Unit 6 Classwork

The purpose of this in-class notebook is for you to gain some experience with bootstrap simulations. You are expected to complete all exercises and turn in your work on Canvas (due date will be on the Unit #6 Classwork Canvas assignment).

Problem #1

Let's compare the "normal theory" confidence interval to boostrap confidence intervals.

1.(a) Generate a sample of size $n = 45$ from an exponential distribution with rate $\lambda = 1$. Calculate $\bar{X}$.

In [ ]:
```
n <- 45
lambda <- 1
samples <- rexp(n = n, rate = lambda)

mean <- mean(samples)

cat("Mean:", mean, "\n")
```

```
Mean: 1.096497
```

1.(b) Calculate $B = 500$ bootstrap samples, each of size n. You might do this in a $B \times n$ matrix, for example, where each row is a boostrap sample. Then, find the mean of each bootstrap sample, denoted $\bar{X}_i^*$, for $i = 1, \ldots, B$.

In [ ]:
```
B <- 500
bootstrap_samples <- matrix(0, nrow = B, ncol = n)

for (i in 1:B) {
  bootstrap_samples[i,] <- sample(sample, replace = TRUE)
}

bootstrap_means <- apply(bootstrap_samples, 1, mean)

cat("Few bootstrap means:\n", head(bootstrap_means))
```

```
Few bootstrap means:
 1.28478 1.359263 1.530449 1.193002 0.8565522 1.349215
```

1.(c) Use the `quantile()` function to find the 2.5th and 97.5th percentile of the distribution of each $\bar{X}_i^*$. Use these values to calculate the

95% pivot boostrap confidence interval for μ.

In [ ]:
```
lower_quantile <- quantile(bootstrap_means, 0.025)
upper_quantile <- quantile(bootstrap_means, 0.975)

confidence_interval <- c(2 * mean - upper_quantile, 2 * mean - lower_quantil

cat("95% Pivot Bootstrap Confidence Interval:", confidence_interval, "\n")
```

```
95% Pivot Bootstrap Confidence Interval: 0.6060012 1.365633
```

1.(d) Compute the appropriate "normal theory" confidence interval for μ, and the bootstrap percentile confidence interval for μ.

In [ ]:
```
sample_sd <- sd(samples)

z <- qnorm(0.975)  # for a 95% confidence interval
normal_theory_interval <- c(mean - z * (sample_sd / sqrt(n)),
                            mean + z * (sample_sd / sqrt(n)))

bootstrap_percentile_interval <- quantile(bootstrap_means, c(0.025, 0.975))

cat("Normal Theory Confidence Interval for mu:", normal_theory_interval, "\r
cat("Bootstrap Percentile Confidence Interval for mu:", bootstrap_percentile
```

```
Normal Theory Confidence Interval for mu: 0.823168 1.369826
Bootstrap Percentile Confidence Interval for mu: 0.8273615 1.586993
Bootstrap Percentile Confidence Interval for mu: 0.8273615 1.586993
```

1.(e) What values can you change above to make these interval estimates closer to each other?

1. Increasing the sample size tends to make the normal approximation more accurate and, in turn, may bring the normal theory and bootstrap intervals closer.
2. Increasing the number of bootstrap samples can lead to more stable estimates and reduce the variability in the bootstrap confidence interval.
3. You can choose a narrower confidence level. For example, instead of a 95% confidence level, you can use a 90% confidence level. A smaller confidence level generally results in a narrower interval.

Problem #2

Suppose that $X_1, \ldots, X_8 \overset{iid}{\sim} Exp(\lambda)$. Let's use the pivot bootstrap to compute a 90% confidence interval for the population variance: $\text{Var}[X_i] = 1/\lambda^2 = \theta$.

2.(a) Generate a sample of size $n = 8$ from the distribution $Exp(\lambda = 3)$ and calculate the population variance, θ (in this example, we are

generating data so that we can see how well our estimation procedure (the confidence interval) will do).

```
In [ ]: sample_size <- 8
        lambda_parameter <- 3
        sample <- rexp(n = sample_size, rate = lambda_parameter)

        population_variance <- var(sample)

        cat("Population Variance (theta):", population_variance, "\n")
```

```
Population Variance (theta): 0.0609199
```

(b) Generate $B = 200$ bootstrap samples from the above sample.

Again, use `replicate()` and `sample()` ...

```
In [ ]: B <- 200

        bootstrap_samples <- replicate(B, sample(sample, replace = TRUE))
```

(c) Calculate the MLE of θ for the original sample. Denote this as $\hat{\theta}$. Then, calculate the MLE of θ for each bootstrap sample. Denote this as $\hat{\theta}_i^*$, for $i = 1, \ldots, B$. Avoid loops! (HINT: use the apply() function.)

```
In [ ]: theta_hat <- 1 / mean(sample)
        theta_hat_star <- 1 / apply(bootstrap_samples, 2, mean)

        cat("MLE of theta for the original sample (theta_hat):", theta_hat, "\n")
        cat("MLE of theta for first few bootstrap samples (theta_hat_star):\n", heac
```

```
MLE of theta for the original sample (theta_hat): 3.084314
MLE of theta for first few bootstrap samples (theta_hat_star):
 3.083255 3.60622 2.214154 2.551361 3.418719 2.603301MLE of theta for first
few bootstrap samples (theta_hat_star):
 3.083255 3.60622 2.214154 2.551361 3.418719 2.603301
```

(d) Use the `quantile()` function to find the 5th and 95th percentile of the distribution of $\hat{\theta}_i^*$. Use these values to calculate the 90% pivot bootstrap confidence interval for θ.

```
In [ ]: lower_quantile_theta_hat_star <- quantile(theta_hat_star, 0.05)
        upper_quantile_theta_hat_star <- quantile(theta_hat_star, 0.95)

        pivot_bootstrap_interval <- c(2 * theta_hat - upper_quantile_theta_hat_star,

        cat("90% Pivot Bootstrap Confidence Interval:", pivot_bootstrap_interval, "\
```

```
90% Pivot Bootstrap Confidence Interval: 0.9887228 4.035769
```

(e) Interpret this confidence interval.

The 90% pivot bootstrap confidence interval for theta provides a range of plausible values for the population variance based on the bootstrap resampling procedure.

Basically, we are 90% confident that the true population variance theta lies within the interval [A,B], where A and B are the lower and upper bounds of the interval.