Néhémy Lim

Bivariate Distributions

1 Distributions of Two Random Variables

Definition 1.1. *Let X and Y be two rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$. A **two-dimensional random variable** (X, Y) is a function mapping $(X, Y) : \Omega \to \mathbb{R}^2$, such that for any numbers $x, y \in \mathbb{R}$:*

$$\{\omega \in \Omega \mid X(\omega) \leq x \text{ and } Y(\omega) \leq y\} \in \mathcal{A} \tag{1}$$

Definition 1.2 (Joint cumulative distribution function)**.** *Let X and Y be two rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The **joint (cumulative) distribution function** (joint c.d.f.) of X and Y is the function F_{XY} given by*

$$F_{XY}(x, y) = \mathbb{P}\left(\{X \leq x\} \cap \{Y \leq y\}\right) \triangleq \mathbb{P}\left(X \leq x, Y \leq y\right), \tag{2}$$

for $x, y \in \mathbb{R}$

We first consider the case of two discrete rrvs.

1.1 Distributions of Two Discrete Random Variables

Example 1. Here is an excerpt of an article published in the *Time* and entitled *Why Gen Y Loves Restaurants – And Restaurants Love Them Even More* :

“According to a new report from the research firm Technomic, 42% of millennials say they visit “upscale casual-dining restaurants” at least once a month. That’s a higher percentage than Gen X (33%) and Baby Boomers (24%) who go to such restaurants once or more monthly.”

Time, Aug. 15, 2012, By Brad Tuttle

We can find the populations of each category provided by the US Census Bureau :

Age	Population
0-19	83,267,556
20-34 (Millenials)	**62,649,947**
35-49 (Gen X)	**63,779,197**
50-69 (Baby Boomers)	**71,216,117**
70+	27,832,721

Table 1: Demography in the US by age group [*US Census data*]

Let X and Y be two random variables defined as follows :

- $X = 1$: a person between 20 and 69 visits upscale restaurants at least once a month and $X = 0$ otherwise
- $Y = 1$: a person between 20 and 69 is a millenial
- $Y = 2$: a person between 20 and 69 is a Gen X
- $Y = 3$: a person between 20 and 69 is a Baby Boomer

We can translate the statement of the article in the form of a contingency table shown below :

$X \backslash Y$	1	2	3
0	36,337	42,732	47,003
1	26,313	21,047	24,213

Table 2: Count Data ($\times 1000$)

The probability that the couple (X, Y) takes on particular values can be found by dividing each cell by the total population of people between 20 and 69.

$X \backslash Y$	1	2	3
0	0.184	0.216	0.238
1	0.133	0.106	0.123

Table 3: Joint Probability Mass Function $p_{XY}(x, y)$

Now, let us define formally the joint probability mass function of two discrete random variables X and Y.

Definition 1.3. *Let X and Y be two discrete rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The **joint pmf** of X and Y, denoted by p_{XY}, is defined as follows :*

$$p_{XY}(x, y) = \mathbb{P}\left(\{X = x\} \cap \{Y = y\}\right) \triangleq \mathbb{P}\left(X = x, Y = y\right), \tag{3}$$

for $x \in X(\Omega)$ and $y \in Y(\Omega)$

Property 1.1. *Let X and Y be two discrete rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with joint pmf p_{XY}, then the following holds :*

- $p_{XY}(x, y) \geq 0$*, for $x \in X(\Omega)$ and $y \in Y(\Omega)$*
- $\sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} p_{XY}(x, y) = 1$

Property 1.2. *Let X and Y be two disctete rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with joint pmf p_{XY}. Then, for any subset $S \subset X(\Omega) \times Y(\Omega)$*

$$\mathbb{P}\left((X, Y) \in S\right) = \sum_{(x,y) \in S} p_{XY}(x, y)$$

The above property tells us that in order to determine the probability of event $\{(X, Y) \in S\}$, you simply sum up the probabilities of the events $\{X = x, Y = y\}$ with values (x, y) in S.

Definition 1.4 (Marginal distributions)**.** *Let X and Y be two discrete rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with joint pmf p_{XY}. Then the pmf of X alone is called the **marginal probability mass function** of X and is defined by:*

$$p_X(x) = \mathbb{P}(X = x) = \sum_{y \in Y(\Omega)} p_{XY}(x, y), \quad \textit{for } x \in X(\Omega) \tag{4}$$

*Similarly, the pmf of Y alone is called the **marginal probability mass function** of Y and is defined by:*

$$p_Y(y) = \mathbb{P}(Y = y) = \sum_{x \in X(\Omega)} p_{XY}(x, y), \quad \textit{for } y \in Y(\Omega) \tag{5}$$

Definition 1.5 (Independence)**.** *Let X and Y be two discrete rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with joint pmf p_{XY}. Let p_X and p_Y be the respective marginal pmfs of X and Y. Then X and Y are said to be **independent** if and only if :*

$$p_{XY}(x, y) = p_X(x) p_Y(y), \quad \textit{for all } x \in X(\Omega) \textit{ and } y \in Y(\Omega) \tag{6}$$

Definition 1.6 (Expected Value)**.** *Let X and Y be two discrete rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with joint pmf p_{XY} and let $g : \mathbb{R}^2 \to \mathbb{R}$ be a bounded piecewise continuous function. Then, the mathematical expectation of $g(X, Y)$, if it exists, is denoted by $\mathbb{E}[g(X, Y)]$ and is defined as follows :*

$$\mathbb{E}[g(X, Y)] = \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} g(x, y) p_{XY}(x, y) \tag{7}$$

Example 2. Consider the following joint probability mass function :

$$p_{XY}(x,y) = \frac{xy^2}{13}\mathbb{1}_S(x,y)$$

with $S = \{(1,1),(1,2),(2,2)\}$

1. Show that p_{XY} is a valid joint probability mass function.

   Answer. The joint probability mass function of X and Y is given by the following table :

$X \backslash Y$	1	2
1	$\frac{1}{13}$	$\frac{4}{13}$
2	0	$\frac{8}{13}$

   We first note that $p_{XY}(x,y) \geq 0$ for all $x,y = 1,2$. Second,

$$\sum_{x=1}^{2}\sum_{y=1}^{2} p_{XY}(x,y) = \frac{1}{13} + \frac{4}{13} + 0 + \frac{8}{13} = 1$$

   Hence, p_{XY} is indeed a valid joint probability mass function.

2. What is $\mathbb{P}(X+Y \leq 3)$?

   Answer. Let us denote B the set of values such that $X+Y \leq 3$:

$$B = \{(1,1),(1,2),(2,1)\}$$

   Therefore,

$$\begin{aligned}
\mathbb{P}(X+Y \leq 3) &= \sum_{(x,y)\in B} p_{XY}(x,y) \\
&= p_{XY}(1,1) + p_{XY}(1,2) + p_{XY}(2,1) \\
&= \frac{1}{13} + \frac{4}{13} + 0 \\
&= \frac{5}{13}
\end{aligned}$$

3. Give the marginal probability mass functions of X and Y.

   Answer. The marginal probability mass function of X is given by :

$$p_X(1) = \sum_{y=1}^{2} p_{XY}(1,y) = p_{XY}(1,1) + p_{XY}(1,2) = \frac{1}{13} + \frac{4}{13} = \frac{5}{13}$$

$$p_X(2) = \sum_{y=1}^{2} p_{XY}(2,y) = p_{XY}(2,2) + p_{XY}(2,2) = 0 + \frac{8}{13} = \frac{8}{13}$$

Similarly, the marginal probability mass function of Y is given by :

$$p_Y(1) = \sum_{x=1}^{2} p_{XY}(x,1) = p_{XY}(1,1) + p_{XY}(2,1) = \frac{1}{13} + 0 = \frac{1}{13}$$
$$p_Y(2) = \sum_{x=1}^{2} p_{XY}(x,2) = p_{XY}(1,2) + p_{XY}(2,2) = \frac{4}{13} + \frac{8}{13} = \frac{12}{13}$$

4. What are the expected values of X and Y?

   Answer. The expected value of X is given by :

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=1}^{2}\sum_{y=1}^{2} x p_{XY}(x,y) \\
&= \sum_{x=1}^{2} x \left\{ \sum_{y=1}^{2} p_{XY}(x,y) \right\} \\
&= \sum_{x=1}^{2} x p_X(x) \\
&= 1 \cdot \frac{5}{13} + 2 \cdot \frac{8}{13} \\
&= \frac{21}{13}
\end{aligned}$$

   Similarly, the expected value of Y is given by :

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_{x=1}^{2}\sum_{y=1}^{2} y p_{XY}(x,y) \\
&= \sum_{y=1}^{2} y \left\{ \sum_{x=1}^{2} p_{XY}(x,y) \right\} \\
&= \sum_{x=1}^{2} y p_Y(y) \\
&= 1 \cdot \frac{1}{13} + 2 \cdot \frac{12}{13} \\
&= \frac{25}{13}
\end{aligned}$$

5. Are X and Y independent?

   Answer. It is easy to see that X and Y are not independent, since, for example:

$$p_X(2)p_Y(1) = \frac{8}{13} \cdot \frac{1}{13} = \frac{8}{169} \neq 0 = p_{XY}(2,1)$$

1.2 The Trinomial Distribution (Optional section)

Example 3. Suppose $n = 20$ students are selected at random:

- Let A be the event that a randomly selected student went to the football game on Saturday. Let $\mathbb{P}(A) = 0.2 = p_1$
- Let B be the event that a randomly selected student watched the football game on TV on Saturday. Let $\mathbb{P}(B) = 0.5 = p_2$
- Let C be the event that a randomly selected student completely ignored the football game on Saturday. Let $\mathbb{P}(C) = 0.3 = 1 - p_1 - p_2$

One possible outcome is:

$$BBCABBAACABBBCCBCBCB$$

That is, the first two students watched the game on TV, the third student ignored the game, the fourth student went to the game, and so on. By independence, the probability of observing that particular outcome is :

$$p_1^4 \, p_2^{10} \, (1 - p_1 - p_2)^{20-4-10}$$

Now, let us denote the three following random variables :

- X is the number in the sample who went to the football game on Saturday,
- Y is the number in the sample who watched the football game on TV on Saturday,
- Z the number in the sample who completely ignored the football game

Then in this case: $X = 4\,, Y = 10$ and thus $Z = 20 - X - Y = 6$. One way of having $(X = 4, Y = 10)$ is given by the above outcome. Actually, there are many ways of observing $(X = 4, Y = 10)$. We find all the possible outcomes by giving all the permutations of the A's, B's and C's of the above ordered sequence. The number of ways we can divide 20 items into 3 distinct groups of respective sizes 4, 10 and 6 is the multinomial coefficient :

$$\binom{20}{4, 10, 6} = \frac{20!}{4! \, 10! \, 6!}$$

Thus, the joint probability of getting $(X = 4, Y = 10)$ is :

$$\frac{20!}{4! \, 10! \, (20 - 4 - 10)!} p_1^4 \, p_2^{10} \, (1 - p_1 - p_2)^{20-4-10}$$

Let us generalize our findings.

Definition 1.7 (Trinomial process)**.** *A **trinomial process** has the following features :*

1. *We repeat $n \in \mathbb{N}^*$ identical trials*

2. *A trial can result in exactly one of three* ***mutually exclusive and exhaustive*** *outcomes, that is, events E_1, E_2 and E_3 occur with respective probabilities p_1, p_2 and $p_3 = 1 - p_1 - p_2$. In other words, E_1, E_2 and E_3 form a partition of Ω.*

3. *p_1, p_2 (thus p_3) remain constant trial after trial. In this case, the process is said to be* ***stationary****.*

4. *The trials are mutually independent.*

Definition 1.8 (Trinomial Distribution)**.** *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let E_1, E_2 and E_3 be three mutually exclusive and exhaustive events that occur with respective probabilities p_1, p_2 and $p_3 = 1-p_1-p_2$. Assume $n \in \mathbb{N}^*$ trials are performed according to a trinomial process. Let X, Y and Z denote the numbers of times events E_1, E_2 and E_3 occur respectively. Then random variables X, Y are said to follow a* ***trinomial distribution*** *$Trin(n, p_1, p_2)$ and their joint probability mass function is given by :*

$$p_{XY}(x, y) = \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y (1 - p_1 - p_2)^{n-x-y} \tag{8}$$

Remark. Note that the joint distribution of X and Y is exactly the same as the joint distribution of X and Z or the joint distribution of Y and Z since $X + Y + Z = n$.

1.3 Distributions of Two Continuous Random Variables

Definition 1.9 (Joint probability density function)**.** *Let X and Y be two continuous rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$. X and Y are said to be* ***jointly continuous*** *if there exists a function f_{XY} such that, for any Borel set on $\mathbb{R}^2$:*

$$\mathbb{P}((X, Y) \in B) = \iint_B f_{XY}(x, y)\, dx\, dy \tag{9}$$

Then function f_{XY} is called the ***joint probability density function*** *of X and Y.*

Moreover, if the joint distribution function F_{XY} is of class $\mathcal{C}^2$, then the joint pdf of X and Y can be expressed in terms of partial derivatives :

$$f_{XY}(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} \tag{10}$$

Property 1.3. *Let X and Y be two continuous rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with joint pdf f_{XY}, then the following holds :*

- *$f_{XY}(x, y) \geq 0$, for $x, y \in \mathbb{R}$*

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

Definition 1.10 (Marginal distributions)**.** *Let X and Y be two continuous rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with joint pdf f_{XY}. Then the pdf of X alone is called the* ***marginal probability density function*** *of X and is defined by:*

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy, \quad \text{for } x \in \mathbb{R} \tag{11}$$

Similarly, the pdf of Y alone is called the ***marginal probability density function*** *of Y and is defined by:*

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx, \quad \text{for } y \in \mathbb{R} \tag{12}$$

Definition 1.11 (Independence)**.** *Let X and Y be two continuous rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with joint pdf f_{XY}. Let f_X and f_Y be the respective marginal pdfs of X and Y. Then X and Y are said to be* ***independent*** *if and only if :*

$$f_{XY}(x,y) = f_X(x) f_Y(y), \quad \text{for all } x, y \in \mathbb{R} \tag{13}$$

Definition 1.12 (Expected Value)**.** *Let X and Y be two continuous rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with joint pdf f_{XY} and let $g : \mathbb{R}^2 \to \mathbb{R}$ be a bounded piecewise continuous function. Then, the mathematical expectation of $g(X,Y)$, if it exists, is denoted by $\mathbb{E}[g(X,Y)]$ and is defined as follows :*

$$\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y)\,dx\,dy \tag{14}$$

Example 4. Let X and Y be two continuous random variables with joint probability density function :

$$f_{XY}(x,y) = 4xy\,\mathbb{1}_{[0,1]^2}(x,y)$$

1. Verify that f_{XY} is a valid joint probability density function.

   Answer. Note that f_{XY} can be rewritten as follows :

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

   It is now clear that $f_{XY}(x,y) \geq 0$ for all $x, y \in \mathbb{R}$. Second, let us verify that the integral of f_{XY} over $\mathbb{R}^2$ is 1 :

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy &= \int_0^1\int_0^1 4xy\,dx\,dy \\
&= \int_0^1 4x\left\{\int_0^1 y\,dy\right\}\,dx \\
&= \int_0^1 4x\left[\frac{y^2}{2}\right]_0^1\,dx \\
&= \int_0^1 2x\,dx \\
&= x^2\Big|_0^1 \\
&= 1
\end{aligned}$$

Hence, f_{XY} is a valid joint probability density function.

2. What is $\mathbb{P}(Y < X)$?

   Answer. The set B in the xy-plane such that $(x,y) \in [0,1]^2$ and $y < x$ is :

$$B = \{(x,y)\,|\,0 \le y < x \le 1\}$$

   Then,

$$\begin{aligned}
\iint_B f_{XY}(x,y)\,dx\,dy &= \int_{x=0}^1\int_{y=0}^x 4xy\,dx\,dy \\
&= \int_0^1 4x\left\{\int_0^x y\,dy\right\}\,dx \\
&= \int_0^1 4x\left[\frac{y^2}{2}\right]_0^x\,dx \\
&= \int_0^1 2x^3\,dx \\
&= \frac{x^4}{2}\Big|_0^1 \\
&= \frac{1}{2}
\end{aligned}$$

3. What are the marginal probability density functions of X and Y?

   Answer. The marginal probability density function of X, for $x \in [0,1]$, is given by :

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy \\
&= \int_0^1 4xy\,dy \\
&= 4x\left[\frac{y^2}{2}\right]_0^1 \\
&= 2x
\end{aligned}
$$

And $f_X(x) = 0$ otherwise. In a nutshell,

$$f_X(x) = 2x\,\mathbb{1}_{[0,1]}(x)$$

Similarly, we find that the marginal probability density function of Y is :

$$f_Y(y) = 2y\,\mathbb{1}_{[0,1]}(y)$$

4. Are X and Y independent?

   Answer. Yes, since for all $x, y \in \mathbb{R}$:

$$f_X(x)f_Y(y) = 4xy\,\mathbb{1}_{[0,1]}(x)\,\mathbb{1}_{[0,1]}(y) = f_{XY}(x,y)$$

5. What are the expected values of X and Y?

   Answer. The expected value of X can be computed as follows :

$$
\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{XY}(x,y)\,dx\,dy \\
&= \int_{-\infty}^{\infty} x\left\{\int_{-\infty}^{\infty} f_{XY}(x,y)\,dy\right\}dx \\
&= \int_{-\infty}^{\infty} x f_X(x)\,dx \\
&= \int_0^1 2x^2\,dx \\
&= \frac{2x^3}{3}\bigg|_0^1 \\
&= \frac{2}{3}
\end{aligned}
$$

   Similarly, we find that the expected value of Y is $2/3$.

Example 5. The joint probability density function of X and Y is given by:

$$f_{XY}(x,y) = \frac{6}{7}\left(x^2 + \frac{xy}{2}\right)\mathbb{1}_S(x,y)$$

with $S = \{(x,y)\,|\,0 < x < 1,\, 0 < y < 2\}$

1. Show that f_{XY} is a valid joint probability density function.
2. Compute the marginal probability density function of X.
3. Find $\mathbb{P}(X > Y)$.
4. What are the expected values of X and Y?
5. Are X and Y independent?

2 The Correlation Coefficient

2.1 Covariance

Definition 2.1 (Covariance)**.** *Let X and Y be two rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The* ***covariance*** *of X and Y, denoted by $Cov(X,Y)$, is defined as follows :*

$$Cov(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \tag{15}$$

upon existence of the above expression

- *If X and Y are discrete rrv with joint pmf p_{XY}, then the covariance of X and Y is :*

$$Cov(X,Y) = \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} (x - \mathbb{E}[X])(y - \mathbb{E}[Y]) p_{XY}(x,y) \tag{16}$$

- *If X and Y are continuous rrv with joint pdf f_{XY}, then the covariance of X and Y is :*

$$Cov(X,Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mathbb{E}[X])(y - \mathbb{E}[Y]) f_{XY}(x,y)\, dx\, dy \tag{17}$$

Theorem 2.1. *Let X and Y be two rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The covariance of X and Y can be calculated as :*

$$Cov(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \tag{18}$$

Example 6. Suppose that X and Y have the following joint probability mass function:

$X \backslash Y$	1	2	3
1	0.25	0.25	0
2	0	0.25	0.25

What is the covariance of X and Y?

Property 2.1. *Here are some properties of the covariance. For any random variables X and Y, we have :*

1. $Cov(X, Y) = Cov(Y, X)$
2. $Cov(X, X) = Var(X)$
3. $Cov(aX, Y) = a\,Cov(X, Y)$ *for* $a \in \mathbb{R}$
4. *Let $X_1, \ldots, X_n$ be n random variables and $Y_1, \ldots, Y_m$ be m random variables. Then :*

$$Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$$

2.2 Correlation

Definition 2.2 (Correlation)**.** *Let X and Y be two rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with respective standard deviations $\sigma_X = \sqrt{Var(X)}$ and $\sigma_Y = \sqrt{Var(Y)}$. The **correlation** of X and Y, denoted by ρ_{XY}, is defined as follows :*

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \tag{19}$$

upon existence of the above expression

Property 2.2. *Let X and Y be two rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$.*

$$-1 \leq \rho_{XY} \leq 1 \tag{20}$$

Interpretation of Correlation The correlation coefficient of X and Y is interpreted as follows:

1. If $\rho_{XY} = 1$, then X and Y are perfectly, positively, linearly correlated.
2. If $\rho_{XY} = -1$, then X and Y are perfectly, negatively, linearly correlated. X and Y are also said to be perfectly linearly anticorrelated
3. If $\rho_{XY} = 0$, then X and Y are completely, linearly uncorrelated. That is, X and Y may be perfectly correlated in some other manner, in a parabolic manner, perhaps, but not in a linear manner.
4. If $0 < \rho_{XY} < 1$, then X and Y are positively, linearly correlated, but not perfectly so.
5. If $-1 < \rho_{XY} < 0$, then X and Y are negatively, linearly correlated, but not perfectly so. X and Y are also said to be linearly anticorrelated

Example. Remember Example 1.1 on dining habits where we defined discrete random variables X and Y as follows :

- $X = 1$: a person between 20 and 69 visits upscale restaurants at least once a month and $X = 0$ otherwise
- $Y = 1$: a person between 20 and 69 is a millenial
- $Y = 2$: a person between 20 and 69 is a Gen X
- $Y = 3$: a person between 20 and 69 is a Baby Boomer

The joint probability mass function of X and Y is given below :

$X \backslash Y$	1	2	3
0	0.184	0.216	0.238
1	0.133	0.106	0.123

Compute the covariance and the correlation of X and Y. What can you say about the relationship between X and Y?

Theorem 2.2. *Let X and Y be two rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and let g_1 and g_2 be two bounded piecewise continuous functions. If X and Y are independent, then*

$$\mathbb{E}[g_1(X)g_2(Y)] = \mathbb{E}[g_1(X)]\,\mathbb{E}[g_2(Y)] \tag{21}$$

provided that the expectations exist.

Proof. Assume that X and Y are discrete random variables with joint probability mass function p_{XY} and respective marginal probability mass functions p_X and p_Y. Then

$$\begin{aligned}
\mathbb{E}[g_1(X)g_2(Y)] &= \sum_{x\in X(\Omega)} \sum_{y\in Y(\Omega)} g_1(x)g_2(y)p_{XY}(x,y) \\
&= \sum_{x\in X(\Omega)} \sum_{y\in Y(\Omega)} g_1(x)g_2(y)p_X(x)p_Y(y) \quad \text{since } X \text{ and } Y \text{ are independent} \\
&= \left\{ \sum_{x\in X(\Omega)} g_1(x)p_X(x) \right\} \left\{ \sum_{y\in Y(\Omega)} g_2(y)p_Y(y) \right\} \\
&= \mathbb{E}[g_1(X)]\,\mathbb{E}[g_2(Y)]
\end{aligned}$$

The result still holds when X and Y are continuous random variables (see Homework 5). $\square$

Theorem 2.3. *Let X and Y be two rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$. If X and Y are independent, then*

$$Cov(X, Y) = \rho_{XY} = 0 \tag{22}$$

Note, however, that the converse of the theorem is not necessarily true! That is, zero correlation does not imply independence. Let's take a look at an example that illustrates this claim.

Example 7. Let X and Y be discrete random variables with the following joint probability mass function:

$X \backslash Y$	-1	0	1
-1	0.2	0	0.2
0	0	0.2	0
1	0.2	0	0.2

1. What is the correlation between X and Y?
2. Are X and Y independent?

3 Conditional Distributions

3.1 Discrete case

Example. An international TV network company is interested in the relationship between the region of citizenship of its customers and their favorite sport.

Sports \ Citizenship	Africa	America	Asia	Europe
Tennis	0.02	0.07	0.04	0.12
Basketball	0.03	0.11	0.01	0.04
Soccer	0.08	0.05	0.04	0.16
Football	0.01	0.17	0.02	0.03

The company is interested in answering the following questions :

- What is the probability that a randomly selected person is African and his/her favorite sport is basketball?
- If a randomly selected person's favorite sport is soccer, what is the probability that he/she is European?
- If a randomly selected person is American, what is the probability that his/her favorite sport is tennis?

Definition 3.1. *Let X and Y be two discrete rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with joint pmf p_{XY} and respective marginal pmfs p_X and p_Y. Then, for $y \in Y(\Omega)$, the* ***conditional probability mass function*** *of X given $Y = y$ is defined by :*

$$p_{X|Y}(x|y) = \mathbb{P}(X = x|Y = y) = \frac{p_{XY}(x, y)}{p_Y(y)} \tag{23}$$

provided that $p_Y(y) \neq 0$.

Definition 3.2 (Conditional Expectation)**.** *Let* X *and* Y *be two discrete rrvs on probability space* $(\Omega, \mathcal{A}, \mathbb{P})$*. Then, for* $y \in Y(\Omega)$*, the* ***conditional expectation*** *of* X *given* $Y = y$ *is defined as follows :*

$$\mathbb{E}[X|Y = y] = \sum_{x \in X(\Omega)} x p_{X|Y}(x|y) \tag{24}$$

Property 3.1 (Linearity of conditional expectation)**.** *Let* X *and* Y *be two discrete rrvs on probability space* $(\Omega, \mathcal{A}, \mathbb{P})$*. If* $c_1, c_2 \in \mathbb{R}$ *and* $g_1 : X(\Omega) \to \mathbb{R}$ *and* $g_2 : X(\Omega) \to \mathbb{R}$ *are piecewise continuous functions. Then, for* $y \in Y(\Omega)$*, we have :*

$$\mathbb{E}[c_1 g_1(X) + c_2 g_2(X)|Y = y] = c_1 \mathbb{E}[g_1(X)|Y = y] + c_2 \mathbb{E}[g_2(X)|Y = y] \tag{25}$$

Remark : The above property is also true for continuous rrvs.

Property 3.2. *Let* X *and* Y *be two discrete rrvs on probability space* $(\Omega, \mathcal{A}, \mathbb{P})$*. If* X *and* Y *are* ***independent****, then, for* $y \in Y(\Omega)$*, we have :*

$$p_{X|Y}(x|y) = p_X(x) \quad \textit{for all } x \in X(\Omega) \tag{26}$$

Similarly, for $x \in X(\Omega)$*, we have :*

$$p_{Y|X}(y|x) = p_Y(y) \quad \textit{for all } y \in Y(\Omega) \tag{27}$$

Remark : The above property is also true for continuous rrvs.

Property 3.3. *Let* X *and* Y *be two discrete rrvs on probability space* $(\Omega, \mathcal{A}, \mathbb{P})$*. If* X *and* Y *are* ***independent****, then, for* $y \in Y(\Omega)$*, we have :*

$$\mathbb{E}[X|Y = y] = \mathbb{E}[X] \tag{28}$$

Similarly, for $x \in X(\Omega)$*, we have :*

$$\mathbb{E}[Y|X = x] = \mathbb{E}[Y] \tag{29}$$

Remark : The above property is also true for continuous rrvs.

Definition 3.3 (Conditional Variance)**.** *Let* X *and* Y *be two discrete rrvs on probability space* $(\Omega, \mathcal{A}, \mathbb{P})$*. Then, for* $y \in Y(\Omega)$*, the* ***conditional variance*** *of* X *given* $Y = y$ *is defined as follows :*

$$Var(X|Y = y) = \mathbb{E}[(X - \mathbb{E}[X|Y = y])^2|Y = y] = \sum_{x \in X(\Omega)} (x - \mathbb{E}[X|Y = y])^2 p_{X|Y}(x|y) \tag{30}$$

Property 3.4. *Let* X *and* Y *be two discrete rrvs on probability space* $(\Omega, \mathcal{A}, \mathbb{P})$*. Then, for* $y \in Y(\Omega)$*, the* ***conditional variance*** *of* X *given* $Y = y$ *can be calculated as follows :*

$$Var(X|Y = y) = \mathbb{E}[X^2|Y = y] - (\mathbb{E}[X|Y = y])^2 = \sum_{x \in X(\Omega)} x^2 p_{X|Y}(x|y) - \left(\sum_{x \in X(\Omega)} x p_{X|Y}(x|y) \right)^2 \tag{31}$$

3.2 Continuous case

Definition 3.4. *Let X and Y be two continuous rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with joint pdf f_{XY} and respective marginal pdfs f_X and f_Y. Then, for $y \in \mathbb{R}$, the* ***conditional probability density function*** *of X given $Y = y$ is defined by :*

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} \tag{32}$$

provided that $f_Y(y) \neq 0$.

Definition 3.5 (Conditional Expectation)**.** *Let X and Y be two continuous rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then, for $y \in \mathbb{R}$, the* ***conditional expectation*** *of X given $Y = y$ is defined as follows :*

$$\mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx \tag{33}$$

Example. Let X and Y be two continuous random variables with joint probability density function :

$$f_{XY}(x,y) = \frac{3}{2}\mathbb{1}_S(x,y)$$

with $S = \{(x,y) \,|\, x^2 < y < 1,\, 0 < x < 1\}$

1. Compute the conditional probability density function of Y given $X = x$ for $x \in (0,1)$.

   Answer. To compute the conditional pdf, we need to compute the marginal pdf of X given by : for $x \in (0,1)$

   $$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(xy)\, dy \\ &= \int_{x^2}^{1} \frac{3}{2}\, dy \\ &= \frac{3}{2}(1 - x^2) \end{aligned}$$

   Hence,

   $$f_X(x) = \frac{3}{2}(1 - x^2)\mathbb{1}_{(0,1)}(x)$$

Therefore the conditional pdf of Y given $X = x$ is thus :

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{f_{XY}(x,y)}{f_X(x)} \\ &= \frac{\frac{3}{2}\mathbb{1}_{(0,1)}(x)\mathbb{1}_{(x^2,1)}(y)}{\frac{3}{2}(1-x^2)\mathbb{1}_{(0,1)}(x)} \\ &= \frac{1}{1-x^2}\mathbb{1}_{(x^2,1)}(y) \end{aligned}$$

for $x \in (0,1)$. Note that we recognize that $f_{Y|X}$ is the pdf of a uniform distribution on $(x^2, 1)$.

2. Find the expected value of Y given $X = x$ for $x \in (0,1)$.

   Answer. The conditional expectation is computed as follows : for $x \in (0,1)$

$$\begin{aligned} \mathbb{E}[Y|X=x] &= \int_{-\infty}^{\infty} y f_{Y|X}(y|x)\,dy \\ &= \int_{x^2}^{1} y\frac{1}{1-x^2}\,dy \\ &= \frac{1}{1-x^2}\left[\frac{y^2}{2}\right]_{x^2}^{1} \\ &= \frac{1-x^4}{2(1-x^2)} \\ &= \frac{x^2+1}{2} \end{aligned}$$

   The result should not be that surprising since the expected value of a uniform distribution is the average of the bounds of the interval.

Definition 3.6 (Conditional Variance)**.** *Let X and Y be two continuous rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then, for $y \in \mathbb{R}$, the **conditional variance** of X given $Y = y$ is defined as follows :*

$$Var(X|Y=y) = \mathbb{E}[(X-\mathbb{E}[X|Y=y])^2|Y=y] = \int_{-\infty}^{\infty}(x-\mathbb{E}[X|Y=y])^2 f_{X|Y}(x|y)\,dx \tag{34}$$

Property 3.5. *Let X and Y be two continuous rrvs on probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then, for $y \in \mathbb{R}$, the **conditional variance** of X given $Y = y$ can be calculated as follows :*

$$Var(X|Y=y) = \mathbb{E}[X^2|Y=y]-(\mathbb{E}[X|Y=y])^2 = \int_{-\infty}^{\infty} x^2 f_{X|Y}(x|y)\,dx - \left(\int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx\right)^2 \tag{35}$$