IPDA1005 Introduction to Probability and Data Analysis

Worksheet 10

Solution

1. A light fixture has two lightbulbs, each of which has a lifetime distribution (measured in thousands of hours) that is independent of the other and that can be described by an exponential distribution with parameter $\lambda = 1$. Let X and Y denote the lifetimes of the first and second bulb, respectively.
   (a) What is the joint pdf of X and Y?
   (b) What is the probability that each bulb lasts at most 1000 h (i.e., $X \leq 1$ and $Y \leq 1$)?

Solution: See Workshop Teaching Week 9 to remind yourself of the properties of the exponential distribution.

(a) Because X and Y are independent, the joint pdf is the product of the two marginal pdfs; hence, $f(x, y) = f_X(x) \cdot f_Y(y) = e^{-x-y}$, $x \geq 0$, $y \geq 0$.

(b) $P(X \leq 1, Y \leq 1) = P(X \leq 1) \cdot P(Y \leq 1) = (1 - e^{-1})(1 - e^{-1}) = 0.4$

2. Two components of a computer have the following joint pdf for their useful lifetimes X and Y (in years):
$$f(x, y) = xe^{-x(1+y)} \qquad x \geq 0 \text{ and } y \geq 0$$
   (a) What is the probability that the lifetime X of the first component exceeds 3?
   (b) What are the marginal pdfs of X and Y? Are the two lifetimes independent? Explain.
   (c) What is the probability that the lifetime of at least one component exceeds 3?

Solution:

(a)
$$P(X > 3) = \int_3^\infty \left(\int_0^\infty xe^{-x(1+y)}\, dy \right) dx = \int_3^\infty e^{-x}\, dx = e^{-3} = 0.05$$

(b) The marginal pdf of X is given by

$$f_X(x) = \int_0^\infty xe^{-x(1+y)}\, dy = e^{-x} \text{ for } x \geq 0$$

The marginal pdf of Y is

$$f_Y(y) = \int_0^\infty xe^{-x(1+y)}\, dx = \frac{1}{(1+y)^2} \text{ for } y \geq 0$$

Clearly, the joint pdf $f(x, y)$ is not a product of the marginal pdfs, so the two random variables are not independent.

(c) The probability that the lifetime of at least one component exceeds 3 is

$$\begin{aligned} P(X > 3 \,\text{or}\, Y > 3) &= 1 - P(X \leq 3, Y \leq 3) \\ &= 1 - \int_0^3 \left(\int_0^3 xe^{-x(1+y)}\, dy \right) dx = 1 - \int_0^3 e^{-x}(1 - e^{-3x})\, dx \\ &= = e^{-3} + \frac{1}{4} - \frac{1}{4}e^{-12} = 0.3 \end{aligned}$$

3. Each front tire of a vehicle is supposed to be filled to a pressure of 26 psi (180 kPa). Suppose the actual air pressure in each tire is a random variable—X for the right tire and Y for the left tire, with joint pdf

$$f(x, y) = \begin{cases} k(x^2 + y^2) & 20 \leq x \leq 30,\ \ 20 \leq y \leq 30 \\ 0 & \text{otherwise} \end{cases}$$

(a) What is the value of k?

(b) What is the probability that both tires are underfilled?

(c) Determine the (marginal) distribution of air pressure in the right tire alone.

(d) Are X and Y independent random variables?

Solution:

(a) To find the value of k, we integrate the joint pdf over the rectangular region specified above, equate it to one, and then solve for k.

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy \\
&= \int_{20}^{30}\int_{20}^{30} k(x^2+y^2)\,dx\,dy \\
&= k\int_{20}^{30}\int_{20}^{30} x^2\,dx\,dy + k\int_{20}^{30}\int_{20}^{30} y^2\,dx\,dy \\
&= 10k\int_{20}^{30} x^2\,dx + 10k\int_{20}^{30} y^2\,dy \\
&= 20k\left(\frac{19000}{3}\right) \implies k = \frac{3}{380000}
\end{aligned}$$

(b)

$$\begin{aligned}
P(X<26, Y<26) &= \int_{20}^{26}\int_{20}^{26} k(x^2+y^2)\,dx\,dy \\
&= k\int_{20}^{26}\left[x^2y+\frac{y^3}{3}\right]_{20}^{26} dx \\
&= k\int_{20}^{26}(6x^2+3192)\,dx \\
&= k\cdot(38304) \\
&= 0.3024
\end{aligned}$$

(c)

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f(x,y)\,dx\,dy \\
&= \int_{20}^{30} k(x^2+y^2)\,dy \\
&= k\left[x^2y+\frac{y^3}{3}\right]_{20}^{30} \\
&= 10kx^2+0.05
\end{aligned}$$

(d) We can obtain $f_Y(y)$ by substituting y for x above; clearly, $f(x,y) \neq f(x)\cdot f(y)$, and hence X and Y are not independent.

4. Annie and Alvie have agreed to meet between 6:00 and 7:00 p.m. for dinner at a local health-food restaurant. Let X be Annie's arrival time and Y be Alvie's arrival time. Suppose X and Y are independent and are each uniformly distributed on the interval $[6,7]$.

(a) What is joint pdf of X and Y?

(b) What is the probability that they both arrive between 6:15 and 6:45 p.m.?

(c) If the first one to arrive will wait only 10 minutes before leaving to eat elsewhere, what is the probability that they'll end up having dinner at the health-food restaurant? [**Hint**: The event of interest is $A = \{(x, y) : |x - y| \leq 1/6\}$. Sketch out this region and calculate its area—it should be straightforward.]

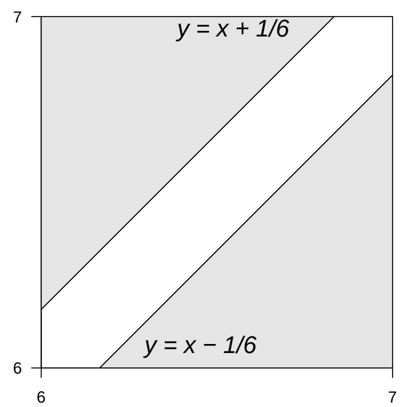
Solution:

(a) Since $f_X(x) = 1/(7-6) = 1$ for $6 \leq x \leq 7$, similarly $f_Y(y) = 1$ for $6 \leq x \leq 7$. Because X and Y are independent,

$$f(x, y) = f_X(x) \cdot f_Y(y) = \begin{cases} 1 & 6 \leq x \leq 7, \ \ 6 \leq y \leq 7 \\ 0 & \text{otherwise} \end{cases}$$

(b) By independence, $P(6.25 \leq X \leq 6.75, 6.25 \leq Y \leq 6.75) = P(6.25 \leq X \leq 6.75) \cdot P(6.25 \leq Y \leq 6.75)$, and hence $P(6.25 \leq X \leq 6.75, 6.25 \leq Y \leq 6.75) = (0.5)(0.5) = 0.25$.

(c) The region A is shown as the white diagonal stripe below.



Thus, $P((X, Y) \in A) = \iint_A 1\, dx\, dy = 1 - (\text{shaded areas}) = 1 - 2\left(\frac{1}{2} \cdot \frac{5}{6} \cdot \frac{5}{6}\right) = \frac{11}{36}$.

5. You and a friend have agreed to meet for lunch between noon (0:00 p.m.) and 1:00 p.m. Denote your arrival time by X, your friend's by Y, and suppose X and Y are independent with pdfs

$$\begin{aligned} f_X(x) &= 3x^2 & 0 \leq x \leq 1 \\ f_Y(y) &= 2y & 0 \leq y \leq 1 \end{aligned}$$

What is the expected amount of time that the one who arrives first must wait for the other person? [Hint: You're being asked to find the expected value of some function of X and Y. First work out what this function is.]

Solution: If you arrive first, the time you have to wait is $Y - X$, but if your friend arrives first, the time s/he has to wait is $X - Y$. Thus, the amount of time the first person has to wait for the second is $h(X, Y) = |X - Y|$. We are told that X and Y are independent, so the joint distribution is given by

$$f(x, y) = f_X(x) \cdot f_Y(y) = (3x^2)(2y) = 6x^2 y$$

Hence, the expected waiting time is

$$\begin{aligned} E[h(X, Y)] &= \int_0^1 \int_0^1 |x - y| \cdot 6x^2 y \, dy \, dx \\ &= \int_0^1 \int_0^x (x - y) \cdot 6x^2 y \, dy \, dx + \int_0^1 \int_x^1 (y - x) \cdot 6x^2 y \, dy \, dx \\ &= \frac{1}{6} + \frac{1}{12} = \frac{1}{4} \text{ hours, or 15 minutes} \end{aligned}$$

6. A surveyor wishes to lay out a square region with each side having length L. However, because of measurement error, she instead lays out a rectangle in which the north–south sides both have length X and the east–west sides both have length Y. Suppose that X and Y are independent and that each is uniformly distributed on the interval $[L-A, L+A]$ (where $0 < A < L$). What is the expected area of the resulting rectangle? [Hint: This questions is easier than it first appears.]

Solution: Both X and Y are uniform on $[L - A, L + A]$, and hence their expected values are $E(X) = E(Y) = L$. Furthermore, X and Y are independent, and so the expected area of the resulting rectangle is $E(XY) = E(X) \cdot E(Y) = L^2$.

7. Suppose that X and Y represent the proportion of marks that students obtained in Test 1 and Test 2, respectively. Their joint distribution is given by

$$f(x, y) = \frac{12}{7} x(x + y) \qquad 0 \le x \le 1, \qquad 0 \le y \le 1$$

(See the Lecture 12 slides for a plot of the $f(x, y)$ surface.)

(a) Calculate $E(X)$ and $E(Y)$. What do you need to obtain first in order to calculate these quantities?

(b) Calculate $E(XY)$ and then $\text{Cov}(X, Y)$. What information does the sign of the covariance provide? Comment on the *magnitude* of the covariance—does it provide any information?

(c) Calculate the correlation coefficient ρ_{XY} between the two sets of scores. What information does it provide beyond what the covariance provides?

(d) What is $E(X+Y)$, the expected value of the sum of the two proportions?

Solution:

(a) Since $E(X) = \int_0^1 x f_X(x)\,dx$, where $f_X(x)$ is the marginal distribution of X (and similarly for Y) we need to first calculate $f_X(x)$ and $f_Y(y)$. It is straightforward to show that $f_X(x) = \int_0^1 f(x,y)\,dy = (12x^2+6x)/7$ and that $f_Y(y) = \int_0^1 f(x,y)\,dx = (4+6y)/7$. Hence,

$$\begin{aligned} E(X) &= \int_0^1 \frac{x}{7}(12x^2+6x)\,dx \\ &= \frac{1}{7}\left[3x^4+2x^3\right]_0^1 \\ &= \frac{5}{7} \end{aligned}$$

and in the same way, we can show that $E(Y) = \int_0^1 y f_Y(y)\,dy = 4/7$.

(b) Recall that $\mathrm{Cov}(X,Y) = E(XY) - E(X)\cdot E(Y)$. Thus,

$$\begin{aligned} E(XY) &= \frac{12}{7}\int_0^1\int_0^1 xy(x^2+xy)\,dx\,dy \\ &= \frac{12}{7}\int_0^1 \left[\frac{x^4}{4}y + \frac{x^3}{3}y^2\right]_0^1 dy \\ &= \frac{12}{7}\int_0^1 \left(\frac{y}{4}+\frac{y^2}{3}\right) dy \\ &= \ldots \\ &= \frac{17}{42} \end{aligned}$$

and hence,

$$\mathrm{Cov}(X,Y) = \frac{17}{42} - \left(\frac{5}{7}\cdot\frac{4}{7}\right) = -\frac{1}{294}$$

The covariance is only slightly negative, but the magnitude carries little information because it is not scale-free. For example, had we expressed the scores on the test out of 10 or 100 instead of 1, the magnitude of the covariance would be rather different.

(c) Using expressions we have come across earlier, it is straightforward (if tedious!) to show that $\mathrm{Var}(X) = 23/490$ and $\mathrm{Var}(Y) = 23/294$. Hence.

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\cdot\sqrt{\mathrm{Var}(Y)}} = \frac{-1/294}{\sqrt{23/490}\cdot\sqrt{23/294}} = -\frac{1}{23}\cdot\sqrt{\frac{5}{3}} = -0.05613$$

In addition to the sign of the relationship, the correlation coefficient also provides a measure of the strength of the relationship, and it is independent of the units of measurement of X and Y.

(d) To calculate $E(X+Y)$, we could evaluate the integral

$$E(X+Y) = \int_0^1 \int_0^1 (x+y) f(x,y)\, dx\, dy$$

but it is straightforward to show that this reduces to $E(X+Y) = E(X) + E(Y) = 4/7 + 5/7 = 9/7$.

(*Sources: All questions adapted from Devore and Berk (2012) and Larsen and Marx (2014).*)

Bibliography

1. Devore, J.L. and Berk, K.N. (2012) *Modern Mathematical Statistics with Applications.* Springer: New York.

2. Marx, R.J. and Marx, M.L. (2014) *An Introduction to Mathematical Statistics and Its Applications*, Fifth Edition. Pearson Education, Inc.: Boston, MA.