In many real life situations, we are often interested in two or more random variables at the same time, and how they are correlated. Typical examples are brain size and IQ, shoe size and height, level of air pollution and fertility, or the number of TV shows watched per week and the age of the watcher.

To understand how random variables are correlated (or are independent from one another) mathematically, we construct their *joint distributions*, which are extensions of the distributions we have seen for single random variables. They are the topic of this lecture.

1 Discrete random variables

1.1 Joint probability mass function

Definition: Consider a discrete random variable X which takes values $\{x_1, x_2, \dots, x_n\}$ and a discrete random variable Y which takes values $\{y_1, y_2, \dots, y_n\}$. The pair (X, Y) of random variables takes values $\{(x_1, y_1), (x_2, y_1), \dots, (x_n, y_1), (x_1, y_2), \dots, (x_n, y_2), \dots, (x_n, y_m)\}$.

The **joint probability mass function** (joint p.m.f.) of X and Y is the function $p_{X,Y}(x_i, y_j)$ giving the probability of the joint outcome $\{X = x_i\} \cap \{Y = y_j\}$, which we write $X = x_i, Y = y_j$:

$$\boxed{p_{X,Y}(x_i, y_j) = P(X = x_i, Y = y_j) \qquad -\infty < x_i, y_j < +\infty} \tag{1}$$

The joint probability mass function for discrete random variables is usually presented as a table:

X \ Y	y_1	y_2	$\dots$	y_m
x_1	$p_{X,Y}(x_1, y_1)$	$p_{X,Y}(x_1, y_2)$	$\dots$	$p_{X,Y}(x_1, y_m)$
x_2	$p_{X,Y}(x_2, y_1)$	$p_{X,Y}(x_2, y_2)$	$\dots$	$p_{X,Y}(x_2, y_m)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	$p_{X,Y}(x_n, y_1)$	$p_{X,Y}(x_n, y_2)$	$\dots$	$p_{X,Y}(x_n, y_m)$

Example: A canonical example for this is the situation in which one rolls two dice, and X corresponds to the value on the first dice, and Y the total of both dice. X takes on the values $\{1, 2, 3, 4, 5, 6\}$, and Y takes on the values $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. The joint probability mass function is shown below. The column with label $p_X(x)$ can be ignored at this point, and will be explained in the next section. The same applies to the row $p_Y(y)$.

X \ Y	2	3	4	5	6	7	8	9	10	11	12		$p_X(x)$
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	0	0	0	0	0		$\frac{1}{6}$
2	0	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	0	0	0	0		$\frac{1}{6}$
3	0	0	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	0	0	0		$\frac{1}{6}$
4	0	0	0	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	0	0		$\frac{1}{6}$
5	0	0	0	0	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	0		$\frac{1}{6}$
6	0	0	0	0	0	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$		$\frac{1}{6}$
$p_Y(y)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$		

1.2 Marginal distributions

We observe in the previous example that the probability mass function of X can be obtained by summing all probabilities along a row. In this context, we say that p_X is the **marginal distribution** of X. In the table above, we added a column and a row for the marginal distributions of p_X and p_Y of X and Y respectively.

Observe that we know that, like any other probability mass function,

$$\sum_i p_X(x_i) = 1 \qquad , \qquad \sum_j p_Y(y_j) = 1$$

Hence

$$\underline{\sum_i \sum_j p_{X,Y}(x_i, y_j) = 1}$$

This is a general result for joint probability mass functions.

Furthermore, since $p_{X,Y}(x_i, y_j) = P(X = x_i, Y = y_j)$, we have

$$\underline{0 \le p_{X,Y}(x_i, y_j) \le 1}$$

1.3 Joint distribution function

We now discuss the generalization of the cumulative distribution function for the case of multiple random variables, called the *joint distribution function*.

Definition: The **joint distribution function** $F_{X,Y}$ of two random variables X and Y is the function $F_{X,Y} : \mathbb{R}^2 \to [0,1]$ defined by

$$\boxed{F_{X,Y}(a,b) = P(X \le a, Y \le b) \qquad a \in \mathbb{R}\,, b \in \mathbb{R}} \tag{2}$$

Let us emphasize again that the comma in $P(X \le a, Y \le b)$ must be understood as an "and", in the sense of the intersection of two events.

Example: For the case of the two dice rolled discussed in the previous page

$$\begin{aligned} F_{X,Y}(3,4) &= P(X \le 3, Y \le 4) \\ &= p_{X,Y}(1,2) + p_{X,Y}(1,3) + p_{X,Y}(1,4) \\ &\qquad + p_{X,Y}(2,3) + p_{X,Y}(2,4) \\ &\qquad\qquad + p_{X,Y}(3,4) \\ &= \frac{6}{36} = \frac{1}{6} \end{aligned}$$

Marginal distribution function

Just as was the case for the joint probability mass function, the knowledge of the joint distribution function $F_{X,Y}$ is sufficient to reconstruct the cumulative distribution functions F_X and F_Y for the random variables X and Y. In that context, F_X and F_Y are called **marginal distribution functions**.

The idea, as before, is to sum along each full row to obtain F_X:

$$\boxed{F_X(x_i) = P(X \le x_i) = F_{X,Y}(x_i, +\infty) = \lim_{b\to\infty} F_{X,Y}(x_i, b)}$$

Likewise, we sum along each full column to obtain F_Y:

$$\boxed{F_Y(y_j) = P(Y \le y_j) = F_{X,Y}(+\infty, y_j) = \lim_{a\to\infty} F_{X,Y}(a, y_j)}$$

Returning to our two dice, we for example have

$$F_Y(5) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} = \frac{5}{18}$$

2 Continuous random variables

2.1 Joint probability density function

Consider a continuous random variable X which takes values in the interval $[a,b]$ and a continuous random variable Y which takes values in the interval $[c,d]$. The pair (X,Y) of random variables takes values in $[a,b] \times [c,d]$, i.e. a rectangle, as shown in Figure 1.

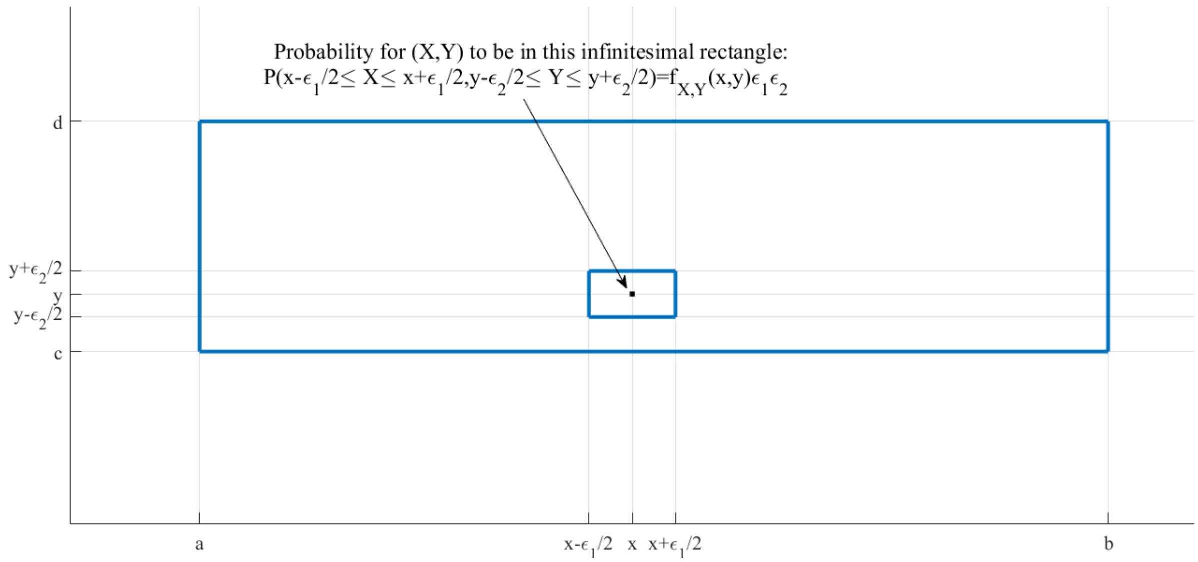

Figure 1: Rectangle $[a, b] \times [c, d]$ in which the pair of continuous random variables (X, Y) can take values, and an infinitesimal rectangle with infinitesimal length ϵ_1 and width ϵ_2 and centered on (x, y). The probability that (X, Y) takes values in this infinitesimal rectangle centered on (x, y) is $P(x - \frac{\epsilon_1}{2} \leq X \leq x + \frac{\epsilon_1}{2}, y - \frac{\epsilon_2}{2} \leq Y \leq y + \frac{\epsilon_2}{2}) = f_{X,Y}(x, y)\epsilon_1\epsilon_2$.

The counterpart of the joint probability mass function for continuous random variables is the *joint probability density function* $f_{X,Y}(x, y)$, defined as follows: the probability that (X, Y) is in the infinitesimal rectangle centered on (x, y) with sides ϵ_1 and ϵ_2 (see Figure 1) is

$$P(x - \frac{\epsilon_1}{2} \leq X \leq x + \frac{\epsilon_1}{2}, y - \frac{\epsilon_2}{2} \leq Y \leq y + \frac{\epsilon_2}{2}) = f_{X,Y}(x, y)\epsilon_1\epsilon_2$$

If we are interested in the probability that X and Y are in a larger rectangle than the infinitesimal rectangle $[x - \frac{\epsilon_1}{2}, x + \frac{\epsilon_1}{2}] \times [y - \frac{\epsilon_2}{2}, y + \frac{\epsilon_2}{2}]$, we need to sum over several infinitesimally small rectangles. This amounts to evaluating integrals. This motivates the following definition:

Definition: Continuous random variables X and Y have a joint continuous distribution if for some function $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ and for all numbers a_1, a_2 and b_1, b_2 such that $a_1 \leq b_1$ and $a_2 \leq b_2$,

$$\boxed{P(a_1 \leq X \leq b_1, a_2 \leq Y \leq b_2) = \int_{a_1}^{b_1}\int_{a_2}^{b_2} f_{X,Y}(x, y)dxdy} \tag{3}$$

where the function $f_{X,Y}$ satisfies:

- $f_{X,Y}(x, y) \geq 0$ for all x and y
- $\int_{-\infty}^{+\infty} f_{X,Y}(x, y)dxdy = 1$

$f_{X,Y}$ is called the **joint probability density function** of X and Y (or sometimes called the bivariate probability density).

Making sense of the definition

$f_{X,Y}$ is a function of *two variables*: it takes x and y as two separate inputs, and outputs a third real number $f_{X,Y}(x, y)$. We need two dimensions to plot all the possible inputs in $f_{X,Y}$ (e.g. the rectangle in Figure), and thus a *third* dimension for the graph of $f_{X,Y}$, corresponding to all the possible outputs of $f_{X,Y}$. This is illustrated in Figure 2.

We then have the following geometric interpretation: the double integral $\int_a^b \int_c^d f_{X,Y}(x, y)dydx$ is the *volume* between the graph $z = f_{X,Y}(x, y)$ of $f_{X,Y}$ over R and the rectangle R in the $x - y$ plane, as shown in Figure 2.

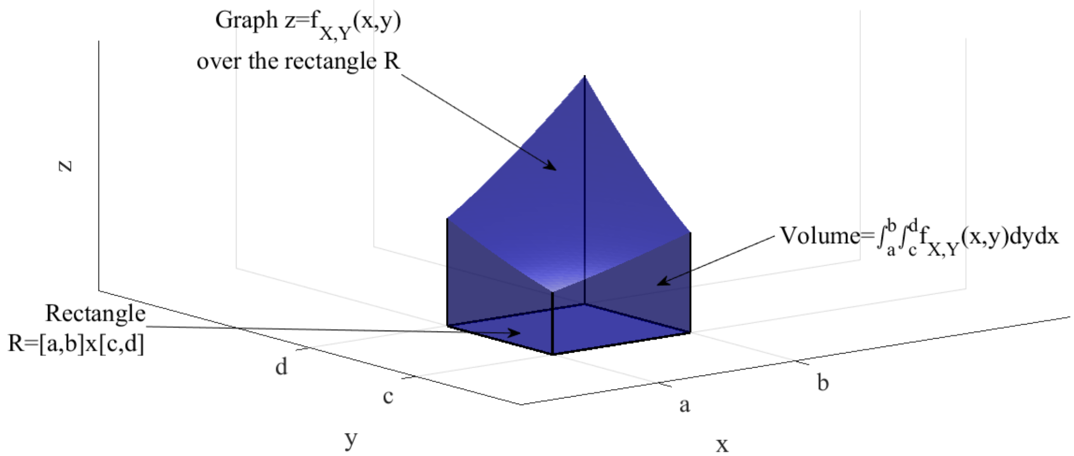

Figure 2: Graph of a function $f_{X,Y}(x,y)$ over the rectangle $[a,b]\times[c,d]$, and geometric interpretation of the double integral of this function over the rectangle R. The blue region between the graph of the function $f_{X,Y}$ and the $x-y$ plane, over the rectangle R, is given by $\int_a^b\int_c^d f_{X,Y}(x,y)dydx$.

2.2 Joint distribution function

The joint distribution function $F_{X,Y}$ is defined in a way that is analogous to the situation with discrete random variables:

For any $a\in\mathbb{R}$ and $b\in\mathbb{R}$,

$$\boxed{F_{X,Y}(a,b)=P(X\leq a,Y\leq b)=P(-\infty\leq X\leq a,-\infty\leq Y\leq b)=\int_{-\infty}^{a}\int_{-\infty}^{b}f_{X,Y}(x,y)dydx} \tag{4}$$

2.3 Marginal distribution functions

As for the discrete random variables, one can recover the individual cumulative distribution functions F_X and F_Y by summing over all y and over all x, respectively:

For any $a\in\mathbb{R}$,

$$\boxed{F_X(a)=P(X\leq a,Y\leq+\infty)=\int_{-\infty}^{a}\int_{-\infty}^{+\infty}f_{X,Y}(x,y)dydx=F_{X,Y}(a,+\infty)=\lim_{b\to+\infty}F_{X,Y}(a,b)}$$

For any $b\in\mathbb{R}$,

$$\boxed{F_Y(b)=P(X\leq+\infty,Y\leq b)=\int_{-\infty}^{b}\int_{-\infty}^{+\infty}f_{X,Y}(x,y)dxdy=F_{X,Y}(+\infty,b)=\lim_{a\to+\infty}F_{X,Y}(a,b)}$$

2.4 Marginal density functions

The results above give us a method for constructing the individual probability density functions f_X and f_Y from $f_{X,Y}$:

- For any $a\in\mathbb{R}$,

$$\boxed{f_X(a)=\frac{d}{da}\left(F_X(a)\right)}=\int_{-\infty}^{+\infty}f_{X,Y}(a,y)dy$$

Intuitively, this makes sense: we are summing over all y values, as we did for discrete random variables. Since we are dealing with *continuous* random variables, the summation becomes an integral.

- Likewise, for any $b \in \mathbb{R}$,

$$\boxed{f_Y(b) = \frac{d}{db}\left(F_Y(b)\right)} = \int_{-\infty}^{+\infty} f_{X,Y}(x,b)dx$$

  In other words, to get f_Y, we "sum" over all values of x (i.e. integrate).

In this context, f_X and f_Y are called **marginal probability density functions**.

2.5 From the joint distribution function to the joint probability density function

We have just seen that the joint distribution $F_{X,Y}$ is linked to the joint probability density function $f_{X,Y}$ through a double integral.

In order to find $f_{X,Y}$ from $F_{X,Y}$, we thus need to differentiate $F_{X,Y}$. Since $F_{X,Y}$ is a function of 2 variables and since $F_{X,Y}$ is the double integral of $f_{X,Y}$ with respect to x and y, we need to generalize the concept of derivative and use *partial derivatives*.

The partial derivative with respect to x is written $\frac{\partial}{\partial x}$. The partial derivative with respect to y is written $\frac{\partial}{\partial y}$. Note the rounded ∂ as opposed to the standard d, to remind us that we are using *partial* derivatives.

We have

$$\boxed{f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial F_{X,Y}}{\partial y}\right) = \frac{\partial}{\partial y}\left(\frac{\partial F_{X,Y}}{\partial x}\right)} \tag{5}$$

What are partial derivatives, and how to compute them ?

The partial derivative $\frac{\partial}{\partial x}$ means that we differentiate the function with respect to x, **treating y as a constant**.

<u>*Example*</u>: Consider the joint distribution function $F_{X,Y}$ given by

$$F_{X,Y}(x,y) = \begin{cases} 0 & \text{if } x \leq 0 \text{ or } y \leq 0 \\ x^2y^2 & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ x^2 & \text{if } 0 \leq x \leq 1 \text{ and } y \geq 1 \\ y^2 & \text{if } x \geq 1 \text{ and } 0 \leq y \leq 1 \\ 1 & \text{if } x \geq 1 \text{ and } y \geq 1 \end{cases}$$

Let us compute $f_{X,Y}(x,y) = \frac{\partial}{\partial x}\left(\frac{\partial F_{X,Y}}{\partial y}\right)$

- If $x \leq 0$ and $y \leq 0$,

$$\frac{\partial F_{X,Y}}{\partial y} = 0 \;\Rightarrow\; f_{X,Y}(x,y) = 0$$

- If $0 \leq x \leq 1$ and $0 \leq y \leq 1$,

$$\frac{\partial F_{X,Y}}{\partial y} = 2x^2y$$

  Hence

$$\frac{\partial}{\partial x}\left(\frac{\partial F_{X,Y}}{\partial y}\right) = \frac{\partial}{\partial x}(2x^2y) = 4xy = f_{X,Y}(x,y)$$

- If $0 \leq x \leq 1$ and $y \geq 1$,

$$\frac{\partial F_{X,Y}}{\partial y} = 0 \;\Rightarrow\; f_{X,Y}(x,y) = 0$$

- If $x \geq 1$ and $0 \leq y \leq 1$,

$$\frac{\partial F_{X,Y}}{\partial y} = 2y$$

  Therefore,

$$\frac{\partial}{\partial x}\left(\frac{\partial F_{X,Y}}{\partial y}\right) = \frac{\partial}{\partial x}(2y) = 0 \;\Rightarrow\; f_{X,Y}(x,y) = 0$$

- If $x \geq 1$ and $y \geq 1$,

$$\frac{\partial F_{X,Y}}{\partial y} = 0 \;\Rightarrow\; f_{X,Y}(x,y) = 0$$

We conclude that

$$f_{X,Y}(x,y) = \begin{cases} 4xy & \text{if } 0 \leq x \leq 1 \ , \ 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$