18.05 Problem Set 5, Spring 2022 Solutions

Problem 1. (35: 5,10,5,10,5 pts.) **Aching joints**
Suppose X and Y have joint pdf $f(x,y) = c(x^2 + xy)$ on $[0,1] \times [0,1]$.

(a) *Find c and the joint cdf $F(x,y)$.*

Solution: We have

$$1 = \int_0^1 \int_0^1 c(x^2 + xy) dy dx = c \int_0^1 x^2 + \frac{x}{2} dx = c\left(\frac{1}{3} + \frac{1}{4}\right) = \frac{7c}{12}.$$

Thus, $\boxed{c = \frac{12}{7}}$. We have:

$$\begin{aligned} F(x,y) = P(X \le x, Y \le y) &= \frac{12}{7} \int_0^x \int_0^y u^2 + uv \, dy \, dx \\ &= \frac{12}{7} \int_0^x u^2 y + \frac{uy^2}{2} \, du \\ &= \boxed{\frac{12}{7}\left(\frac{x^3 y}{3} + \frac{x^2 y^2}{4}\right)}. \end{aligned}$$

(b) *Find the marginal cumulative distribution functions F_X and F_y and the marginal pdf f_X and f_Y.*

Solution: The marginal cdf's are:

$$\begin{aligned} F_X(x) = F(x,1) &= \boxed{\frac{12}{7}\left(\frac{x^3}{3} + \frac{x^2}{4}\right)} \\ F_Y(y) = F(1,y) &= \boxed{\frac{12}{7}\left(\frac{y}{3} + \frac{y^2}{4}\right)}. \end{aligned}$$

The marginal pdf's are found by differentiating the marginal cdf:

$$f_X(x) = \frac{12}{7}\left(x^2 + \frac{x}{2}\right) \qquad f_Y(y) = \frac{12}{7}\left(\frac{1}{3} + \frac{y}{2}\right).$$

We could also have found them by integrating the joint pdf:

$$f_X(x) = \int_0^1 f(x,y)\, dy = \boxed{\frac{12}{7}\left(x^2 + \frac{x}{2}\right)} \qquad f_Y(y) = \int_0^1 f(x,y)\, dx = \boxed{\frac{12}{7}\left(\frac{1}{3} + \frac{y}{2}\right)}.$$

(c) *Find $E[X]$ and $\mathrm{Var}(X)$.*

Solution: The computation is slightly easier if we use the formula $\mathrm{Var}(X) = E[X^2] - E[X]^2$.

$$\begin{aligned} E[X] &= \int_0^1 x f_X(x)\, dx \;=\; \frac{12}{7} \int_0^1 x\left(x^2 + \frac{x}{2}\right) dx \;=\; \frac{12}{7}\left(\frac{1}{4} + \frac{1}{6}\right) = \boxed{\frac{5}{7} \approx 0.7143} \\ E[X^2] &= \int_0^1 x^2 f_X(x)\, dx = \frac{39}{70} \approx 0.5571. \end{aligned}$$

Thus $\boxed{\text{Var}(X) = E[X^2] - E[X]^2 \approx 0.0469}$.

(d) *Find the covariance and correlation of X and Y.*

Solution: First we'll need $E[Y]$ and $\text{Var}[Y]$. The computations are similar to those in part (c).

$$\begin{aligned}
E[Y] &= \int_0^1 y f_Y(y)\,dy = \frac{12}{7}\int_0^1 y\left(\frac{1}{3}+\frac{y}{2}\right)\,dy = \frac{4}{7} \approx 0.5714 \\
E[Y^2] &= \int_0^1 y^2 f_Y(y) = \frac{12}{7}\int_0^1 y^2\left(\frac{1}{3}+\frac{y}{2}\right)\,dy = \frac{17}{42} \approx 0.4048 \\
\text{Var}(Y) &= E[Y^2] - E[Y]^2 \approx 0.0782
\end{aligned}$$

Now, covariance is defined as $\text{Cov}(X,Y) = E[(X-\mu_x)(Y-\mu_Y)]$. We could compute this directly, but it's slightly easier to use the formula $\text{Cov}(X,Y) = E[XY] - E[X]E[Y]$.

$$\begin{aligned}
E[XY] &= \int_0^1\int_0^1 xyf(x,y)\,dy\,dx = \frac{12}{7}\int_0^1\int_0^1 x^3y + x^2y^2\,dy\,dx = \frac{17}{42} \approx 0.4048 \\
\text{Cov}(X,Y) &= E[XY] - E[X]E[Y] \approx \boxed{-0.0034} \\
\text{Cor}(X,Y) &= \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y} = \boxed{-0.0561}
\end{aligned}$$

(e) *Are X and Y independent?*

Solution: $\boxed{\text{No they are not independent.}}$ We can see this in two ways. First, their joint pdf is not the product of the marginal pdfs. Second, their covariance is not 0.

Problem 2. (10 pts.) **Independence**

Suppose X and Y are random variables with the following joint pmf. Are X and Y independent?

$X\backslash Y$	1	2	3
1	1/18	1/9	1/6
2	1/9	1/6	1/18
3	1/6	1/18	1/9

Solution: To check independence we have to check if all the cell probabilities are the product of marginal probabilities. So, first we compute the marginal probabilities by summing along rows and columns.

$X\backslash Y$	1	2	3	$p(x)$
1	1/18	1/9	1/6	1/3
2	1/9	1/6	1/18	1/3
3	1/6	1/18	1/9	1/3
$p(y)$	1/3	1/3	1/3	1

Now, we can easily check that the joint distribution is not the product of the marginals. For example,

$$P(X=1, Y=1) = \frac{1}{18}, \text{ but } P(X=1)P(Y=1) = \frac{1}{9}.$$

So, X and Y are not independent.

Problem 3. (20: 10,10 pts.) **Correlation**
Suppose X and Y are random variables with

$$P(X = 1) = P(X = -1) = \frac{1}{2}; \qquad P(Y = 1) = P(Y = -1) = \frac{1}{2}.$$

Let $c = P(X = 1 \text{ and } Y = 1)$.

(a) *Determine the joint distribution of X and Y, $\text{Cov}(X, Y)$, and $Cor(X, Y)$.*

Solution: We make the joint distribution table by starting with the marginal distributions and putting c in the $X = 1$, $Y = 1$ cell. The other three cells in the table are then determined.

$X \backslash Y$	1	-1	$p(x)$
1	c	$0.5 - c$	0.5
-1	$0.5 - c$	c	0.5
$p(y)$	0.5	0.5	1

We easily compute: $E[X] = 0$, $E[Y] = 0$, $\text{Var}(X) = 1$, $\text{Var}(Y) = 1$. Computing directly:

$$\begin{aligned} E[XY] &= (1 \cdot 1)c + (-1 \cdot 1)(0.5 - c) + (1 \cdot -1)(0.5 - c) + (-1 \cdot -1)c \\ &= 4c - 1 \end{aligned}$$

Thus,

$$\begin{aligned} \text{Cov}(X, Y) &= E[XY] - E[X]E[Y] = 4c - 1 \\ \text{Cor}(X, Y) &= \frac{\text{Cor}(X, Y)}{\sigma_X \sigma_Y} = 4c - 1. \end{aligned}$$

(b) *For what value(s) of c are X and Y independent? For what value(s) of c are X and Y 100% correlated?*

Solution: Note that the correlation runs from -1 to 1 as c runs from 0 to 0.5.

If X and Y are independent then we must have $\text{Cov}(X, Y) = 0$. This only happens when $c = \frac{1}{4}$. Covariance equal 0 does not guarantee independence, but for this value of c, it is easy to check that all four probabilities in the table are 0.25 and X and Y are, indeed, independent.

When $c = 0$ the correlation is -1, which means X and Y are fully correlated (sometimes called fully anti-correlated). When $c = 0.5$ the correlation is 1.0 and X and Y are fully correlated.

Problem 4. (40: 5,5,10,10,10 pts.) **Don't be late!**
Alicia and Bernardo are trying to meet for lunch and both will arrive, independently of each other, uniformly and at random between noon and 1pm. Let A and B be the number of minutes after noon at which Alicia and Bernardo arrive, respectively. Then A and B are independent uniformly distributed random variables on $[0, 60]$.

Hint: *For parts (c-e) you might find it easiest to find the fraction of the square* $[0,60]\times[0,60]$ *filled by the event.*

(a) *Find the joint pdf* $f(a,b)$ *and joint cdf* $F(a,b)$*.*

Solution: The joint probability density function is $f(a,b) = \frac{1}{3600}$ and the joint cumulative distribution function is

$$F(a,b) = \int_0^a \int_0^b f(s,t)\,ds\,dt = \frac{ab}{3600}$$

(b) *Find the probability that Alicia arrives before 12:30.*

Solution: Since A is uniformly distributed on $[0,60]$, $P(A \le 30) = \frac{1}{2}$.

(c) *Find the probability that Alicia arrives before 12:15 and Bernardo arrives between 12:30 and 12:45 in two ways:*

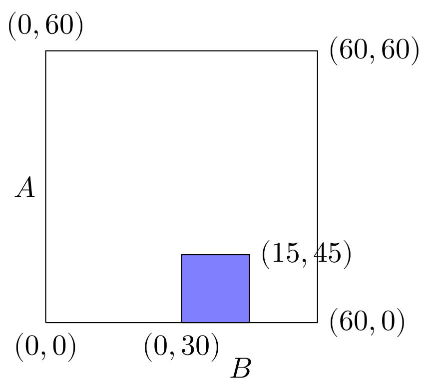
(i) By using the fact that A *and* B *are independent.*

(ii) By shading the corresponding area of the square $[0,60]\times[0,60]$ *and finding what proportion of the square is shaded.*

Solution: (i) $P(A \le 15, 30 \le B \le 45) = P(A \le 15)P(30 \le B \le 45) = 0.0625$

(ii) The range of (A,B) is the square $[0,60]\times[0,60]$. The event 'Alicia arrives before 12:15 and Bernardo arrives between 12:30 and 12:45' is represented by the solid blue rectangle. Since the probability distribution is uniform the probability of the blue rectangle is just the fraction of the entire square that it covers.

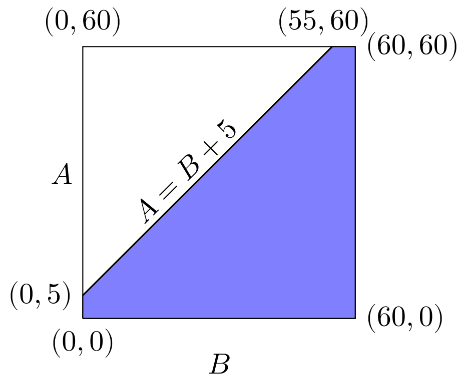
Area of blue rectangle $= 15\times 15 = 225$. Fraction of the entire square $= 225/3600 = 0.0625$.



(d) *Find the probability that Alicia arrives less than five minutes after Bernardo. (Hint: use method (ii) from part (c).)*

Solution: The shaded area in the figure below corresponds to the event '$A \le B+5$'. (Note: if Alicia arrives before Bernardo then she arrives less than 5 minutes after him.) That is, it corresponds to all pairs of arrival times (a,b) such that $a \le b+5$. $P(A \le B+5)$ is then just the area of the blue region divided by the area of the entire square. The area of the blue region is the area of the full square minus the area of the unshaded triangle. The area of the white region is $\dfrac{55^2}{2}$. So,

$$P(A \le B+5) = \frac{1}{3600}\left(3600 - \frac{55^2}{2}\right) = \boxed{0.5799}.$$



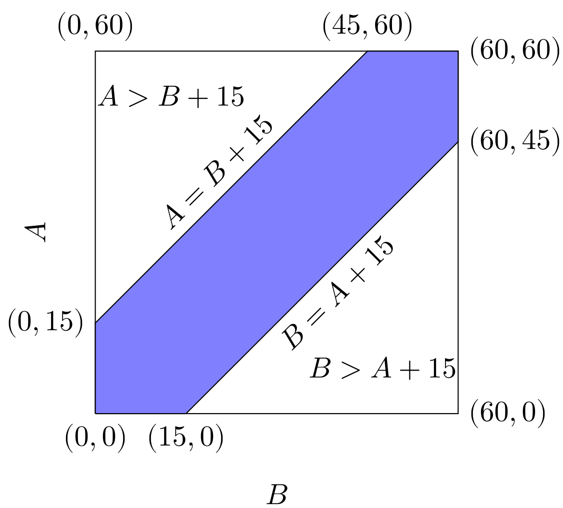
(e) *Now suppose that Alicia and Bernardo are both rather impatient and will leave if they have to wait more than 15 minutes for the other to arrive. What is the probability that Alicia and Bernardo will have lunch together?*

Solution: Alicia and Bernardo arrive within 15 minutes of each other is event

$$E = B - 15 \leq A \leq B + 15.$$

This is the blue shaded region in the figure below. We see that the area of each white triangle is $\frac{45^2}{2}$. So, the combined white area is 45^2 and

$$P(E) = \frac{3600 - 45^2}{3600} = \boxed{\frac{7}{16}}.$$



Problem 5. (10 pts.) **Overlapping sums**
Suppose $X_1, X_2, \ldots$ are independent exponential(2) random variables. Suppose also that X is the sum of the first n and Y is the sum of X_{n-7} to X_{2n-8}. Compute $\mathrm{Cov}(X, Y)$ *and Cor(X, Y). You should assume that $n \geq 8$.*

Hints: The variance of an exponential(λ) random variable is $1/\lambda^2$. Use the linearity rules for covariance. What is the size of the overlap?

Solution: The problem states that

$$X = \sum_{i=1}^{n} X_i \qquad Y = \sum_{i=n-7}^{2n-8} X_i$$

Notice that X and Y have an overlap of 8 terms. We can use the linearity rules for covariance:

$$\mathrm{Cov}(X, Y) \;=\; \mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=n-7}^{2n-8} X_j\right) \;=\; \sum_{i=1}^{n}\sum_{j=n-7}^{2n-8} \mathrm{Cov}(X_i, X_j).$$

That is, $\mathrm{Cov}(X, Y)$ is the sum of all pairs consisting of a term from X and a term from Y. Since the different terms are independent we know that $\mathrm{Cov}(X_i, X_j) = 0$ if $i \neq j$. So only the overlap contributes to the covariance:

$$\mathrm{Cov}(X, Y) = \sum_{i=n-7}^{n} \mathrm{Cov}(X_i, X_i) = \sum_{i=n-7}^{n} \mathrm{Var}(X_i) = \sum_{i=n-7}^{n} \frac{1}{4}$$

There are 8 such terms, all with the variance 1/4, so $\boxed{\mathrm{Cov}(X, Y) = 8 \cdot 1/4 = 2.}$

Since X is the sum of n independent variables with variance 1/4 we have $\mathrm{Var}(X) = n/4$. Likewise $\mathrm{Var}(Y) = n/4$. So

$$\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{2}{n/4} = \boxed{\frac{8}{n}.}$$

MIT OpenCourseWare
https://ocw.mit.edu

18.05 Introduction to Probability and Statistics
Spring 2022

For information about citing these materials or our Terms of Use, visit: https://ocw.mit.edu/terms.