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:

$\bar{Y} = \frac{Y}{n} = \hat{p}$ Where, $Y_i \sim Bern(p)$

$\bar{Y} \sim N(E[y_i], \frac{Var[Y_i]}{n})$

$\hat{p} = \bar{Y} = Y/n = \sum Y_i/n \ \sum Y_i \sim N(np, np(1-p))$

$ln(\frac{\hat{p}}{1-\hat{p}})$ estimates $ln(\frac{p}{1-p})$ $\hat{\lambda} = g(\hat{p}) \sim N(g(p), g^{'}(p)^2 * \frac{p(1-p)}{n})$

So,

$\hat{\lambda} \sim N(ln(p/(1-p)), (\frac{1}{p} + \frac{1}{1-p})^2 * \frac{p(1-p)}{n})$

$V(\hat{\lambda}) = (\frac{1}{p} + \frac{1}{1-p})^2 * \frac{p(1-p)}{n} = (\frac{1}{.3} + \frac{1}{.7})^2 * \frac{.3*.7}{100} = \frac{1}{21}$

---

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:

$\bar{Y} = \frac{\sum Y_i}{n}$ $E(Y_i) = \int_0^1 y * 3y^2 dy = \frac{3}{4}y^4|_0^1 = \frac{3}{4}$ $E(\bar{Y}) = E(\frac{\sum Y_i}{n}) = \frac{1}{n}E(\sum Y_i) = \frac{1}{n}\sum E(Y_i) = \frac{1}{40}(40)(.75) = 0.75$

Since,

$Var(Y_i) = E(Y_i^2) - E(Y_i)^2 = \int_0^1 y^2 3y^2 dy - 0.75^2 = 0.0375$

$Var(\bar{Y}) = \sigma^2/n = 0.0375/40$

So, $\bar{Y} \sim N(3/4, 3/3200)$

And we are looking for $P(\bar{Y} > 0.7)$ After normalizing, $P(Z > -1.6329) = 0.9484$

---

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: This would have taken me over an hour just to write up, so here is my handwritten work.

3) Let $Y_1, \dots, Y_n$ 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

$$\hat{\theta} = \max(Y_1, \dots, Y_n)$$

Show that $\hat{\theta}$ is a biased estimator for θ.

Show $E(\hat{\theta}) = \theta$ for unbiased

$B(\hat{\theta}) = E(\hat{\theta}) - \theta$

$E(\hat{\theta}) =$

1. Find cdf of Y
2. Find pdf

max involves CDF

$P(\max(Y_1, \dots, Y_n) \le y) = F(y)$

$F(\hat{\theta}) = P(\max(Y_1, \dots, Y_n) \le y)$

$F(\hat{\theta}) = P(\max(Y_1, \dots Y_n) \le y)$

so

$$F(y) = \int_{-\infty}^{y} \frac{\alpha x^{\alpha-1}}{\theta^{\alpha}} dx = \frac{1}{\theta^{\alpha}} \int_0^y \alpha x^{\alpha-1} dx = \frac{x^{\alpha}}{\theta^{\alpha}} \Big|_0^y = \frac{y^{\alpha}}{\theta^{\alpha}}$$

$$\begin{aligned} F(y) &= P(\max\{Y_1, \dots, Y_n\} \le y) \\ &= P(Y_1 \le y, Y_2 \le y, \dots, Y_n \le y) \\ &= P(Y_1 \le y) \dots P(Y_n \le y) \\ &= \left(\frac{y^{\alpha}}{\theta^{\alpha}}\right)^n \end{aligned}$$

$$f(y) = \frac{d}{dy} F(y) = \frac{d}{dy} \left(\frac{y^{\alpha}}{\theta^{\alpha}}\right)^n = \cancel{n\left(\frac{y^{\alpha}}{\theta^{\alpha}}\right)^{n-1}\left(\frac{\alpha y^{\alpha-1}}{\theta^{\alpha}}\right)}$$

$$= \frac{\alpha n y^{\alpha n}}{y \theta^{\alpha n}} = \frac{\alpha n y^{\alpha n - 1}}{\theta^{\alpha n}}$$

$$E(\hat{\theta}) = \int_{-\infty}^{\infty} y f(y) dy = \int_0^{\theta} y \cdot f(y) dy = \frac{\alpha n}{\theta^{\alpha n}} \int_0^{\theta} y^{\alpha n} dy$$

$$E(\hat{\theta}) = \frac{n\alpha}{\theta^{\alpha n}} \frac{y^{\alpha n+1}}{\alpha n+1}\Big|_0^{\theta} = \underbrace{\frac{n\alpha}{\alpha n+1}}_{\text{Bias}}\theta \qquad B(\hat{\theta}) = E(\hat{\theta}) - \theta$$

$$E(\hat{\theta}) = E(\max\{Y_1, \dots Y_n\})$$

$$=$$

Since max
Need the cdf

$$E(X) = \int x f(x)\,dx$$

CDF of max

$$F(\theta) = P(\max\{Y_1, \dots Y_n\} \le y)$$

$$= P(Y_1 \le y, Y_2 \le y, \dots Y_n \le y)$$

Since if the max $\le y$ then all of y_i is $\le y$

$$\text{(indep)} = P(Y_1 \le y) P(Y_2 \le y) \cdots P(Y_n \le y)$$

$$= P(Y_1 \le y)^n$$

$$= (\text{CDF})^n = (F(y))^n$$

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

$$F(y) = \int_0^y \alpha t^{\alpha-1}/\theta^{\alpha}\,dt = \frac{y^{\alpha}}{\theta^{\alpha}}$$

θ is β $\quad F(\hat{\theta}) = \left(\frac{y^{\alpha}}{\theta^{\alpha}}\right)^n = \left(\frac{y^{\alpha}}{\beta^{\alpha}}\right)^n$

$$\frac{d}{dy} F(\hat{\theta}) = f(\hat{\theta}) = n\alpha \frac{y^{\alpha n-1}}{\theta^{\alpha n}} = n\alpha \frac{y^{\alpha n-1}}{\beta^{\alpha n}}$$

$$E(\hat{\theta}) = \int_0^{\theta} y \cdot n\alpha \frac{y^{\alpha n-1}}{\theta^{\alpha n}}\,dy = \underbrace{\frac{n\alpha}{\alpha n+1}}_{\text{Bias}}\theta$$

So $\hat{\theta}$ is a biased estimator for θ

****

(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:

b) $V(\hat{\theta})$ and $MSE(\hat{\theta})$

$$V(\hat{\theta}) = E(\hat{\theta}^2) - E(\hat{\theta})^2$$

$$E[\hat{\theta}^2] = \int_0^{\theta} y^2 \, n\alpha \frac{y^{\alpha n-1}}{\theta^{\alpha n}} \, dy$$

$$= \frac{n\alpha}{\theta^{\alpha n}} \int_0^{\theta} y^{\alpha n+1} \, dy$$

$$= \frac{n\alpha}{\alpha n+2} \theta^2$$

$$V(\hat{\theta}) = \frac{n\alpha}{\alpha n+2} \hat{\theta}^2 - \left(\frac{n\alpha}{n\alpha+1} \theta\right)^2$$

$$MSE = V(\hat{\theta}) + B(\hat{\theta})^2$$

$$= \left(\frac{n\alpha}{n\alpha+2} \hat{\theta}^2 - \left(\frac{n\alpha}{n\alpha+1} \theta\right)^2\right) + \left(\frac{n\alpha}{n\alpha+1}\right)^2$$

(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).

```r
rsim <- function(n, alpha, theta){
  u <- runif(n)
  Y <- theta * u ^ (1 / alpha)
  return(Y)
}
nreps <- 10000
n <- 6
alpha <- 2
```

```r
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:

a)

Since $\hat{\theta}$ is unbiased, $B(\hat{\theta}) = E(\hat{\theta}) - \theta = 0$

b)

If $B(\hat{\theta}) = 5$ then,

$5 = E(\hat{\theta}) - \theta \; E(\hat{\theta}) = 5 + \theta$

---

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:

a)

Show $B(\hat{\theta}_3) = 0$

$B(\hat{\theta}_3) = E(\hat{\theta}_3) - \theta \; B(\hat{\theta}_3) = E(a\hat{\theta}_1 + (1-a)\hat{\theta}_2) - \theta$ Using properties of Expected Values,

$B(\hat{\theta}_3) = a\theta + (\theta - a\theta) - \theta = 0$

b)

$Var(\hat{\theta}_3) = Var(a\hat{\theta}_1 + (1-a)\hat{\theta}_2) \; Var(\hat{\theta}_3) = a^2 Var(\hat{\theta}_1) + (1-a)^2 Var(\hat{\theta}_2) \; Var(\hat{\theta}_3) = a^2\sigma_1^2 + (1-a^2)\sigma_2^2$

We would want a to be 0.5 to minimize the variance.

---

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:

6) (8.8 in text)

Y_1, Y_2, Y_3 denote a random sample from an exponential distribution with density function

$$f(y) = \begin{cases} \left(\frac{1}{\theta}\right) e^{-y/\theta}, & y > 0 \\ 0, & \text{elsewhere} \end{cases}$$

$$\hat{\theta}_1 = Y_1, \quad \hat{\theta}_2 = \frac{Y_1 + Y_2}{2}, \quad \hat{\theta}_3 = \frac{Y_1 + 2Y_2}{3}$$

$$\hat{\theta}_4 = \min(Y_1, Y_2, Y_3), \quad \hat{\theta}_5 = \bar{Y} = \frac{\Sigma Y_i}{n}$$

a) Which ~~one~~ of these Estimators are unbiased?

b) Among the unbiased Estimators, which has the smallest Variance?

a) i) $B(\hat{\theta}_1) = E(\hat{\theta}_1) - \theta$.

Since $f(y)$ is the density of an exponential distribution

$$E(\hat{\theta}_1) = E(Y_1) = \theta$$

So $B(\hat{\theta}_1) = 0$

ii)

$$B(\hat{\theta}_2) = E\left(\frac{Y_1 + Y_2}{2}\right) - \theta$$
$$= \tfrac{1}{2}(E(Y_1 + Y_2) - \theta$$
$$= \tfrac{1}{2}E(Y_1) + \tfrac{1}{2}E(Y_2) - \theta$$
$$= \tfrac{1}{2}\theta + \tfrac{1}{2}\theta - \theta = 0$$

iii) $B(\hat{\theta}_3) = E\left(\frac{Y_1 + 2Y_2}{3}\right) - \theta$

$$= \tfrac{1}{3}\theta + \tfrac{2}{3}\theta - \theta = 0$$

IV) $B(\hat{\theta}_4) = E(\min(Y_1, Y_2, Y_3)) - \theta$

$$\min(Y_1, Y_2, Y_3) \sim \text{Exp}(\theta/n)$$
$$= \frac{\theta}{3} - \theta = \frac{-2\theta}{3}$$

v) $B(\hat{\theta}_5) = E(\bar{Y}) - \theta$

$$= E\left(\frac{\Sigma Y_i}{n}\right) - \theta$$
$$= \theta - \theta = 0$$

$V(\hat{\theta}) = E[(\hat{\theta} - E(\hat{\theta}))^2]$

(6) b)

i) $Var(\hat{\theta}_1) = \cancel{E(\hat{\theta}_1^2) - E(\hat{\theta}_1)^2} = \theta^2$ since $Y_1 \sim Exp(\theta)$

$= E((\hat{\theta}_1 - E(\hat{\theta}_1))^2)$

$= E((\hat{\theta}_1 - \theta)^2)$

$= E(\hat{\theta}_1^2 + \theta^2 - 2\hat{\theta}_1\theta)$

$= E(\hat{\theta}_1^2) + E(\theta^2) - 2E(\hat{\theta}_1\theta)$

Var of $exp(\theta)$ is θ^2

ii) $Var(\hat{\theta}_2) = Var(\frac{Y_1 + Y_2}{2})$

$= \frac{1}{4}(Var(Y_1) + Var(Y_2))$

$= \frac{1}{4}(2\theta^2) = \frac{1}{2}\theta^2$

iii) $Var(\hat{\theta}_3) = Var(\frac{Y_1 + 2Y_2}{3})$

$= \frac{1}{9}(Var(Y_1) + 4\,Var(Y_2))$

$= \frac{1}{9}(\theta^2 + 4\theta^2)$

$= \frac{5}{9}\theta^2$

iv) $Var(\hat{\theta}_5) = Var(\bar{Y}) = Var(\frac{Y_1 + Y_2 + Y_3}{3})$

$= \frac{1}{9}(Var(Y_1) + V(Y_2) + V(Y_3))$

$= \frac{1}{2}\theta^2$

The estimators with the smallest variance is $\hat{\theta}_2$ and $\hat{\theta}_5$

****

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:

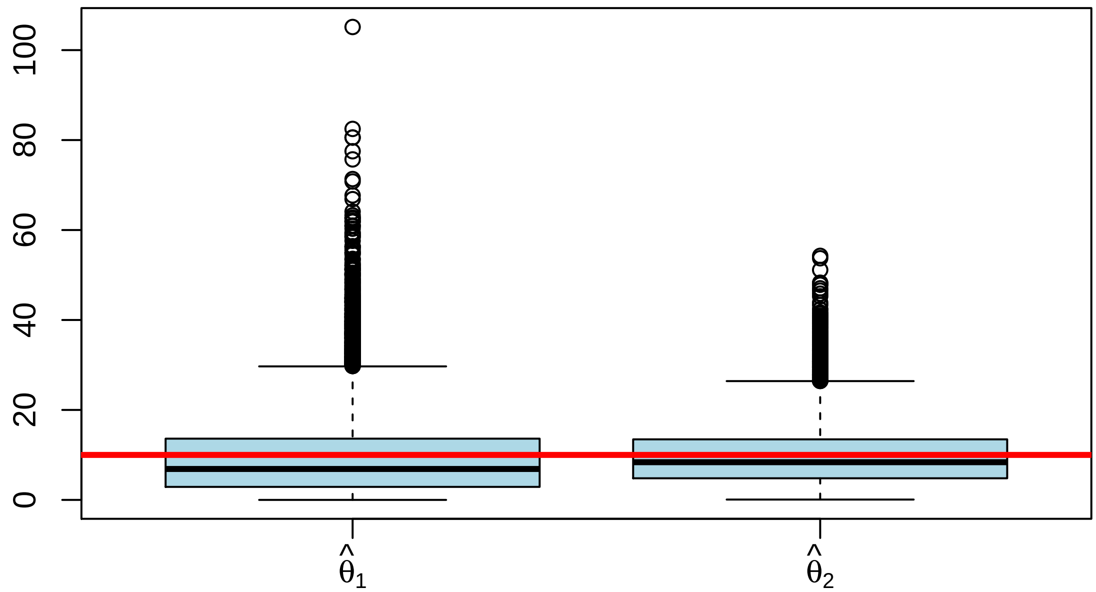
```r
theta <- 10
set.seed(4302019)
Y <- rexp(3 * 10000, rate = 1 / theta)
YY <- matrix(Y, 3, 10000)
theta_hat1 <- YY[1, ]                    # Y_1 only
theta_hat2 <- (YY[1, ] + YY[2, ]) / 2    # (Y_1 + Y_2) / 2
var(theta_hat1)
```

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

```r
var(theta_hat2)
```

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

```r
# 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:

```r
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
theta_hat3 <- (YY[1, ] + 2*YY[2, ]) / 3
theta_hat4 <- apply(YY, MARGIN = 2, FUN = min)
theta_hat5 <- apply(YY, MARGIN = 2, FUN = mean)
var(theta_hat1)
```

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

```r
var(theta_hat2)
```

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

```r
var(theta_hat3)
```

```
## [1] 55.7437
```

```r
var(theta_hat4)
```

```
## [1] 11.14563
```

```r
var(theta_hat5)
```

```
## [1] 32.94742
```

```r
# Side-by-side boxplots to compare estimators (you can add more estimators, separated by commas):
boxplot(theta_hat1, theta_hat2, theta_hat3, theta_hat4, theta_hat5, col = "LightBlue", names = c(express
abline(h = theta, lwd = 3, col = "red")
```