

STAT 422: Homework #2

Due, 2/8/22, 11:59pm

Problem 1

Statement

Let the random variable X denote the lifetime (in days) of cut roses delivered by a florist. Suppose X has the following pdf:

$$f_X(x) = \frac{2\theta^2}{x^3} I_{[\theta, \infty)}$$

In this case, the parameter θ depends on the amount of time between when the roses were harvested and when they were delivered. Suppose you receive a random sample of 12 roses for Valentine's Day, and you are interested in studying the lifetime of your flowers statistically.

- If you are interested in studying the lifetimes of your flowers statistically, why is it advantageous that your bouquet be stored in 12 separate vases, rather than a single vase? Explain.
- What is the probability *all* of the roses you receive will live $(7 + \theta)$ days or less?
- Find the pdf of $X_{(1)}$.
- Set up, **but do not evaluate**, the expression to find the expected value of the minimum lifetime (in days) of a delivery of cut roses.
- Without doing any calculations**, Explain, in words, why you might be more interested in estimating the quantity in part d) than $E(X)$. Your explanation should include the meanings of $E(X_{(1)})$ and $E(X)$ in the context of the scenario.

Solution

- The flowers being in different vases allows us to claim that the points are sampled iid. If they were all in one vase, then the life of roses could influence each other (i.e if a rose dies there may be more nutrients, or something bad in the shared water, etc.) and they are not independent.
- Lets first compute $F_X(x)$:

$$F_X(x) = \int_{\theta}^x 2\theta^2 y^{-3} dy = \left(-4\theta^2 y^{-2} \right)_{\theta}^x = -4\theta^2 x^{-2} + 4\theta^2 \theta^{-2} = \frac{-4\theta^2}{x^2} + 4$$

Now, we can easily compute the probability that all the roses will live $(7 + \theta)$ days. For one flower, we have $P(X \geq 7 + \theta) = 1 - P(X \leq 7 + \theta) = 1 - F_X(7 + \theta) = -3 + \frac{4\theta^2}{(7+\theta)^2}$, then since each flower is sampled iid: we can compute the probability of all the flowers by: $(1 - F_X(7 + \theta))^{12} = (-3 + \frac{4\theta^2}{(7+\theta)^2})^{12}$

c. We have from theorem 6.5: $f_{X_{(1)}}(x) = n(1 - F_X(x))^{n-1}f_X(x)$. Thus:

$$f_{X_{(1)}}(x) = n \left(\frac{-4\theta^2}{x^2} + 4 \right)^{n-1} \frac{2\theta^2}{x^3} I_{[\theta, \infty)}$$

d. We know that $\mathbb{E}[X_{(1)}] = \int_{\theta}^{\infty} x f_{X_{(1)}}(x) dx$:

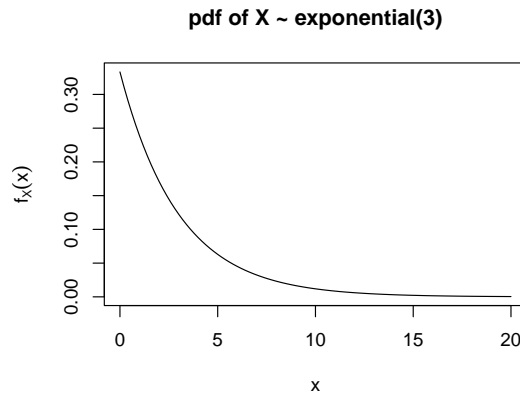
$$\mathbb{E}[X_{(1)}] = \int_{\theta}^{\infty} x f_{X_{(1)}}(x) dx = \int_{\theta}^{\infty} x n \left(\frac{-4\theta^2}{x^2} + 4 \right)^{n-1} \frac{2\theta^2}{x^3} dx = \int_{\theta}^{\infty} n \left(\frac{-4\theta^2}{x^2} + 4 \right)^{n-1} \frac{2\theta^2}{x^2} dx$$

e. The $\mathbb{E}[X_{(1)}]$ is more useful than $\mathbb{E}[X]$ because we want to know when the first rose is going to die. I do not want to give my significant other a dead rose, but want to give 12 living roses. So I want to know $\mathbb{E}[X_{(1)}]$, which is the expected number of days before the first rose dies. While $\mathbb{E}[X]$ is the expected number of days for any rose to live, if I waited $\mathbb{E}[X]$ days, I would most likely have only 6 living roses, which is not what I want.

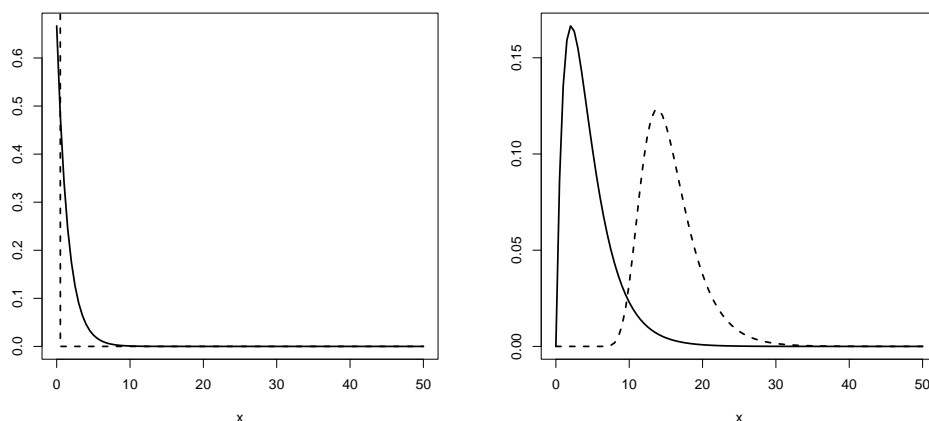
Problem 2

Statement

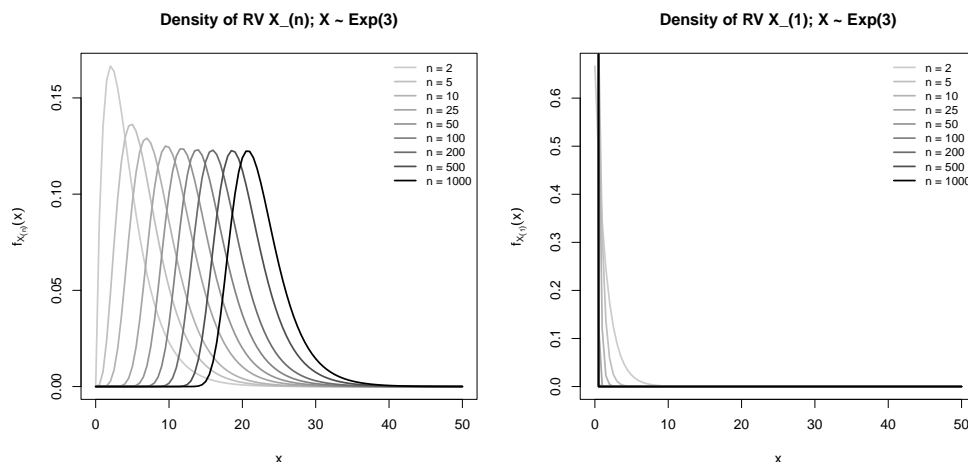
We derived the exact sampling distribution for two order statistics, $X_{(1)}$ and $X_{(n)}$ for a random sample of size n , X_1, \dots, X_n , where $X_i \stackrel{iid}{\sim} \text{Uniform}(0,1)$. We compared how these distributions changed as n was varied between 2 and 100. Additionally, you should try deriving the same two distributions (i.e., $f_{X_{(1)}}, f_{X_{(n)}}$) for a general n when the parent distribution for the RS is Exponential with a mean of 3 units. Below, the parent distribution is plotted.



- a. The following plots show the distributions of $X_{(1)}$ and $X_{(n)}$, each for $n = 2$ and $n = 100$. The plotting window separates the type of order statistic (i.e., the max and the min) and the linetype distinguishes between the sample sizes. Provide titles, y-axis labels, and legends for these two figures. Explain your reasoning for your choices in a few sentences.



- b. The following figures were generated by `plot_max_exp` and `plot_min_exp` (found on the R file provided). For each order statistic ($X_{(1)}$ and $X_{(n)}$), write a sentence about how the sample size affects the sampling distribution.

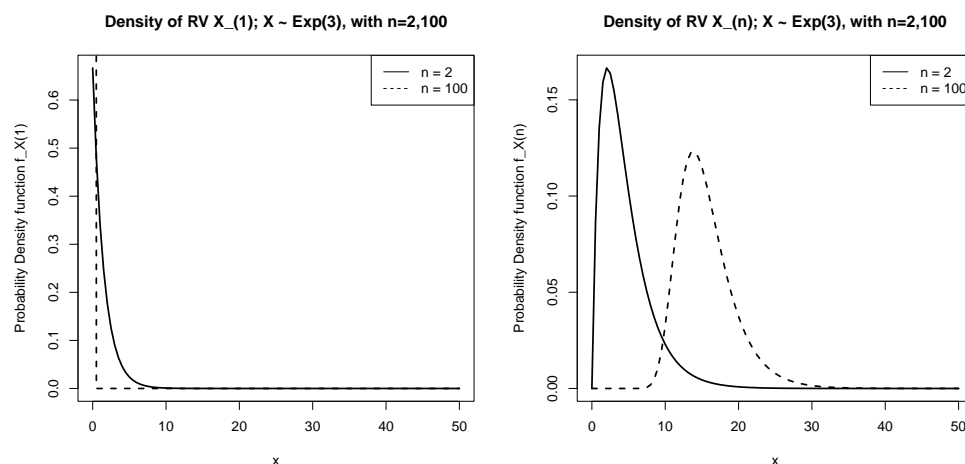


- c. The code posted on the website, for this assignment, provides functions for computing the sampling distribution $X_{(1)}$ and $X_{(n)}$ for a RS of size n `max_exp`, `min_exp`, and for comparing random samples of different sizes `plot_max_exp` and `plot_min_exp`. *Carefully* walk through this code, reading the comments and pausing after executing each line to think about what the code did. Explore the sampling distributions for each of the order statistics for sample sizes you are interested in. *Write down at least one question you have about sampling distributions for order statistics or about any piece of the R code.*

- d. Describe the steps you would take to generate an approximate sampling distribution for $X_{(n)}$, in words. *Note: you do not have to write any code.*

Solution

- a. The following plots show the distributions of $X_{(1)}$ and $X_{(n)}$, each for $n = 2$ and $n = 100$. The plotting window separates the type of order statistic (i.e., the max and the min) and the linetype distinguishes between the sample sizes. Provide titles, y-axis labels, and legends for these two figures. Explain your reasoning for your choices in a few sentences.



- b. For $X_{(1)}$, as n increases, the density shifts strongly towards zero. This makes sense since as we have more and more samples, the minimum is probably going to be small. Because $X \sim \text{Exp}(3)$ we see above that the ‘most likely’ value (the value with most density) is zero. So as we sample many times, we should be sampling close to zero, which explains why as n goes big, the density of $X_{(1)}$ becomes very high near zero. For $X_{(n)}$, as n increases we see distributions with steadily increasing means. As n increases, we expect $X_{(n)}$ to also increase.
- c. One thing I learned was using the expression function in R to format the axis labels nicely. I am curious as to why the distributions look like they do. For the $X_{(n)}$ case they look almost normal, while for $X_{(1)}$ they look almost like $1/kx$ for some constants k . I understand that Gamma distribution can do this, but it is just a very interesting shape and am curious as to why they look like they do.
- d. I would sample n points from an exponential distribution (or any that I was querying) using `rexp` function. Then I would compute the maximum from this sample. I would then repeat this process a ton of times (10000 or so, depending on n and how accurate I want to be). Then I would plot a histogram of the results.