LECTURE 3: MULTIVARIATE RANDOM VARIABLES

MECO 7312.
INSTRUCTOR: DR. KHAI CHIONG
SEPTEMBER 4, 2024

Previously, we looked at univariate random variables, that is, the variable of interest is a scalar. Most of the time however, we are interested in the behavior of a vector. For instance, the behavior of (i) quantities and prices, (ii) employment and GDP, (iii) customer shopping frequency and spending, (iv) temperature and rainfall, (v) prices of multiple assets etc.

An n-dimensional random vector is a function from a sample space Ω into $\mathbb{R}^n$, the n-dimensional Euclidean space.

1. Pdf and pmf of bivariate random variables

1.1. Discrete case

Consider the experiment of tossing two fair dice. The sample space of this experiment is the set of all the possible outcomes. $\Omega = \{(1,1),(1,2),\dots,(2,1),\dots\}$, where $|\Omega| = 36$.

Define X = sum of the two dice, Y = |difference of the two dice|. In this way, we have defined the bivariate random vector (X, Y).

1.) What is $P(X = 6, Y = 0)$? The event $X = 6$ and $Y = 0$ occurs if and only if the two dice are 3. Hence, $P(X = 6, Y = 0) = \frac{1}{36}$.

2.) How about $P(X = 8, Y = 2)$?

$$P(X = 8, Y = 2) = \frac{1}{18}$$

3.) How about $P(X = 7, Y \leq 4)$?

$$P(X = 7, Y \leq 4) = \sum_{y=0}^{4} P(X = 7, Y = y) = \frac{4}{36} = \frac{1}{9}$$

First die, second die	1	2	3	4	5	6
1	$(2,0)$	$(3,1)$	$(4,2)$	$(5,3)$	$(6,4)$	$(7,5)$
2	$(3,1)$	$(4,0)$	$(5,1)$	$(6,2)$	$(7,3)$	$(8,4)$
3	$(4,2)$	$(5,1)$	$(6,0)$	$(7,1)$	$(8,2)$	$(9,3)$
4	$(5,3)$	$(6,2)$	$(7,1)$	$(8,0)$	$(9,1)$	$(10,2)$
5	$(6,4)$	$(7,3)$	$(8,2)$	$(9,1)$	$(10,0)$	$(11,1)$
6	$(7,5)$	$(8,4)$	$(9,3)$	$(10,2)$	$(11,1)$	$(12,0)$

TABLE 1. All possible outcomes of (X, Y). Each of the realization per cell is equally likely.

Let (X, Y) be a discrete bivariate random vector. Then the function $f(x, y)$ from $\mathbb{R}^2$ to $\mathbb{R}$ defined by $f(x, y) = P(X = x, Y = y)$ is called the joint probability mass function of (X, Y). The notation $f_{X,Y}(x, y)$ will also be used.

1.1.1. Marginal pmf

Given the joint pmf $f_{X,Y}(x, y)$, the marginal pmf of X denoted by $f_X(x)$ is given by:

$$f_X(x) = \sum_{y \in \mathbb{R}} f_{X,Y}(x, y)$$

Similarly, the marginal pmf of Y denoted by $f_Y(y)$ is given by:

$$f_Y(y) = \sum_{x \in \mathbb{R}} f_{X,Y}(x, y)$$

Consider the dice experiment above, what is $f_X(3) = P(X = 3)$?

$$f_X(3) = \sum_{y \in \mathbb{R}} f_{X,Y}(3, y) = \sum_{y} P(X = 3, Y = y) = P(X = 3, Y = 1) = \frac{1}{18}$$

1.2. Continuous case

A function $f(x, y)$ from $\mathbb{R}^2$ to $\mathbb{R}$ is called a joint probability density function or joint pdf of the continuous bivariate random vector (X, Y) if for every $A \subseteq \mathbb{R}^2$:

$$P((X, Y) \in A) = \int \int_A f_{X,Y}(x, y) \, dx \, dy$$

Any function $f(x,y)$ satisfying $f(x,y) \geq 0$ for all $(x,y) \in \mathbb{R}^2$ and

$$1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\,dx\,dy$$

is the joint pdf of some continuous bivariate random vector (X,Y).

Example: consider the following function.

$$f(x,y) = \begin{cases} 6xy^2 & 0 \leq x \leq 1, \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

The support of (X,Y) is the unit square. We check that $P((X,Y) \in \mathbb{R}^2) = 1$.

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\,dx\,dy &= \int_0^1 \int_0^1 6xy^2\,dx\,dy \\ &= \int_0^1 3y^2\,dy \\ &= 1 \end{aligned}$$

What is $P((X,Y) \in A)$, where A is the region defined by $A = \{(x,y) \in \mathbb{R}^2 : x \leq \frac{1}{2}, y \leq \frac{1}{2}\}$?

$$\begin{aligned} P((X,Y) \in A) &= P(X \leq \frac{1}{2}, Y \leq \frac{1}{2}) \\ &= \int_{-\infty}^{1/2} \int_{-\infty}^{1/2} f(x,y)\,dx\,dy \\ &= \int_0^{1/2} \int_0^{1/2} 6xy^2\,dx\,dy \\ &= \int_0^{1/2} \frac{3}{4} y^2\,dy \\ &= \frac{1}{32} \end{aligned}$$

We can visualize the joint pdf using Mathematica. We will see that geometric intuitions can be useful sometimes – we interpret $P((X,Y) \in A)$ as the volume underneath the curve $f(x,y)$ with respect to the region A.

Example: consider again the pdf $f(x,y) = 6xy^2$ with the support on the unit square. What is $P(X+Y \geq 1)$?

Let A be the region in 2-dimensional Euclidean space such that $A = \{(x,y) \in \mathbb{R}^2 : x+y \geq 1, 0 < x < 1, 0 < y < 1\}$. Essentially we are asking $P((X,Y) \in A)$. Graphically, A is the upper-right triangle of the unit square.

$$\begin{aligned} A &= \{(x,y) \in \mathbb{R}^2 : 1 \leq x+y, 0 < x < 1, 0 < y < 1\} \\ &= \{(x,y) \in \mathbb{R}^2 : 1-y \leq x < 1, 0 < y < 1\} \end{aligned}$$

Therefore,

$$\begin{aligned} P(X+Y \geq 1) = \int\int_A f(x,y)\,dx\,dy &= \int_0^1 \int_{1-y}^1 6xy^2\,dx\,dy \\ &= \int_0^1 [3x^2y^2]_{1-y}^1\,dy \\ &= \int_0^1 3y^2 - 3(1-y)^2y^2\,dy \\ &= \int_0^1 3y^2 - 3y^2 + 6y^3 - 3y^4\,dy \\ &= \left[\frac{3}{2}y^4 - \frac{3}{5}y^5\right]_0^1 \\ &= \frac{9}{10} \end{aligned}$$

Example: consider the following function.

$$f(x,y) = \begin{cases} 1 & 0 \leq x \leq 1, \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

This volume of this pdf is just the unit cube. Calculate $P(X^2+Y^2 \leq 1)$. First, we show using brute-force algebra that $P(X^2+Y^2 \leq 1) = \frac{\pi}{4}$, then we use a simple geometric argument that $P(X^2+Y^2 \leq 1) = \frac{\pi}{4}$.

$P(X^2+Y^2 \leq 1)$ equals to $P((X,Y) \in A)$ where $A = \{(x,y) \in \mathbb{R}^2 : x^2+y^2 \leq 1, x \in [0,1], y \in [0,1]\}$.

$$
\begin{aligned}
P(X^2+Y^2 \leq 1) &= \int\int_A f(x,y)\,dx\,dy \\
&= \int_0^1 \int_0^{\sqrt{1-y^2}} 1\,dx\,dy \\
&= \int_0^1 \sqrt{1-y^2}\,dy \\
&= \left[\frac{1}{2}\left(y\sqrt{1-y^2}+\sin^{-1}(y)\right)\right]_0^1 \\
&= \frac{\pi}{4}
\end{aligned}
$$

However, because the pdf has a uniform height of one with the support on the unit square, $P(X^2+Y^2 \leq 1)$ is just the volume of a cylinder split into 4 equal parts. Specifically, this cylinder has a height of one, and a radius of one.

1.2.1. Marginal pdf

The marginal pdf of X is defined as:

$$
f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy, \qquad \text{for } x \in \mathbb{R}
$$

The marginal pdf of Y is defined as:

$$
f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx, \qquad \text{for } y \in \mathbb{R}
$$

Example: consider again the pdf $f(x,y) = 6xy^2$ with the support on the unit square.

Derive the marginal pdf of X. Then, calculate $P(\frac{1}{2} < X < \frac{3}{4})$.

$$
f_X(x) = \int_0^1 6xy^2\,dy = \left[2xy^3\right]_0^1 = 2x, \quad \text{for } x \in [0,1]
$$

$$
P\left(\frac{1}{2} < X < \frac{3}{4}\right) = \int_{\frac{1}{2}}^{\frac{3}{4}} f_X(x)dx = \int_{\frac{1}{2}}^{\frac{3}{4}} 2x\,dx = \frac{5}{16}
$$

2. Joint cdf

The joint cdf of (X, Y) is defined as:

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$$

When (X, Y) is a continuous random vector, then

$$F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(s, t)\, ds\, dt$$

From the fundamental theorem of calculus, this implies that

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

The marginal cdf $F_X(x)$ can be obtained from $\lim_{y\to\infty} F(x, y) = F_X(x)$.

Example: consider the cdf:

$$F(x, y) = \begin{cases} 0 & x < 0 \text{ or } y < 0 \\ xy & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ x & 0 \leq x \leq 1, y > 1 \\ y & 0 \leq y \leq 1, x > 1 \\ 1 & x > 1, y > 1 \end{cases}$$

Therefore by calculating $f(x, y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$:

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

Also check that the marginal cdf $F_X(x)$ can be obtained as:

$$F_X(x) = \lim_{y\to\infty} F(x, y) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

3. Expectation

Let g be a function from $\mathbb{R}^2$ to $\mathbb{R}$. For the discrete case,

$$\mathbb{E}[g(X,Y)] = \sum_{(x,y)\in\mathbb{R}^2} g(x,y)P(X=x,Y=y)$$

Take $g(X,Y) = XY$. What is $\mathbb{E}[XY]$ in the dice experiment above?

$$\mathbb{E}[XY] = g(1,0)P(X=1,Y=0) + g(1,1)P(X=1,Y=1) + \dots$$

For the continuous case, we have:

$$\mathbb{E}[g(X,Y)] = \int\int_{(x,y)\in\mathbb{R}^2} g(x,y)f_{X,Y}(x,y)\,dx\,dy$$

Example:

Throw darts randomly at a unit square, record the x-coordinates and y-coordinates, and multiply them together. On average, what value would you expect?

In another words, consider the pdf $f(x,y) = 1$ with the support on the unit square. What is $\mathbb{E}[XY]$?

$$\begin{aligned}
\mathbb{E}[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\,dx\,dy \\
&= \int_0^1\int_0^1 xy\,dx\,dy \\
&= \int_0^1 y/2\,dy \\
&= \frac{1}{4}
\end{aligned}$$

What about $\mathbb{E}[X]$ or $\mathbb{E}[Y]$? Calculate the marginals first.

Example: What if we don't throw darts uniformly but try to aim away from the origin? Consider again the pdf $f(x,y) = 6xy^2$ with the support on the unit square.

$$\mathbb{E}[XY] = \int_0^1 \int_0^1 xy\, 6xy^2\, dx\, dy$$
$$= \int_0^1 2y^3\, dy$$
$$= \frac{1}{2}$$

How would we calculate $\mathbb{E}[X^2Y]$?

$$\mathbb{E}[X^2Y] = \int_0^1 \int_0^1 x^2y\, 6xy^2\, dx\, dy$$
$$= \frac{3}{8}$$

4. Conditional probabilities

Consider the bivariate random variables (X, Y). The random variable Y conditional on $X = x$ is denoted by $Y|X = x$. Now, $Y|X = x$ is another random variable , but it is a *scalar* random variable. The density of $Y|X = x$ is given by:

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

Example:

Consider the random variables (X, Y) that has the joint pdf $f(x, y) = 6xy^2$ for $(x, y) \in [0, 1]^2$. Consider the random variable $Y|X = 0.5$. This random variable is a scalar random variable. The pdf of $Y|X = 0.5$ is in terms of y only:

$$f_{Y|X=0.5}(y) = \frac{f_{X,Y}(0.5, y)}{f_X(0.5)}$$
$$= \frac{6(0.5)y^2}{2(0.5)}$$
$$= 3y^2 \quad \text{for } y \in [0, 1]$$

Now consider $Y|X$, which is a *bivariate* random variable, unlike $Y|X = x$, which is a scalar random variable. In particular, the joint density of $Y|X$ is:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

The two pdfs look identical, except in the pdf of $Y|X = x$, we treat x as fixed and as such, $f_{Y|X=x}(y)$ is a one-dimensional function. On the other hand, the joint pdf of $Y|X$ is a function of both x and y, and as such it is two-dimensional. That is, $f_{Y|X=x} : \mathbb{R} \to \mathbb{R}$, but $f_{Y|X} : \mathbb{R}^2 \to \mathbb{R}$.

Example:

Consider again the bivariate random variable (X, Y) that has the joint pdf $f(x, y) = 6xy^2$ for $(x, y) \in [0, 1]^2$. The joint density of $Y|X$ is given by:

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{f_{X,Y}(x,y)}{f_X(x)} \\ &= \frac{6xy^2}{2x} \\ &= \begin{cases} 3y^2 & \text{for } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Note: $f_{Y|X} : \mathbb{R}^2 \to \mathbb{R}$ is a function of both x and y, as the support of the function explicitly depends on x.

Example: Consider the joint density $f(x, y) = x+y$, with support on $(x, y) \in [0, 1]^2$. What is the joint density of $Y|X$?

First show that the marginal density of X is $f_X(x) = \frac{1}{2} + x$, for $x \in [0, 1]$. Therefore the conditional density is:

$$f_{Y|X}(y|x) = \begin{cases} \frac{2(x+y)}{1+2x} & (x, y) \in [0, 1]^2 \\ 0 & \text{otherwise} \end{cases}$$

4.1. Conditional expectation

Consider the random variable $Y|X = x$. The expectation $\mathbb{E}[Y|X = x]$ is defined as $\mathbb{E}[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X=x}(y)\, dy$. Note that $\mathbb{E}[Y|X = x]$ is a constant. In general, we have $\mathbb{E}[g(Y)|X = x] = \int_{-\infty}^{\infty} g(y) f_{Y|X=x}(y)\, dy$, for some function g.

Example: consider again the joint pdf $f(x, y) = x + y$ with the support given by $\{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 1, 0 \leq y \leq 1\}$. From the previous derivation, the conditional density is $f_{Y|X=x}f(y) = \frac{2(x+y)}{1+2x}$ for $(x, y) \in [0, 1]^2$.

$$\begin{aligned}
\mathbb{E}[Y|X=x] &= \int_0^1 y f_{Y|X=x} f(y)\, dy \\
&= \int_0^1 y \frac{2(x+y)}{1+2x}\, dy \\
&= \int_0^1 \frac{2xy}{1+2x} + \frac{2y^2}{1+2x}\, dy \\
&= \frac{x}{1+2x} + \frac{2}{3(1+2x)} = \frac{2+3x}{3+6x}
\end{aligned}$$

$E[Y|X=x]$ is treated as a constant. We check that $\mathbb{E}[Y|X=0]=2/3$, $\mathbb{E}[Y|X=1]=5/9$. $\mathbb{E}[Y|X=x]$ is decreasing in x, what is the geometric intuition behind this?

Now let $\mathbb{E}[Y|X=x]=g(x)$. Then we define $\mathbb{E}[X|Y]$ to be the random variable Z obtained by the transformation $Z=g(X)$. As such, $\mathbb{E}[Y|X]$ is a scalar random variable that has the same probability space as X. For this example, $\mathbb{E}[Y|X]$ is the random variable defined by the transformation $Z=\frac{2+3X}{3+6X}$. We can then derive the pdf of $Z \equiv \mathbb{E}[Y|X]$. In particular, the inverse of the transformation is $g^{-1}(z)=\frac{2-3z}{-3+6z}$, with $\frac{dg^{-1}(z)}{dz} = -\frac{1}{3(1-2z)^2}$. Therefore, $f_Z(z) = \left|\frac{dg^{-1}(z)}{dz}\right| f_X(g^{-1}(z)) = \frac{1}{3(1-2z)^2}(\frac{1}{2}+\frac{2-3z}{-3+6z})$ for $z \in [\frac{4}{9}, \frac{2}{3}]$.

Conditional expectation is important and useful later on. Suppose Y is an outcome variable of interest, and X is a variable that can be used to predict Y. An excellent predictor of Y as a function of $X=x$ is $\mathbb{E}[Y|X=x]$. This is optimal in a formal way.[1] For instance, Y is the transaction price of a house in the neighborhood and X is the square footage of the house. Then we can predict the price of a house when the square footage is 1000 as $\mathbb{E}[Y|X=1000]$.

5. Independence

If $X \sim f_X(x)$ and $Y \sim f_Y(y)$ are independent, then the joint pdf of (X,Y) is:

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

Furthermore, if a joint pdf $f_{X,Y}(x,y)$ can be factored as:

[1] If your loss function is a mean-squared error. That is, let $f(X)=\mathbb{E}[Y|X]$, then $f(X)$ minimizes the mean-squared error $\mathbb{E}[(Y-h(X))^2]$ among all possible functions $h(X)$.

$$f_{X,Y}(x,y) = g(x)h(y)$$

Then X and Y are independent random variables.

Example: consider again the joint pdf $f(x,y) = 6xy^2$ with the support on the unit square. Are X and Y independent? What about $f(x,y) = 1$ with the support on the unit square?

Consider the pdf $f(x,y) = 2$ with support on the triangle $\{(x,y) \in [0,1]^2 : x + y \leq 1\}$. Are X and Y independent?

6. Covariance and correlation

The covariance between X and Y is:

$$\begin{aligned} \text{Cov}(X,Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] \end{aligned}$$

Remember $\mathbb{E}[XY] = \int\int xyf(x,y)\,dx\,dy$.

The correlation between X and Y is:

$$\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

Which is bounded between $[-1,1]$.

A useful result is:

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X,Y)$$

It is also easy to show that if a is a constant, then $\text{Cov}(aX,Y) = a\text{Cov}(X,Y)$ and $\text{Cov}(X,a) = 0$. Further, $\text{Cov}(X,X) = \text{Var}(X)$. Moreover, $\text{Cov}(X + Z, Y) = \text{Cov}(X,Y) + \text{Cov}(Z,Y)$, which implies that if a is a constant, then $\text{Cov}(X + a, Y) = \text{Cov}(X,Y)$.

Show that when X and Y are independent, then $\text{Cov}(X,Y) = 0$. However the converse is not necessarily true! Zero covariance does not imply independence. Covariance only measures a linear relationship between X and Y. For example, consider a random variable X such that its first and third moments are zero. Now, if $Y = X^2$, then $\mathrm{Cov}(X, Y) = 0$. This means that covariance cannot capture non-linear relationship between random variables. Instead, it is a good idea to always plot the scatterplot and inspect any non-linearity in the scatterplots.

Example:

Consider the joint pdf $f(x, y) = 6xy^2$ with the support on the unit square. Recall that $\mathbb{E}[XY] = \int_0^1 \int_0^1 xy\, 6xy^2 \, dx \, dy = \frac{1}{2}$. Moreover, $\mathbb{E}[X] = \frac{2}{3}$ and $\mathbb{E}[Y] = \frac{3}{4}$. Therefore, $\mathrm{Cov}(X, Y) = 0$.

Similar calculations can be done for the discrete case:

$$\mathbb{E}[XY] = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} xyP(X = x, Y = y)$$

Finally, recall the joint pdf $f(x, y) = x + y$ with the support on $\{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 1, 0 \leq y \leq 1\}$.

Previously, we found that the marginal density of Y is $f_Y(y) = \frac{1}{2} + y$ for $y \in [0, 1]$. As such, $\mathbb{E}[Y] = \frac{1}{4} + \frac{1}{3} = \frac{7}{12}$.

$$\begin{aligned}
\mathbb{E}[XY] &= \int_0^1 \int_0^1 xyf(x, y) \, dx \\
&= \int_0^1 \int_0^1 xy(x + y) \, dx \, dy \\
&= \frac{1}{3}
\end{aligned}$$

Therefore the covariance between X and Y is $-\frac{1}{144}$, which is negative. This number seems small, because it has not been normalized with the scale of (X, Y). We can also show that $\mathrm{Var}(Y) = 11/144$, and $\mathrm{Var}(X) = 11/144$.[2] Hence, the correlation between (X, Y) is $-\frac{1}{11}$. Does this make geometric sense?

[2] Note that the pdf is symmetric in x and y.