Bootstrap In-class assignment

Example 1: F_n and bootstrap samples

Suppose that we have observed the sample

$$x = \{2, 2, 1, 1, 5, 4, 4, 3, 1, 2\}.$$

Resampling from x we select 1,2,3,4, or 5 with probabilities 0.3,0.3,0.1,0.2, and 0.1, respectively, What is the cdf F_{X^*} of a randomly selected replicate, which is exactly the ecdf $F_n(x)$?

$$F_{X^*}(x) = F_n(x) = \begin{cases} 0, & x < 1; \\ 0.3, & 1 \le x < 2; \\ 0.6, & 2 \le x < 3; \\ 0.7, & 3 \le x < 4; \\ 0.9, & 4 \le x < 5; \\ 1, & x \ge 5 \end{cases} .$$

Example 2: Bootstrap estimate of standard error

The law school data set `law` is available in the `bootstrap` package. The data frame contains LSAT (average score on law school admission test score) and GPA (average undergraduate grade-point average) for 15 law schools.

This data set is a random sample from the universe of 82 law schools in `law82 (bootstrap)`. Estimate the correlation between LSAT and GPA scores, and compute the bootstrap estimate of the standard error of the sample correlation.

```
#library(bootstrap)    #for the law data
LSAT<-c(576, 635, 558, 578, 666, 580, 555, 661, 651, 605, 653, 575, 545, 572, 594)
GPA<-c(339, 330, 281, 303, 344, 307, 300, 343, 336, 313, 312, 274, 276, 288, 296)
law=data.frame(LSAT=LSAT, GPA=GPA)
print(cor(law$LSAT, law$GPA))
```

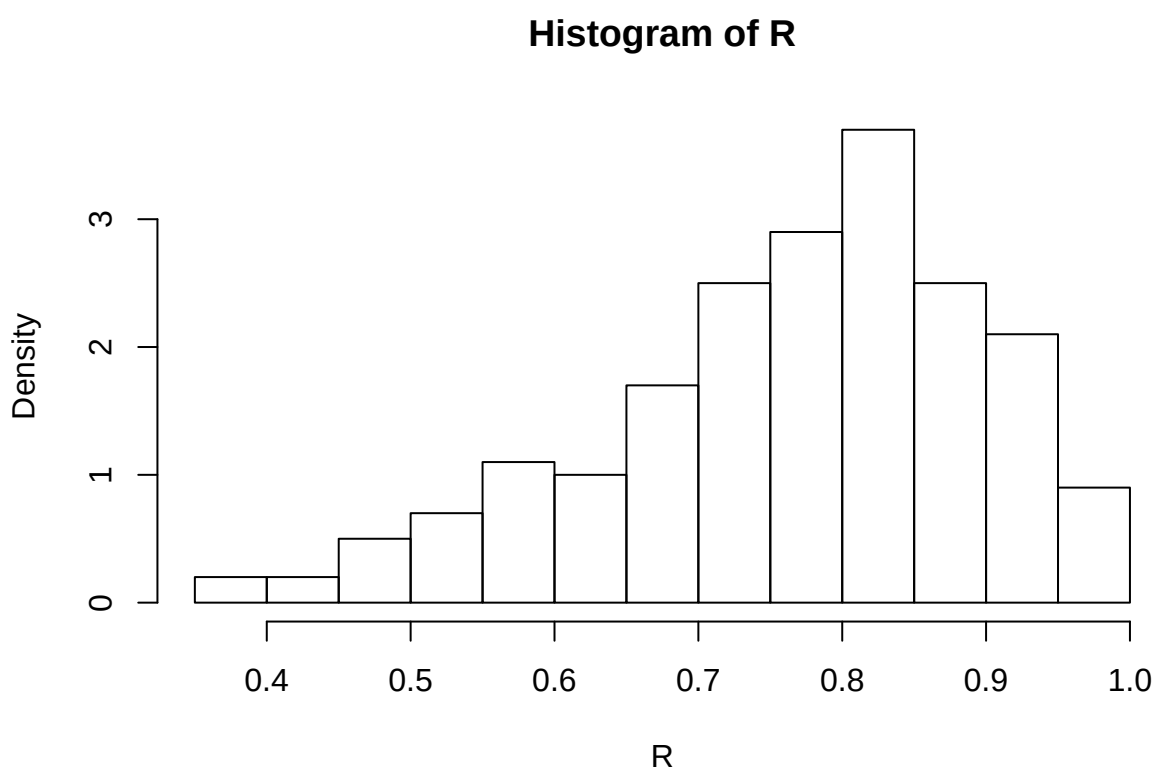
```
## [1] 0.7763745
```

```
#set up the bootstrap
    B <- 200            #number of replicates
    n <- nrow(law)      #sample size
    R <- numeric(B)     #storage for replicates

#bootstrap estimate of standard error of R
    for (b in 1:B) {
        #randomly select the indices
        i <- sample(1:n, size = n, replace = TRUE)
        LSAT <- law$LSAT[i]       #i is a vector of indices
        GPA <- law$GPA[i]
        R[b] <- cor(LSAT, GPA)
    }
#output
print(se.R <- sd(R))
```

```
## [1] 0.1316945
```

```
hist(R, prob = TRUE)
```



Example 3: Bootstrap estimate of bias

Estimate the correlation between LSAT and GPA scores, and compute the bootstrap estimate of bias in the sample correlation.

```
#sample estimate for n=15
theta.hat <- cor(law$LSAT, law$GPA)

#bootstrap estimate of bias
B <- 2000   #larger for estimating bias
n <- nrow(law)
theta.b <- numeric(B)

for (b in 1:B) {
i <- sample(1:n, size = n, replace = TRUE)
LSAT <- law$LSAT[i]
GPA <- law$GPA[i]
theta.b[b] <- cor(LSAT, GPA)
    }
bias <- mean(theta.b - theta.hat)
bias
```

```
## [1] -0.002323304
```

Example 4: Boostrap estimate of bias of a ratio estimate

The `patch(bootstrap)` data contains measurements of a certain hormone in the bloodstream of eight subjects after wearing a medical patch. The parameter of interest is

$$\theta = \frac{E(new) - E(old)}{E(old) - E(placebo)}.$$

If $|\theta| \leq 0.20$, this indicates bioequivalence of the old and new batches. The statistic is $\bar{Y}/\bar{Z}$. Compute a bootstrap estimate of bias in the bioequivalence ratio statistic.

```
#data(patch, package="bootstrap")
#patch
#if you can't install
patch=read.csv("patch.csv")
n<-nrow(patch)
B<-2000
theta.b<-numeric (B)
theta.hat<-mean(patch$y)/mean(patch$z)

for (b in 1:B){
  i<-sample(1:n, size=n, replace=T)
  y<- patch$y[i]
  z<-patch$z[i]
  theta.b[b]<-mean(y)/mean(z)
}
bias<-mean(theta.b)-theta.hat
print(list(theta.hat=theta.hat,bias=bias))
```

```
## $theta.hat
## [1] -0.0713061
##
## $bias
## [1] 0.008551326
```