170S Week 7 Discussion Notes

Colin Ni

February 17, 2025

Problem 1. Suppose the time it takes a company to fill an advertised open position is exponentially distributed with mean 20 (days), and suppose that the times are independent even when there are multiple advertised open positions.

(i) The company needs at least a 90% chance that someone will be hired in the next 5 days. What is the smallest number n of open positions the company should advertise?

(ii) On the other hand, the company is worried that all n positions will be filled in the next 40 days. What is the probability of this happening?

(iii) If the company advertises three open positions, how long will it take on average for all of them to be filled?

Hint. The time it takes to fill an advertised open position is exponential with rate $\frac{1}{20}$, *i.e.* it follows the distribution $\mathrm{Exp}(\frac{1}{20})$. In part ($i$), let n be the number of advertised open positions, and let $X_1, \ldots, X_n$ be the times it takes to fill each one. By our assumption these are i.i.d., so consider the first order statistic $X_{(1)}$. What does this represent? What is its pdf / cdf? For parts (ii) and (iii), use the same setup but consider a different order statistic. For (ii) specifically you may want to recall that it is easy to write down the cdf of an order statistic. $\square$

Solution. Let us continue from the hint.

For (i), the first order statistic $X_{(1)}$ is the time it takes for the first person to be hired, and we want $\mathbb{P}(X_{(1)} \leq 5) = 0.9$ or in other words $F_{X_{(1)}}(5) = 0.9$. A big-brain way to determine the cdf of $X_{(1)}$ is to use something we discussed in week 3, namely that the first order statistic of i.i.d. exponentials is exponential with an n times larger rate, so here $X_{(1)} \sim \mathrm{Exp}(\frac{n}{20})$. Another way to do this is by inspecting the pdf $f_{X_{(1)}}(x)$ and observing that it is the pdf for $\mathrm{Exp}(\frac{n}{20})$. Either way, we have

$$F_{\mathrm{Exp}(\frac{n}{20})}(5) = 1 - e^{-\frac{n}{4}} = 0.9 \quad \text{if and only if} \quad n = -4\ln(0.1) \approx 9.2,$$

so the company should advertise 10 positions.

For (ii), we are now interested in the probability $\mathbb{P}(X_{(10)} \leq 40)$ or in other words the cdf $F_{X_{(10)}}(40)$. As the hint alludes to, recall from our week 1 discussion section that the cdf of the rth order statistic $X_{(r)}$ is in general

$$F_{X_{(r)}}(x) = \sum_{j=r}^{n} \binom{n}{j} F_X(x)^j (1 - F_X(x))^{n-j}.$$

In our case it is

$$F_{X_{(10)}}(x) = (1 - e^{-\frac{1}{20}x})^{10},$$

so $F_{X_{(10)}}(40) = (1 - e^{-2})^{10} \approx 0.23$.

For (iii), we are now interested in the expected value $\mathbb{E}(X_{(3)})$ of the third order statistic $X_{(3)}$. A big-brain way to find this is to again use what we did in week 3, namely that the other order statistics of i.i.d. exponentials are sums of exponentials with rates that decrease in a predictable way:

$$\mathbb{E}(X_{(3)}) = \left(\frac{1}{3} + \frac{1}{2} + \frac{1}{1}\right) = \frac{110}{3}.$$

Another way to do this is by using the pdf

$$f_{X_{(3)}}(x) = 3 \cdot (1 - e^{-\frac{1}{20}x})^2 \cdot \frac{1}{20} e^{-\frac{1}{20}x}$$

to compute by hand (trust but verify) the expected value

$$\int_0^\infty \frac{3}{20}(1 - e^{-\frac{1}{20}x})^2 e^{-\frac{1}{20}x} x \, \mathrm{d}x = \frac{110}{3}.$$

□

Problem 2. Let $X \sim \mathrm{Unif}(0, \theta)$ be a uniform random variable where $\theta > 0$ is an unknown parameter, and let $x_1, \ldots, x_n$ be a sample drawn from X. In this problem, we will explore some estimators for θ.

(i) Find the MLE $\hat{\theta}$ for the parameter θ, and determine whether it is unbiased.

(ii) Find the method of moments estimator $\tilde{\theta}$ for the parameter θ, and determine whether it is unbiased.

(iii) Show, by using an indicator function, that the sample maximum $x_{(n)}$ is a sufficient statistic for θ.

(iv) Show that the sample mean $\bar{x}$ is not a sufficient statistic for θ and hence that the sum of the samples $n\bar{x}$ is not either.

(v) However, in the case $n = 2$, show that the sum $n\bar{x} = x_1 + x_2$ along with the difference $x_1 - x_2$ are jointly sufficient statistics for θ.

Hint. For this entire problem, be mindful of the support of X. In other words, keep in mind that the pdf is zero outside of the interval $[0, \theta]$. For (i), you should think about the sample maximum $x_{(n)}$, and for (iii) the indicator function you may find helpful is the function $1\{x_{(n)} \leq \theta\}$ which is defined to be 1 when $x_{(n)} \leq \theta$ and zero otherwise. □

Solution. For (i), following the hint, let us write down carefully that

$$f_X(x_i \mid \theta) = \begin{cases} \frac{1}{\theta} & x_i \leq \theta \\ 0 & \text{otherwise,} \end{cases}$$

where by default $0 \leq x_i$ because x_i is drawn from $\mathrm{Unif}(0, \theta)$. Thus the likelihood function is

$$\begin{aligned} L(x_1, \ldots, x_n \mid \theta) &= f_X(x_1) \cdots f_X(x_n) \\ &= \begin{cases} \frac{1}{\theta^n} & x_1, \ldots, x_n \leq \theta \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} \frac{1}{\theta^n} & x_{(n)} \leq \theta \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

where in the last equality we use the nth order statistic $x_{(n)}$. To find $\hat{\theta}$ that maximizes this likelihood function, observe that this likelihood function is positive precisely when $x_{(n)} \leq \theta$ and that here smaller values of θ increase the likelihood function. Thus there is a unique $\hat{\theta} = x_{(n)}$. To see that this is a biased estimator, recall that the order statistics of uniform are beta by what we discussed in our week 1 discussion section, or more precisely

$$x_{(n)} \sim \theta \cdot \mathrm{Beta}(n, 1),$$

which has mean $\mathbb{E}(x_{(n)}) = \frac{n}{n+1} \cdot \theta$.

For (ii), the first moment is $\mathbb{E}(X) = \frac{\theta}{2}$, and our estimator for the first moment is $\bar{x}$. Plugging in $\tilde{\theta}$ into the first moment, equating with our estimator, and solving for $\tilde{\theta}$ gives

$$\tilde{\theta} = 2\bar{x}.$$

This is unbiased because

$$\mathbb{E}(2\bar{x}) = 2\mathbb{E}(\bar{x}) = 2\mathbb{E}(X) = \theta,$$

where in the middle equality we use that the sample mean is an unbiased estimator of the mean, as we showed in our week 2 discussion section.

For (iii), following the hint, let us write the likelihood function further as

$$L(x_1, \ldots, x_n \mid \theta) = \frac{1}{\theta^n} \cdot 1\{x_{(n)} \leq \theta\}.$$

Since the only way in which the sample $x_1, \ldots, x_n$ occurs in this expression is through the sample maximum $x_{(n)}$, this shows that the sample maximum $x_{(n)}$ is a sufficient statistic for θ.

For (iv), consider the case $n = 2$, and suppose for the sake of contradiction that

$$\frac{1}{\theta^2} \cdot 1\{x_{(2)} \leq \theta\} = L(x_1, x_2 \mid \theta) = h(x_1, x_2) g(\theta, \bar{x})$$

for some functions h and g, using Wikipedia's notation. Note

$$0 = L\left(0, 1 \;\middle|\; \frac{3}{4}\right) = h(0,1)g\left(\frac{3}{4}, \frac{1}{2}\right),$$

so $h(0,1) = 0$ or $g(\frac{3}{4}, \frac{1}{2}) = 0$. But

$$\frac{1}{\left(\frac{3}{4}\right)^2} = L\left(\frac{1}{2}, \frac{1}{2} \;\middle|\; \frac{3}{4}\right) = h\left(\frac{1}{2}, \frac{1}{2}\right) g\left(\frac{3}{4}, \frac{1}{2}\right),$$

so $g(\frac{3}{4}, \frac{1}{2}) \neq 0$, which means it must have been $h(0,1) = 0$. Finally,

$$\frac{1}{4} = L\left(0, 1 \mid 2\right) = h(0,1)g\left(2, \frac{1}{2}\right) = 0$$

is a contradiction. We conclude that such a factorization cannot exist.

For (v), the key point is that

$$x_{(2)} = \frac{(x_1 + x_2) + |x_1 - x_2|}{2}.$$

Now, since $x_{(2)}$ is sufficient for θ and can be written as a function of $x_1 + x_2$ and $x_1 - x_2$, it follows that $x_1 + x_2$ and $x_1 - x_2$ are jointly sufficient for θ. $\square$