Bias Correction

Since we can estimate the bias of an estimator, we can use this to correct for it.

$$
\begin{aligned}
Bias_F &= E_F\left[T(\mathbf{x}) - t(F)\right] \\
&= E_F\left[T(\mathbf{x})\right] - t(F)
\end{aligned}
$$

Can approximate this by

$$
\begin{aligned}
Bias_{F_n^*} &= E_{F_n^*}\left[T\left(\mathbf{x}^*\right) - t\left(F_n^*\right)\right] \\
&= E_{F_n^*}\left[T\left(\mathbf{x}^*\right)\right] - t\left(F_n^*\right)
\end{aligned}
$$

One approach to correcting for bias is based on figuring out the expectation of an estimator.

For example, lets estimate μ_1^2 by $\bar{x}^2$

$$
\begin{aligned}
E\left[\bar{x}^2\right] &= E\left[\left(\mu_1 + \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_1)\right)^2\right] \\
&= \mu_1^2 + 2\mu_1 E\left[x - \mu_1\right] + \frac{\omega_2}{n} \\
&= \mu_1^2 + \frac{\omega_2}{n}
\end{aligned}
$$

Thus a less biased estimator is

$$\bar{x}^2 - \frac{\hat{\omega}_2}{n} = \bar{x}^2 - \frac{s^2}{n}$$

This approach assumes that the expectations can be determined.

As this is often difficult to do, the bootstrap gives us an easy way to approximate this as

$$\begin{aligned}\widehat{Bias}_B &= \hat{E}\left[T\right]^* - t\left(F_n^*\right) \\ &= \hat{E}\left[T\right]^* - T\left(\mathbf{x}\right)\end{aligned}$$

So the estimator

$$T\left(\mathbf{x}\right) - \widehat{Bias}_B$$

usually will have lower bias than $T\left(\mathbf{x}\right)$

Parametric Bootstrap

Assume that the data comes from some parametric family F_θ.

For example, with the Law School example, we could assume that the data is bivariate normal ($\theta = (\mu, \Sigma)$).

The common approach for determining standard errors in the parametric setting is to use the delta rule or some other asymptotic approximation.

For example

$$se(r) = \frac{1 - r^2}{\sqrt{n-3}}$$

So for the Law School example, r = 0.776 which gives $se(r)$ = 0.115 (which is similar to the nonparmetric bootstrap value of 0.130 (B = 1000)).

Instead of using the textbook asymptotic formula, we can use the parametric bootstrap instead.

Parametric Bootstrap for Estimating Standard Errors and Bias

1) Estimate the parameter given the assumed distributional form (call it $\hat{\theta}$)

2) Select B independent parametric bootstrap samples $\mathbf{x}^{*1}, \mathbf{x}^{*2}, \ldots, \mathbf{x}^{*B}$, each consisting of n data values drawn from the distribution $F_{\hat{\theta}}$.

3) Evaluate the bootstrap replication corresponding to each bootstrap sample,

$$T\left(\mathbf{x}^{*b}\right); \quad b = 1, \ldots, B$$

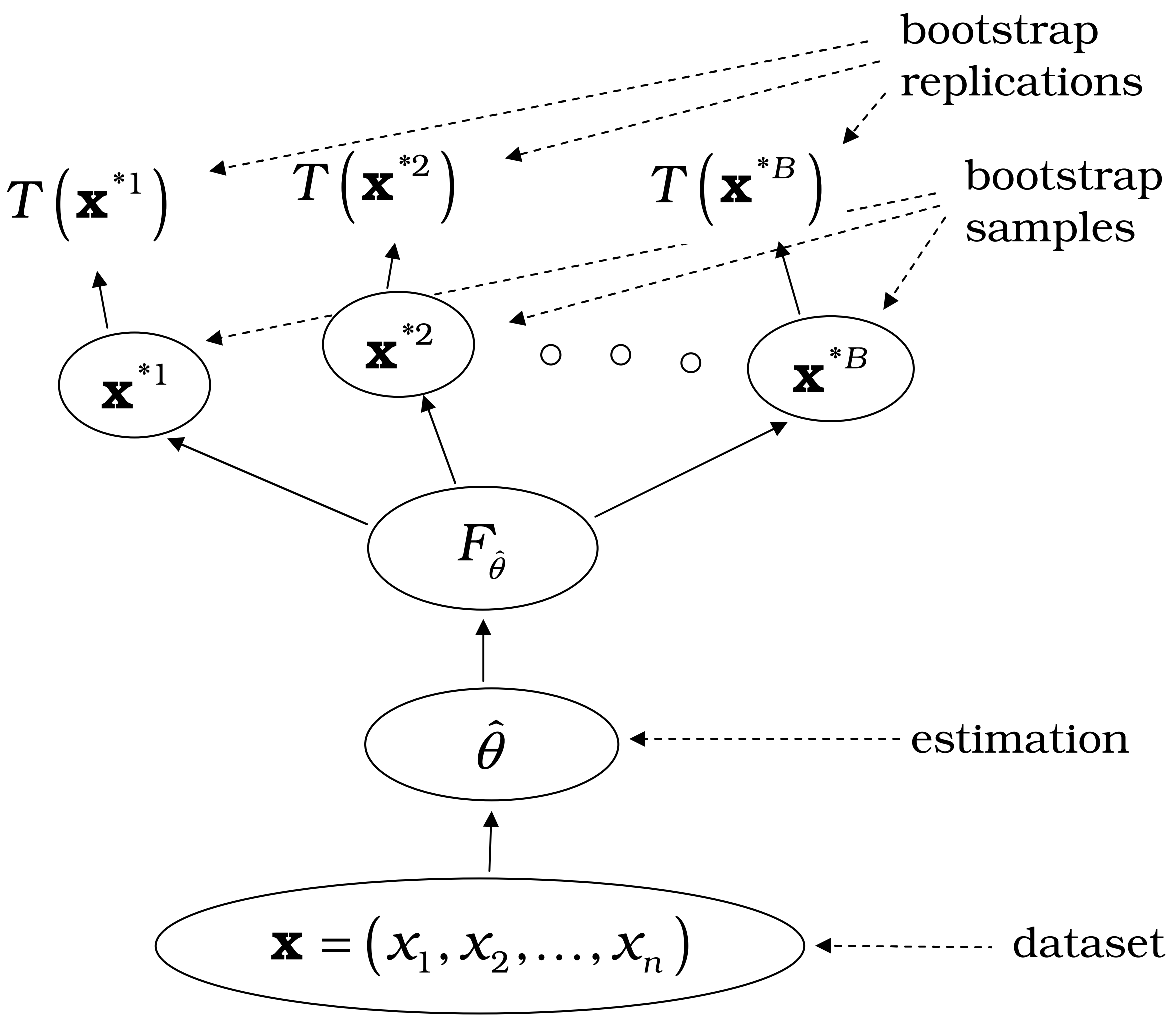
4) Evaluate the standard error, $se(T)^*$ by

$$\widehat{se}(T)^* = \sqrt{\frac{\sum_{b=1}^{B}\left(T\left(\mathbf{x}^{*b}\right) - \hat{E}[T]^*\right)^2}{B-1}}$$

where $\hat{E}[T]^* = \sum_{b=1}^{B} T\left(\mathbf{x}^{*b}\right) \Big/ B$

5) Evaluate the bias $\widehat{Bias}_B$ by

$$\widehat{Bias}_B = \hat{E}[T]^* - T(\mathbf{x})$$



For the Law School example, lets set $\hat{\theta} = \left(\hat{\mu}, \hat{\Sigma}\right)$ where

$$\hat{\mu} = [\bar{y} \quad \bar{z}]^T$$

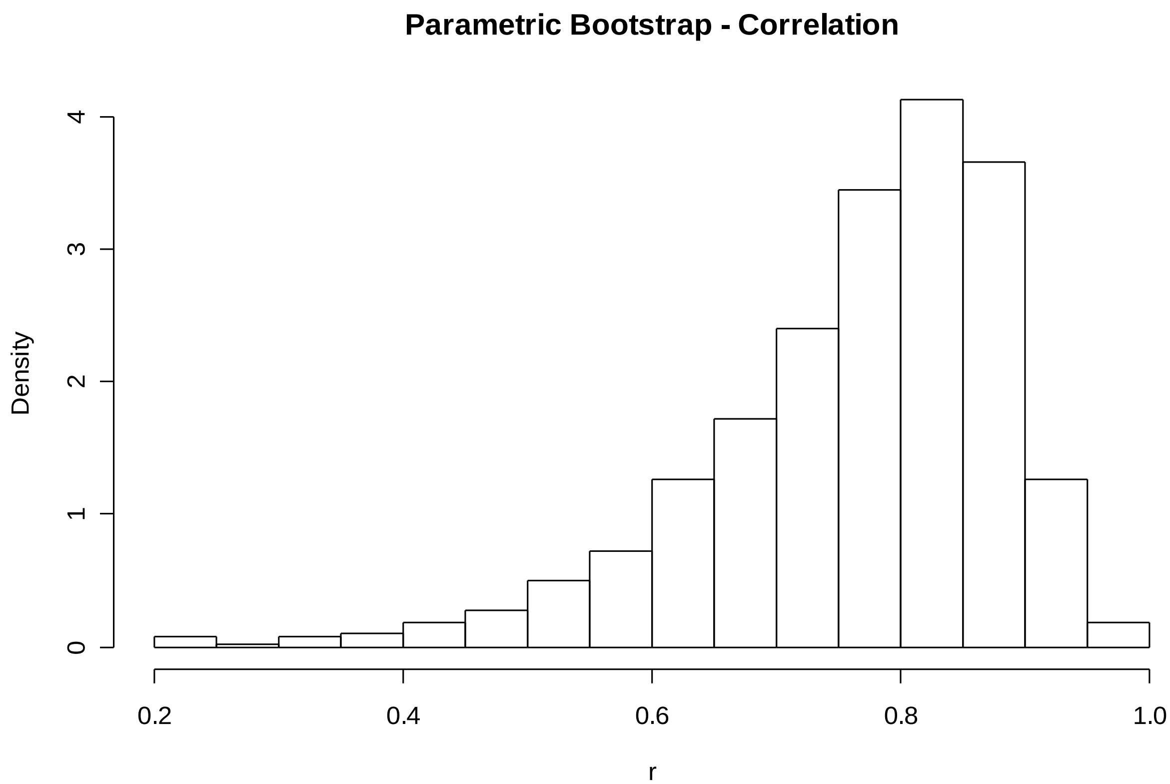
$$\hat{\Sigma} = \frac{1}{14}\begin{bmatrix} \sum (y_i - \bar{y})^2 & \sum (y_i - \bar{y})(z_i - \bar{z}) \\ \sum (y_i - \bar{y})(z_i - \bar{z}) & \sum (z_i - \bar{z})^2 \end{bmatrix}$$

are the usual parameter estimates (y: LSAT, z: GPA). For the example, these are

$$\hat{\mu} = [600.267 \quad 3.095]^T$$

$$\hat{\Sigma} = \begin{bmatrix} 1746.781 & 7.902 \\ 7.902 & 0.0593 \end{bmatrix}$$

$$r = 0.776$$



$se(r)$ = 0.115 (asymptotic formula)

$\widehat{se}(r)$ = 0.122 (parametric bootstrap)

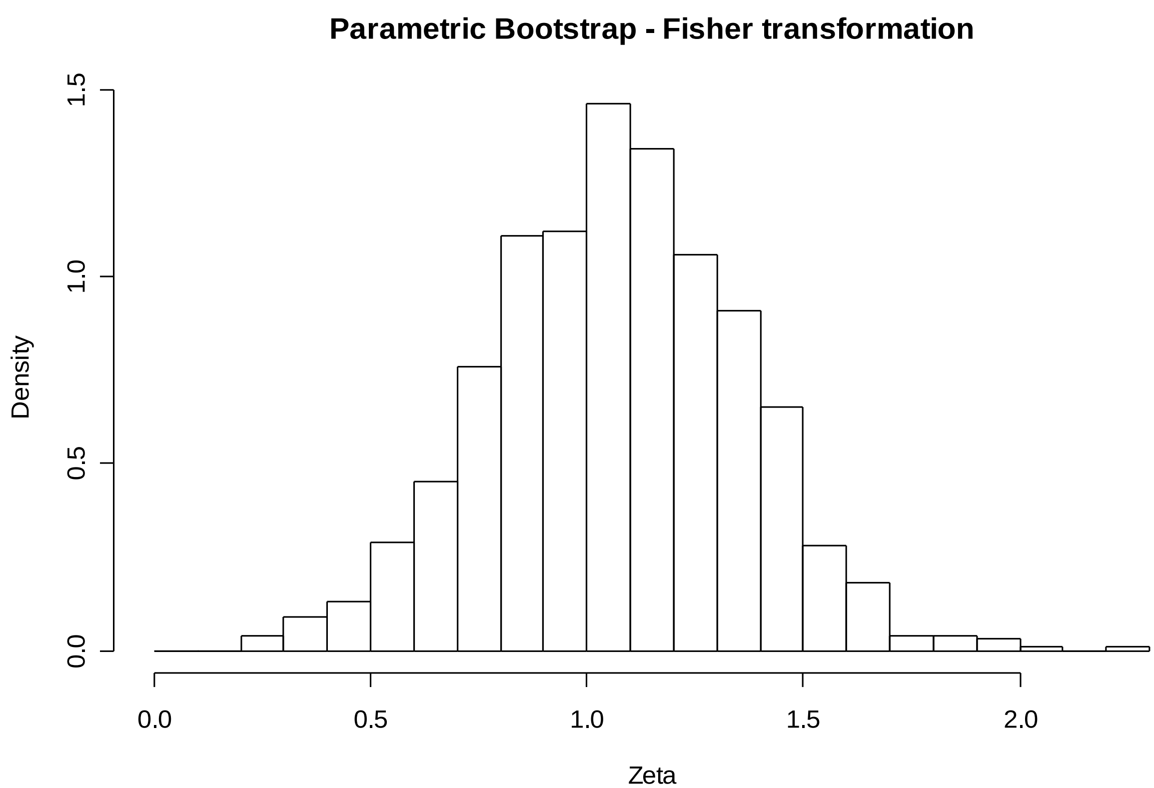
$\widehat{Bias}$ = -0.0125

For the correlation, Fisher's transformation is often used since it better distributional properties. Fisher showed that

$$\xi = \frac{1}{2}\frac{1+r}{1-r} \sim N\left(\frac{1}{2}\frac{1+\rho}{1-\rho}, \frac{1}{n-3}\right)$$

approximately.

Lets transform our bootstrap sample to see how well this works



$$\frac{1}{2}\frac{1+r}{1-r} = 1.036$$

$se(\xi)$ = 0.289 (asymptotic formula)

$\widehat{se}(\xi)$ = 0.292 (parametric bootstrap)

$\widehat{Bias}$ = 0.0319

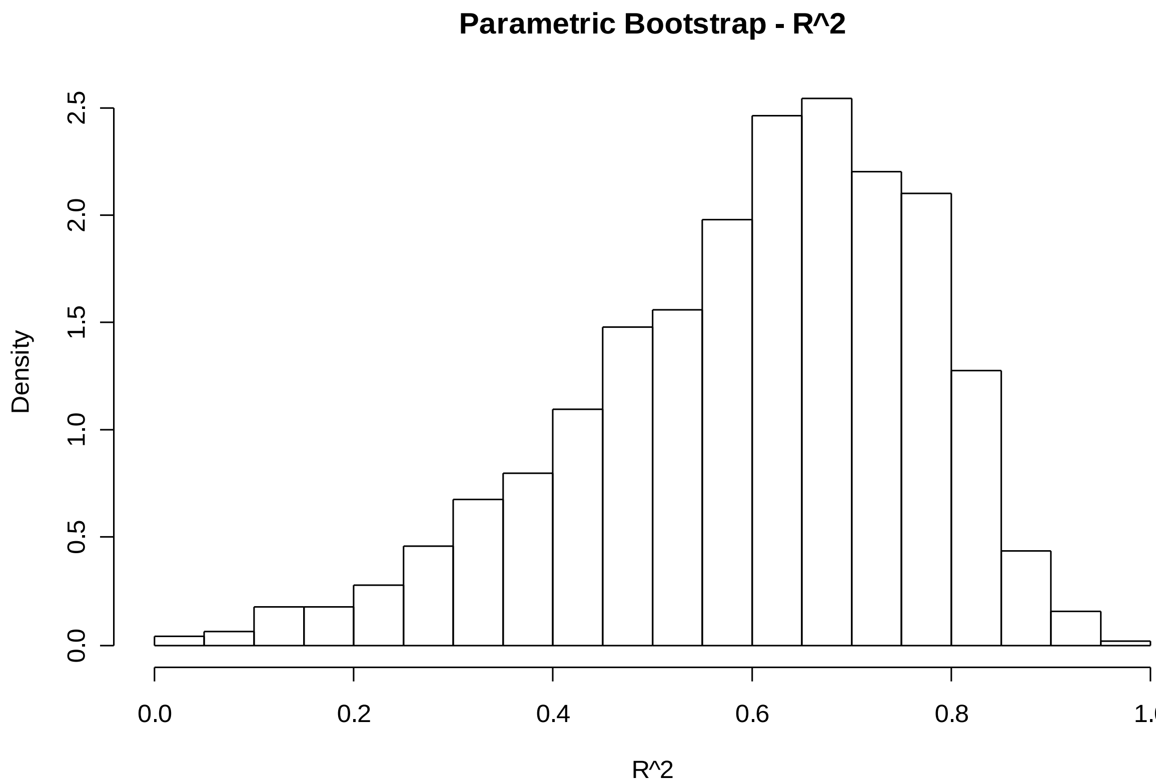
For both of these examples, the parametric bootstrap estimates of the standard error agree well with the asymptotic formula.

This is usually the case. So why bother with the parametric bootstrap?

It can provide more accurate answers. Some of the text book formulas only work well when the sample size is large. An example of this is $se(RR)$ from a 2x2 table.

It can be used when asymptotic formulas for standard errors are unknown.

For example, what is $se\left(r^2\right)$?



r^2 = 0.603

$se\left(r^2\right)$ = ??? (asymptotic formula)

$\widehat{se}\left(r^2\right)$ = 0.169 (parametric bootstrap)

$\widehat{Bias}$ = -0.00445

Actually an asymptotic formula for $se\left(r^2\right)$ is probably known since

$$r^2 = \frac{F}{F + n - 2}$$

where F is the usual ANOVA F-test for $\rho = 0$

Confidence Intervals

As the bootstrap is used to approximate the sampling distribution, it can be used to generate confidence intervals (and for hypothesis testing as well).

There is a wide range of bootstrapping approaches to this problem. Which of these to use depends on the form of the bootstrap distribution.

In the examples, I'll use a nonparametric bootstrap, but parametric bootstrap can also be used.

Also in all the examples, I'll be using 1 - 2α percentile intervals.

Notation: $\hat{\theta}^*(b) = T\left(\mathbf{x}^{*b}\right)$

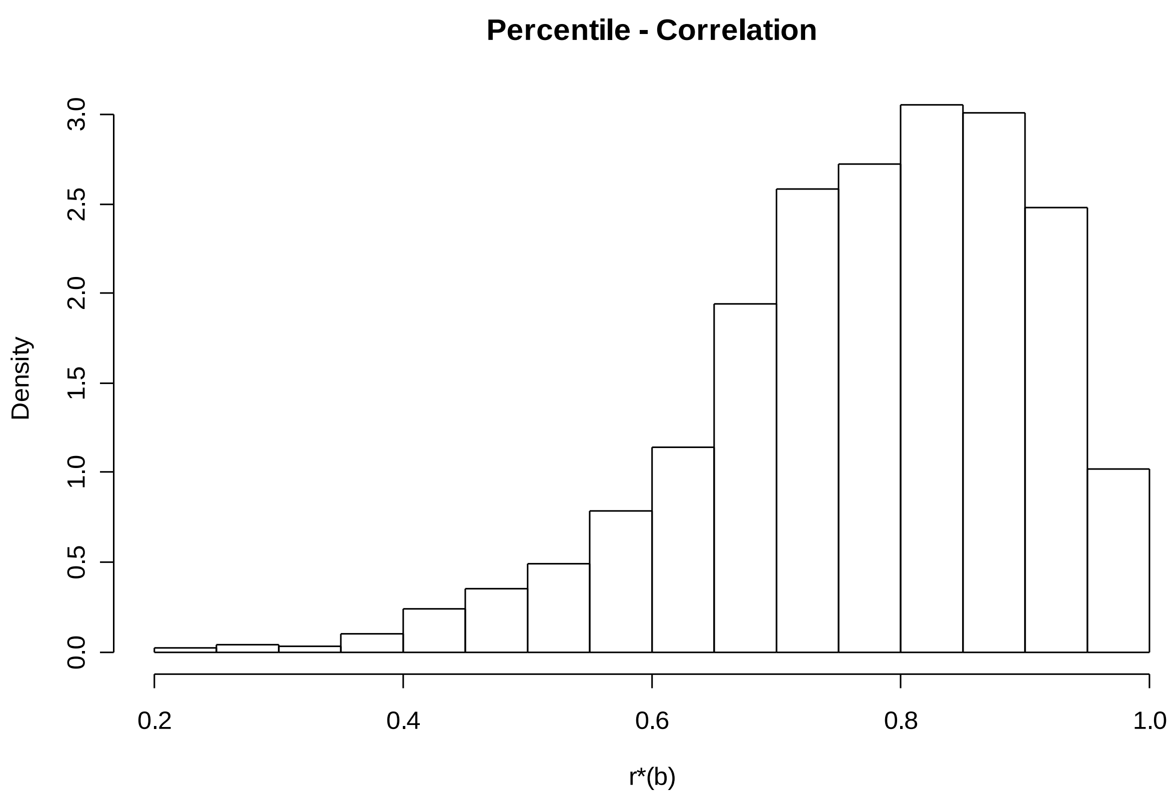
Percentile Interval

Easy interval to generate

$$\left[\hat{\theta}_{lo}, \hat{\theta}_{up}\right] = \left[\hat{\theta}_B^{*(\alpha)}, \hat{\theta}_B^{*(1-\alpha)}\right]$$

where $\hat{\theta}_B^{*(\alpha)}$ is the 100αth empirical percentile of the $\hat{\theta}^*(b)$ values.

So if B = 2000 and α = 0.05, we need the 100th and 900th ordered values of the $\hat{\theta}^*(b)$'s



For the correlation example, confidence intervals for different confidence levels are

Level	Lower	Upper
90%	0.525	0.951
95%	0.458	0.964
99%	0.367	0.980

To get these confidence interval, B needs to be fairly large since we need to determine the tail properties of the sampling distribution. Efron and Tibshirani recommend B being at least 1000 for reasonable choices of α.

Bootstrap-t interval

Based on the standard t based confidence interval

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

which is based on

$$z = \frac{\hat{\theta} - \theta}{se\left(\hat{\theta}\right)}$$

having an approximate t distribution.

The idea behind the bootstrap-t interval is to use the bootstrap to approximate the distribution of z.

$$\text{Let } Z^*\left(b\right) = \frac{\hat{\theta}^*\left(b\right) - \hat{\theta}}{\widehat{se}^*\left(b\right)}$$

Then the $\hat{t}^{(\alpha)}$ is the 100αth empirical percential of the $Z^*\left(b\right)$.

The bootstrap-t interval is

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

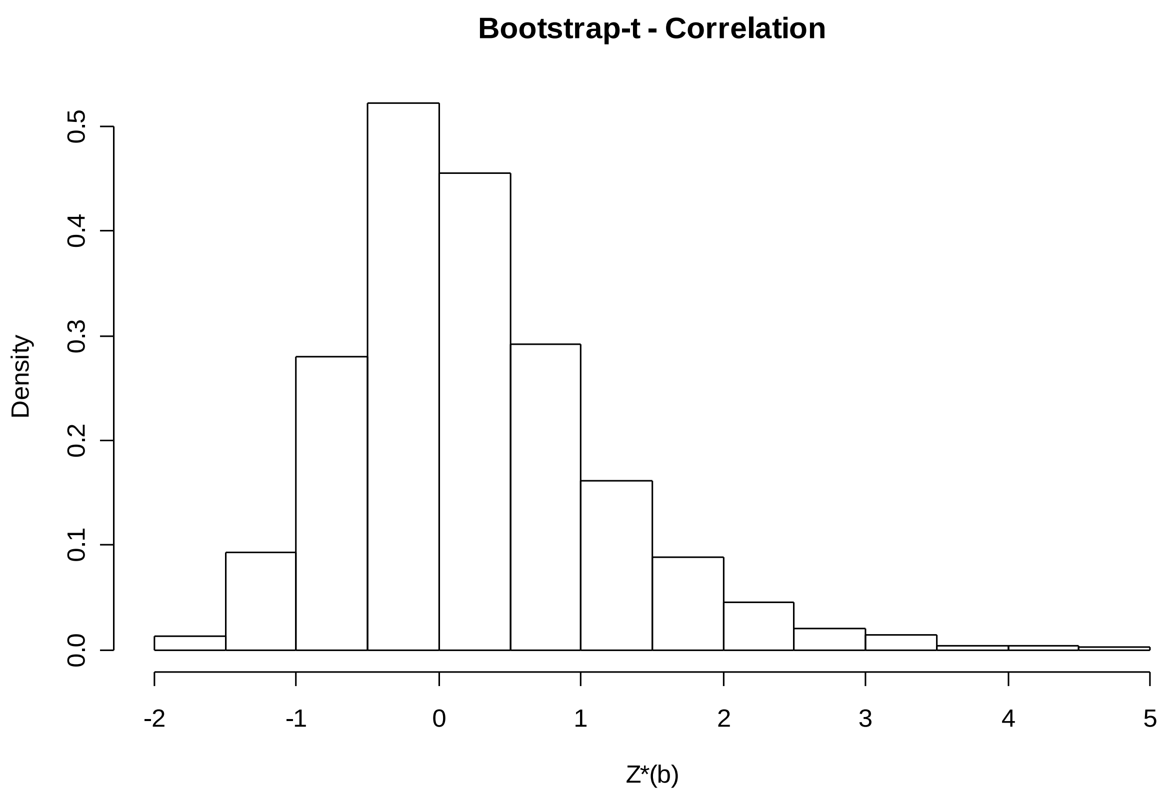
One complication to this procedure is that a standard error, $\widehat{se}^*\left(b\right)$ is needed for each bootstrap sample.

For the correlation example, you could use the textbook formula assuming normality

$$se(r) = \frac{1 - r^2}{\sqrt{n-3}}$$

Another approach, not used here, is to use a second level of bootstrapping to estimate the standard error.

This requires $B_1 B_2$ total bootstrap samples, where B_1 is the number of bootstrap samples to get the distribution of $Z^*(b)$ and B_2 is the number of bootstrap samples for each B_1 sample to get the standard error.



Level	Lower	Upper
90%	0.522	0.910
95%	0.459	0.938
99%	0.311	0.975

This approach often works better when the distribution of $Z^*(b)$ is roughly pivotal (the distribution doesn't depend on the parameters of interest.

For the correlation example, Fisher's transformation can be used.

Get a CI for ξ and then transform back to get one for ρ.

In this case the back transformation is

$$\rho = \frac{e^{2\xi} - 1}{e^{2\xi} + 1}$$

So for this case, if a confidence interval for ξ is $\left[\xi_{lo}, \xi_{up}\right]$, a confidence interval for ρ is

$$\left[\frac{e^{2\xi_{lo}}-1}{e^{2\xi_{lo}}+1}, \frac{e^{2\xi_{up}}-1}{e^{2\xi_{up}}+1}\right]$$

For the example, the intervals are

Level	Lower	Upper
90%	-0.020	0.926
95%	-0.221	0.941
99%	-0.555	0.955

These intervals are based on the asymptotic variance formula for ξ and can be replaced by a bootstrap estimate. This gives much different intervals than the other 2 procedures.

While it didn't occur with this example, it is possible for an end point, to be outside the range [-1, 1] with either of the bootstrap-*t* approaches.

Bootstrap-*t* based intervals are not transformation respecting.

However the Percentile intervals are, assuming of course you aren't doing something stupid with your estimation procedures.

There are other bootstrapping approaches to confidence intervals.

The most common 2 are

- BC_a: Bias-corrected and accelerated
- ABC: Approximate bootstrap confidence.

ABC is an approximation to BC_a which reduces the number of bootstrap samples.

These two approaches tend to give better intervals than those discussed earlier.

They are both transformation respecting and tend to have more accurate coverage probabilities.