Homework 02: Due 9/20

Stat061-F23

Prof Amanda Luby

---

1. Define a new estimator $\hat{\theta}_4$ for the Uniform$(0, \theta)$ distribution as follows. Is $\hat{\theta}_4$ unbiased? You can give an intuitive justification or a mathematical one, and are welcome to use any results from class.

$$\hat{\theta}_4 = \begin{cases} 2\bar{X} & \text{if } \max\{X_i\} < 2\bar{X} \\ \max\{X_i\} & \text{otherwise} \end{cases}$$

2. Let $Y_1, Y_2, \ldots Y_n$ be a random sample of size n from the pdf $f_y(y) = \frac{1}{\theta} e^{-y/\theta}, y > 0$.

   (a) Show that $\hat{\theta}_1 = Y_1$, $\hat{\theta}_2 = \bar{Y}$, and $\hat{\theta}_3 = nY_{\text{min}}$ are all unbiased estimators for θ. You may use the general formula for the pdf of a minimum: $f_{Y_{min}}(y) = n(1 - F_y(y))^{n-1} f_y(y)$, where f_y is the pdf of Y and F_y is the cdf of Y.
   (b) Find an expression for the probability that $\hat{\theta}_1$ is within 0.1 of θ. (*Hint:* use the pdf of $\hat{\theta}_1$)
   (c) What is the probability from (b) if $\theta = .5$? what if $\theta = 2$?
   (d) Find the variances of $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$. Comment on which estimator is most efficient.

3. Suppose $X_1, X_2, \ldots, X_n$ are iid from a Gamma(α, λ) distribution. That is, $f_x(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$. You can also use $E(X) = \frac{\alpha}{\lambda}$ and $V(X) = \frac{\alpha}{\lambda^2}$.

   (a) Write out the likelihood function, and show that it depends on the data values only through $\bar{X}$ and $\bar{X}_g = (\prod X_i)^{1/n}$ ($\bar{X}_g$ is the *geometric mean*).
   (b) When α and λ are both unknown, the MLE does not have a closed form solution. Instead, find the MoM estimates $\hat{\alpha}$ and $\hat{\lambda}$.
   (c) ~~Are the MoM estimates unbiased?~~
   (d) Gamma random variables are the waiting times for Poisson occurrences. In sports, goals are often assumed to follow a Poisson process, which means that the waiting time for the 1st goal can be assumed to be a Gamma random variable. For the Swarthmore women's soccer team so far this year, the first goal in $n = 4$ games has occurred at $X_i = 11.01667, 3.05, 76.65, 24.1333$ minutes. Report $\hat{\alpha}$ and $\hat{\lambda}$ for these data. (Note that $\alpha = 1$ implies an exponential distribution, which would be the case if goals occur as a Poisson process)

4. The following R code simulates $n = 25$ draws from a Uniform(0,10) distribution 10,000 times. Here we know that $\theta = 10$, but we want to investigate how our estimators behave. For each sample, we estimate (1) the MoM estimate $\hat{\theta}_1 = 2\bar{X}$, (2) the MLE $\hat{\theta}_2 = X_{max}$, and (3) our $\hat{\theta}_4$ estimate from Q1.

```r
library(tidyverse)
set.seed(091523) # set a random seed for reproducibility
sim_results = tibble(   # sim_results is a data frame that
  MOM = rep(NA, 10000), # will store our 10,000 estimates x3
  MLE = rep(NA, 10000),
  Est3 = rep(NA, 10000)
)
```

```r
for(ii in 1:10000){
  x = runif(25, 0, 10) # draw 25 uniform(0,10) RV's
  sim_results$MOM[ii] = 2 * mean(x) # compute MoM and store
  sim_results$MLE[ii] = max(x) # compute MLE and store
  sim_results$Est3[ii] = max(2*mean(x), max(x)) # compute Est3 and store
}
```

(a) Construct a histogram for each estimate (e.g. one for the distribution of MoM, one for MLE, one for Est3)

(b) Compute the mean and variance of the MoM, MLE, and Est3 estimates. Comment on the bias and efficiency of each. Do the results surprise you given your answer to Question 1 and our previous results from class?

5. Let f_y be a continuous pdf with median M. If $Y_1, ..., Y_n \sim f_y$, the sample median $\hat{M} = \text{Median}(Y_1, ..., Y_n)$ has an approximate $N(M, \frac{1}{4n(f_y(M))^2})$ distribution.

   (a) Suppose $Y_1, ..., Y_n \sim N(\mu, \sigma^2)$. Show that $\hat{M}$ has an approximate $N(\mu, \frac{\pi\sigma^2}{2n})$.

   (b) Find the relative efficiency of $\bar{Y}$ to $\hat{M}$.

   (c) Using your result from (b), show that $\bar{Y}$ achieves the same standard error as $\hat{M}$ with only 63.7% as much data

   (d) Find the relative efficiency of $\bar{Y}$ to $\hat{M}$ if $Y_1, ..., Y_n \sim \text{Unif}(0, \theta)$. What do you conclude?