APMA 1650

Homework 8 - Solutions

Problem 1. (20 pts)

a. (10 pts) Let $X \sim \text{Gamma}(n, \lambda)$ and $Y \sim \text{Exponential}(\lambda)$ be independent. Use the convolution to show that

$$Z = X + Y \sim \text{Gamma}(n+1, \lambda).$$

b. (10 pts) Let $X \sim \text{Pascal}(n, p)$ and $Y \sim \text{Geometric}(p)$ be independent. Use the convolution to show that

$$Z = X + Y \sim \text{Pascal}(n+1, p).$$

Hint: You may find useful the following "hockey stick" identity

$$\sum_{k=n}^{m} \binom{k}{n} = \binom{m+1}{n+1}$$

Solution:

a. The PDF of Z is given by the convolution

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(z-x) f_X(x) \mathrm{d}x = \int_0^z f_Y(z-x) f_X(x)$$

Note that the bounds are actually from 0 to z since $f_X(x) = 0$ when $x \leq 0$ and $f_Y(z-x) = 0$ when $x \geq z$. Using the PDFs

$$f_X(x) = \frac{1}{\Gamma(n)} \lambda^n x^{n-1} e^{-\lambda x}, \quad f_Y(y) = \lambda e^{-\lambda x},$$

gives for $z \in [0, \infty)$

$$\begin{aligned} f_Z(z) &= \frac{1}{\Gamma(n)} \lambda^{n+1} \int_0^z x^{n-1} e^{-\lambda(z-x)} e^{-\lambda x} \mathrm{d}x \\ &= \frac{1}{\Gamma(n)} \lambda^{n+1} e^{-\lambda z} \int_0^z x^{n-1} \mathrm{d}x \\ &= \frac{1}{\Gamma(n) n} \lambda^{n+1} z^n e^{-\lambda z} \\ &= \frac{1}{\Gamma(n+1)} \lambda^{n+1} z^n e^{-\lambda z}, \quad \text{where we used} \quad \Gamma(n+1) = \Gamma(n) n. \end{aligned}$$

Therefore $Z \sim \Gamma(n+1, \lambda)$.

b. Note that the PMF of $Z = X + Y$ is given by the convolution of the two distributions for each $z \in R_Z = \{n+1, \ldots\}$

$$P_Z(z) = \sum_{x \in R_X} P_Y(z-x)P_X(x) = \sum_{x=n}^{z-1} P_Y(z-x)P_X(x).$$

Note that the sum is from n to $z-1$ since $P_Y(z-x) = 0$ when $x > z-1$ and $P_X(x) = 0$ when $x < n$. Using the definitions

$$P_X(x) = \binom{x-1}{n-1} p^{n-1} q^{x-n}, \quad P_Y(y) = pq^{y-1},$$

gives

$$\begin{aligned}
P_Z(z) &= \sum_{x=m}^{z-1} pq^{z-x-1} \binom{x-1}{n-1} p^{n-1} q^{x-n} \\
&= p^n q^{z-(n+1)} \sum_{x=n}^{z-1} \binom{x-1}{n-1} \\
&= \binom{z-1}{n} p^n q^{z-(n+1)} \quad \text{by the hockey stick identity.}
\end{aligned}$$

Therefore $Z \sim \text{Pascal}(n+1, p)$.

Problem 2. (50 pts) A certain hospital is stocking COVID vaccines at the beginning of the week. The hospital has only a certain amount of freezer space to store the vaccine and the supply varies from week to week. Let Y_1 denote the supply of vaccines at the beginning of the week, measured in terms of the proportion of the capacity of the freezer (assumed to be a number from 0 to 1). Assume the PDF of Y_1 is given by

$$f_{Y_1}(y_1) = \begin{cases} 2y_1 & 0 \le y_1 \le 1 \\ 0 & \text{otherwise} \end{cases}.$$

Let Y_2 be the amount of vaccines actually administered during the week (measured in terms of proportion of freezer capacity). Assume that the conditional PDF for Y_2 given Y_1 is given for each $0 < y_1 \le 1$ by

$$f_{Y_2|Y_1}(y_2|y_1) = \begin{cases} \frac{3y_2^2}{y_1^3} & 0 \le y_2 \le y_1 \\ 0 & \text{otherwise} \end{cases}.$$

a. (10 pts) Find EY_1 and EY_2

b. (10 pts) Find $\text{Var}(Y_1)$ and $\text{Var}(Y_2)$.

c. (20 pts) Find the covariance and correlation of Y_1 and Y_2. Explain in words what the sign your correlation means here.

d. (10 pts) The amount of surplus vaccines at the end of the week is $Y_2 - Y_1$. Find $E(Y_1 - Y_2)$ and $\mathrm{Var}(Y_1 - Y_2)$ using your answers to parts a-c.

Solution: First lets note that the joint PDF is given by

$$f_{Y_1,Y_2}(y_1, y_2) = f_{Y_2|Y_1}(y_2|y_1) f_{Y_1}(y_1) = \begin{cases} 6\frac{y_2^2}{y_1^2} & 0 \le y_1 \le 1,\ 0 \le y_2 \le y_1 \\ 0 & \text{otherwise.} \end{cases}$$

a. To compute the expectations we have

$$EY_1 = \int_0^1 2y_1^2 \mathrm{d}y_1 = \frac{2}{3},$$

and

$$EY_2 = \int_0^1 \int_0^{y_1} 6\frac{y_2^3}{y_1^2}\,\mathrm{d}y_2\mathrm{d}y_1 = \int_0^1 \frac{3}{2} y_1^2 \mathrm{d}y_1 = \frac{1}{2}.$$

b. Similarly for the variance

$$EY_1^2 = \int_0^1 2y_1^3 \mathrm{d}y_1 = \frac{1}{2},$$

therefore

$$\mathrm{Var}(Y_1) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}.$$

Similarly

$$EY_2^2 = \int_0^1 \int_0^{y_1} \frac{6y_2^4}{y_1^2} \mathrm{d}y_2\mathrm{d}y_1 = \int_0^1 \frac{6}{5} y_1^3 \mathrm{d}y_1 = \frac{3}{10},$$

therefore

$$\mathrm{Var}(Y_2) = \frac{3}{10} - \left(\frac{1}{2}\right)^2 = \frac{1}{20}$$

c. To find the covariance, we use the formula

$$\mathrm{Cov}(Y_1, Y_2) = E(Y_1Y_2) - (EY_1)(EY_2).$$

First we compute

$$E(Y_1Y_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 y_2 f_{Y_1,Y_2}(y_1, y_2) \mathrm{d}y_1\mathrm{d}y_2 = 6 \int_0^1 \int_0^{y_1} \frac{y_2^3}{y_1} \mathrm{d}y_2\mathrm{d}y_1 = \frac{3}{4} \int_0^1 y_1^3 \mathrm{d}y_1 = \frac{3}{8}.$$

Therefore

$$\mathrm{Cov}(Y_1, Y_2) = \frac{3}{8} - \left(\frac{2}{3}\right)\left(\frac{1}{2}\right) = \frac{1}{24}.$$

Likewise the correlation is given by

$$\rho(Y_1, Y_2) = \frac{\mathrm{Cov}(Y_1, Y_2)}{\sqrt{\mathrm{Var}(Y_1)\mathrm{Var}(Y_2)}} = \frac{\frac{1}{24}}{\sqrt{\left(\frac{3}{10}\right)\left(\frac{1}{20}\right)}} = \frac{5}{6\sqrt{6}} \approx 0.3402.$$

Note that the correlation is positive but not that close to 1. This means there is a somewhat positive linear relationship between Y_1 and Y_2.

d. The average surplus is by linearity of expectation

$$E(Y_1 - Y_2) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}$$

While the variance is given using the properties of covariance

$$\begin{aligned}\text{Var}(Y_1 - Y_2) &= \text{Var}(Y_1) + \text{Var}(Y_2) - 2\text{Cov}(Y_1, Y_2) \\ &= \frac{1}{18} + \frac{1}{20} - 2\left(\frac{1}{24}\right) \\ &= \frac{1}{45}.\end{aligned}$$

Problem 3. (30 pts) The average weight of a golden retriever is 70 pounds.

a. (10 pts) Give an upper bound on the probability that a certain golden retriever is at least 80 pounds.

b. (10 pts) Suppose you know the standard deviation for the weight distribution is 5 pounds. Find a lower bound on the probability that a certain golden retriever is between 82 and 58 pounds.

c. (10 pts) Suppose you know that the weight distribution is normal (with the same mean and standard deviation given above). Repeat part b using the CDF calculator for the normal distribution. How close was your estimate in part b?

Solution:

a. To give an upper bound, we use the fact that weight is non-negative and use Markov's inequality

$$P(X \geq 80) \leq \frac{EX}{80} = \frac{70}{80} = \frac{7}{8}.$$

b. Now that we know the standard deviation, we can use Chebyshev

$$\begin{aligned}P(58 \leq X \leq 82) = P(-12 \leq X - 70 \leq 12) &= P(|X - 70| \leq 12) \\ \geq 1 - \frac{5^2}{12^2} = 1 - \frac{25}{144} &=\approx 0.8264.\end{aligned}$$

c. If the distribution is normal, we see that

$$Z = \frac{x - \mu_X}{\sigma_X} = \frac{X - 70}{5} \sim N(0, 1)$$

and therefore

$$P(58 \leq X \leq 82) = P\left(-\frac{12}{5} \leq Z \leq \frac{12}{5}\right) = 2\Phi(12/5) - 1 \approx 2(0.9918) - 1 = 0.9836$$

This is a good amound bigger than the probability that you get from Chebyshev, but Chebyshev wasn't too bad.