ISLR | Chapter 4 Exercises

Marshall McQuillen

6/29/2018

Conceptual

1

$$f(\alpha) = Var(\alpha X + (1 - \alpha)Y)$$

Using the statistical property that $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$, the above equation can be rewritten as:

$$f(\alpha) = Var(\alpha X) + Var((1 - \alpha)Y) + 2Cov(\alpha X, (1 - \alpha)Y)$$

Then, using the statistical property that $Var(cX) = c^2 Var(X)$ and $Cov(aX, bY) = abCov(X, Y)$, the equation can once again be rewritten as:

$$f(\alpha) = \alpha^2 Var(X) + (1 - \alpha)^2 Var(Y) + 2\alpha(1 - \alpha)Cov(X, Y)$$

Multiplying the $\alpha(1 - \alpha)$ comes out to:

$$f(\alpha) = \alpha^2 Var(X) + (1 - \alpha)^2 Var(Y) + 2(\alpha - \alpha^2)Cov(X, Y)$$

By then taking the partial derivative of $f(\alpha)$ with respect to α, the slope of the function at a given alpha can be obtained:

$$\frac{\partial f(\alpha)}{\partial \alpha} = 2\alpha\sigma_X^2 + 2(1 - \alpha)(-1)\sigma_Y^2 + 2(1 - 2\alpha)\sigma_{XY}$$

Divide by 2:

$$\frac{\partial f(\alpha)}{\partial \alpha} = \alpha\sigma_X^2 + (-1 + \alpha)\sigma_Y^2 + (1 - 2\alpha)\sigma_{XY}$$

Expand the second and third terms in the equation:

$$\frac{\partial f(\alpha)}{\partial \alpha} = \alpha\sigma_X^2 + -\sigma_Y^2 + \alpha\sigma_Y^2 + \sigma_{XY} - 2\alpha\sigma_{XY}$$

Factor α out of all possible terms:

$$\frac{\partial f(\alpha)}{\partial \alpha} = \alpha(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}) - \sigma_Y^2 + \sigma_{XY}$$

Divide each term by $(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY})$:

$$\frac{\partial f(\alpha)}{\partial \alpha} = \alpha - \frac{\sigma_Y^2 + \sigma_{XY}}{(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY})}$$

Since the goal is to minimize the equation, setting the partial derivative to zero will return an equation that is a minimum.

$$0 = \alpha - \frac{\sigma_Y^2 + \sigma_{XY}}{(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY})}$$

Subtract α

$$-\alpha = - \frac{\sigma_Y^2 + \sigma_{XY}}{(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY})}$$

Multiply by -1:

$$\alpha = \frac{\sigma_Y^2 - \sigma_{XY}}{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}}$$

2

- **A**. Since a bootstrapped sample contains N observations of the original sample of the population, each sample being chosen at random with replacement, the probability that the first observation in a bootstrapped sample is *not* the jth observation is $\frac{n - 1}{n}$.
- **B**. The probability that the second bootstrap observation is *not* the jth observation is $\left(\frac{n - 1}{n}\right)^2$.
- **C**. Since a boostrapped sample contains N observations, the probability that the jth observation (x_j) is *not* in the bootstapped sample ($S - b$) is:

$$P(x_j \ not \ in \ S_b) = \left(\frac{n - 1}{n}\right)^n$$

  Which can be simplified to:

$$P(x_j \ not \ in \ S_b) = \left(1 - \frac{1}{n}\right)^n$$

- **D**. Since the probability that the jth observation is *not* in the boostrap sample is $\left(1 - \frac{1}{n}\right)^n$, the probability that the jth observation *is* in the bootstrap sample would be the complement, $1 - \left(1 - \frac{1}{n}\right)^n$. When $n = 5$, this comes out to $1 - \left(1 - \frac{1}{5}\right)^5 = 0.67232 = 67.23\%$
- **E**.

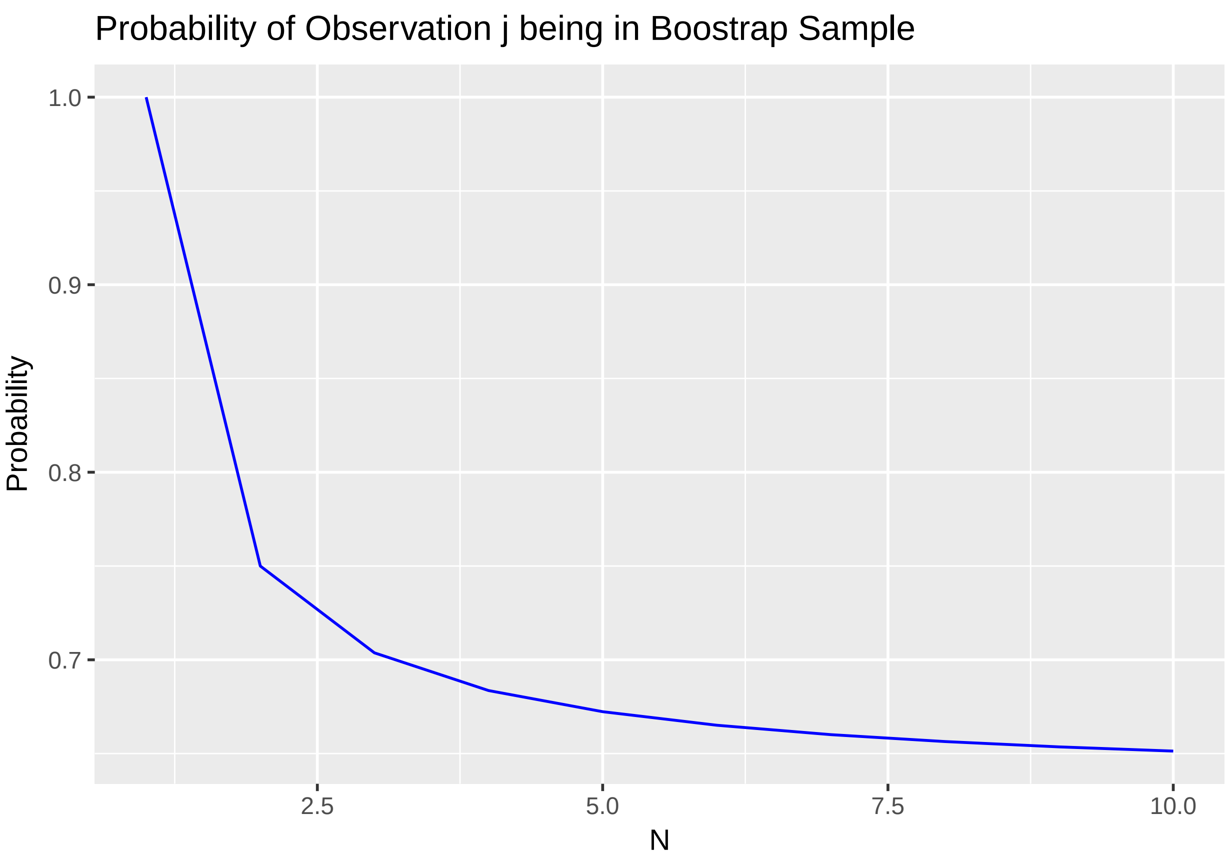
$$1 - \left(1 - \frac{1}{100}\right)^{100} = 0.6339677 = 63.40\%$$

- **F**.

$$1 - \left(1 - \frac{1}{100}\right)^{100} = 0.632139 = 63.21\%$$

- **G**. It is clear that as N increases the probability that the jth observation is in the bootstrap sample asymptotically approaches 0.632. The below plot illustrates this phenomenon (only displaying 1 to 10 for illustration purposes)

```
library(ggplot2)
x <- 1:100000
y <- 1 - (1 -(1/x))^x
df <- data.frame(x, y)
display_df <- df[1:10,]
ggplot(display_df, aes(x = x, y = y)) +
    geom_line(color = 'blue') +
    labs(x = "N", y = "Probability",
         title = "Probability of Observation j being in Boostrap Sample")
```



- **H**. The below code is showing mathematically what the plot above shows; that the limit of the function $1 - \left(1 - \frac{1}{x}\right)^x$ as x approaches infinity is 0.632.

```
store <- rep(NA, 10000)
for (i in 1:10000) {
    store[i] <- sum(sample(1:100, replace = TRUE)==4) > 0
}
mean(store)
```

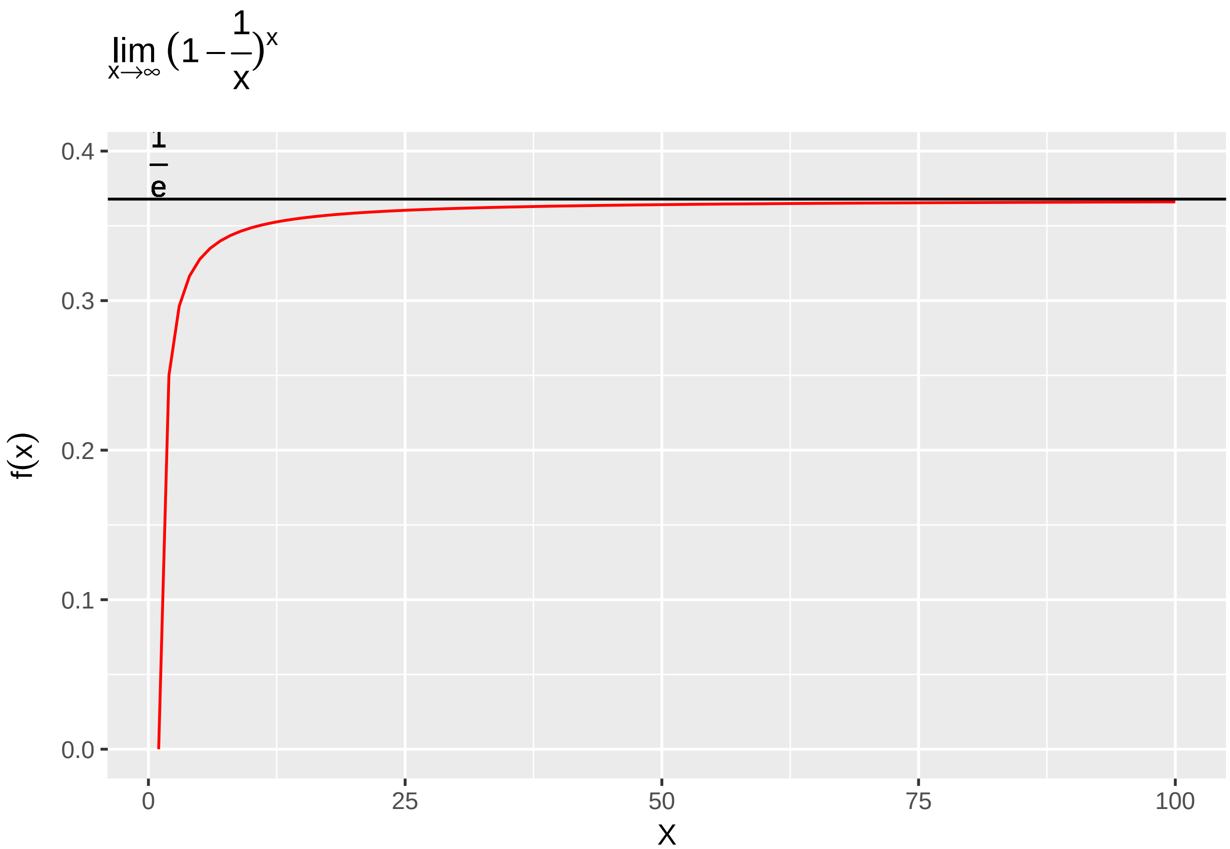
```
## [1] 0.6386
```

This can be written as:

$$\lim_{x \to \infty} \left(1 - \left(1 - \frac{1}{x}\right)^x\right) = 0.632$$

However, the inner part of that equation, $\lim_{x \to \infty} \left(1 - \frac{1}{x}\right)^x$, simplifies to $\frac{1}{\epsilon}$, proven by plot below:

```r
x <- 1:100
y <- (1 - (1/x))^x
asymptote <- rep(0.3678794, 100)
df <- data.frame(x, y, asymptote)
ggplot(df, aes(x = x, y = y)) +
    geom_line(color = 'red') +
    geom_hline(aes(yintercept = asymptote)) +
    annotate('text', x = 1, y = df$asymptote + 0.025,
             label = 'frac(1, e)', parse = TRUE) +
    labs(x = "X", y = expression(f(x)),
         title = expression(lim((1 - over(1, "x"))^"x", x %->% infinity)))
```



Therefore:

$$\lim_{x \to \infty} \left(1 - \left(1 - \frac{1}{x}\right)^x\right) = 0.632 = 1 - \frac{1}{\epsilon}$$