MATH 3215 FALL 2019
FINAL EXAM SOLUTIONS

1. (a) **I flip a fair coin 3 times. Let X be the number of flips that are immediately preceded by a heads. For example, $X(THH) = 1$ because flip 1 is not preceded by a heads, flip 2 also is not preceded by a heads, but flip 3 is.**

   i. **Find the pmf for X.**
   Solution. First the range of X is $\{0, 1, 2\}$ because flip 1 is never preceded by a heads. To find the PMF, we can write out all the possibilities:

$$f(0) = \mathbb{P}(\{TTT\}, \{TTH\}) = \frac{1}{4}$$
$$f(1) = \mathbb{P}(\{HTT\}, \{HTH\}, \{THT\}, \{THH\}) = \frac{1}{2}$$
$$f(2) = \mathbb{P}(\{HHT\}, \{HHH\}) = \frac{1}{4}.$$

   (Alternatively we can notice that X is equal to the number of heads among flips 1 and 2.)

   ii. **Compute Var(X^2).**
   Solution. Since

$$\mathrm{Var}(X^2) = \mathbb{E}((X^2)^2) - (\mathbb{E}(X^2))^2 = \mathbb{E}X^4 - (\mathbb{E}X^2)^2,$$

   we first compute

$$\mathbb{E}X^4 = 0^4 \cdot \frac{1}{4} + 1^4 \cdot \frac{1}{2} + 2^4 \cdot \frac{1}{4} = \frac{9}{2},$$

   and

$$\mathbb{E}X^2 = 0^2 \cdot \frac{1}{4} + 1^2 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{4} = \frac{3}{2}.$$

   Therefore $\mathrm{Var}(X^2) = 9/2 - (3/2)^2 = 9/4$.

(b) **I now flip the coin 100 times but the coin is switched for one that comes up heads with only probability 0.02. What is the approximate probability that no more than two heads appear?**
Solution. We use the poisson approximation, which states that if $\lambda > 0$ is fixed, then as $n \to \infty$, one has

$$\mathrm{Bin}(n, \lambda/n) \text{ is approximately Poisson}(\lambda).$$

In our experiment, the number of heads that come up is distributed as $\mathrm{Bin}(100, 2/100)$, which is approximately Poisson(2), and this distribution has pmf $f(x) = e^{-2}2^x/x!$ for $x = 0, 1, 2, \ldots$. Therefore

$$\mathbb{P}(X \leq 2) \text{ is approximately } \sum_{x=0}^{2} f(x) = e^{-2} \sum_{x=0}^{2} \frac{2^x}{x!},$$

which equals

$$e^{-2}(1+2+2) = 5e^{-2}.$$

2. **Let X and Y be independent variables that have exponential distribution with parameter $\theta = 1$.**

(a) **Let $Z = \min\{X, Y\}$. Compute $\mathbb{P}(Z \geq z)$ for $z \geq 0$ and then find the CDF of Z.**

Solution. By independence,

$$\begin{aligned}\mathbb{P}(Z \geq z) = \mathbb{P}(\min\{X, Y\} \geq z) = \mathbb{P}(X \geq z \text{ and } Y \geq z) &= \mathbb{P}(X \geq z)\mathbb{P}(Y \geq z) \\ &= \mathbb{P}(X \geq z)^2.\end{aligned}$$

For an exponential distribution with $\theta = 1$, the pdf is $f(x) = e^{-x}$ for $x \geq 0$, and so

$$\mathbb{P}(X \geq z) = \int_z^\infty e^{-x} \, \mathrm{d}x = -e^{-x}\Big|_z^\infty = e^{-z}.$$

Therefore for $z \geq 0$, we have

$$\mathbb{P}(Z \geq z) = e^{-2z}.$$

This means that the CDF of Z is

$$F(z) = \begin{cases} 0 & \text{if } z < 0 \\ 1 - e^{-2z} & \text{if } z \geq 0. \end{cases}$$

(b) **Compute the PDF for Z and then compute $\mathbb{E}Z$.**

Solution. Since the PDF is the derivative of the CDF, we obtain

$$f(z) = \frac{\mathrm{d}}{\mathrm{d}z}F(z) = \begin{cases} 0 & \text{if } z < 0 \\ 2e^{-2z} & \text{if } z \geq 0. \end{cases}$$

(Technically the derivative is not defined at 0, but we will not worry since this is only one point, and we will integrate this, so it will not matter.) Note that this is the pdf for exponential with parameter $\theta = 1/2$.

We can then compute

$$\mathbb{E}Z = \int_{-\infty}^\infty z f(z) \, \mathrm{d}z = \int_0^\infty 2ze^{-2z} \, \mathrm{d}z.$$

Since we know that Z is exponential with parameter $\theta = 1/2$, this mean is $1/2$.

(c) **Thinking of X and Y as associated two two independent Poisson processes, justify your above answer for $\mathbb{E}Z$ heuristically. Similarly, what should the value of $\mathbb{E}\min\{X_1, X_2, X_3, X_4\}$ be, where the X_i's are i.i.d. with the same distribution as that of X and Y?**

Solution. Imagine being at a call center where calls arrive with rate 1/min. Then the first arrival time of a call has the same distribution as does X. Take another independent call center whose calls arrive with rate 1/min, so that the first arrival time of a call has the same distribution as does Y. It is reasonable to believe then that if we mark all the calls for both call centers (simultaneously) on the same time line, we get a Poisson process with rate 2/min (that is, with $\lambda = 2$. Then $\min\{X, Y\}$ is the first arrival time in this new process, and it should be exponential with $\theta = 1/2$. This would give $\mathbb{E}\min\{X, Y\} = 1/2$.

In the case of $X_1, \ldots, X_4$, we simply take four independent Poisson processes with rate 1/min and mark them all on the same timeline, getting a process with rate 4/min. Then $\min\{X_1, X_2, X_3, X_4\}$ is the first arrival time in this new process, and should have expected value 1/4.

3. **You are a member of a class of 18 students. A bowl contains 18 chips: 1 blue and 17 red. Each student is to take one chip from the bowl without replacement. The student who draws the blue chip is guaranteed as A for the course.**

(a) **If you choose to draw first, what is your probability of getting the blue chip?**

Solution. This was a homework problem.

(b) **If the bowl has 2 blue chips and 16 red, and you choose to draw last, what is your probability of getting a blue chip? Justify your answer.**

Solution. This was a homework problem.

(c) **Again with 2 blue chips and 16 red, if you choose to draw fifth, what is your probability of getting a blue chip? Justify your answer.**

Solution. This was a homework problem.

4. **We roll two fair four-sided die. Let X be the number of ones and Y be the number of twos and threes.**

(a) **Plot the joint PMF for X and Y.**

Solution. This was done in class.

(b) **Compute $\mathbb{E}X$, $\mathbb{E}Y$, and Cov(X, Y). Are X and Y positively or negatively correlated?**

Solution. This was done in class.

(c) **Given the the variance of X is $3/8$ and the variance of Y is $1/2$, what is the least-squares regression line for (X, Y)? Plot it on the same graph you used for the joint PMF.**

Solution. This was done in class.

5. **We take n samples from a distribution with PDF**

$$f(x) = (1/\theta^2)xe^{-x/\theta}, \text{ for } x > 0,$$

where $\theta > 0$ is a parameter.

(a) **Find the maximum likelihood estimator for θ.**

Solution. The likelihood function is simply the joint pdf: for $x_1, \dots, x_n > 0$,

$$L(x_1, \dots, x_n) = \frac{1}{\theta^2}x_1 e^{-x_1/\theta} \cdot \dots \cdot \frac{1}{\theta^2}x_n e^{-x_n/\theta},$$

and so the log-likelihood function is (for $x_1, \dots, x_n > 0$)

$$\begin{aligned}\log L(x_1, \dots, x_n) &= \log\left(\frac{1}{\theta^{2n}} \cdot (x_1 \cdots x_n) \cdot e^{-\frac{x_1+\cdots+x_n}{\theta}}\right) \\ &= -2n\log\theta + \sum_{i=1}^{n}\log x_i - \frac{x_1 + \cdots + x_n}{\theta}.\end{aligned}$$

Taking the first derivative in θ, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\log L(x_1, \dots, x_n) = \frac{-2n}{\theta} + \frac{x_1 + \cdots + x_n}{\theta^2}.$$

We then find the critical points, where this expression is zero and end up with

$$\frac{2n}{\theta} = \frac{n\bar{x}}{\theta^2},$$

or

$$\hat{\theta} = \frac{\bar{x}}{2}.$$

To check that this is a maximum, we can perform the second derivative test:

$$\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}L(x_1, \dots, x_n) = \frac{2n}{\theta^2} - \frac{2n\bar{x}}{\theta^3}$$

Putting $\bar{x} = 2\hat{\theta}$, we obtain

$$\frac{2n}{(\hat{\theta})^2} - \frac{4n\hat{\theta}}{(\hat{\theta})^3} = -\frac{2n}{(\hat{\theta})^2} < 0.$$

(b) **We now take 10 samples from this distribution and obtain the data 3, 10, 7, 22, 1, 4, 3, 7, 9, 41. Find the first, second, and third quartiles.**

Solution. We first line up the data in order:

$$1, 3, 3, 4, 7, 7, 9, 10, 22, 41.$$

To compute the first quartile, $\tilde{\pi}_{0.25}$, we first compute $p(n+1) = 11/4 = 2 + 3/4$ where $p = 0.25$ and $n = 10$. Then we write

$$\tilde{\pi}_{0.25} = y_2 + \frac{3}{4}(y_3 - y_2) = 3.$$

For $p = 0.5$, we write $p(n+1) = 11/2 = 5 + 1/2$, and so

$$\tilde{\pi}_{0.5} = y_5 + \frac{1}{2}(y_6 - y_5) = 7.$$

Last, for $p = 0.75$, we write $p(n+1) = 33/4 = 8 + 1/4$, and so

$$\tilde{\pi}_{0.75} = y_8 + \frac{1}{4}(y_9 - y_8) = 10 + \frac{1}{4}(22 - 10) = 13.$$

(c) **Draw a box plot showing suspected outliers and outliers.**

Solution. I cannot draw it here, but you should draw a box with left line at 3, middle line at 7, and right line at 13. Therefore the IQR is 10, and suspected outliers will be between (3/2)IQR (that is, 15) and 3IQR (that is 30) from the left and right sides of the box. This means that only 41 is an outlier, and it is a suspected outlier. The whiskers should extend from the box (to the left) down to 1, and from the box (to the right) up to 22, and 41 should appear only as a point.

6. (a) (3 pts) **We take $n = 36$ (assumed to be large) samples of exam scores from Hootman University's physics department and find $\bar{x} = 82$ and $s = 5$. Find an approximate 95% confidence interval for the mean exam score of a test from that department.**

Solution. For an approximate confidence interval with σ unknown, but n large, we have the formula

$$\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}.$$

Here $\alpha = 0.05$, so the approximate confidence interval is

$$82 \pm 1.96 \cdot \frac{5}{6}.$$

(b) **Suppose we redo the experiment, already knowing now that $\sigma = 6$. Solve for the number of samples n we would need to be sure with approximately 90% confidence that the mean exam score is within two points of $\bar{x}$.**

Solution. For known σ and $\alpha = 0.1$, we have an approximate confidence interval

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}},$$

or

$$82 \pm 1.645 \frac{6}{\sqrt{n}}.$$

We would like the radius of this interval to be no larger than 2, so we obtain

$$1.645 \frac{6}{\sqrt{n}} \leq 2,$$

giving

$$n \geq \left(\frac{6 \cdot 1.645}{2} \right)^2 = 9 \cdot 1.645^2.$$

(c) **We are now taking Hootman U to honor court because we believe that the students are getting unethical help from their instructors. The court requests a hypothesis test with hypotheses $H_0 : \mu = 80$ and $H_a : \mu = 83$, with mandated critical region $\{\bar{x} \geq 82\}$ and $n = 9$ samples. Compute the probability of type-2 error for this test. You may assume that the test scores are normally distributed with $\sigma = 6$.**

Solution. The type-2 error is the probability of making a mistake when hypothesis 2 (the alternative hypothesis) is true. Making a mistake in this case would be accepting the null hypothesis, and this means that our data does not fall in the critical region. Therefore

$$\beta = \mathbb{P}(\text{type-2 error}) = \mathbb{P}(\bar{x} < 82; \mu = 83).$$

We now standardize, using that $\bar{x} \sim N(83, 36/9) = N(83, 4)$. This gives

$$\mathbb{P}\left(\frac{\bar{x} - 83}{2} < \frac{82 - 83}{2}; \mu = 83 \right) = \Phi(-0.5).$$

From the table, using that $\Phi(-0.5) = 1 - \Phi(0.5) \sim 1 - 0.6915 = 0.3085$, we obtain

$$\mathbb{P}(\text{type-2 error}) \sim 0.3085 = 30.85\%.$$

This is a bad statistical test and we probably need more samples.