APMA 1650 - Spring 2021

Final Exam - Solutions

Problem 2 (Aptitude) (9 pts) Employees in a firm are given an aptitude test when first employed. Experience has shown that of the 60 percent who passed the test, 70 percent of them were rated as good workers, whereas of the 40 percent who failed the test only 50 percent were rated as good workers.

a. (4 pts) What is the probability that an employee, selected at random will be a good worker?

b. (5 pts) What is the probability that a good worker failed the test?

Solution: Let T be the event that an employee passed test and G be the event than an employee was rated as good. We know that

$$P(T) = 0.6, \quad P(T^c) = 0.4, \quad P(G|T) = 0.7, \quad P(G|T^c) = 0.5.$$

a. Using the law of total probability

$$P(G) = P(G|T)P(T) + P(G|T^c)P(T^c) = 0.7(0.6) + 0.5(0.4) = 0.62$$

b. We want to find $P(T^c|G)$. Using Bayes rule

$$P(T^c|G) = \frac{P(G|T^c)P(T^c)}{P(G)} = \frac{.5(.4)}{0.62} = \frac{10}{31} \approx 0.323$$

Problem 3 (Pivotal) (17 pts) Let $X_1, X_2, \ldots X_n$ be a random sample from a distribution with PDF

$$f(x;\theta) = \begin{cases} \frac{2x}{\theta^2} & 0 \leq x \leq \theta \\ 0 & \text{otherwise.} \end{cases}$$

a. (4 pts) Show that $\hat{\Theta} = \max\{X_1, X_2, \ldots, X_n\}$ is a maximum likelihood estimator (MLE) for θ.

b. (4 pts) Find the CDF $F_{\hat{\Theta}}$ for $\hat{\Theta}$.

c. (4 pts) Show that $\hat{\Theta}/\theta$ is a pivotal quantity.

d. (5 pts) Use the pivotal quantity from part (c) to find the general form of a $100(1-\alpha)\%$ confidence interval for θ. State everything as explicitly as possible.

Solution:

a. For each $x_1, x_2, \dots x_n$, the likelihood function is

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \begin{cases} (\frac{2}{\theta^2})^n \prod_{i=1}^{n} x_i & 0 \leq x_1, x_2, \dots x_n \leq \theta, \\ 0 & \text{otherwise} \end{cases}$$

Following the same argument as in HW9 problem 4a, we see that the maximum is obtained at

$$\hat{\Theta} = \max\{X_1, X_2, \dots, X_n\}$$

b. The CDF of $f(x;\theta)$ is

$$F_X(x) = \int_0^x f(x;\theta)\mathrm{d}x = \begin{cases} 0 & x < 0 \\ \left(\frac{x}{\theta}\right)^2 & 0 \leq x \leq \theta \\ 1 & \text{otherwise} \end{cases}$$

Using the fact that $\{\hat{\Theta} \leq x\} = \bigcap_{i=1}^{n}\{X_i \leq x\}$, and the independence of X_i gives

$$F_{\hat{\Theta}}(x) = F_X(x)^n = \begin{cases} 0 & x < 0 \\ \left(\frac{x}{\theta}\right)^{2n} & 0 \leq x \leq \theta\,. \\ 1 & \text{otherwise} \end{cases}$$

c. To show that $Q = \hat{\Theta}/\theta$ is pivotal, we need to show that the distribution of $\hat{\Theta}$ doesn't depend on θ. By the CDF method, we see that

$$F_Q(x) = P(\hat{\Theta}/\theta \leq x) = P(\hat{\Theta} \leq x\theta) = F_{\hat{\Theta}}(x\theta) = \begin{cases} 0 & x < 0 \\ x^{2n} & 0 \leq x \leq 1\,, \\ 1 & \text{otherwise} \end{cases}$$

which clearly doesn't depend on θ.

d. To find the confidence interval, we first compute probabilities $q_{\alpha/2,n}$ and $q_{1-\alpha/2,n}$ that satisfy

$$q_{\alpha/2,n} = F_Q^{-1}(1-\alpha/2) = \left(1 - \frac{\alpha}{2}\right)^{1/2n}$$

and

$$P(q_{1-\alpha/2,n} \leq \hat{\Theta}/\theta \leq q_{\alpha/2,n}) = 1 - \alpha$$

Using algebra we deduce that

$$P\left(\frac{\hat{\Theta}}{q_{\alpha/2,n}} \leq \theta \leq \frac{\hat{\Theta}}{q_{1-\alpha/2,n}}\right) = 1 - \alpha$$

This means that

$$\left[\frac{\max\{X_1,\dots,X_n\}}{\left(1-\frac{\alpha}{2}\right)^{1/2n}}, \frac{\max\{X_1,\dots,X_n\}}{\left(\frac{\alpha}{2}\right)^{1/2n}}\right]$$

is a $(1-\alpha)\%$ confidence interval for θ.

Problem 4 (Vaccine Party) (12 pts) You are celebrating the end of the semester by throwing huge party with all your vaccinated friends (perhaps a bit prematurely...). You have an epic cooler set up with hundreds of your favorite IPA on ice. To make sure they are cold enough you randomly sample 12 beers and measure their temperatures (in Fahrenheit) and obtain a sample mean of 48 with a standard deviation of 3. Assume that the distribution of the temperatures follow a *normal distribution* and the samples are independent.

a. (4 pts) Your friends are very picky about their beer temperature. Give them a 99% confidence interval for the mean temperature of the IPAs.

b. (4 pts) Going one step further, give an 80% confidence interval for variance of the temperature of the IPAs.

c. (4 pts) In general, let $\overline{X}$ and S^2 be the sample mean and sample variance of 12 independent temperature measurements of the IPAs. Find a value a such that

$$P(\overline{X} - \mu \leq aS) = 0.01,$$

where μ is the true mean of the temperature in your collection of IPAs. State your answer to three decimal places.

Solution:

a. We use a t-distribution since the sample is normal and small. The general confidence interval is of the form

$$\left[\overline{X} - \frac{t_{\alpha/2,n-1}S}{\sqrt{n}}, \overline{X} + \frac{t_{\alpha/2,n-1}S}{\sqrt{n}}\right]$$

In our case we have

$$\overline{X} = 48, \quad S = 3, \quad \alpha = 0.01, \quad n = 12.$$

Using the t-table for $n - 1 = 11$ degrees of freedom gives $t_{0.005,11} = 3.106$. Therefore the 99% confidence interval is

$$\left[48 - \frac{3.106(3)}{\sqrt{12}}, 48 + \frac{3.106(3)}{\sqrt{12}}\right] \approx [45.3101, 50.6899]$$

b. For the variance, the general form for the confidence interval is

$$\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right]$$

Looking up the χ^2 values for $\alpha = 0.2$ and $n = 12$ gives

$$\chi^2_{0.1,11} = 17.275, \quad \chi^2_{0.9,11} = 5.578.$$

Therefore a 80% confidence interval is

$$\left[\frac{11(9)}{17.275}, \frac{11(9)}{5.578}\right] \approx [5.7308, 17.7483].$$

c. Note that the t-distribution with $n-1$ degrees of freedom is given by

$$T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$$

when $n = 12$ we have

$$T = \frac{\overline{X}-\mu}{S/\sqrt{12}}$$

is a t-distribution with 11 degrees of freedom. Therefore

$$P(\overline{X}-\mu \leq aS) = P\left(T \leq a\sqrt{12}\right) = F_{T(11)}(a\sqrt{12}) = 0.01$$

It follows that

$$a = \frac{1}{\sqrt{12}} F^{-1}_{T(11)}(0.01) = -t_{0.02,11}/\sqrt{12} \approx -0.7846.$$

Here we used symmetry of the t-distribution

$$F^{-1}_{T(11)}(0.01) = -F^{-1}_{T(11)}(1-0.01) = -t_{0.02,11} = -2.718.$$

Problem 5 (Study Buddies) (15 pts) Two students, conspicuously named I and II, are studying for an exam. Let X_1 and X_2 be the proportion of time that I and II spend studying during the exam "study week". Assume the joint PDF of X_1 and X_2 is given by

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} x_1 + x_2, & 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

a. (3 pts) Find the marginal densities of X_1 and X_2.

b. (3 pts) These two students are roommates. Does the amount of time one of them studies affect the amount of time the other does? Namely, are X_1 and X_2 independent or dependent?

c. (4 pts) Find $P(X_1 \geq 1/2 | X_2 \geq 1/2)$.

d. (5 pts) Find $\rho(X_1, X_2)$. Interpret your answer. Are the roommates a good or bad influence on each other?

Solution:

a. The marginals are given by

$$f_{X_1}(x_1) = \int_0^1 (x_1+x_2)\mathrm{d}x_2 = x_1 + \frac{1}{2},$$

and

$$f_{X_2}(x_2) = \int_0^1 (x_1+x_2)\mathrm{d}x_1 = x_2 + \frac{1}{2}.$$

b. They are not independent since

$$f_{X_1,X_2}(x_1,x_2) = x_1 + x_2 \neq \left(x_1 + \frac{1}{2}\right)\left(x_2 + \frac{1}{2}\right) = f_{X_1}(x_1) f_{X_2}(x_2)$$

which are not equal (for instance take $x_1 = x_2 = 0$).

c.

$$\begin{aligned}
P(X_1 \geq 1/2 | X_2 \geq 1/2) &= \frac{P(X_1 \geq 1/2, X_2 \geq 1/2)}{P(X_2 \geq 1/2)} \\
&= \frac{\int_{1/2}^{1}\int_{1/2}^{1}(x_1+x_2)\mathrm{d}x_1\mathrm{d}x_2}{\int_{1/2}^{1} x_2 + \frac{1}{2}\mathrm{d}x_2} \\
&= \frac{\frac{1}{4}(x_1^2|_{1/2}^{1} + x_2^2|_{1/2}^{1})}{\frac{1}{2}\left(x_2^2|_{1/2}^{1} + 1/2\right)} \\
&= \frac{1-\frac{1}{4}}{1+\frac{1}{4}} = \frac{3}{5}.
\end{aligned}$$

d. To compute $\rho(X_1, X_2)$, we first compute the variances

$$\begin{aligned}
\mathrm{Var}(X_1) &= EX_1^2 - (EX_1)^2 = \int_0^1 x_1^2(x_1 + 1/2)\mathrm{d}x_1 - \left(\int_0^1 x_1(x_1+1/2)\mathrm{d}x_1\right)^2 \\
&= \frac{1}{4} + \frac{1}{6} - \left(\frac{1}{3} + \frac{1}{4}\right)^2 \\
&= \frac{11}{144}
\end{aligned}$$

Since X_2 has the same marginal distribution as X_1 we also have $\mathrm{Var}(X_2) = \frac{11}{144}$. Next we compute the covariance

$$\begin{aligned}
\mathrm{Cov}(X_1,X_2) &= E(X_1X_2) - (EX_1)(EX_2) \\
&= \int_0^1\int_0^1 x_1x_2(x_1+x_2)\mathrm{d}x_1\mathrm{d}x_2 - \left(\int_0^1 x_1\left(x_1+\tfrac{1}{2}\right)\mathrm{d}x_1\right)\left(\int_0^1 x_2\left(x_2+\tfrac{1}{2}\right)\mathrm{d}x_2\right) \\
&= \frac{1}{3} - \left(\frac{1}{3}+\frac{1}{4}\right)^2 = \frac{-1}{144}
\end{aligned}$$

Therefore the correlation is given by

$$\rho(X_1,X_2) = \frac{\mathrm{Cov}(X_1,X_2)}{\sqrt{\mathrm{Var}(X_1)}\sqrt{\mathrm{Var}(X_2)}} = \frac{\frac{-1}{144}}{\frac{11}{144}} = \frac{-1}{11}$$

Since the correlation is negative, it appears that the roommates are not a good influence on each other (although the effect is weak).

Problem 6 (Light bulbs) (12 pts) The lifetimes of two light bulbs are given by two independent Exponential(λ) random variables X_1 and X_2.

a. (3 pts) What is the joint PDF of X_1 and X_2?

b. (4 pts) What is the $P(X_2 \le X_1)$?

c. (5 pts) Let $Y = X_2/X_1$, what is the CDF and PDF of Y? Does it depend on λ?

Solution:

a. Since X_1 and X_2 are independent the joint PDF is given by

$$f_{X_1,X_2}(x_1, x_2) = \begin{cases} \lambda^2 e^{-\lambda(x_1+x_2)} & 0 \le x_1, x_2 < \infty \\ 0 & \text{otherwise} \end{cases}.$$

b. We calculate the probability by

$$\begin{aligned} P(X_2 \le X_1) &= \iint_{\{x_2 \le x_1\}} \lambda^2 e^{-\lambda(x_1+x_2)} \mathrm{d}x_1 \mathrm{d}x_2 \\ &= \int_0^\infty \int_0^{x_1} \lambda^2 e^{-\lambda(x_1+x_2)} \mathrm{d}x_2 \mathrm{d}x_1 \\ &= \int_0^\infty \lambda e^{-\lambda x_1}(1 - e^{-\lambda x_1}) \mathrm{d}x_1 \\ &= 1 - \frac{1}{2} = \frac{1}{2} \end{aligned}$$

c. To determine the distribution of $Y = X_2/X_1$, we compute the CDF

$$F_Y(y) = P(Y \le y) = P(X_2 \le yX_1) = \int_0^\infty \int_0^{yx_1} \lambda^2 e^{-\lambda(x_1+x_2)} \mathrm{d}x_2 \mathrm{d}x_1$$
$$= \int_0^\infty \lambda e^{-\lambda}(1 - e^{-\lambda y x_1}) \mathrm{d}x_1 = 1 - \frac{1}{y+1} = \frac{y}{1+y}.$$

Therefore

$$F_Y(y) = \begin{cases} 0 & y \le 0 \\ \frac{y}{1+y} & 0 \le y < \infty \end{cases}$$

The PDF is therefore

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \begin{cases} \frac{1}{(1+y)^2} & 0 \le y < \infty \\ 0 & \text{otherwise} \end{cases}.$$

It is easy to see that this distribution doesn't depend on λ.

Problem 7 (Estimators) (13 pts) Let $Y_1, Y_2, \ldots Y_n$ be a random sample from a Poisson(θ) distribution.

a. (5 pts) Find the maximum likelihood estimator $\hat{\Theta}_{MLE}$ for θ.

b. (3 pts) Show that $\hat{\Theta}_{MLE}$ is an unbiased and consistent estimator for θ.

c. (5 pts) Show that the sample variance S^2 is also an unbiased and consistent estimator for θ.

Solution:

a. The likelihood function for a given sample $y_1, y_2, \ldots, y_n$ is given by

$$L(\theta) = \prod_{i=1}^{n} \frac{\theta^{y_i} e^{-\theta}}{y_i!}$$

Taking the logarithm gives

$$\ln(L(\theta)) = \sum_{i=1}^{n} y_i \ln(\theta) - \theta - \ln(y_i!)$$

and the derivative is

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \ln(L(\theta)) = \sum_{i=1}^{n} y_i \frac{1}{\theta} - n = 0$$

Therefore the maximum is achieved at

$$\theta = \frac{1}{n} \sum_{i=1}^{n} y_i,$$

and so the maximum likelihood estimator $\hat{\Theta}_{MLE}$ is given by

$$\hat{\Theta}_{MLE} = \overline{Y}$$

b. $\overline{Y}$ is unbiased since it is the sample mean, and it is consistent by the WLLN (or since all MLEs are consistent)

c. The sample variance is given by

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

Easy way We know from class that this is an unbiased and consistent estimator for the variance of any distribution (as long as the variance is finite). This means that since $\mathrm{Var}(Y) = \theta$ for the Poisson distribution that S^2 is consistent and unbiased.

Harder way If you want to show this explicitly, we first write the sample variance as

$$S^2 = \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2\right]$$

Taking expected value gives

$$\begin{aligned} ES^2 &= \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n} EY_i^2 - E\overline{Y}^2\right] = \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n}(\mathrm{Var}(Y_i)+\theta^2) - \mathrm{Var}(\overline{X}) - \theta^2\right] \\ &= \frac{n}{n-1}\left(\theta - \frac{\theta}{n}\right) = \theta \end{aligned}$$

Similarly to show consistency, we note that

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \frac{n}{n-1}\left(\overline{Y^2} - \overline{Y}\right)$$

by the WLLN (since the variance of Y_i and Y_i^2 is finite), we have

$$\overline{Y^2} \xrightarrow{p} E[Y^2] = \theta + \theta^2$$

and similarly

$$\overline{Y} \xrightarrow{p} E[Y] = \theta.$$

Using the continuous mapping theorem (for $\overline{Y}^2$) and the fact that the sum of two random variables converging in probability converge again in probability and that $\frac{n}{n-1} \to 1$ gives

$$S^2 \xrightarrow{p} \left(\theta + \theta^2 - \theta^2\right) = \theta.$$

Therefore S^2 is consistent.

Problem 8 (Strings) (11 pts) Let X be the number of occurrences of the string 'EXAM' in a random string of length 10 (there are 26 letters in the alphabet). 'EXAM' needs to appear as 4 consecutive letters like in TMDEXAMOPA.

a. (3 pts) How many strings are there with exactly 10 letters?

b. (3 pts) What is the probability of the string 'EXAM' appearing in the first four letters?

c. (5 pts) What is EX, the expected number occurrences of the string 'EXAM'?

Solution:

a. There are $26^1 0$ strings length 10

b. There are 26^6 strings with exam appearing in the first four letters and therefore the probability is

$$\frac{26^6}{26^{10}} = \frac{1}{26^4}.$$

c. Let X_i be a Bernoulli random variable defined by

$$X_i = \begin{cases} 1 & \text{if "EXAM" appears with letter "E" i starting in the ith place of the string} \\ 0 & \text{otherwise} \end{cases}$$

and we have $EX_i = \frac{1}{26^4}$. Note that i ranges from $i = 1$ to $i = 10 - 3 = 7$. So that the total number of occurances of "EXAM" is

$$X = \sum_{i=1}^{7} X_i.$$

Additionally, X_i cannot be 1 for all i since if "EXAM" appears, the the next three letters are occupied. Therefore X_i are not independent. However by linearity of expectation

$$EX = \sum_{i=1}^{7} EX_i = \frac{7}{26^4}.$$

Problem 9 (CLTree) (11 pts) You own a struggling tree farm with 100 trees ready to be sold for the year. You found that the heights of the trees (in feet) are independent Exponential(1/6) random variables $X_1, X_2, \ldots X_{100}$. You sell each tree at a price \$10 per foot. Each tree costs you \$57 (to grow and maintain) for the year.

a. (3 pts) Assuming you sell all 100 trees, give an expression I for the total dollar amount earned per tree.

b. (3 pts) Find $E[I]$ and $\mathrm{Var}(I)$. Is the average dollar earned per tree more than your yearly cost per tree?

c. (5 pts) Use the central limit theorem to estimate the probability that you lose money this year.

Solution:

a. The total dollar amount per tree is

$$I = \frac{1}{100} \sum_{i=1}^{100} 10 X_i = \frac{1}{10} \sum_{i=1}^{100} X_i.$$

b. Since $X_i \sim \text{Exponential}(1/6)$, $EX_i = 6$ and $\text{Var}(X_i) = 36$. Therefore

$$EI = \frac{1}{10}\sum_{i=1}^{100} EX_i = \frac{100}{10}6 = 60.$$

and by independence

$$\text{Var}(I) = \frac{1}{100}\text{Var}\left(\sum_{i=1}^{100} X_i\right) = \frac{100}{100}\text{Var}(X_i) = 36$$

Since $60 > 57$, you earn more money than it costs to grow on average.

c. The probability that you lose money this year is

$$P\left(I < 57\right).$$

To apply the central limit theorem we first standardize

$$\overline{Z} = \frac{I - 60}{6}$$

and since $\overline{Z}$ is approximately $N(0,1)$, we deduce

$$P(I < 57) = P\left(\overline{Z} < \frac{57-60}{6} = -1/2\right) \approx \Phi(-1/2) \approx 0.3085.$$

Therefore you have roughly a 31% chance of losing money this year.