Exercise 1: Bootstrapping

Suppose that a sample consists of the following observations:

$$78, \ 86, \ 97, \ 91, \ 83, \ 89, \ 92, \ 88, \ 79, \ 68$$

Then the α-trimmed mean is the average of the inner $(1 - 2\alpha)$ values in the sample. For example, if $\alpha = 0.2$, then the trimmed mean is the average of the inner 60% of the observations.

$$68, \ 78, \ \mathbf{79}, \ \mathbf{83}, \ \mathbf{86}, \ \mathbf{88}, \ \mathbf{89}, \ \mathbf{91}, \ 92, \ 97$$

Why might we prefer to use the trimmed mean over the usual sample mean in this case?

```
# generate a vector containing the data
x <- c(78, 86, 97, 91, 83, 89, 92, 88, 79, 68)
x <- sort(x)

# the sample mean is given by
mean(x)

## [1] 85.1

# the trimmed mean is given by
mean(x, trim = 0.2)

## [1] 86
```

Note that the trimmed mean is significantly more robust to the outlier, 3.

We can compute the $\alpha = 0.2$ trimmed mean in R for the above sample using the command `mean(x, trim = 0.2)`.

Suppose that we don't know a general formula for the standard error of the trimmed mean. Our goal is to estimate it using bootstrapping.

1. Suppose that we were fairly confident that the data comes from a $Normal(\mu, 8^2)$ distribution (but we don't know μ). Use parametric bootstrap to estimate the standard error of the trimmed mean.

To perform parametric bootstrap, we

- estimate mu by mu.hat

- sample $X_1^*, ..., X_n^*$ from $N(\hat{\mu}, 8^2)$

- calculate our estimate for the estimate, $\hat{\theta}^*$, the trimmed mean for each bootstrapped sample

```r
# estimate mu to be the mean of our sample
mu.hat <- mean(x)

# we will sample from N(mu.hat, 30) B = 1000 times
B <- 1000
n <- length(x)

# store the samples in the matrix x.pboot
x.pboot <- matrix(NA, ncol = B, nrow = n)
for(i in 1:B){
  # store the sample in column i
  x.pboot[,i] <- rnorm(n, mu.hat, 8)
}

# calculate the trimmed mean for each of our pbootstrapped samples
trimmed.pboot <- apply(x.pboot, 2, function(x) mean(x, trim = 0.2))
SD.trimmed.pboot <- sd(trimmed.pboot)

SD.trimmed.pboot

## [1] 2.706538
```

2. Suppose we have absolutely no idea what the underlying distribution of the data is. Use non-parametric bootstrap to estimate the standard error of the trimmed mean.

```r
# estimate mu to be the mean of our sample
mu.hat <- mean(x)

# we will sample with replacement from our original sample B = 1000 times
B <- 1000
n <- length(x)

# store the samples in the matrix x.npboot
x.npboot <- matrix(NA, ncol = B, nrow = n)
for(i in 1:B){
  # store the sample in column i
  x.npboot[,i] <- sample(x, replace=TRUE)
}
```

```
# calculate the trimmed mean for each of our npbootstrapped samples
trimmed.npboot <- apply(x.npboot, 2, function(x) mean(x, trim = 0.2))
SD.trimmed.npboot <- sd(trimmed.npboot)

SD.trimmed.npboot

## [1] 2.77128
```

The SD estimate is fairly similar using both methods.

3. Calculate 95% pivot confidence intervals for each of these estimates.

Recall from lectures that the Pivot confidence interval for θ is given by

$$(2\hat{\theta} - \hat{\theta}^*_{1-\alpha/2}, 2\hat{\theta} - \hat{\theta}^*_{\alpha/2})$$

where $\hat{\theta}^*_{1-\alpha/2}$ is the $(1 - \alpha/2)$-percentile of the bootstrapped estimates $\hat{\theta}^*$ and similarly for $\hat{\theta}^*_{\alpha/2}$.

```
############### parametric CI ################
pboot.percentiles <- quantile(trimmed.pboot , probs = c(0.025, 0.975))


# calcualte the CI
pboot.CI <- c(2*mu.hat - pboot.percentiles[2],
              2*mu.hat - pboot.percentiles[1])
names(pboot.CI) <- c("lower bound","upper bound")
pboot.CI

## lower bound upper bound
##    80.08836    90.47615

############### non parametric CI ##############
npboot.percentiles <- quantile(trimmed.npboot , probs = c(0.025, 0.975))

# calculate the estimate for the trimmed mean
npboot.hat.trim <- mean(trimmed.npboot)
# calcualte the CI
npboot.CI <- c(2*mu.hat - npboot.percentiles[2],
               2*mu.hat - npboot.percentiles[1])
names(npboot.CI) <- c("lower bound","upper bound")
npboot.CI

## lower bound upper bound
##    79.70000    90.36667
```

Exercise 2: Hypothesis testing

Suppose that you had read an article which claimed that the average weight of red pandas was 6.3kg. However, suppose that you had recently weighed $n = 121$ red pandas and found that these pandas had an average weight of 5.1kg.
Assume that it is well-known that the standard deviation of the weights of red pandas was 1.4.
Do you have enough evidence to claim with 95% probability that the true average weight of red pandas is in fact less than 6.3kg?

Note that in this case our null hypothesis is that

$$H_0 : \mu = 6.3$$

and our alternative hypothesis is

$$H_1 : \mu < 6.3$$

Since we are interested in the **mean**, our test statistic is given by

$$T(X_1, ..., X_n) = \frac{\text{observed } - \text{ expected}}{\text{SE of estimator}} = \frac{\bar{X}_n - \mu_0}{SD(\bar{X})}$$

so

$$T(X_1, ..., X_n) = \frac{5.1 - 6.3}{1.4/\sqrt{121}} = -9.43$$

Our p-value is thus

$$P(Z < -9.43) = 2.05 \times 10^{-21}$$

which is negligibly small. This implies that the probability of observing an average weight of 5.1 or something more extreme (an even smaller average) given that the true mean is 6.3 (given H_0 true) is so small that it is extremely unlikely that the deviation was due to chance. That is, we have enough evidence to reject H_0 in favour of H_1 that the true mean is less than 6.3kg.

Exercise 3: Neyman-Pearson

Suppose we have a sample $X_1, ..., X_n$ such that $X_i \sim Normal(\mu, 16)$ where μ is unknown. We want to test the null hypothesis

$$H_0 : \mu = 10$$

against the alternative hypothesis

$$H_1 : \mu = 15$$

Find the best test with sample size $n = 16$ and significance level $\alpha = 0.05$. What is the power of this test?

Since both hypotheses are simple, the Neyman-Pearson lemma tells us tha the best test is one of the form that rejects H_0 in favour of H_1 when

$$\frac{f_0(X)}{f_1(X)} < c(\alpha)$$

In this case,

$$f_i(x) = \frac{1}{\sqrt{32\pi}} e^{-\frac{(x - \mu_i)^2}{32}}$$

so that the likelihood ratio is given by

$$\frac{f_0(X_1, ..., X_n)}{f_1(X_1, ..., X_n)} = \frac{\left(\frac{1}{\sqrt{32\pi}}\right)^{16} e^{-\frac{\sum_{i=1}^{16}(X_i - 10)^2}{32}}}{\left(\frac{1}{\sqrt{32\pi}}\right)^{16} e^{-\frac{\sum_{i=1}^{16}(x - 15)^2}{32}}}$$

so the most powerful test must be of the form

$$\exp\left[-\frac{1}{32}\left(\sum_{i=1}^{16}(X_i - 10)^2 - \sum_{i=1}^{16}(X_i - 15)^2\right)\right] \leq c(\alpha)$$

Simplifying, we get

$$\exp\left[-\frac{1}{32}\left(\sum_{i=1}^{16}X_i^2 - 20\sum_{i=1}^{16}X_i + 1600 - \sum_{i=1}^{16}X_i^2 + 30\sum_{i=1}^{16}X_i - 3600\right)\right] \leq c(\alpha)$$

taking the log of both sides of the inequality, collecting like terms and multiplying through by 32, we get:

$$-10\sum_{i=1}^{16}X_i + 2000 \leq 32\log(c(\alpha))$$

This implies that

$$\sum_{i=1}^{16}X_i \geq -\frac{1}{10}(32\log(c(\alpha)) - 2000)$$

So the Neyman-Pearson lemma tells us that the rejection region for the most powerful test for testing $H_0 : \mu = 10$ against $H_1 : \mu = 15$ under the normal probability model, is of the form

$$\sum_{i=1}^{16} X_i \geq k$$

where k is selected so that the size of the critical region is $\alpha = 0.05$. Note that, under the null hypothesis $\sum_{i=1}^{16} X_i \sim N(10 \times 16, 16 \times 16)$, so since we want

$$P_{H_0}\left(\sum_{i=1}^{16} X_i \geq k\right) = 0.05$$

we get that $k = 186.3$

```
qnorm(0.95, 10*16, sqrt(16*16))

## [1] 186.3177

# just to make sure:
1 - pnorm(186.3, 10*16 ,sqrt(16*16))

## [1] 0.05011393
```

That is the Neyman-Pearson lemma tells us that the rejection region for the most powerful test for testing $H_0 : \mu = 10$ against $H_1 : \mu = 15$ under the normal probability model, is

$$\sum_{i=1}^{16} X_i > 186.3$$

The power (probability of rejecting H_0 when H_0 is false, so $\mu = 15$ rather than $\mu = 10$) of such a test is

$$
\begin{aligned}
P\left(\sum_{i=1}^{16} X_i > 186.3 | H_1\right) &= P\left(\sum_{i=1}^{16} X_i > 186.3 \text{ when } \mu = 15\right) \\
&= P\left(Z > \frac{186.3 - 15 \times 16}{\sqrt{16 \times 16}}\right) \\
&= P(Z > -3.36) \\
&= 0.9996
\end{aligned}
$$

since under H_1 (H_0 false), $\sum_{i=1}^{16} X_i \sim N(15 \times 16, 16 \times 16)$, so it follows that $\frac{\sum_{i=1}^{16} -15 \times 16}{\sqrt{16 \times 16}} \sim N(0, 1)$

```r
1 - pnorm(-3.36)
```

```
## [1] 0.9996103
```

The power can't get much better than that, and the Neyman-Pearson Lemma
tells us that we shouldn't expect it to get better!.