1.3 Properties of Estimators

We can use the bootstrap to estimate different properties of estimators.

1.3.1 Standard Error

Recall $se(\hat{\theta}) = \sqrt{Var(\hat{\theta})}$. We can get a **bootstrap** estimate of the standard error:

$$\hat{se}(\hat{\theta}) = \sqrt{\frac{1}{B}\sum_{b=1}^{B}\left(\hat{\theta}^{*(b)} - \bar{\theta}^{*}\right)^2} \qquad \bar{\theta}^{*} = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^{*(b)}$$

1.3.2 Bias

Recall bias$(\hat{\theta}) = E[\hat{\theta} - \theta] = E[\hat{\theta}] - \theta$. We can get a **bootstrap** estimate of the bias:

$$\hat{\text{bias}}(\hat{\theta}) = \underbrace{\bar{\hat{\theta}}^{*}}_{\text{computed from the bootstrap samples}} - \underbrace{\hat{\theta}}_{\text{computed from original data}} = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\theta}^{(b)} - \hat{\theta}\right)$$

Overall, we seek statistics with small se and small bias.

$$MSE = \text{Variance} + \text{bias}^2$$
$$= E[(\hat{\theta} - \theta)^2].$$

$\Rightarrow$ Bootstrap procedure to estimate MSE:

1. Compute $\hat{\theta}$ from original sample $\underline{Y} = (Y_1, \dots, Y_n)$
2. Take a number (B) of bootstrap samples of size n from data $\underline{Y}^{*(1)}, \dots, \underline{Y}^{*(B)}$
3. Compute $\hat{\theta}^{*(b)}$ estimate of θ obtained from bs sample.
4. $\hat{MSE} = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\theta}^{*(b)} - \hat{\theta}\right)^2$

1.4 Sample Size and # Bootstrap Samples

$$n = \text{sample size} \quad \& \quad B = \text{\# bootstap samples}$$

If n is too small, or sample isn't representative of the population,

bootstrap procedure will be poor no matter how large B is.

Guidelines for B –

$B \approx 1000$ for se + bias

$B \approx 2000$ for CI's (depends on α : small $\alpha \Rightarrow \uparrow B$).

Best approach –

Repeat bootstrap w/ different seeds. If estimates very different, $\uparrow B$.

Your Turn

In this example, we explore bootstrapping in the rare case where we know the values for the entire population. If you have all the data from the population, you don't need to bootstrap (or really, inference). It is useful to learn about bootstrapping by comparing to the truth in this example.

In the package `bootstrap` is contained the average LSAT and GPA for admission to the population of 82 USA Law schools (an old data set – there are now over 200 law schools). This package also contains a random sample of size $n = 15$ from this dataset.

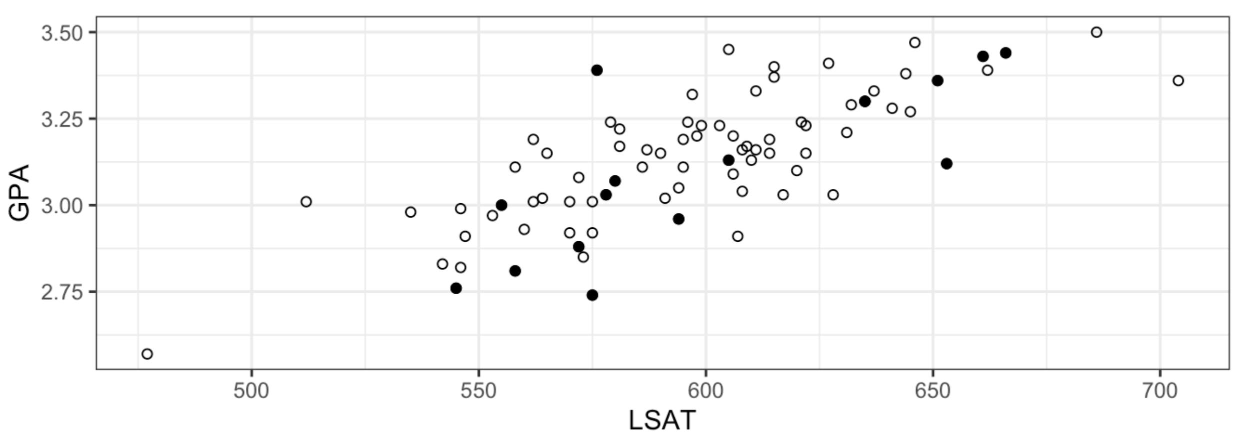
```
library(bootstrap)

head(law)
```

random sample
$n = 15$

```
##   LSAT  GPA
## 1  576 3.39
## 2  635 3.30
## 3  558 2.81
## 4  578 3.03
## 5  666 3.44
## 6  580 3.07
```

```
ggplot() +
  geom_point(aes(LSAT, GPA), data = law) +
  geom_point(aes(LSAT, GPA), data = law82, pch = 1)
```

full data set



correlation

We will estimate the correlation $\theta = \rho(\text{LSAT}, \text{GPA})$ between these two variables and use a bootstrap to estimate the sample distribution of $\hat{\theta}$.

```
# sample correlation
cor(law$LSAT, law$GPA)
```

n=15

$\hat{\theta}$

```
## [1] 0.7763745
```

```
# population correlation
cor(law82$LSAT, law82$GPA)
```

θ

```
## [1] 0.7599979
```

```
# set up the bootstrap
B <- 200
n <- nrow(law)
r <- numeric(B) # storage

for(b in B) {
  ## Your Turn: Do the bootstrap!
}
```

1. Plot the sample distribution of $\hat{\theta}$. Add vertical lines for the true value θ and the sample estimate $\hat{\theta}$.

2. Estimate $sd(\hat{\theta})$.

3. Estimate the bias of $\hat{\theta}$

1.5 Bootstrap CIs

We will look at ~~five~~ four different ways to create confidence intervals using the boostrap and discuss which to use when.

1. Percentile Bootstrap CI
2. Basic Bootstrap CI (adjusted for bias)
3. ~~Standard Normal Bootstrap CI~~
4. Bootstrap t (studentized).
5. Accelerated Bias-Corrected (BCa)

adjusted for skewness

Key ideas:

1. When you say and "bootstrap CI" you need to say which one.
2. (For now) whatever you are bootstrapping needs to be iid.
3. bootstrapping is an attempt to simulate replication (think about interpretation of CI).

1.5.1 Percentile Bootstrap CI (Probably the one you're thinking of).

Let $\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)}$ be bootstrap replicates and let $\hat{\theta}_{\alpha/2}$ be the $\alpha/2$ quantile of $\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)}$.

two-sided

Then, the $100(1-\alpha)\%$ Percentile Bootstrap CI for θ is

$$\left(\hat{\theta}_{\alpha/2}, \hat{\theta}_{1-\alpha/2}\right)$$

In R, if `bootstrap.reps = c(`$\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)}$`)`, the percentile CI is

vector of bootstrap samples of $\hat{\theta}^{(b)}$.

```
quantile(bootstrap.reps, c(alpha/2, 1 - alpha/2))
```

Assumptions/usage

- Widely used because simple to implement and explain.
- Drawback: CI's usually too narrow, leading to lower coverage.
- Can use when little bias and skewness in bootstrap dsn.

Justification (Efron):

dsn function of a r.v. symmetric about 0. $H^{-1}(\alpha) = -H^{-1}(1-\alpha)$ for $0<\alpha<1$.

Assume the existence of an increasing transformation g s.t. $P\left[g(\hat{\theta}) - g(\theta) \le x\right] = H(x)$

$\Rightarrow$ "bootstrap" World: $P^*\left[g(\hat{\theta}^*) - g(\hat{\theta}) \le x\right] \approx H(x)$ $(*)$.

Suppose g is known. Then $P\left[g^{-1}(g(\hat{\theta}) - x) \le \theta\right] = H(x)$. Then set this prob. equal to $1-\alpha$.

$= 1-\alpha \Rightarrow x = H^{-1}(1-\alpha)$.

$P\left[g^{-1}(g(\hat{\theta}) - H^{-1}(1-\alpha)) \le \theta\right] = 1-\alpha \Rightarrow$ If g is known this gives a one-sided CI w/ coverage probability $1-\alpha$.

$= \left(g^{-1}(g(\hat{\theta}) - H^{-1}(1-\alpha)), \infty\right)$ (upper-bound version similar).

goal: show $g^{-1}(g(\hat{\theta}) - H^{-1}(1-\alpha))$ is estimated by the left endpoint of percentile bootstrap interval, $\hat{\theta}_\alpha$.

$$\alpha = P^*(\hat{\theta}^* \le \hat{\theta}_\alpha) = P^*(g(\hat{\theta}^*) \le g(\hat{\theta}_\alpha))$$
$$= P^*(g(\hat{\theta}^*) - g(\hat{\theta}) \le g(\hat{\theta}_\alpha) - g(\hat{\theta})) \approx H(g(\hat{\theta}_\alpha) - g(\hat{\theta})) \quad \text{from } (*).$$

$$\Rightarrow H^{-1}(\alpha) = g(\hat{\theta}_\alpha) - g(\hat{\theta}) \Rightarrow g^{-1}(H^{-1}(\alpha) + g(\hat{\theta})) \approx \hat{\theta}_\alpha$$

$\overset{''}{=} g^{-1}(g(\hat{\theta}) - H^{-1}(1-\alpha))$ Same lower bound as above!

Note: we do not need to know g to calculate $\hat{\theta}_\alpha$!