"Always be a little improbable."

Oscar Wilde

10
Pairs of random variables

General facts

• in the previous tutorials we analyzed experiments in which an outcome is one number.

• we'll start to analyze experiments in which an outcome is a collection of numbers; each number is a sample value of a random variable; thus we analyze experiments that produce two random variables X and Y.

• the results presented here can be generalized for a system of n random variables $X_1, X_2, \dots X_n$

For a pair of discrete random variables X and Y

• the joint cumulative distribution function of a pair of two discrete random variables X and Y is

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

• it has similar properties with the CDF corresponding to a single random varaiable:

i) $0 \leq F_{X,Y}(x,y) \leq 1$

ii) $F_X(x) = \lim_{y\to\infty} F_{X,Y}(x,y)$

iii) $F_Y(y) = \lim_{x\to\infty} F_{X,Y}(x,y)$

iv) $\lim_{x\to-\infty} F_{X,Y}(x,y) = \lim_{x\to-\infty} F_{X,Y}(x,y) = 0$

v) if $x \leq x_1$ and $y \leq y_1$, then $F_{X,Y}(x,y) \leq F_{X,Y}(x_1,y_1)$

vi) $\lim_{x,y\to\infty} F_{X,Y}(x,y) = 1$

• the joint probability mass function of a pair of two discrete random variables X and Y is

$$P_{X,Y}(x,y) = P\left(X=x,\ Y=y\right)$$

for independent random variables X, Y one has

$$P_{X,Y}(x,y) = P\left(X=x,\ Y=y\right) = P(X=x)\cdot P(Y=y)$$

• keep in mind that $\{X=x,\ Y=y\}$ is an event in an experiment, the sample space becomes now

$$S_{X,Y} = \{(x,y):\quad P_{X,Y}(x,y) > 0\}$$

• for discrete random variables X, Y the characteristic property of the joint PMF is

$$\sum_{x\in S_X}\sum_{y\in S_Y} P_{X,Y}(x,y) = 1$$

• the probability of the event A is

$$P(A) = \sum_{(x,y)\in A} P_{X,Y}(x,y)$$

• the marginal probability mass functions are

$$P_X(x) = \sum_{y\in S_Y} P_{X,Y}(x,y), \qquad P_Y(y) = \sum_{x\in S_X} P_{X,Y}(x,y)$$

For a pair of continuous random variables X, Y

• the joint probability density function of the continuous random variables X and Y is a function $f_{X,Y}(x,y)$ with the properties

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

• the joint cummulative distribution function is defined by

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

and

$$P(a \leq X \leq b,\quad c \leq Y \leq d) = F_{X.Y}(b,d) - F_{X.Y}(b,c) - F_{X.Y}(a,d) + F_{X.Y}(a,c)$$

• for independent random variables

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

• the probability of an event A is

$$P(A) = \iint_A f_{X,Y}(x,y)dxdy$$

- the marginal probability density functions are

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

An Example

Mark and Lisa are two real estate agents. Let X and Y be the respective numbers of houses sold by them in a month. Based on past sales, we estimated the following joint probabilities for X and Y

		X = 0	X = 1	X = 2
Y	0	0.12	0.42	0.06
	1	0.21	0.06	0.03
	2	0.07	0.02	0.01

Thus, for example $P(0,1) = 0.21$, meaning that the joint probability for Mark and Lisa to sell 0 and 1 houses, respectively, is 0.21. Other entries in the table are interpreted similarly. Note that the sum of all entries must equal to 1.

The marginal probabilities are calculated by summing across rows and down columns

		X = 0	X = 1	X = 2	
Y	0	0.12	0.42	0.06	0.6
	1	0.21	0.06	0.03	0.3
	2	0.07	0.02	0.01	0.1
		0.4	0.5	0.1	1.0

This gives us the probability mass functions for X and Y individually:

X

x	$P(x)$
0	0.4
1	0.5
2	0.1

Y

y	$P(y)$
0	0.6
1	0.3
2	0.1

Thus, for example, the marginal probability for Mark to sell 1 house is 0.5. We have

$$P(X = 0 \text{ and } Y = 2) = 0.07,$$

but $P(X = 0) = 0.4$, and $P(Y = 2) = 0.1$, hence, X and Y are not independent

$$P(X = 0 \text{ and } Y = 2) \neq P(X = 0) \cdot P(Y = 2)$$

We could be interested in the probability for having two houses sold (by either Mark or Lisa) in a month. This can be computed by adding the probabilities for all combinations of (x, y) pairs that result in a sum of 2:

$$P(X + Y = 2) = P(0, 2) + P(1, 1) + P(2, 0) = 0.19$$

Using this method, we can derive the probability mass function for the variable $X + Y$

$x + y$	$P(x + y)$
0	0.12
1	0.63
2	0.19
3	0.05
4	0.01

Solved problems

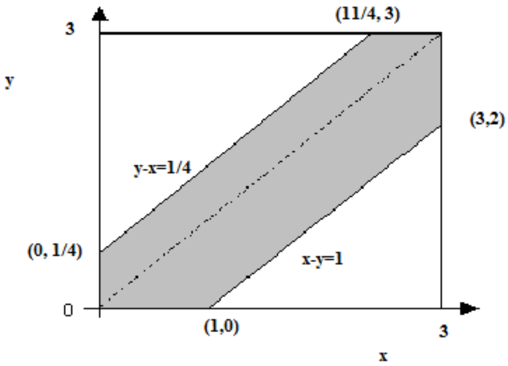
Problem 1. ***(Probability of meeting, revisited)***
A man and a woman decide to meet in a restaurant after 21 o'clock. The restaurant closes at 24 o'clock. Because of their busy schedule they decide that whoever arrives first at the restaurant will wait, for a while, for the other one. The man would be ready to wait an hour and the woman only 15 minutes! What's the probability that they will meet?

Solution: Let us denoted by X the continuous random variable that measures the time when the woman arrives and by Y the time when the man arrives. Anyone can arrive at any moment, between 21 o'clock and 24 o'clock, with equal probability. Hence X and Y will be uniformly distributed. Moreover, these random variables are independent with values in the interval $[0, 3]$ since one can consider 0 being 21 and 3 being 24 o'clock.

The random pair (X, Y) will model mathematically the problem. We have to denote also by E the event: they meet in that restaurant. The probability of E will be computed according to the formula

$$P(E) = \iint_E f_{X,Y}(x,y)\, dxdy$$

where E is the gray region in this illustration.



Since X are Y independent their joint probability density function will satisfy the identity

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

where

$$f_X(x) = \begin{cases} \frac{1}{3-0}, & x \in [0,3] \\ 0, & \text{otherwise} \end{cases} \qquad \text{and} \qquad f_Y(y) = \begin{cases} \frac{1}{3-0}, & y \in [0,3] \\ 0, & \text{otherwise} \end{cases}$$

are the PDF's of X and Y (uniformly distributed r.v.). Thus

$$P(E) = \iint_E \frac{1}{9}\, dxdy.$$

We compute this integral denoting by Δ_1 and Δ_2 the white triangles from above. It is easy to see

$$\Delta_1 \cup A \cup \Delta_2 = [0,3] \times [0,3]$$

Hence

$$\iint_{[0,3]\times[0,3]} \frac{1}{9}\, dxdy = \iint_{\Delta_1} \frac{1}{9}\, dxdy + \iint_A \frac{1}{9}\, dxdy + \iint_{\Delta_2} \frac{1}{9}\, dxdy.$$

One can represent these two triangles as type I regions

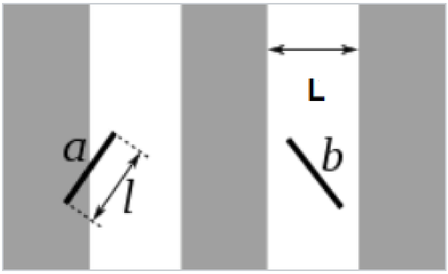
$$\int_0^3 \int_0^3 \frac{1}{9}\, dxdy = \int_0^{\frac{11}{4}} \int_{x+\frac{1}{4}}^3 \frac{1}{9}\, dydx + \iint_A \frac{1}{9}\, dxdy + \int_1^3 \int_0^{x-1} \frac{1}{9}\, dydx$$

$$\iff \quad 1 = \frac{121}{288} + \iint_A \frac{1}{9}\, dxdy + \frac{4}{18}$$

$$\implies P(E) = \iint_A \frac{1}{9}\, dxdy = 0.35 = 35\%$$

Problem 2. ***(Buffon's needle problem)***
Suppose we have a floor made of parallel strips of wood, each the same width L, and we drop a needle of length ℓ onto the floor. What is the probability that the needle will lie across a line between two strips?

Solution: This problem has an interesting history, it was proposed by count de Buffon in the XVIII-th century. It's solution is $p = \frac{2}{\pi}\frac{\ell}{L}$, thus it can be used to design a Monte Carlo method for approximating the number π, although that was not the original motivation for de Buffon's question.



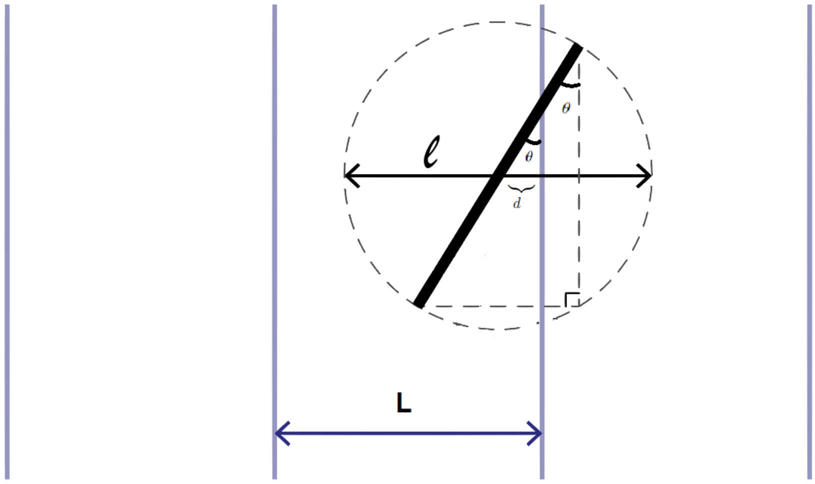
We want to keep things as simple as possible, hence we simplify the problem by treating only the case $\ell < L$. First of all, we have to observe that the acute angle θ, made between the needle and the parallel strips of the floor, influences the chance the needle has the to touch one of these lines.

In order to model mathematically the problem, it is not enough to know the angle θ, as one can observe in the above illustration. The final position of a needle depends also on the distance between the needle and the parallel strips of the floor. We'll consider the distance to the closest strip (after the needle is dropped on the floor). The most elegant way for estimating this distance is to measuring the distance from the middle of the needle to this strip. Let us denoted with d this distance. It is obvious that d can take only values between 0 (when the middle lies on the strip) and $\frac{L}{2}$ (when the middle at equal distances between two strips). Any value in the interval $[0, \frac{L}{2}]$ can be taken by d with equal probability, hence d will be one of the values of a uniformly distributed random variable D. The same conclusion holds for the angle θ which belongs to a uniformly distributed random variable Θ with values in the interval $[0, \frac{\pi}{2}]$.

Consequently, we can use the pair of random variables (D, Θ) in order to model mathematically the position of the needle on the floor. By their very definition the probability density functions, belonging to these r.v., are

$$f_D(d) = \begin{cases} \frac{1}{\frac{L}{2}} & 0 \le d \le \frac{L}{2} \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad f_\Theta(\theta) = \begin{cases} \frac{1}{\frac{\pi}{2}} & 0 \le \theta \le \frac{\pi}{2} \\ 0 & \text{otherwise} \end{cases}$$

Let us now define the event E: the needle lies accros a line. It is a good idea to draw the problem. The event occurs if $d \le \frac{\ell}{2}\sin\theta$, see the illustration



Have a look at those two right triangles with acute angles θ and opposite legs d, and $\frac{\ell}{2}\sin\theta$ respectively. The needle will intersect the blue line if $d \le \frac{\ell}{2}\sin\theta$, otherwise the blue line is out of the zone with which the needle can have points of intersection. Hence the event E can be defined as

$$E = \left\{ (d,\theta) \in \mathbb{R}^2 \;:\; d \le \frac{\ell}{2}\sin\theta \right\}$$

so

$$P(E) = \iint_E f_{D,\Theta}(d,\theta)\; dd d\theta.$$

We have an unpleasant notation dd, which means integration with respect to d :)). Since D and Θ are independent, their joint probability density function will be

$$f_{D,\Theta}(d,\theta) = f_D(d)\cdot f_\Theta(\theta) = \begin{cases} \frac{4}{L\pi}, & 0 \le d \le \frac{L}{2},\quad 0 \le \theta \le \frac{\pi}{2} \\ 0, & \text{in rest} \end{cases}$$

In order to evaluate this double integral, the set E must be written as a type I or II region

$$E = \left\{ (d,\theta) \in \mathbb{R}^2 \;: 0 \le \theta \le \frac{\pi}{2},\quad 0 \le \; d \le \frac{\ell}{2}\sin\theta \right\}$$

Eventually

$$P(E) = \iint_E f_{D,\Theta}(d,\theta)\; dd d\theta = \int_0^{\frac{\pi}{2}} \left(\int_0^{\frac{\ell}{2}\sin\theta} \frac{4}{L\pi}\; dd \right) d\theta$$
$$= \int_0^{\frac{\pi}{2}} \frac{4}{L\pi}\frac{\ell}{2}\sin\theta\; d\theta = -\frac{4}{L\pi}\frac{\ell}{2}\cos\theta\Big|_0^{\frac{\pi}{2}} = \frac{2\ell}{\pi L}$$

Problem 3. *Two discret random variables X and Y have the joint PMF*

$$P_{X,Y}(x,y) = \begin{cases} cxy, & x = 1,2,4 \quad y = 1,3 \\ 0, & \textit{otherwise} \end{cases}$$

i) What is $P(Y < X)$?

ii) What is $P(Y = 3)$?

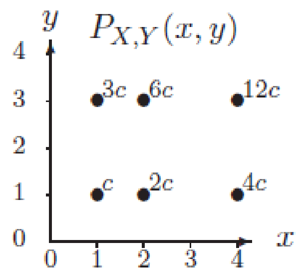
iii) Find the marginal PMF's of X and Y

iv) Find the expected values $E(X)$ and $E(Y)$

v) Compute the standard deviation σ_X

Solution:

i) In this problem, it is helpful to label points with nonzero probability on the X, Y plane



The characteristic property of the joint PMF is

$$\sum_{x \in S_X} \sum_{y \in S_Y} P_{X,Y}(x,y) = 1$$

thus

$$\sum_{x=1,2,4} \sum_{y=1,3} cxy = 28c = 1 \qquad \Longrightarrow \; c = \frac{1}{28}$$

The event $Y < X$ can be described by $A = \{(x,y) \in S_{X,Y} : y < x\} = \{(2,1),(4,1),(4,3)\}$ and its probability will be

$$P(A) = \sum_{(x,y) \in A} P_{X,Y}(x,y) = \frac{2 \cdot 1 + 4 \cdot 1 + 4 \cdot 3}{28} = \frac{18}{28}$$

ii) In order to find the probability $P(Y = 3)$ one can either define the event $Y = 3$ which can described by $B = \{(x,y) \in S_{X,Y} : y = 3\} = \{(1,3),(2,3),(4,3)\}$ or one can compute the marginal probability mass function according to the formula

$$P_Y(y) = \sum_{x \in S_X} P_{X,Y}(x,y) = \sum_{x=1,2,4} \frac{1}{28} x \cdot 3 = \frac{21}{28}$$

iii) The complete version of the marginal PMF of Y is obtained in the same way

$$P_Y(y) = \sum_{x \in S_X} P_{X,Y}(x, y) = \sum_{x=1,2,4} P_{X,Y}(x, y) = \begin{cases} \frac{7}{28}, & y = 1 \\ \frac{21}{28}, & y = 3 \\ 0, & \text{otherwise} \end{cases}$$

and the corresponding marginal PMF of X

$$P_X(x) = \sum_{y \in S_Y} P_{X,Y}(x, y) = \sum_{y=1,3} P_{X,Y}(x, y) = \begin{cases} \frac{4}{28}, & x = 1 \\ \frac{8}{28}, & x = 2 \\ \frac{16}{28}, & x = 4 \\ 0, & \text{otherwise} \end{cases}$$

iv) The expected value of X can now be computed using the marginal PMF of X

$$E(X) = \sum_{x \in S_X} x \cdot P_X(x) = 1 \cdot \frac{4}{28} + 2 \cdot \frac{8}{28} + 4 \cdot \frac{16}{28} = 3$$

v) For the standard deviation $\sigma_X = \sqrt{var(X)}$ we have to compute the variance

$$var(X) = E(X^2) - E(X)^2 = \sum_{x \in S_X} x^2 \cdot P_X(x) - \left(\sum_{x \in S_X} x \cdot P_X(x) \right)^2 = 10/7$$

Eventually $\sigma_X = \sqrt{\frac{10}{7}}$

Problem 4. *Two continuous random variables X and Y have the joint density function*

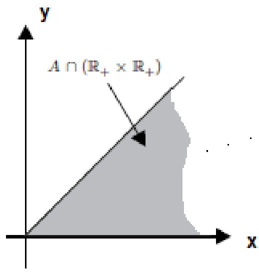
$$f_{X,Y}(x, y) = \begin{cases} 6e^{-2x-3y}, & x \geq 0, \ y \geq 0 \\ 0, & \textit{otherwise} \end{cases}$$

Evaluate the probabilities $P(X \geq Y)$ and $P(X + Y \leq 1)$.

Solution: Let us consider the event $A = \{X \geq Y\} = \{(x, y) \in \mathbb{R}^2 : x \geq y\}$, since $f_{X,Y}$ is nonzero only for $x \geq 0$ and $y \geq$ one can restrict to the first quadrant $\mathbb{R}_+ \times \mathbb{R}_+$ when computing probabilities

$$P(A) = \iint_A f_{X,Y}(x, y) \ dxdy = \iint_{A \cap (\mathbb{R}_+ \times \mathbb{R}_+)} f_{X,Y}(x, y) \ dxdy$$

Everytime one has to deal with double integrals it's worth drawing the set. In our case we have to integrate on $A \cap (\mathbb{R}_+ \times \mathbb{R}_+)$ which is the region enclosed between the line $y = x$ and the Ox axis



The set $A\cap(\mathbb{R}_+\times\mathbb{R}_+)$ can be written as a type I region

$$A\cap(\mathbb{R}_+\times\mathbb{R}_+)=\{(x,y)\in\mathbb{R}_+\times\mathbb{R}_+:\ 0\le x<\infty,\ 0\le y\le x\}$$

have a look in our previous tutorial about the double integral. As a consequence

$$\begin{aligned} P(A) &= \iint_A f_{X,Y}(x,y)\ dxdy = \iint_{A\cap(\mathbb{R}_+\times\mathbb{R}_+)} 6e^{-2x-3y}\ dxdy \\ &= \int_0^\infty \left(\int_0^x 6e^{-2x-3y}\ tcdy\right) dx = \int_0^\infty 2e^{-2x}\left(-e^{-3y}\Big|_0^x\right)\ dx \\ &= \int_0^\infty (2e^{-2x}-2e^{-5x})\ dx = \frac{3}{5} \end{aligned}$$

The generalized Riemann integral $\int_a^\infty f(x)\ dx$ is naturally defined as a limit of Riemann integrals

$$\int_a^\infty f(x)\ dx = \lim_{b\to\infty}\int_a^b f(x)\ dx$$

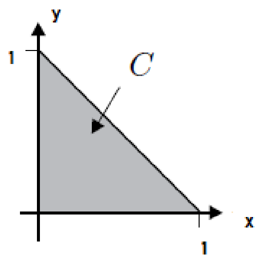
That's the reason why, for example

$$\int_0^\infty 2e^{-2x}\ dx = \lim_{b\to\infty}\int_0^b 2e^{-2x}\ dx = \lim_{b\to\infty} -e^{-2x}\Big|_0^b = \lim_{b\to\infty}(-e^{-2b}+1)=1$$

For the second part of the problem we define the event

$$B=\{(x,y)\in\mathbb{R}^2: x+y\le 1\}$$

and again, since f is nonvanishing only for pozitive values of x,y, it is enough to compute the probability of $C=B\cap(\mathbb{R}_+\times\mathbb{R}_+)$



Now, C can be written as a type I region or a type II region. As a type II region will be

$$C = \{(x, y) \in \mathbb{R}_+ \times \mathbb{R}_+ : \ 0 \le y \le 1, \ 0 \le x \le 1 - y\}$$

$$P(C) = \iint_C f_{X,Y}(x, y)\ dxdy = \int_0^1 \left(\int_0^{1-y} 6e^{-2x-3y} dx \right)\ dy = 1 - 2e^{-3} - 3e^{-2}$$

Remark: Always verify if your probabilities represent subunit numbers! One can check easily $0 < 1 - 2e^{-3} - 3e^{-2} < 1$.

Proposed problems

Problem 1. *The joint probability function for the random variables X and Y is given in the table.*

(a) Find the marginal probability functions of X and Y.
(b) Find $P(1 \le X < 3, Y \ge 1)$.
(c) Determine whether X and Y are independent

X \ Y	0	1	2
0	1/18	1/9	1/6
1	1/9	1/18	1/9
2	1/6	1/6	1/18

Problem 2. *The joint density function of two continuous random variables X and Y is:*

$$f(x, y) = \begin{cases} cxy, & 0 < x < 4, 1 < y < 5 \\ 0, & \text{otherwise} \end{cases}$$

a) Find the value of c
b) Find $P(1 < X < 2, 2 < Y < 3)$
c) Find the marginal distribution functions of X and of Y

Problem 3. *Let the random variable X be the portion of a flood insurance claim for flooding damage to the house and Y the portion of the claim for flooding damage to the rest of the property. The joint density function of X and Y is given by $f(x, y) = 3 - 2x - y$ for $0 < x, y < 1$ and $x + y < 1$. What are the marginal densities of X and Y ?*

Problem 4. *A business trip is equally likely to take* $2, 3$, *or* 4 *days. After a d-day trip, the change in the traveler's weight, measured as an integer number of pounds, is uniformly distributed between* $-d$ *and d pounds. For one such trip, denote the number of days by* D *and the change in weight by* W. *Find the joint PMF* $P_{D,W}(d, w)$.

Problem 5. *Over the circle* $X^2 + Y^2 \leq r^2$, *random variables* X *and* Y *have the uniform PDF*

$$\begin{cases} 1/(\pi r^2), & x^2 + y^2 \leq r^2 \\ 0, & \textit{otherwise} \end{cases}$$

Find the marginal PDF's $f_X(x)$ *and* $f_Y(y)$.

Problem 6. *Flip a coin twice. On each flip, the probability of heads equals p. Let* X_i *equal the number of heads (either* 0 *or* 1*) on flip* i. *Let* $Y = X_1 - X_2$ *and* $Z = X_1 + X_2$. *Find* $P_{Y,Z}(y, z)$.

Bibliography

[1] R. Yates and D. Goodman. *Probability and Stochastic processes*, Wiley&Sons, 2005.

[2] D. Montgomery and G. Runger. *Applied Statistics and Probability for Engineers*, Wiley, 2014.

[3] C. Ariesanu. *Lecture Notes on Special Mathematics*, 2020.