Chapter 5. Multiple Random Variables

5.2: Joint Continuous Distributions

Slides (Google Drive) Alex Tsun Video (YouTube)

5.2.1 Joint PDFs and Expectation

The joint continuous distribution is the continuous counterpart of a joint discrete distribution. Therefore, conceptual ideas and formulas will be roughly similar to that of discrete ones, and the transition will be much like how we went from single variable discrete RVs to continuous ones.

To think intuitively about joint continuous distributions, consider throwing darts at a dart board. A dart board is two-dimensional and a certain 2D position on the dart board is (x, y). Because x and y positions are continuous, we want to think about the joint distribution between two continuous random variables X and Y representing the location of the dart. What is the joint density function describing this scenario?

Definition 5.2.1: Joint PDFs

Let X, Y be continuous random variables. The joint PDF of X and Y is:

$$f_{X,Y}(a, b) \geq 0$$

The joint range is the set of pairs (c, d) that have nonzero density:

$$\Omega_{X,Y} = \{(c, d) : f_{X,Y}(c, d) > 0\} \subseteq \Omega_X \times \Omega_Y$$

Note that the double integral over all values must be 1:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(u, v)dudv = 1$$

Further, note that if $g : \mathbb{R}^2 \to \mathbb{R}$ is a function, then LOTUS extends to the multidimensional case:

$$\mathbb{E}[g(X, Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(s, t)f_{X,Y}(s, t)dsdt$$

The joint PDF must satisfy the following (similar to univariate PDFs):

$$\mathbb{P}(a \leq X < b, c \leq Y \leq d) = \int_a^b\int_c^d f_{X,Y}(x, y)dydx$$

Example(s)

Let X and Y be two jointly continuous random variables with the following joint PDF:

$$f_{X,Y}(x, y) = \begin{cases} x + cy^2 & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find and sketch the joint range $\Omega_{X,Y}$.

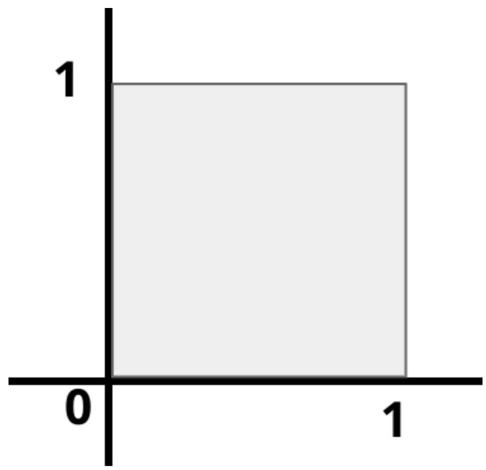
(b) Find the constant c that makes $f_{X,Y}$ a valid joint PDF.

(c) Find $\mathbb{P}\left(0 \leq X \leq \frac{1}{2}, 0 \leq Y \leq \frac{1}{2}\right)$.

Solution

(a)

$$\Omega_{X,Y} = \left\{(x,y) \in \mathbb{R}^2 : 0 \leq x \leq 1, 0 \leq y \leq 1\right\}$$



(b) To find c, the following condition has to be satisfied:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$$

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy \\
&= \int_{0}^{1}\int_{0}^{1} (x + cy^2)dxdy \\
&= \int_{0}^{1} \left[\frac{1}{2}x^2 + cy^2x\right]_{x=0}^{x=1} dy \\
&= \int_{0}^{1} \left(\frac{1}{2} + cy^2\right) dy \\
&= \left[\frac{1}{2}y + \frac{1}{3}cy^3\right]_{y=0}^{y=1} \\
&= \frac{1}{2} + \frac{1}{3}c
\end{aligned}$$

Thus, $c = \frac{3}{2}$.

(c)

$$\begin{aligned}
\mathbb{P}\left(0 \leq X \leq \frac{1}{2}, 0 \leq Y \leq \frac{1}{2}\right) &= \int_0^{1/2} \int_0^{1/2} \left(x + \frac{3}{2}y^2\right) dxdy \\
&= \int_0^{1/2} \left[\frac{1}{2}x^2 + \frac{3}{2}y^2 x\right]_{x=0}^{x=1/2} dy \\
&= \int_0^{1/2} \left(\frac{1}{8} + \frac{3}{4}y^2\right) dy \\
&= \frac{3}{32}
\end{aligned}$$

□

Example(s)

Let X and Y be two jointly continuous random variables with the following PDF:

$$f_{X,Y}(x,y) = \begin{cases} x+y & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Find $\mathbb{E}\left[XY^2\right]$.

Solution By LOTUS,

$$\begin{aligned}
\mathbb{E}\left[XY^2\right] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (xy^2) f_{X,Y}(x,y) dxdy \\
&= \int_0^1 \int_0^1 xy^2(x+y) dxdy \\
&= \int_0^1 \left(\frac{1}{3}y^2 + \frac{1}{2}y^3\right) dy \\
&= \frac{17}{72}
\end{aligned}$$

□

5.2.2 Marginal PDFs

Definition 5.2.2: Marginal PDFs

Suppose that X and Y are jointly distributed continuous random variables with joint PDF $f_{X,Y}(x,y)$. The marginal PDFs of X and Y are respectively given by the following:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx$$

Note this is exactly like for joint discrete random variables, with integrals instead of sums.

(Extension): If Z is also a continuous random variable, then the marginal PDF of Z is:

$$f_Z(z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)dxdy$$

Solution

Example(s)

Find the marginal PDFs $f_X(x)$ and $f_Y(y)$ given the joint PDF:

$$f_{X,Y}(x,y) = \begin{cases} x + \dfrac{3}{2}y^2 & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Then, compute $\mathbb{E}[X]$. (This is the same joint density as the first example, plugging in $c = 3/2$).

For $0 \leq x \leq 1$:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \\ &= \int_0^1 \left(x + \frac{3}{2}y^2\right) dy \\ &= \left[xy + \frac{1}{2}y^3\right]_{y=0}^{y=1} \\ &= x + \frac{1}{2} \end{aligned}$$

Thus, the marginal PDF $f_X(x)$ is:

$$f_X(x) = \begin{cases} x + \frac{1}{2} & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

For $0 \leq y \leq 1$:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx \\ &= \int_0^1 \left(x + \frac{3}{2}y^2\right) dx \\ &= \left[\frac{1}{2}x^2 + \frac{3}{2}y^2 x\right]_{x=0}^{x=1} \\ &= \frac{3}{2}y^2 + \frac{1}{2} \end{aligned}$$

Thus, the marginal PDF $f_Y(y)$ is:

$$f_Y(y) = \begin{cases} \frac{3}{2}y^2 + \frac{1}{2} & 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Note that to compute $\mathbb{E}[X]$ for example, we can either use LOTUS, or just the marginal PDF $f_X(x)$. These methods are equivalent. By LOTUS (taking $g(X,Y)=X$),

$$\mathbb{E}[X]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)dxdy=\int_0^1\int_0^1 x\left(x+\frac{3}{2}y^2\right)dxdy$$

Alternatively, by definition of expectation for a single RV,

$$\mathbb{E}[X]=\int_{-\infty}^{\infty} x f_X(x)dx=\int_0^1 x\left(x+\frac{1}{2}\right)dx$$

It only takes two lines or so of algebra to show they are equal! □

5.2.3 Independence of Continuous Random Variables

Definition 5.2.3: Independence of Continuous Random Variables

Continuous random variables X,Y are independent, written $X\perp Y$, if for all $x\in\Omega_X$ and $y\in\Omega_Y$,

$$f_{X,Y}(x,y)=f_X(x)f_Y(y)$$

Recall $\Omega_{X,Y}=\{(x,y):f_{X,Y}(x,y)>0\}\subseteq\Omega_Y\times\Omega_Y$. A necessary but not sufficient condition for independence is that $\Omega_{X,Y}=\Omega_X\times\Omega_Y$. That is, if $\Omega_{X,Y}=\Omega_X\times\Omega_Y$, then we have to check the condition, but if not, then we know they are not independent.
This is because if there is some $(a,b)\in\Omega_X\times\Omega_Y$ but not in $\Omega_{X,Y}$, then $f_{X,Y}(a,b)=0$ but $f_X(a)>0$ and $f_Y(b)>0$, which violates independence. (This is very similar to independence for discrete RVs).

5.2.4 Multivariate: From Discrete to Continuous

The following table tells us the relationships between discrete and continuous joint distributions.

	Discrete	**Continuous**
Joint Dist	$p_{X,Y}(x,y)=\mathbb{P}(X=x,Y=y)$	$f_{X,Y}(x,y)\neq\mathbb{P}(X=x,Y=y)$
Joint CDF	$F_{X,Y}(x,y)=\sum_{t\le x,s\le y}p_{X,Y}(t,s)$	$F_{X,Y}(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f_{X,Y}(t,s)\,dsdt$
Normalization	$\sum_{x,y}p_{X,Y}(x,y)=1$	$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_{X,Y}(x,y)\,dxdy=1$
Marginal Dist	$p_X(x)=\sum_y p_{X,Y}(x,y)$	$f_X(x)=\int_{-\infty}^{\infty}f_{X,Y}(x,y)dy$
Expectation	$\mathbb{E}[g(X,Y)]=\sum_{x,y}g(x,y)p_{X,Y}(x,y)$	$\mathbb{E}[g(X,Y)]=\int\int g(x,y)f_{X,Y}(x,y)dxdy$
Conditional Dist	$p_{X\|Y}(x\|y)=\frac{p_{X,Y}(x,y)}{p_Y(y)}$	$f_{X\|Y}(x\|y)=\frac{f_{X,Y}(x,y)}{f_Y(y)}$
Conditional Exp	$\mathbb{E}[X\|Y=y]=\sum_x xp_{X\|Y}(x\|y)$	$\mathbb{E}[X\|Y=y]=\int_{-\infty}^{\infty}xf_{X\|Y}(x\|y)dx$
Independence	$\forall x,y,p_{X,Y}(x,y)=p_X(x)p_Y(y)$	$\forall x,y,f_{X,Y}(x,y)=f_X(x)f_Y(y)$

We'll explore the two conditional rows (second and third last rows) in the next section more, but you can guess that $p_{X|Y}(x\mid y)=\mathbb{P}(X=x\mid Y=y)$, and use the definition of conditional probability to see that it is $\mathbb{P}(X=x,Y=y)/\mathbb{P}(Y=y)$, as stated!

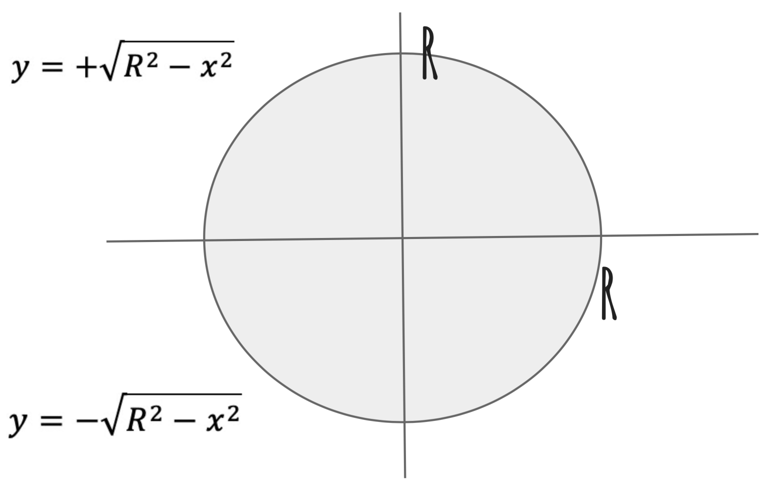
Example(s)

Let's return to our dart example. Suppose (X, Y) are jointly and uniformly distributed on the circle of radius R centered at the origin (example a dart throw).

1. First find and sketch the joint range $\Omega_{X,Y}$.
2. Now, write an expression for the joint PDF $f_{X,Y}(x, y)$ and carefully define it for all $x, y \in \mathbb{R}$.
3. Now, solve for the range of X and write an expression we can evaluate to find $f_X(x)$, the marginal PDF for X.
4. Now, let Z be the distance from the center that the dart falls. Find Ω_Z and write an expression for $\mathbb{E}[Z]$.
5. Finally, determine using the definition of independence whether X and Y are independent.

Solution

1. The joint range is $\Omega_{X,Y} = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \le R^2\}$ since the values must be within the circle of radius R. We can sketch the range as follows, with the semi-circles below and above the y-axis labeled with their respective equations.



2. The height of the density function is constant, say h, since it is uniform. The double integral over all x and y must equal one ($\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y)dxdy = 1$), meaning the volume of this cylinder must be 1. The volume is base times height, which is $\pi R^2 \cdot h$, and setting it equal to 1 gives $h = \frac{1}{\pi R^2}$. This gives us the following joint PDF:

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{\pi R^2} & x, y \in \Omega_{X,Y} \\ 0 & \text{otherwise} \end{cases}$$

3. Well, X can range from $-R$ to R, since there are points on the circle with x values in this range. So the range of X is:

$$\Omega_X = [-R, R]$$

Setting up this integral will be trickier than in the earlier examples, because when finding $f_X(x)$ and integrating out the y, the limits of integration actually depend on x. Imagine making a tick mark at some $x \in [-R, R]$ (on the x-axis) and drawing a vertical line through x: where does y enter and leave (like summing a column in a joint PMF)? Based on the equations we had earlier for y in terms of x (see the sketch above), this give us:

$$f_X(x) = \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} f_{X,Y}(x,y)dy$$

Again, this is different from the previous examples, and you MUST sketch/plot the joint range to figure this out. If you learned how to do double integrals, this is exactly the same idea.

4. Well, the distance will be given by $Z = \sqrt{X^2+Y^2}$, which is the definition of distance. We can further see that Z will take on any value from 0 to R, since the point could be at the origin and as far as R. This gives, $\Omega_Z = [0, R]$.

   Then, to solve for the expected value of Z, we can use LOTUS, and only integrate over the joint range of X and Y (since the joint PDF is 0 elsewhere). We have to be careful in setting up the bounds of our integral. X will range from $-R$ to R as we discussed earlier. But as X ranges across these values, Y will range from $-\sqrt{R^2-x^2}$ to $\sqrt{R^2-x^2}$. We had $Z = \sqrt{X^2+Y^2}$, so for the expected value we have:

$$\mathbb{E}[Z] = \mathbb{E}\left[\sqrt{X^2+Y^2}\right] = \int_{-R}^{R}\int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \sqrt{x^2+y^2} f_{X,Y}(x,y)dydx$$

   Note that we could've set up this integral $dxdy$ instead - what would the limits of integration have been? It would've been

$$\mathbb{E}[Z] = \mathbb{E}\left[\sqrt{X^2+Y^2}\right] = \int_{-R}^{R}\int_{-\sqrt{R^2-y^2}}^{\sqrt{R^2-y^2}} \sqrt{x^2+y^2} f_{X,Y}(x,y)dxdy$$

   Your outer limits must be just the range of Y (both constants), and your inner limits may depend on the outer variable of integration.

5. No, they are not independent. We can see this with the test: $\Omega_{X,Y} \neq \Omega_X \times \Omega_Y$. This is because X and Y both have marginal range from $-R$ to R, but the joint range is not a rectangle of this region (it is a circle). More explicitly, take a point $(0.99R, 0.99R)$ which is basically the top right of the square (R, R)). We get $0 = f_{X,Y}(0.99R, 0.99R) \neq f_X(0.99R)f_Y(0.99R) > 0$. This is because the joint PDF is defined to be 0 at $(0.99R, 0.99R)$ (not in the circle), but the marginal PDFs of both X and Y are nonzero at $0.99R$ (since $0.99R$ is in the marginal range of both).

□

Example(s)

Now let's consider another example where we have a continuous joint distribution (X, Y), where $X \in [0, 1]$ is the proportion of the time until the midterm that you actually spend studying for it and $Y \in [0, 1]$ is your percentage score on the exam.
Suppose the joint PDF is:

$$f_{X,Y}(x,y) = \begin{cases} ce^{-(y-x)} & x, y \in [0,1] \text{ and } y \geq x \\ 0 & \text{otherwise} \end{cases}$$

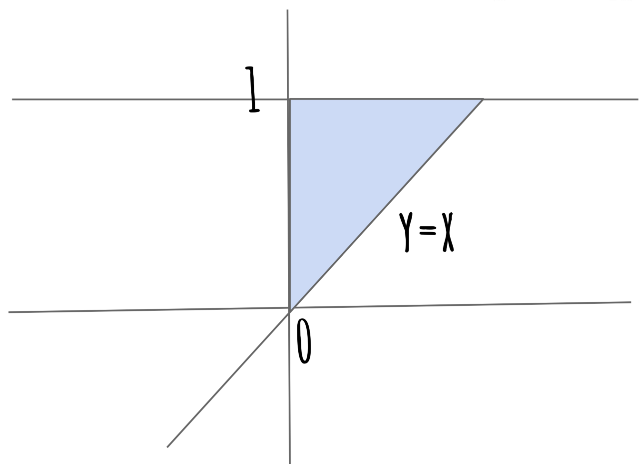
1. First, consider the joint range and sketch it. Then, interpret it in English in the context of the

problem.

2. Now, write an expression for c in the PDF above.
3. Now, find Ω_Y and write an expression that we could evaluate to find $f_Y(y)$.
4. Now, write an expression that we could evaluate to find $\mathbb{P}(Y \geq 0.9)$.
5. Now, write an expression that we can evaluate to find $\mathbb{E}[Y]$, the expected score on the exam.
6. Finally, consider whether X and Y are independent.

Solution

1. X can range from any value in $[0, 1]$ without conditions. Then Y will only be bounded in that it must be less than or equal to X. We can first draw the line $x = y$, and then the region above this line for which x, y are less than 1 will be our range. That gives us the following:



In English, this means that your score is at least the percentage of time that you studied, as your score will be that proportion or more.

2. To solve for c, we should find the volume above this triangle on the x-y plane and invert it, since $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y)dxdy = 1$. To find the area we can integrate in terms of x or y first, which will give us the following two equivalent expressions:

$$c = \frac{1}{\int_0^1 \int_x^1 e^{-(y-x)}dydx} = \frac{1}{\int_0^1 \int_0^y e^{-(y-x)}dxdy}$$

We'll explain the first equality using the $dydx$ ordering. Since dx is the outer integral, the limits must be just the range of X, which is $[0, 1]$. For each value of x (draw a vertical line through x on the x-axis), y goes between x and 1, so those are the inner limits of integration.

Now, for the second equality using $dxdy$ ordering, the outer integral is dy, so the limits are the range of Y, also $[0, 1]$. Then, for each value of y (draw a horizontal line through y on the y-axis), x goes between 0 and y, and so those are the inner limits of integration.

3. Well, $\Omega_Y = [0, 1]$ as we can see in our graph above that Y takes on values in this range. For the marginal PDF we have to integrate in respect to X, which will take on values in the range 0 to y based on our graph. So, we have:

$$f_Y(y) = \int_0^y ce^{-(y-x)}dx$$

4. We can integrate from 0.9 to 1 to solve for this, using the marginal PDF that we solved for above. This takes us back to the univariate case essentially, and gives us the following:

$$\mathbb{P}(Y \geq 0.9) = \int_{0.9}^1 f_Y(y)dy = \int_{0.9}^1 \int_0^y ce^{-(y-x)}dxdy$$

5. By definition of expectation (univariate), or LOTUS, we have:

$$\mathbb{E}[Y] = \int_0^1 yf_Y(y)dy = \int_0^1 \int_0^y cye^{-(y-x)}dxdy$$

6. $\Omega_{X,Y} \neq \Omega_X \times \Omega_Y$ since the sketch of the range is not a rectangle. The joint range is not equal to the cartesian product of the marginal ranges. To be concrete, consider the point $(x = 0.99, y = 0.01)$ (basically the corner $(1, 0)$). I chose this point because it was in the Cartesian product $\Omega_X \times \Omega_Y = [0, 1] \times [0, 1]$, but not in the joint range (see the picture from the first part). Since it's not in the joint range (shaded region), we have $f_{X,Y}(0.99, 0.01) = 0$, but since $0.99 \in \Omega_X$ and $0.01 \in \Omega_Y$, $f_X(0.99) > 0$ and $f_Y(0.01) > 0$. Hence, I've found a pair of points (x, y) where the joint density isn't equal to the product of the marginal densities, violating independence.

$\square$