Introduction to Statistics I: Homework 8

Due on December 7, 2025 at 23:59

Liliana Pazdan-Siudeja

Hashem A. Damrah

UO ID: 952102243

Problem 1. The joint probability density function of X and Y is given by

$$f(x,y) = \begin{cases} \frac{6}{7}\left(x^2 + \frac{xy}{2}\right) & 0 < x < 1, 0 < y < 2 \\ 0 & \text{otherwise} \end{cases}.$$

(i) Verify that f is a joint density function.

(ii) Compute the density function of X.

(iii) Find $P\{X > Y\}$.

(iv) Find $P\{Y > 1/2 | X < 1/2\}$.

(v) Find $E[X]$.

Solution to (i). If f is a valid joint density function, then the double integral of $f(x,y)$ over its entire support must equal 1. Thus, we get

$$\iint_R f(x,y)\, \mathrm{d}A = \int_0^1 \int_0^2 \frac{6}{7}\left(x^2 + \frac{xy}{2}\right)\, \mathrm{d}y\, \mathrm{d}x$$

$$= \frac{6}{7}\int_0^1 x^2 y + \frac{xy^2}{4}\Big|_{y=0}^{y=2}\, \mathrm{d}x$$

$$= \frac{6}{7}\int_0^1 2x^2 + x\, \mathrm{d}x$$

$$= \frac{6}{7}\left[\frac{2x^3}{3} + \frac{x^2}{2}\right]\Big|_0^1$$

$$= \frac{6}{7}\left(\frac{2}{3} + \frac{1}{2}\right) = \frac{6}{7} \cdot \frac{7}{6} = 1.$$

Thus, $f(x,y)$ is a valid joint density function. □

Solution to (ii). The marginal density function of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, \mathrm{d}y$$

$$= \int_0^2 \frac{6}{7}\left(x^2 + \frac{xy}{2}\right)\, \mathrm{d}y$$

$$= \frac{6}{7}\left(x^2 y + \frac{xy^2}{4}\right)\Big|_{y=0}^{y=2}$$

$$= \frac{6}{7}(2x^2 + x) = \frac{6}{7}x(2x + 1).$$

for $0 < x < 1$, and $f_X(x) = 0$ otherwise. □

Solution to (iii). The probability of $P\{X > Y\}$ is given by

$$P\{X > Y\} = \iint_{x > y} f(x,y)\, \mathrm{d}A$$

$$= \int_0^1 \int_0^x \frac{6}{7}\left(x^2 + \frac{xy}{2}\right)\, \mathrm{d}y\, \mathrm{d}x$$

$$= \frac{6}{7}\int_0^1 \left(x^2 y + \frac{xy^2}{4}\right)\Big|_{y=0}^{y=x}\, \mathrm{d}x$$

$$= \frac{6}{7} \int_0^1 \left(x^3 + \frac{x^3}{4} \right) dx$$

$$= \frac{6}{7} \int_0^1 \frac{5x^3}{4} \, dx = \frac{15}{14} \left. \frac{x^4}{4} \right|_0^1 = \frac{15}{56}. \qquad \square$$

Solution to (iv). Using the definition of conditional probability, we have

$$P\{Y > 1/2 | X < 1/2\} = \frac{P\{Y > 1/2, X < 1/2\}}{P\{X < 1/2\}}.$$

Computing the numerator, we get

$$P\{Y > 1/2, X < 1/2\} = \iint_{Y>1/2, X<1/2} f(x, y) \, dA$$

$$= \int_0^{1/2} \int_{1/2}^2 \frac{6}{7} \left(x^2 + \frac{xy}{2} \right) dy \, dx$$

$$= \frac{6}{7} \int_0^{1/2} x^2 \left(2 - \frac{1}{2} \right) + \frac{x}{2} \left(2 - \frac{1}{8} \right) dx$$

$$= \frac{6}{7} \int_0^{1/2} \frac{3}{2} x^2 + \frac{15}{16} x \, dx$$

$$= \frac{6}{7} \left(\frac{3x^3}{6} + \frac{15x^2}{32} \right) \Big|_0^{1/2} = \frac{6}{7} \left(\frac{1}{16} + \frac{15}{128} \right) = \frac{69}{448}.$$

Now, computing the denominator, we have

$$P\{X < 1/2\} = \int_0^{1/2} f_X(x) \, dx$$

$$= \int_0^{1/2} \frac{6}{7} x(2x + 1) \, dx$$

$$= \frac{6}{7} \int_0^{1/2} (2x^2 + x) \, dx$$

$$= \frac{6}{7} \left(\frac{2x^3}{3} + \frac{x^2}{2} \right) \Big|_0^{1/2} = \frac{6}{7} \left(\frac{1}{12} + \frac{1}{8} \right) = \frac{6}{7} \cdot \frac{5}{24} = \frac{5}{28}.$$

Thus, we have

$$P\{Y > 1/2 | X < 1/2\} = \frac{69/448}{5/28} = \frac{69}{80}. \qquad \square$$

Solution to (v). The expected value of X is given by

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) \, dx$$

$$= \int_0^1 x \cdot \frac{6}{7} x(2x + 1) \, dx$$

$$= \frac{6}{7} \int_0^1 (2x^3 + x^2) \, dx$$

$$= \frac{6}{7} \left(\frac{2x^4}{4} + \frac{x^3}{3} \right) \Big|_0^1 = \frac{6}{7} \left(\frac{1}{2} + \frac{1}{3} \right) = \frac{6}{7} \cdot \frac{5}{6} = \frac{5}{7}. \qquad \square$$

Problem 2. A fair die is rolled 20 times. Calculate the expected value and variance of the sum of the 20 rolls.

Solution. Let X_i be the outcome of the i-th roll, for $i = 1, 2, \ldots, 20$. The sum of the 20 rolls is given by $S = X_1 + X_2 + \cdots + X_{20}$. Since the die is fair, each X_i has an expected value of

$$E[X_i] = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{21}{6} = 3.5.$$

Therefore, the expected value of the sum S is

$$E[S] = E[X_1] + E[X_2] + \cdots + E[X_{20}] = 20 \cdot 3.5 = 70.$$

Next, we compute the variance of each roll. The variance of a fair die roll is given by

$$\text{Var}(X_i) = E[X_i^2] - E^2[X_i].$$

First, we calculate $E[X_i^2]$ to get

$$E[X_i^2] = \frac{1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2}{6} = \frac{91}{6}.$$

Thus, the variance of each roll is

$$\text{Var}(X_i) = \frac{91}{6} - (3.5)^2 = \frac{91}{6} - \frac{49}{4} = \frac{364 - 294}{24} = \frac{70}{24} = \frac{35}{12}.$$

Since the rolls are independent, the variance of the sum S is

$$\text{Var}(S) = \text{Var}(X_1) + \text{Var}(X_2) + \cdots + \text{Var}(X_{20}) = 20 \cdot \frac{35}{12} = \frac{700}{12} = \frac{175}{3}. \qquad \square$$

Problem 3. A total of n balls, numbered 1 through n, are put into n urns, also numbered 1 through n in such a way that ball i is equally likely to go into any of the urn $1, 2, \cdots, i$. Find the expected number of urns that are empty.

    Hint: Define an indicator variable, X_i.

Solution. Let X be the number of empty urns. We can express X as the sum of indicator variables X_i, where $X_i = 1$ if urn i is empty, and $X_i = 0$ otherwise, for $i = 1, 2, \ldots, n$. Thus,

$$X = X_1 + X_2 + \cdots + X_n.$$

By the linearity of expectation, we have

$$E[X] = E[X_1] + E[X_2] + \cdots + E[X_n] = \sum_{i=1}^{n} E[X_i].$$

To compute $E[X_i]$, we need to find the probability that urn i is empty. For urn i to be empty, none of the balls $1, 2, \ldots, i$ can go into urn i. The probability that ball j (where $1 \le j \le i$) does not go into urn i is $1 - 1/j$. Therefore, the probability that urn i is empty is

$$P(X_i = 1) = \prod_{j=i}^{n} 1 - \frac{1}{j} = \prod_{j=i}^{n} \frac{j-1}{j} = \frac{i-1}{n}.$$

Thus,

$$E[X_i] = P(X_i = 1) = \frac{i-1}{n}.$$

Finally, we have

$$E[X] = \sum_{i=1}^{n} \frac{i-1}{n} = \frac{1}{n} \sum_{i=1}^{n} (i-1) = \frac{1}{n} \cdot \frac{(n-1)n}{2} = \frac{n-1}{2}.$$

Thus, the expected number of empty urns is $(n-1)/2$. $\qquad \square$

Problem 4. For a group of 100 people, compute

   (i) the expected number of days of the year that are birthdays of exactly 3 people.

   (ii) the expected number of distinct birthdays.

Solution to (i). Let X be the random variable representing the number of days with exactly 3 birthdays. We can express X as the sum of indicator variables X_i, where $X_i = 1$ if day i has exactly 3 birthdays, and $X_i = 0$ otherwise, for $i = 1, 2, \ldots, 365$. Thus,

$$X = X_1 + X_2 + \cdots + X_{365}.$$

By the linearity of expectation, we have

$$E[X] = E[X_1] + E[X_2] + \cdots + E[X_{365}] = 365 \cdot E[X_1].$$

To compute $E[X_1]$, we need to find the probability that exactly 3 people have their birthday on day 1. The number of ways to choose 3 people from 100 is $\binom{100}{3}$. Each of these chosen people has a $\frac{1}{365}$ chance of having their birthday on day 1, while the remaining 97 people have a $364/365$ chance of not having their birthday on day 1. Therefore,

$$P(X_1 = 1) = \binom{100}{3} \left(\frac{1}{365}\right)^3 \left(\frac{364}{365}\right)^{97}.$$

Thus,

$$E[X_1] = P(X_1 = 1) = \binom{100}{3} \left(\frac{1}{365}\right)^3 \left(\frac{364}{365}\right)^{97}.$$

Finally, we have

$$E[X] = 365 \cdot \binom{100}{3} \left(\frac{1}{365}\right)^3 \left(\frac{364}{365}\right)^{97} \approx 0.9301. \qquad \square$$

Solution to (ii). Again, we can let Y be the random variable representing the number of distinct birthdays. We can express Y as the sum of indicator variables Y_i, where $Y_i = 1$ if day i has at least one birthday, and $Y_i = 0$ otherwise, for $i = 1, 2, \ldots, 365$. Thus,

$$Y = Y_1 + Y_2 + \cdots + Y_{365}.$$

By the linearity of expectation, we have

$$E[Y] = E[Y_1] + E[Y_2] + \cdots + E[Y_{365}] = 365 \cdot E[Y_1].$$

To compute $E[Y_1]$, we need to find the probability that at least one person has their birthday on day 1. The complement of this event is that no one has their birthday on day 1. The probability that a single person does not have their birthday on day 1 is $364/365$. Therefore, the probability that all 100 people do not have their birthday on day 1 is $(364/365)^{100}$. Thus,

$$P(Y_1 = 1) = 1 - \left(\frac{364}{365}\right)^{100}.$$

Thus,

$$E[Y_1] = P(Y_1 = 1) = 1 - \left(\frac{364}{365}\right)^{100}.$$

Finally, we have

$$E[Y] = 365 \cdot \left(1 - \left(\frac{364}{365}\right)^{100}\right) \approx 87.576. \qquad \square$$

Problem 5. The random variables X and Y have a joint probability density function given by

$$f(x,y) = \begin{cases} \dfrac{2e^{-2x}}{x} & 0 \leq x < \infty, 0 \leq y \leq x \\ 0 & \text{otherwise} \end{cases}.$$

Compute $\mathrm{Cov}(X,Y)$.

Solution. Using the equation $\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y]$, we first compute $E[X]$, $E[Y]$, and $E[XY]$. Starting with $E[X]$, we have

$$E[X] = \int_0^\infty \int_0^x x \cdot \frac{2e^{-2x}}{x} \, dy \, dx \qquad\qquad E[Y] = \int_0^\infty \int_0^x y \cdot \frac{2e^{-2x}}{x} \, dy \, dx$$

$$= \int_0^\infty \int_0^x 2e^{-2x} \, dy \, dx \qquad\qquad = \int_0^\infty \frac{2e^{-2x}}{x} \cdot \frac{y^2}{2} \Big|_0^x \, dx$$

$$= \int_0^\infty 2xe^{-2x} \, dx = \frac{1}{2}. \qquad\qquad = \int_0^\infty xe^{-2x} \, dx = \frac{1}{4}.$$

$$E[XY] = \int_0^\infty \int_0^x xy \cdot \frac{2e^{-2x}}{x} \, dy \, dx$$

$$= \int_0^\infty 2e^{-2x} \cdot \frac{y^2}{2} \Big|_0^x \, dx$$

$$= \int_0^\infty x^2 e^{-2x} \, dx = \frac{1}{4}.$$

Thus, we have

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = \frac{1}{4} - \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}. \qquad \square$$

Problem 6. If X_1, X_2, X_3, and X_4 are (pairwise) uncorrelated random variables, each having mean 0 and variance 1, compute the correlation of

(i) $X_1 + X_2$ and $X_2 + X_3$.

(ii) $X_1 + X_2$ and $X_3 + X_4$.

Solution to (i). The correlation between two random variables A and B is given by

$$\rho(A,B) = \frac{\mathrm{Cov}(A,B)}{\sqrt{\mathrm{Var}(A)\,\mathrm{Var}(B)}}.$$

Letting $A = X_1 + X_2$ and $B = X_2 + X_3$, we first compute $\mathrm{Cov}(A,B)$,

$$\mathrm{Cov}(A,B) = \mathrm{Cov}(X_1 + X_2, X_2 + X_3)$$

$$= \mathrm{Cov}(X_1, X_2) + \mathrm{Cov}(X_1, X_3) + \mathrm{Cov}(X_2, X_2) + \mathrm{Cov}(X_2, X_3)$$

$$= 0 + 0 + 1 + 0 = 1,$$

since they are pairwise uncorrelated. Next, we compute $\mathrm{Var}(A)$ and $\mathrm{Var}(B)$,

$$\mathrm{Var}(A) = \mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = 1 + 1 = 2$$

$$\mathrm{Var}(B) = \mathrm{Var}(X_2 + X_3) = \mathrm{Var}(X_2) + \mathrm{Var}(X_3) = 1 + 1 = 2.$$

Thus, the correlation is

$$\rho(A,B) = \frac{1}{\sqrt{2 \cdot 2}} = \frac{1}{2}. \qquad \square$$

Solution to (ii). Letting $A = X_1 + X_2$ and $B = X_3 + X_4$, we first compute $\text{Cov}(A, B)$,

$$
\begin{aligned}
\text{Cov}(A, B) &= \text{Cov}(X_1 + X_2, X_3 + X_4) \\
&= \text{Cov}(X_1, X_3) + \text{Cov}(X_1, X_4) + \text{Cov}(X_2, X_3) + \text{Cov}(X_2, X_4) \\
&= 0 + 0 + 0 + 0 = 0,
\end{aligned}
$$

again, since they are pairwise uncorrelated. Next, we compute $\text{Var}(A)$ and $\text{Var}(B)$,

$$
\begin{aligned}
\text{Var}(A) &= \text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) = 1 + 1 = 2 \\
\text{Var}(B) &= \text{Var}(X_3 + X_4) = \text{Var}(X_3) + \text{Var}(X_4) = 1 + 1 = 2.
\end{aligned}
$$

Thus, the correlation is

$$
\rho(A, B) = \frac{0}{\sqrt{2 \cdot 2}} = 0. \qquad \square
$$