$
- The percentile points for a 95% confidence interval:

| | | |
|---|---|---|
| Normal | -1.96 | 1.96 |
| $t$ | -2.31 | 2.31 |
| Bootstrap-$t$ | -4.86 | 1.61 |

- This translates into the following confidence intervals:

| | | |
|---|---|---|
| Normal | 28.5 | 84.0 |
| $t$ | 23.6 | 88.9 |
| Bootstrap-$t$ | 33.4 | 125.1 |

# The bootstrap-$t$ interval: R

- If you want to implement this confidence interval in R, your function that you pass to boot will need to return two things: $\hat{\theta}_b^*$ and $\hat{\mathbb{V}}(\hat{\theta}_b^*)$
- For example:

```
mean.boot <- function(x, ind) {
  c(mean(x[ind]), var(x[ind])/length(x))
}
out <- boot(x, mean.boot, 999)
boot.ci(out)
```

- The function `boot.ci` returns the bootstrap-$t$ interval (which it calls the "studentized" interval) along with the normal interval and some other intervals which we will talk about next

# Double bootstrap

- Estimation of $\widehat{SE}_b^*$ is straightforward for the mean (and a few other statistics), but what about in general?
- Once again, the functional delta method and jackknife are options
- The bootstrap is also an option – here we use the bootstrap within a bootstrap replication to estimate the standard error of a bootstrap replication – and is called the *double bootstrap*
- The obvious drawback to the double bootstrap is that it requires $B^2$ bootstrap replications, and is therefore time-consuming to compute

# Confidence intervals and sampling distributions

- The next interval we will discuss is not based on the usual pivoting ideas we encounter when constructing confidence intervals
- Instead, it is based on the observation that, if $\hat{\theta}^* \sim \mathrm{N}(\hat{\theta}, \widehat{\mathrm{SE}}^2)$,

$$\hat{\theta} \pm \widehat{\mathrm{SE}} z_{1-\alpha/2} = (\hat{\theta}^*_{\alpha/2}, \hat{\theta}^*_{1-\alpha/2})$$

  is a $1-\alpha$ level confidence interval for $\theta$, where $\hat{\theta}^*_\alpha$ is the $\alpha$th percentile of the distribution of $\hat{\theta}^*$
- Letting $\hat{G}$ the empirical CDF of $\hat{\theta}^*$, another way of writing the above interval is $[\hat{G}^{-1}(\alpha/2), \hat{G}^{-1}(1-\alpha/2)]$
- Thus, there is a connection (at least, if $\hat{\theta}^* \sim \mathrm{Normal}$) between the resampling distribution of $\hat{\theta}^*$ and confidence intervals for $\theta$

# The percentile interval

- This interval – called the *percentile interval* – is much simpler to construct than the bootstrap-$t$, and turns out to work surprisingly well in a wide range of examples
- The percentile interval has an additional justification beyond approximation to the normal: it is invariant to transformations of $\theta$
- For example, suppose we wish to construct a confidence interval for $\phi = m(\theta)$, where $m$ is a monotone transformation
- We do not need to generate new bootstrap replications: $\hat{\phi}_b^* = m(\hat{\theta}_b^*)$, because

$$\left(\hat{\phi}_{\alpha/2}, \hat{\phi}_{1-\alpha/2}\right) = \left[m(\hat{\theta}_{\alpha/2}), m(\hat{\theta}_{1-\alpha/2})\right]$$

# Percentile interval theorem

- In the 20th century, statisticians explored dozens of transformations of various statistics designed to make the sampling distribution of the statistic more normal, from simple ones like the log-odds to more complex ones like the Fisher and Anscombe transformations
- One compelling justification for the percentile interval is that, if such a transformation exists, the percentile interval will find it automatically
- **Theorem:** Suppose there exists $\phi = m(\theta)$ such that

$$\hat{\phi}^* \sim N\left(\hat{\phi}, \tau^2\right)$$

for some $\tau$. Then the percentile interval $(\hat{\theta}^*_{\alpha/2}, \hat{\theta}^*_{1-\alpha/2})$ is equivalent to the optimal pivot-based interval based on the above relationship.

# Example: Automatic transformation

- As an example, suppose $X_1, \ldots, X_{10} \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$, and we are interested in estimating $\theta = e^{\mu}$
- In this case, there does exist a normalizing transformation $\phi = \log(\theta)$ such that $\hat{\phi} \sim \mathrm{N}(\phi, \tau^2)$
- Suppose, however, that we didn't know any of this and applied the ordinary nonparametric bootstrap:

| | $\hat{\theta}_L$ | $\hat{\theta}_U$ |
|---|---|---|
| Exact | 0.49 | 2.29 |
| Normal bootstrap | 0.15 | 1.81 |
| Percentile interval | 0.50 | 2.10 |

# Percentile intervals in R

- Percentile intervals are also returned by `boot.ci`
- They can also be obtained by applying `quantile` to the output of `boot`
- For the mouse survival data from earlier,

| | | |
|---|---|---|
| Normal | 28.5 | 84.0 |
| $t$ | 23.6 | 88.9 |
| Bootstrap-$t$ | 33.4 | 125.1 |
| Percentile | 32.1 | 87.7 |

# Standard vs. bootstrap approach

- The standard approach is based on taking literally the following asymptotic relationship

$$\frac{\hat{\theta} - \theta}{\sigma} \sim N(0, 1)$$

- The bootstrap percentile interval relaxes that assumption; instead of requiring normality of $\hat{\theta}$, it requires only that $\exists\, m : \phi = m(\theta)$ satisfies

$$\frac{\hat{\phi} - \phi}{\tau} \sim N(0, 1)$$

- However, this still requires that there exists a single transformation that is both normalizing and variance-stabilizing – often, such a transformation does not exist

# Generalizing the bootstrap assumptions

- Thus, in a brilliant 1987 paper, Efron considers a generalization of the bootstrap assumptions
- Suppose that for some monotone increasing transformation $m$, some bias constant $z_0$, and some acceleration constant $a$, the following relationship holds for $\phi = m(\theta)$:

$$\frac{\hat{\phi} - \phi}{\sigma} \sim N(-z_0, 1), \qquad \sigma = 1 + a\phi$$

- Efron named the confidence interval based on this assumption the $\text{BC}_a$ interval, because it corrects for both bias and "acceleration" of the variance

# The BC$_a$ interval

- **Theorem**: If the relationship on the previous slide holds, then the following interval is "correct," in the sense of being the optimal pivot-based interval:

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

where

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

- Note that, in order to construct the $\mathrm{BC}_a$ interval, one has to estimate $z_0$ and $a$, but, like the percentile interval, it is not necessary to know $m$

# Example revisited

- Unfortunately, we don't have time to cover in detail the derivation of this interval and the estimation of $z_0$ and $a$, but it is available from `boot.ci`
- When applied to our mouse survival data, we have $\hat{z}_0 = -0.026$ and $\hat{a} = 0.066$, which gives us the adjusted percentage points $(0.037, 0.986)$ and the resulting interval

| | | |
|---|---|---|
| Normal | 28.5 | 84.0 |
| $t$ | 23.6 | 88.9 |
| Bootstrap-$t$ | 33.4 | 125.1 |
| Percentile | 32.1 | 87.7 |
| $\mathrm{BC}_a$ | 37.3 | 90.9 |

# Asymptotic accuracy of confidence intervals

- Let us denote a confidence interval procedure by its confidence points $\hat{\theta}[\alpha]$, where ideally, $\mathbb{P}(\theta \leq \hat{\theta}[\alpha]) = \alpha$
- A confidence point is called *first-order accurate* if

$$\mathbb{P}(\theta \leq \hat{\theta}[\alpha]) = \alpha + O(n^{-1/2})$$

and *second-order accurate* if

$$\mathbb{P}(\theta \leq \hat{\theta}[\alpha]) = \alpha + O(n^{-1})$$

# Second-order accuracy of bootstrap confidence intervals

- It can be shown (our textbook has some details on this and the regularity conditions that are required) that the standard interval and the percentile interval are first-order accurate, while the bootstrap-$t$ and $\mathrm{BC}_a$ intervals are second-order accurate – regardless of the true distribution, $F$
- This is a powerful theoretical justification for the bootstrap-$t$ and $\mathrm{BC}_a$ intervals
- Arguments can be made for either interval in different situations
- The $\mathrm{BC}_a$ interval is also transformation-invariant and *range-preserving*, meaning that, if it is not possible for a statistic or a function of a statistic to lie outside a certain range $[a, b]$, then the $\mathrm{BC}_a$ interval will be contained in $[a, b]$ (the bootstrap-$t$ intervals are neither transformation-invariant nor range-preserving)

# Bootstrap failure #1

- Bootstrap confidence intervals are unquestionably a tremendous methodological advance, and the $\mathrm{BC}_a$ interval represents the "state of the art" as far as nonparametric confidence intervals goes
- However, bootstrap intervals are limited by the accuracy of $\hat{F}$
- The empirical CDF is generally poorly estimated at the tails of a distribution
- Consequently, it is difficult to produce nonparametric confidence intervals for statistics that are highly influenced by distribution tails

# Bootstrap failure #2

- Furthermore, the bootstrap still requires conditions such as Hadamard differentiability and can fail to produce accurate intervals when those conditions do not hold
- We have already encountered the example of density estimation
- As an additional example, consider $\hat{\theta} = \max(x_i)$
- The $\mathrm{BC}_a$ interval can never exceed $\hat{G}^{-1}(1) = \max(x_i)$, and therefore provides a 0% confidence interval for any continuous distribution

# Homework

- In an earlier homework, we examined the coverage probabilities of parametric and nonparametric confidence intervals for the variance
- Your assignment now is to compare three nonparametric methods: the functional delta method, the bootstrap percentile interval, and the $\mathrm{BC}_a$ interval
- **Homework:** Conduct a simulation study to determine how the coverage probability and average interval width of these two intervals varies with the sample size $n$, when (i) the data is truly normally distributed, and (ii) the data follows an exponential distribution. For each distribution, produce a plot of coverage probability versus sample size, with lines representing the various methods, as well as a corresponding plot for interval width.