1 Nonparametric Bootstrap

Let $Y_1, \ldots, Y_n \sim F$ with pdf $f(y)$. Recall, the empirical cdf is defined as

$$F_n(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(\underline{Y}_i \leq y) \quad y \in \mathbb{R}^d$$

MLE of F and as $n \to \infty$, $F_n \to F$.

Theoretical: Sample $\underline{Y} \sim F$, use $\underline{Y}_1, \ldots, \underline{Y}_n$ to compute F_n

Bootstrap: Sample $\underline{Y}^* \sim F_n$, use $\underline{Y}_1^*, \ldots, \underline{Y}_n^*$ to compute F_n^*

of size n.

The idea behind the nonparametric bootstrap is to sample many data sets from $F_n(y)$, which can be achieved by resampling from the data **with replacement.** (iid case).

How many possible Bootstrap samples? n^n

Are $Y_1^*, \ldots, Y_n^*$ independent?

$$P(Y_1^* = a, Y_2^* = b) = \frac{\sum_{i=1}^{n} \mathbb{I}(Y_i^* = a)}{n} \cdot \frac{\sum_{i=1}^{n} \mathbb{I}(Y_i^* = b)}{n} = P(Y_1^* = a)\, P(Y_2^* = b) \quad \text{Yes}$$

Do we always want this?
(more later...)

```r
# observed data
x <- c(2, 2, 1, 1, 5, 4, 4, 3, 1, 2)

# create 10 bootstrap samples
x_star <- matrix(NA, nrow = length(x), ncol = 10)
for(i in 1:10) {
  x_star[, i] <- sample(x, length(x), replace = TRUE)
}
x_star
```

$x^{*(1)}$ $x^{*(10)}$

```
##       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
##  [1,]    1    2    4    1    2    1    2    3    3     4
##  [2,]    4    4    1    1    1    2    2    1    2     1
##  [3,]    2    2    2    4    5    4    4    5    1     4
##  [4,]    4    4    2    5    2    4    5    5    1     3
##  [5,]    2    1    5    1    3    2    4    2    4     4
##  [6,]    4    4    2    1    4    4    4    3    1     2
##  [7,]    1    1    2    1    2    1    2    2    3     1
##  [8,]    4    4    1    3    3    3    5    1    2     4
##  [9,]    4    1    2    3    2    1    2    1    4     2
## [10,]    3    4    5    1    5    4    5    2    4     1
```

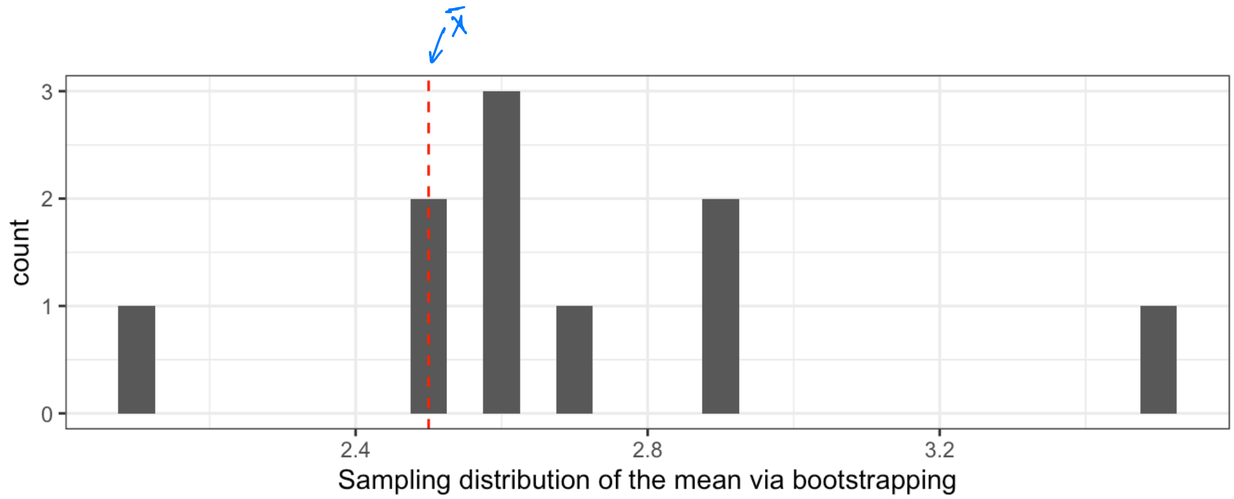
```r
# compare mean of the sample to the means of the bootstrap samples
mean(x)
```

```
## [1] 2.5
```

```r
colMeans(x_star)
```

```
##  [1] 2.9 2.7 2.6 2.1 2.9 2.6 3.5 2.5 2.5 2.6
```

```r
ggplot() +
  geom_histogram(aes(colMeans(x_star)), binwidth = .05) +
  geom_vline(aes(xintercept = mean(x)), lty = 2, colour = "red") +
  xlab("Sampling distribution of the mean via bootstrapping")
```



1.1 Algorithm

Goal: estimate the sampling distribution of a statistic based on observed data $y_1, \ldots, y_n$.

Let θ be the parameter of interest and $\hat{\theta}$ be an estimator of θ. Then,

For $b = 1, \ldots, B$

1. Sample $y^{*(b)} = (y_1^{*(b)}, \ldots, y_n^{*(b)})$ by sampling w/ replacement from $(y_1, \ldots, y_n)$
   (i.e. Sample from F_n)

2. $\hat{\theta}^{(b)} = \hat{\theta}(y^{*(b)})$
   estimate of θ based on b^{th} bootstrap sample.

Using $\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)}$ we can

- estimate the sampling dsn of $\hat{\theta}_n$ (via histogram, density estimator)
- estimate SE of $\hat{\theta}$
- estimate bias of $\hat{\theta}$
- estimate a CI (many ways).

etc.

1.2 Justification for iid data

Suppose $Y_1, \ldots, Y_n$ are iid with $\mathrm{E}Y_1 = \mu \in \mathbb{R}$, $\mathrm{Var}(Y_1) = \sigma^2 \in (0, \infty)$. Let's approximate the distribution of $T_n = \sqrt{n}(\bar{Y}_n - \mu)$ via the bootstrap.

Theorem: If $Y_1, Y_2, \ldots$ are iid with $\mathrm{Var}(Y_1) = \sigma^2 \in (0, \infty)$, then $\sup_{y \in \mathbb{R}} |P(T_n \le y) - P_*(T_n^* \le y)| \equiv \Delta_n \to 0$ as $n \to \infty$ almost surely (a.s).

Given $\underline{Y} = \{Y_1, \ldots, Y_n\}$ draw $Y_1^*, \ldots, Y_n^*$ bootstrap sample. Then,

bootstrap probability

$$P_*(Y_1^* = Y_i) = P(Y_1^* = Y_i \mid \underline{Y}) = \frac{1}{n} \qquad 1 \le i \le n.$$

The bootstrap version of our statistic T_n is $T_n^* = \sqrt{n}\left(\bar{Y}_n^* - E_* Y_1^*\right) = \sqrt{n}\left(\bar{Y}_n^* - \bar{Y}_n\right)$

bootstrap expected value.

where $E_*(Y_1^*) = E[Y_1^* \mid \underline{Y}] = \sum_{i=1}^n \frac{1}{n} Y_i = \bar{Y}_n$ also $E_*(\bar{Y}_n^*) = E_*\left(\frac{1}{n}\sum_{i=1}^n Y_i^*\right) = \frac{1}{n}\sum_{i=1}^n E_* Y_i^* \overset{\text{iid}}{=} \bar{Y}_n$

Also $P_*(T_n^* \le y) = P(T_n^* \le y \mid \underline{Y})$ approximates $P(T_n \le y)$ $y \in \mathbb{R}$ (theorem). ← exists because dsn of $Y_1^*, \ldots, Y_n^*$ exists, but hard to compute directly b/c n^n possible bootstrap samples. $\Rightarrow$ use simulation.

The proof of this theorem requires two facts:

i. (Berry-Esseen Lemma) Let $Y_1, \ldots, Y_n$ be independent with $\mathrm{E}Y_i = 0$ and $\mathrm{E}|Y_i|^3 < \infty$ for $i = 1, \ldots, n$. Let $\sigma_n^2 = n\mathrm{Var}(\bar{Y}_n) = n^{-1}\sum_{i=1}^n \mathrm{E}Y_i^2 > 0$. Then,

$$\sup_{y \in \mathbb{R}} \left| P\left(\frac{\sqrt{n}\bar{Y}_n}{\sigma_n} \le y\right) - \Phi(y) \right| = \sup_{x \in \mathbb{R}} \left| P\left(\sqrt{n}\bar{Y}_n \le x\right) - \Phi\left(\frac{x}{\sigma_n}\right) \right| \le \frac{2.75}{n^{3/2}\sigma_n^3} \sum_{i=1}^n \mathrm{E}|Y_i|^3.$$

standard normal cdf.

M-Z SLLN

ii. (Marcinkiewicz-Zygmund SLLN) Let X_i be a sequence of iid random variables with $\mathrm{E}|X_i|^p < \infty$ for $p \in (0, 2)$. Then, for $S_n = \sum_{i=1}^n X_i$,

$$\frac{1}{n^{1/p}}(S_n - nc) \to 0 \text{ as } n \to \infty \text{ almost surely } (*)$$

for any $c \in \mathbb{R}$ if $p \in (0, 1)$ and for $c = \mathrm{E}X_1$ if $p \in [1, 2)$. If $(*)$ holds for some $c \in \mathbb{R}$, then $\mathrm{E}|X_1|^p < \infty$.

Specifically, we will use that if $\{Y_i\}$ are iid w/ $EY_1^2 < \infty$, then

$$\frac{1}{n^{3/2}} \sum_{i=1}^n |Y_i|^3 \to 0 \quad \text{as } n \to \infty \text{ a.s.}$$

letting $X_i = |Y_i|^3$ because $E|X_i|^p = E|Y_i|^{3p} < \infty$ for $p = 2/3$, we may take $c = 0$.

Proof:

Write $\sup_{y\in\mathbb{R}} \left| P(T_n \le y) - P_*(T_n^* \le y) \right| \le \underbrace{\sup_{y\in\mathbb{R}} \left| P(T_n \le y) - \Phi(y/\sigma) \right|}_{\tilde{\Delta}_n} + \underbrace{\sup_{y\in\mathbb{R}} \left| P_*(T_n^* \le y) - \Phi(y/\sigma) \right|}_{\Delta_{n1}}$

$\tilde{\Delta}_n \to 0$ by CLT since $Y_1, \dots, Y_n$ iid, $EY_1^2 < \infty$.

Note that

$$\sigma_{n*}^2 \equiv n \, Var_*(\bar{Y}_n^*) = n \, Var_*\left(\frac{1}{n} \sum_{i=1}^n Y_i^* \right) \overset{iid}{=} \frac{n}{n^2} \sum_{i=1}^n Var_*(Y_i^*) = Var_* Y_1^*$$

$$= E_*[(Y_1^*)^2] - [E_* Y_1^*]^2 \quad \text{where } Y_1^* = \begin{cases} Y_1 & \text{w.p. } 1/n \\ \vdots & \vdots \\ Y_n & \text{w.p. } 1/n. \end{cases}$$

$$= \frac{1}{n} \sum_{i=1}^n Y_i^2 - (\bar{Y}_n)^2$$

So $\sigma_{n*}^2 \to EY_1^2 - (EY_1)^2 = \sigma^2$ as $n \to \infty$ w.p. 1 by SLLN since $EY_1^2 < \infty$.

By the Berry Esseen Lemma on $T_n^* = \sqrt{n}(\bar{Y}_n^* - E_* Y_1^*)$ and $|a-b| \le 2\max\{|a|, |b|\}$

$$\Rightarrow |a-b|^3 \le 8 \max\{|a|^3, |b|^3\}$$
$$\le 8(|a|^3 + |b|^3)$$

$$\sup_{y\in\mathbb{R}} \left| P_*(\underset{\sqrt{n}(\bar{Y}_n^* - E_* Y_1^*)}{T_n^*} \le y) - \Phi\left(\frac{y}{\sigma_{n*}}\right) \right| \overset{\text{Berry Esseen}}{\le} \frac{2.75}{n^{3/2}\sigma_{n*}^3} \, n \underbrace{E_* |Y_1^* - E_* Y_1^*|^3}_{\sum_{i=1}^n \frac{1}{n} |Y_i - \bar{Y}_n|^3}$$

$$= \frac{2.75}{n^{1/2}\sigma_{n*}^3} \frac{1}{n} \sum_{i=1}^n |Y_i - \bar{Y}_n|^3$$

$$\le \frac{2.75}{n^{1/2}\sigma_{n*}^3} \frac{1}{n} \sum_{i=1}^n 8\left(|Y_i|^3 + |\bar{Y}_n|^3 \right)$$

$\to 0$ as $n \to \infty$ w.p. 1 since $\bar{Y}_n \to \mu < \infty$ w.p. 1.

$$= \frac{8(2.75)}{\sigma_{n*}^3} \left(\boxed{\frac{|\bar{Y}_n|^3}{n^{1/2}}} + \boxed{\frac{1}{n^{3/2}} \sum_{i=1}^n |Y_i|^3} \right)$$

$\to 0$ as $n \to \infty$ w.p. 1 by M-Z SLLN (prev. page)

Finally, use $\sigma_{n*}^2 \to \sigma^2$ w.p. 1 $\Rightarrow \Phi\left(\frac{y}{\sigma_{n*}}\right) \to \Phi\left(\frac{y}{\sigma}\right)$ as $n \to \infty$ for any $y \in \mathbb{R}$ since $\Phi(\cdot)$ is continuous.

Thus $\sup_{y\in\mathbb{R}} \left| \Phi\left(\frac{y}{\sigma_{n*}}\right) - \Phi\left(\frac{y}{\sigma}\right) \right| \to 0$ as $n \to \infty$ w.p. 1 (Polya's theorem).

So, $\overset{\Delta_{n1}}{\overset{\shortparallel}{\sup_{y\in\mathbb{R}}}} \left| P_*(T_n^* \le y) - \Phi\left(\frac{y}{\sigma}\right) \right| \le \sup_{y\in\mathbb{R}} \left| P_*(T_n^* \le y) - \Phi\left(\frac{y}{\sigma_{n*}}\right) \right| + \sup_{y\in\mathbb{R}} \left| \Phi\left(\frac{y}{\sigma_{n*}}\right) - \Phi\left(\frac{y}{\sigma}\right) \right| \to 0$ w.p. 1.