STAT430 Homework #4: Due Friday, March 1, 2019.

Name:

---

0. **Remember that there is no class or office hour on Monday, February 25.** We start Chapter 8 in this homework, with some additional problems related to the end of Chapter 7. Read **Sections 8.1-8.4** on estimation. We have the first midterm February 22. Be sure you are strong so far on distribution of maximum and minimum order statistic; expectation, variance and covariance calculations; normal probability computations; the "origin stories" of the χ^2, t, and F distributions; and the Central Limit Theorem and its applications and extensions (normal approximation for sample mean and sample total; normal approximation to the binomial; delta method; bias and mean squared error). This homework has some good review problems and is worth working on prior to the exam.

1. Let $Y \sim \text{Binomial}(n, p)$. The "odds of success" are defined as

$$\frac{\text{probability of success}}{\text{probability of failure}} = \frac{p}{1-p}$$

and the log-odds are

$$\lambda = \ln\left(\frac{p}{1-p}\right).$$

The standard, unbiased estimator of p is $\hat{p} = Y/n$. Plugging in this estimator, we have the estimated log-odds

$$\hat{\lambda} = \ln\left(\frac{\hat{p}}{1-\hat{p}}\right).$$

Use the delta method as described in class to determine the approximate distribution of $\hat{\lambda}$ for large n. For $p = 0.3$ and $n = 100$, verify that the variance of your approximate distribution is $1/21$.

Answer:

---

2. Suppose $Y_1, \ldots, Y_{40}$ denote a random sample of measurements on the proportion of impurities in iron ore samples. Let each Y_i have probability density function given by

$$f(y) = \begin{cases} 3y^2, & 0 \le y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

The ore is to be rejected by the potential buyers if $\bar{Y}$ exceeds 0.7. Find the approximate probability that $P(\bar{Y} > 0.7)$ for the sample of size $n = 40$.

Answer:

---

3. Let $Y_1, \ldots, Y_n$ denote a random sample from a population whose density is given by

$$f(y) = \begin{cases} \alpha y^{\alpha-1}/\theta^{\alpha}, & 0 \le y \le \theta, \\ 0, & \text{elsewhere,} \end{cases}$$

where $\alpha > 0$ is a known, fixed value, but θ is unknown. Consider the estimator $\hat{\theta} = \max(Y_1, \ldots, Y_n)$.
(a). Show that $\hat{\theta}$ is a biased estimator for θ.

Answer:

(b). Derive expressions for bias, variance and MSE of $\hat{\theta}$ as functions of n, α, and θ. (You do not need to simplify beyond evaluating the integrals.) Use R to evaluate your expressions numerically when $n = 6$, $\alpha = 2$, and $\theta = 5$.

Answer:

(c). Use the following R code to simulate 10000 random samples of size $n = 6$ from the original density with $\alpha = 2$ and $\theta = 5$ and compute $\hat{\theta}$ for each random sample. Then approximate the bias by `mean(theta_hat) - theta`, the variance by `var(theta_hat)`, and the MSE by `mean((theta_hat - theta) ^ 2)`. Compare to your theoretical values in (b).

```
rsim <- function(n, alpha, theta){
  u <- runif(n)
  Y <- theta * u ^ (1 / alpha)
  return(Y)
}
nreps <- 10000
n <- 6
alpha <- 2
theta <- 5
set.seed(4302019)
Y <- rsim(n * nreps, alpha = 2, theta = 5) # simulate 10,000 random samples of size 6
YY <- matrix(Y, n, nreps)                  # arrange into matrix with random sample in each column
```

Answer:

4. Complete **Exercise 8.2** of the text.

Answer:

5. Complete **Exercise 8.6** of the text. If σ_2^2 is much larger than σ_1^2, does your answer make sense? What if σ_2^2 is much smaller than σ_1^2?

Answer:

6. Complete **Exercise 8.8** of the text, giving an explicit expression for the mean and variance of each estimator. You can use the result of **Exercise 6.81**: if $Y_1, Y_2, \ldots, Y_n$ are iid exponential random variables with mean θ, then $\min(Y_1, Y_2, \ldots, Y_n)$ has an exponential distribution with mean θ/n.

Answer:

---

7. Assess your results in problem 6 above via simulation, using $\theta = 10$. Check your theoretical variance expressions against the empirical variances from simulation, and compare the estimators using side-by-side boxplots of their values. Use the following code to get started:

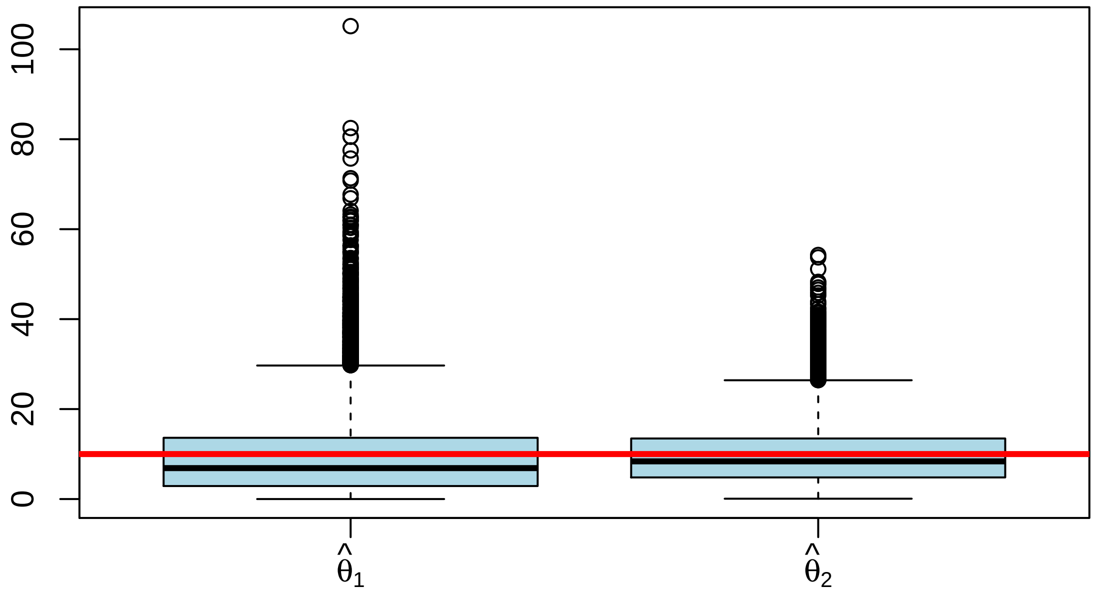
```
theta <- 10
set.seed(4302019)
Y <- rexp(3 * 10000, rate = 1 / theta)
YY <- matrix(Y, 3, 10000)
theta_hat1 <- YY[1, ]                        # Y_1 only
theta_hat2 <- (YY[1, ] + YY[2, ]) / 2  # (Y_1 + Y_2) / 2
var(theta_hat1)
```

```
## [1] 97.36235
```

```
var(theta_hat2)
```

```
## [1] 49.48804
```

```
# Side-by-side boxplots to compare estimators (you can add more estimators, separated by commas):
boxplot(theta_hat1, theta_hat2, col = "LightBlue", names = c(expression(hat(theta)[1]), expression(hat(
abline(h = theta, lwd = 3, col = "red")
```



Answer:
