LECTURE 3: MULTIVARIATE RANDOM VARIABLES

MECO 7312.
INSTRUCTOR: DR. KHAI CHIONG
SEPTEMBER 15, 2021

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. $\Omega = \{(1,1),(1,2),\ldots,(2,1),\ldots\}$, 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, $P(X = 6, Y = 0) = \frac{1}{36}$.

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.

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

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

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

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.

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

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 $f_X(3) = P(X = 3)$?

$$\begin{aligned} f_X(3) &= \sum_{y \in \mathbb{R}} f_{X,Y}(3, y) \\ &= \sum_{y=0}^{5} P(X = 3, Y = y) \\ &= P(X = 3, Y = 1) \\ &= \frac{1}{18} \end{aligned}$$

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.

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

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.

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

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.

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

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 $A = \{X^2 + Y^2 \leq 1\}$.

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

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 $P(\frac{1}{2} < X < \frac{3}{4})$.

$$f_X(x) = \int_0^1 6xy^2\, dy = 2x$$

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

2. Joint cdf

The bivariate 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

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

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 from $\lim_{y\to\infty} F(x, y) = F_X(x) = x$.

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 $\mathbb{E}[XY]$?

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

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.

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

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

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

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:

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

Example:

Consider the random variables (X, Y) that has the joint pdf $f(x, y) = 6xy^2$ for $0 < x < 1$ and $0 < y < 1$. 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:

$$\begin{aligned}f_{Y|X=0.5}(y) &= \frac{f_{X,Y}(0.5, y)}{f_X(0.5)} \\ &= \frac{6(0.5)y^2}{2(0.5)} \\ &= 3y^2 \quad \text{for } 0 < y < 1\end{aligned}$$

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 so $f_{Y|X=x}(y|x)$ is a one-dimensional function. While in the pdf of $Y|X$, we treat x to be an argument of the function, so that the pdf of $Y|X$ 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 $0 < x < 1$ and $0 < y < 1$. The joint density of $Y|X$ is given by:

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

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

Example: consider the problem of throwing darts at a *square* centered at the origin, i.e. $\{(x, y) \in \mathbb{R}^2 : -1 \le x \le 1, -1 \le y \le 1\}$.

Now suppose we tend to throw with a bias, and our darts tend to land closer to the positive diagonal line. We can model this as $f(x, y) \propto 4 - (x - y)^2$ with the support on $\{(x, y) \in \mathbb{R}^2 : -1 \le x \le 1, -1 \le y \le 1\}$. Show that the constant of proportionality is just 40/3, therefore the pdf is $f(x, y) = \frac{3}{40}(4 - (x - y)^2)$.

Show that the marginal density of Y is $f_Y(y) = \frac{1}{20}(11 - 3y^2)$. Therefore the conditional density is:

$$
f_{X|Y}(x|y) = \begin{cases} \frac{3\left((x-y)^2-4\right)}{6y^2-22} & (x, y) \in [-1, 1]^2 \\ 0 & \text{otherwise} \end{cases}
$$

We can also check that the density of $X|Y = 0$ is $f_{X|Y=0}(x) = 3(4 - x^2)/22$ for $x \in [-1, 1]$.

4.1. Conditional expectation

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

Example: consider again the joint pdf $f(x, y) = 6xy^2$ for $0 < x < 1$ and $0 < y < 1$. The density of $Y|X = x$ is $f_{Y|X=x}(y|x) = 3y^2$ for $0 < y < 1$. What is $\mathbb{E}[X|Y = y]$?

$$\begin{aligned}\mathbb{E}[Y|X=x] &= \int_{-\infty}^{\infty} y f_{Y|X=x}(y|x)\,dy \\ &= \int_0^1 3y^3\,dy \\ &= \frac{3}{4}\end{aligned}$$

It just so happened that $\mathbb{E}[Y|X=x]$ above does not depend on what x is. In general, $\mathbb{E}[Y|X=x]$ is parameterized by x (but we cannot have y appearing in the expression).

Now let $\mathbb{E}[Y|X=x]=f(x)$. Then we define $\mathbb{E}[X|Y]$ to be the random variable obtained by the transformation $f(X)$. As such, $\mathbb{E}[Y|X]$ is a scalar random variable that has the same probability space as X.

Example: consider again the joint pdf $f(x,y)=\frac{3}{40}(4-(x-y)^2)$ with the support given by $\{(x,y)\in\mathbb{R}^2 : -1\le x\le 1, -1\le y\le 1\}$. Recall that the conditional density is $f_{X|Y=y}f(x|y)=\frac{3\left((x-y)^2-4\right)}{6y^2-22}$ for $x\in[-1,1]$

$$\begin{aligned}\mathbb{E}[X|Y=y] &= \int_{-1}^{1} x f_{X|Y=y}(x|y)\,dx \\ &= \int_{-1}^{1} x\frac{3\left((x-y)^2-4\right)}{6y^2-22}\,dx \\ &= \frac{2y}{11-3y^2}\end{aligned}$$

$E[X|Y=y]$ is treated as a constant. We check that $\mathbb{E}[X|Y=0]=0$, $\mathbb{E}[X|Y=0.5]\approx 0.1$ and $\mathbb{E}[X|Y=1]=0.25$. Now consider $\mathbb{E}[X|Y]$, which is a random variable defined by the transformation $Z=\frac{2Y}{11-3Y^2}$. We can then derive the pdf of $\mathbb{E}[X|Y]$ from the marginal pdf $f_Y(y)=\frac{1}{20}\left(11-3y^2\right)$, $y\in[-1,1]$.

Conditional expectation is very 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 $\mathbb{E}[Y|X=1000]$.

[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 $h(X)$.

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:

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

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

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:

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

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 (roughly) a linear relationship between X and Y. We can easily construct a random variable X such that $\text{Cov}(X, X^2) = 0$ (for example, any random variable such that the first and the third moments are zero), which 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, $\text{Cov}(X, Y) = 0$.

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 $f(x, y) = \frac{3}{40}(4 - (x - y)^2)$ with the support on $\{(x, y) \in \mathbb{R}^2 : -1 \leq x \leq 1, -1 \leq y \leq 1\}$. We expect the covariance between X and Y to be positive.

Previously, we found that the marginal density of Y is $f_Y(y) = \frac{1}{20}(11 - 3y^2)$. Now it turns out that $\mathbb{E}[Y] = 0$.

$$\begin{aligned}\mathbb{E}[XY] &= \int_{-1}^{1} \int_{-1}^{1} xyf(x, y)\, dx \\ &= \int_{-1}^{1} \int_{-1}^{1} xy \frac{3}{40}(4 - (x - y)^2)\, dx\, dy \\ &= \frac{1}{15}\end{aligned}$$

Therefore the covariance between X and Y is $\frac{1}{15}$, which is strictly positive. This number seems small, because it has not been normalized with the scale of (X, Y). We can also show that $\text{Var}(Y) = 23/75$, and $\text{Var}(X) = 23/75$.[2] Hence, the correlation between (X, Y) is $5/23 \approx 0.217$.

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

7. Important identities

7.1. Law of iterated expectations

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]]$$

Now $\mathbb{E}[Y|X]$ is a scalar random variable, and inhabits the same probability space as X. Therefore, the outer expectation on the right-hand side is taken with respect to $f_X(x)$.

$$\begin{aligned}\mathbb{E}[Y|X = x] &= \int_{-\infty}^{\infty} y f_{Y|X=x}(y|x)\, dy \\ &= g(x)\end{aligned}$$

$$\begin{aligned}\mathbb{E}[\mathbb{E}[Y|X]] &= \mathbb{E}[g(X)] \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} y f_{Y|X}(y|x)\, dy \right) f(x)\, dx\end{aligned}$$

Intuitively, suppose we use realizations of the variable X to predict Y. Then the average of the predicted values over X equals to the average of Y. Note that it is usually more customary to use Y to denote the dependent variable (such as demand/sales), and X to denote the explanatory variables (such as price).

Example:

Recall the pdf $f(x, y) = \frac{3}{40}(4 - (x - y)^2)$ with the support on $\{(x, y) \in \mathbb{R}^2 : -1 \leq x \leq 1, -1 \leq y \leq 1\}$. Previously, we found that:

$$\mathbb{E}[Y|X] = \frac{2X}{11 - 3X^2}$$

Therefore,

$$\begin{aligned}\mathbb{E}[\mathbb{E}[Y|X]] &= \int \frac{2x}{11 - 3x^2} f_X(x)\, dx \\ &= \int_{-1}^{1} \frac{2x}{11 - 3x^2} \frac{1}{20} \left(11 - 3x^2\right)\, dx \\ &= 0 \\ &= \mathbb{E}[Y]\end{aligned}$$

7.2. Important properties of conditional expectations

This section is adapted from Chapter 2 of "Econometric Analysis of Cross Section and Panel Data" by Jeffrey M. Wooldridge.

Let Y, W be random variables. Let X be the random variable such that $X = g(W)$, for some function g.

Comparing $\mathbb{E}[Y|W]$ and $\mathbb{E}[Y|X]$, we can think of $\mathbb{E}[Y|X]$ as conditioning on a set of events that is a subset of the set of events being conditioned on in $\mathbb{E}[Y|W]$. Because if we know the outcome of W, then we would know X, but the converse is not true.

$$\mathbb{E}[\mathbb{E}[Y|X]|W] = \mathbb{E}[Y|X] \tag{1}$$

$$\mathbb{E}[\mathbb{E}[Y|W]|X] = \mathbb{E}[Y|X] \tag{2}$$

A phrase useful for remembering both equations above: "The smaller information set always dominates". This is also known as the Tower Property of conditional expectations, which can be demonstrated more formally with measure-theoretic notations.

Some consequences of this useful property:

$$\mathbb{E}[\mathbb{E}[Y|X]|X^2] = \mathbb{E}[\mathbb{E}[Y|X^2]|X] = \mathbb{E}[Y|X^2] \tag{3}$$

$$\mathbb{E}[\mathbb{E}[Y|X,Z]|X] = \mathbb{E}[\mathbb{E}[Y|X]|X,Z] = \mathbb{E}[Y|X] \tag{4}$$

7.3. Conditional variance identity

$$\text{Var}(Y) = \mathbb{E}[\text{Var}(Y|X)] + \text{Var}(\mathbb{E}[Y|X])$$

$\mathbb{E}[Y|X]$ and $\text{Var}(Y|X)$ are each scalar random variable that is a transformation of X and has the same probability space as X. Therefore, the expectation and variance on the right-hand side is taken with respect to the pdf $f_X(x)$.

Example:

Using the same example as before, we have the pdf $f(x,y) = \frac{3}{40}(4-(x-y)^2)$ with the support on $\{(x,y) \in \mathbb{R}^2 : -1 \leq x \leq 1, -1 \leq y \leq 1\}$.

$$\begin{aligned}
\mathbb{E}[Y|X] &= \frac{2X}{11-3X^2} \\
\text{Var}(\mathbb{E}[Y|X]) &= \mathbb{E}[(\mathbb{E}[Y|X])^2] - (\mathbb{E}[\mathbb{E}[Y|X]])^2 \\
&= \int_{-1}^{1} \left(\frac{2x}{11-3x^2}\right)^2 f_X(x)\,dx - \mathbb{E}[Y]^2 \\
&= \int_{-1}^{1} \left(\frac{2x}{11-3x^2}\right)^2 \frac{1}{20}\left(11-3x^2\right)\,dx \\
&= 0.01459
\end{aligned}$$

We can derive $\text{Var}(Y|X)$ by:

$$\begin{aligned}
\text{Var}[Y|X=x] &= \mathbb{E}[Y^2|X=x] - (\mathbb{E}[Y|X=x])^2 \\
&= \int_{-1}^{1} y^2 f_{Y|X=x}(y|x)\,dy - (\mathbb{E}[Y|X=x])^2 \\
&= \int_{-1}^{1} y^2 \frac{3\left((x-y)^2-4\right)}{6x^2-22}\,dy - \left(\frac{2x}{11-3x^2}\right)^2 \\
&= \frac{15x^4-126x^2+187}{5\left(11-3x^2\right)^2} \\
&= h(x)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[\text{Var}[Y|X]] &= \int_{-1}^{1} h(x) f_X(x)\,dx \\
&= 0.29207
\end{aligned}$$

Therefore, $\mathbb{E}[\text{Var}[Y|X]] + \text{Var}(\mathbb{E}[Y|X]) = 0.29207 + 0.01459 = 0.306667 \approx \text{Var}(Y)$.

8. Example: putting everything together

Suppose X and Y are distributed uniformly on the triangle $(0,0), (0,1), (1,0)$. That is:

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

1.) Is this a valid pdf?

2.) Derive the marginal pdfs.

$$f_X(x) = \int_0^{1-x} 2\,dy = 2(1-x) \text{ for } x \in [0,1]$$

$$f_Y(y) = \int_0^{1-y} 2\,dy = 2(1-y) \text{ for } y \in [0,1]$$

3.) Calculate $\text{Cov}(X,Y)$

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

$$\mathbb{E}[X] = \int_0^1 2(1-x)x\,dx = \frac{1}{3}$$

$$\mathbb{E}[Y] = \int_0^1 2(1-y)y\,dy = \frac{1}{3}$$

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

Hence $\text{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = \frac{1}{12} - (\frac{1}{3})(\frac{1}{3}) = -\frac{1}{36}$

4.) Calculate $P(Y \leq 1-2X)$:

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

5.) Derive $\mathbb{E}[Y|X = x]$ and $\text{Var}(Y|X = x)$:

First, the density of $Y|X$:

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f(x)} = \frac{2}{2(1-x)}, \text{ for } 0 \leq x \leq 1,\ 0 \leq y \leq 1,\ x + y \leq 1$$

Conditional expectation:

$$\mathbb{E}(Y|X = x) = \int_0^{1-x} y f_{Y|X=x}(y|x)\,dy = \int_0^{1-x} \frac{y}{(1-x)}\,dy = \frac{1-x}{2}$$

Conditional variance:

$$\begin{aligned} \text{Var}(Y|X = x) &= \mathbb{E}[Y^2|X = x] - \mathbb{E}[Y|X = x]^2 \\ &= \int_0^{1-x} y^2 f_{Y|X=x}(y|x)\,dy - \left(\frac{1-x}{2}\right)^2 \\ &= \frac{1}{3}(1-x)^2 - \left(\frac{1-x}{2}\right)^2 \\ &= \frac{1}{12}(1-x)^2 \end{aligned}$$

6.) Derive $\text{Var}(\mathbb{E}[Y|X])$ and $\mathbb{E}[\text{Var}(Y|X)]$:

$$\begin{aligned}\operatorname{Var}(\mathbb{E}[Y|X]) &= \mathbb{E}[(\mathbb{E}[Y|X])^2] - \mathbb{E}[\mathbb{E}[Y|X]]^2 \\ &= \int_0^1 \left(\frac{1-x}{2}\right)^2 2(1-x)\,dx - \mathbb{E}[Y] \\ &= \frac{1}{8} - \frac{1}{9} \\ &= \frac{1}{72}\end{aligned}$$

$$\begin{aligned}\mathbb{E}(\operatorname{Var}[Y|X]) &= \int_0^1 \frac{1}{12}(1-x)^2 \cdot 2(1-x)\,dx \\ &= \frac{1}{24}\end{aligned}$$

Indeed, we see that the Conditional Variance Identity holds true here. $\operatorname{Var}(Y) = \mathbb{E}[\operatorname{Var}(Y|X)] + \operatorname{Var}(\mathbb{E}[Y|X])$, where $\operatorname{Var}(Y) = \int_0^1 y^2 2(1-y)\,dy - \frac{1}{9} = \frac{1}{18}$.

9. Transformation of bivariate random variables

Let (X, Y) be a bivariate random vector. Consider a new bivariate random vector (U, V) defined by $U = g_1(X, Y)$, $V = g_2(X, Y)$. What is the probability distribution of (U, V)?

Let $\mathcal{A}$ denote the support of the (X, Y), i.e. $\mathcal{A} = \{(x, y) \in \mathbb{R}^2 : f_{X,Y}(x, y) > 0\}$.

The transformation is $U = g_1(X, Y)$ and $V = g_2(X, Y)$. The support of (U, V) is then $\mathcal{B} = \{(u, v) \in \mathbb{R}^2 : u = g_1(x, y), v = g_2(x, y) \text{ for some } (x, y) \in \mathcal{A}\}$.

Assume that g_1 and g_2 are functions such that the relationship between $\mathcal{A}$ and $\mathcal{B}$ is one-to-one and onto (a bijection). For each $(u, v) \in \mathcal{B}$, there is only one $(x, y) \in \mathcal{A}$ such that $u = g_1(x, y)$ and $v = g_2(x, y)$.

As such, we can solve the equations $u = g_1(x, y)$ and $v = g_2(x, y)$ in terms of x and y. That is, there is an inverse transformation such that $x = h_1(u, v)$ and $y = h_2(u, v)$, where h_1 and h_2 are differentiable functions.

Define the Jacobian matrix:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial h_1}{\partial u} & \frac{\partial h_1}{\partial v} \\ \frac{\partial h_2}{\partial u} & \frac{\partial h_2}{\partial v} \end{bmatrix}$$

The determinant of the Jacobian matrix is:

$$J = \det(\mathbf{J}) = \begin{vmatrix} \frac{\partial h_1}{\partial u} & \frac{\partial h_1}{\partial v} \\ \frac{\partial h_2}{\partial u} & \frac{\partial h_2}{\partial v} \end{vmatrix}$$

That is, $J = \frac{\partial h_1}{\partial u}\frac{\partial h_2}{\partial v} - \frac{\partial h_1}{\partial v}\frac{\partial h_2}{\partial u}$.

The joint pdf of (U, V) is:

$$f_{U,V}(u,v) = \begin{cases} f_{X,Y}(h_1(u,v), h_2(u,v))\,|J| & \text{for } (u,v) \in \mathcal{B} \\ 0 & \text{otherwise} \end{cases}$$

$|\det(\mathbf{J})|$ is often called the Jacobian, or the Jacobian of the transformation, or the Jacobian determinant. Note that $\det(\mathbf{J})$ is a function of u, v. Moreover, $\det(\boldsymbol{J}) \neq 0$ since there is an inverse transformation such that $x = h_1(u,v)$ and $y = h_2(u,v)$, where h_1 and h_2 are differentiable functions. The Jacobian is also used during change-of-variables in multiple integrals.

9.1. Example

Let X and Y be independent, standard Normal random variables.

Consider the transformation $U = X + Y$ and $V = X - Y$. What is the joint pdf of (U, V)?

The joint pdf of (X, Y) is just $f_{X,Y}(x,y) = f_X(x)f_Y(y) = \frac{1}{2\pi}e^{-\frac{x^2}{2}}e^{-\frac{y^2}{2}}$ since X and Y are independent.

The support of (X, Y) is $\mathbb{R}^2$. It follows that U and V can also take any value from $-\infty$ to ∞.

The inverse transformation is $x = h_1(u,v) = \frac{u+v}{2}$ and $y = h_2(u,v) = \frac{u-v}{2}$.

The Jacobian of the transformation is:

$$J = \begin{vmatrix} \frac{\partial h_1}{\partial u} & \frac{\partial h_1}{\partial v} \\ \frac{\partial h_2}{\partial u} & \frac{\partial h_2}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}$$

Hence the joint pdf of (U, V) is:

$$\begin{aligned} f_{U,V}(u,v) &= f_{X,Y}(h_1(u,v), h_2(u,v))\,|J| \\ &= \frac{1}{2\pi} e^{-\frac{(\frac{u+v}{2})^2}{2}} e^{-\frac{(\frac{u-v}{2})^2}{2}} \frac{1}{2} \\ &= \left(\frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{u^2}{4}} \right) \left(\frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{v^2}{4}} \right) \end{aligned}$$

Note that the pdf of $N(\mu, \sigma^2)$ is $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

Hence the joint pdf of (U, V) can be factored into two functions $f_U(u)$ and $f_V(v)$. Moreover, $f_U(u)$ is the pdf of $N(0, 2)$. That is, $U \sim N(0, 2)$ and $V \sim N(0, 2)$. The sum (U) and difference (V) of independent normal random variables are independent normal random variables, as long as $\text{Var}(X) = \text{Var}(Y)$.

We can also consider the ratio and the product of Normal variables. Consider the transformation $U = X/Y$ and $V = X$. What is the joint pdf of (U, V)? What about the product $V = XY$?

9.2. Discrete bivariate random vectors

Let (X, Y) be a a discrete bivariate random vector. Let $\mathcal{A}$ be the support of (X, Y), i.e. the set of points where the joint pmf of (X, Y) takes strictly positive values. Note that $\mathcal{A}$ must be a countable set (either finite or countably infinite).

The joint pmf of (U, V) is:

$$f_{U,V}(u,v) = P(U = v, V = v) = \sum_{(x,y)\in\mathcal{A}: g_1(x,y)=u, g_2(x,y)=v} f_{X,Y}(x,y)$$