Independent Random Variables

March 27, 2020

1 Independent random variables

Now that we have introduced the idea of working with several random variables at the same time, we can generalize the notation of independence from events to random variables. Recall 2 events A and B are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

In other words, the probability that both A and B happen is the same as the probability A happens times the probability B happens. Using conditional probability, we have the equivalent formulation that A and B are independent if

$$\mathbb{P}(A|B) = \mathbb{P}(A)$$

In other words, knowing that the event B happened doesn't effect the probability of A happening.

Random variables X_1 and X_2 are independent if every event that depends on just X_1 is independent from every event that depends on just X_2. Fortunately, if suffices to just check events of the form $a_i < X_i \leq b_i$, meaning if X_1 and X_2 are independent then for any a_1, b_1, a_2, b_2,

$$\mathbb{P}(\{a_1 < X_1 \leq b_1\} \cap \{a_2 < X_2 \leq b_2\}) = \mathbb{P}(\{a_1 < X_1 \leq b_1\})\mathbb{P}(\{a_2 < X_2 \leq b_2\})$$

Since $\mathbb{P}(\{a_1 < X_1 \leq b_1\}) = F_{X_1}(b_1) - F_{X_1}(a_1)$ (where F_{X_1} is the CDF of X_1) it is often more convenient to define independence in terms of CDFs:

Definition 1.1. *The random variables X_1 and X_2 are* ***independent*** *if for every pair of real numbers x_1, x_2*

$$F_{X_1,X_2}(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2),$$

where $F_{X_1,X_2}(x_1, x_2)$ is the joint CDF of X_1 and X_2, $F_{X_1}(x_1)$ is the CDF of X_1, and $F_{X_2}(x_2)$ is the CDF of X_2.

The above definition is convenient, because the CDF makes perfect sense in both the discrete and continuous case, although it is a bit less useful in the discrete case. Since we usually work with pdfs or pmfs, it will be useful to reformulate the definition in terms them.

Theorem 1.2. *Let X_1 and X_2 be discrete random variables, they are* ***independent*** *if and only if for every pair of real numbers x_1, x_2*

$$p_{X_1,X_2}(x_1, x_2) = p_{X_1}(x_1)p_{X_2}(x_2),$$

where $p_{X_1,X_2}(x_1, x_2)$ is the joint pmf of X_1 and X_2, $p_{X_1}(x_1)$ is the pmf of X_1, and $p_{X_2}(x_2)$ is the pmf of X_2.

THEOREM 1.3. *Let X_1 and X_2 be continuous random variables, they are* ***independent*** *if and only if for every pair of real numbers x_1,x_2*

$$f_{X_1,X_2}(x_1,x_2)=f_{X_1}(x_1)f_{X_2}(x_2),$$

where $f_{X_1,X_2}(x_1,x_2)$ is the joint pdf of X_1 and X_2, $f_{X_1}(x_1)$ is the pdf of X_1, and $f_{X_2}(x_2)$ is the pdf of X_2.

Sometimes you will be given a joint distribution of random variables and asked to check if they are independent or not. Other times you will given independent random variables and their marginal distributions, and then asked to compute something about them. In the second case, the first step will often be to compute the joint distribution of the random variables, which is found by simply multiplying the marginal pdfs together.

Let's look at some examples, we can start with the examples from the previous lectures.

EXAMPLE 1.4. *Roll a fair 4-sided dice twice. Let Y_1 be the result of the first roll and Y_2 be the result of the second. Are the random variables Y_1 and Y_2 independent?*

Solution: In the previous notes we computed

$$p_{Y_1,Y_2}(y_1,y_2)=\frac{1}{16} \text{ for } y_1=1,2,3,4;\ y_2=1,2,3,4 \text{ and 0 otherwise },$$

$$p_{Y_1}(y_1)=1/4 \text{ for } y_1=1,2,3,4 \text{ and 0 otherwise },$$

and

$$p_{Y_2}(y_2)=1/4 \text{ for } y_2=1,2,3,4 \text{ and 0 otherwise}$$

Since $1/4*1/4=1/16$ we have

$$p_{Y_1,Y_2}(y_1,y_2)=p_{Y_1}(y_1)p_{Y_2}(y_2)$$

for all y_1 and y_2, so the random variables are independent.

More generally, if random variables are defined in terms of independent trials, then they will be independent.

EXAMPLE 1.5. *There are 20 balls in a box, labeled $1,2,\ldots,20$. You draw 2 balls out of the box without replacement. Let X_1 be the number of the first ball you draw and X_2 be the number on the second ball. Are X_1 and X_2 independent?*

Solution: We should suspect they are not independent, since the r.v.'s are generated from non-independent trials. In order to verify this intuition, we just need to find a pair of values x_1 and x_2 such that $p_{X_1,X_2}(x_1,x_2)\neq p_{X_1}(x_1)p_{X_2}(x_2)$.

Let's try $x_1=1, x_2=1$, from our previous computations $p_{X_1,X_2}(1,1)=0$, but $p_{X_1}(1)=1/20$, $p_{X_2}(1)=1/20$. Since $0\neq 1/20^2$ we see the random variables are not independent.

We could have also tried $x_1=1, x_2=2$, (or really any two number between 1 and 20 that are not equal). Once again from our previous computations we see $p_{X_1,X_2}(1,2)=1/(20*19)$, but $p_{X_1}(1)=1/20$, $p_{X_2}(2)=1/20$. Since $1/(20*19)\neq 1/20^2$ we have again verified that the random variables are not independent.

Example 1.6. *Let X_1 and X_2 be continuous random variables with jpdf:*

$$f_{X_1,X_2}(x_1,x_2)=\begin{cases} 1/8, & \text{if } -1\leq x_1\leq 3 \text{ and } 2\leq x_2\leq 4 \\ 0, & \text{otherwise.} \end{cases}$$

Are X_1 and X_2 independent?

Solution: We've seen when $-1\leq x_1\leq 3$ we have:

$$f_{X_1}(x_1)=\frac{1}{4}.$$

A similar computation shows when $2\leq x_2\leq 4$ we have:

$$f_{X_2}(x_2)=\frac{1}{2}.$$

It is then easy to check that $f_{X_1,X_2}(x_1,x_2)=f_{X_1}(x_1)f_{X_2}(x_2)$ for all x_1,x_2. Therefore the random variables are independent.

In fact we have a theorem that generalizes the last example.

Theorem 1.7. *Let X and Y be random variables with joint pdf $f_{X,Y}(x,y)$ that is positive if and only $a\leq x\leq b$ and $c\leq y\leq d$ for some (possibly infinite) numbers. If there exist functions $g(x)$ and $h(y)$ such that*

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

then X and Y are independent.

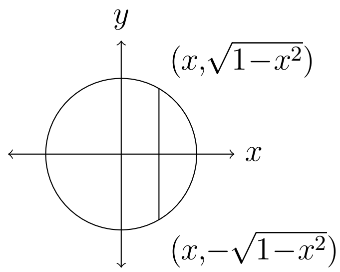
Note that we don't require $g(x)$ and $h(y)$ to be pdfs (but there will exist some constant that can multiply them to make this so).

EXAMPLE 1.8. *Let the X and Y be random variables, taking values in a disk of radius 1 with joint pdf*

$$f_{X,Y}(x,y)=\frac{24}{\pi}x^2y^2 \text{ for } x,y \text{ such that } x^2+y^2\leq 1, \text{ and } 0 \text{ otherwise .}$$

Are X and Y independent?

Solution: It is tempting to applying the above theorem and they are independent, but we need to note the joint pdf is supported on a disk not a rectangle as in the theorem. So we need to compute the marginal pdfs. We'll compute $f_X(x)$ ($f_Y(y)$ is basically the same), we pick an x point between -1 and 1 and then integrate y over the corresponding line.



Where we have the noted the line with x constant intersects the circle at $y=\pm\sqrt{1-x^2}$).

$$f_X(x)=\int f_{X,Y}(x,y)dy=\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{24}{\pi}x^2y^2dy=\frac{24}{\pi}x^2\left(\frac{y^3}{3}\right)_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}=\frac{24}{\pi}\frac{2x^2(1-x^2)^{3/2}}{3}=\frac{16}{\pi}x^2(1-x^2)^{3/2}$$

for $-1\leq x\leq 1$, and $f_X(x)=0$ otherwise. Similarly

$$f_Y(y)=\begin{cases} \frac{16}{\pi}y^2(1-y^2)^{3/2}, & \text{if } -1\leq y\leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Inside the circle clearly

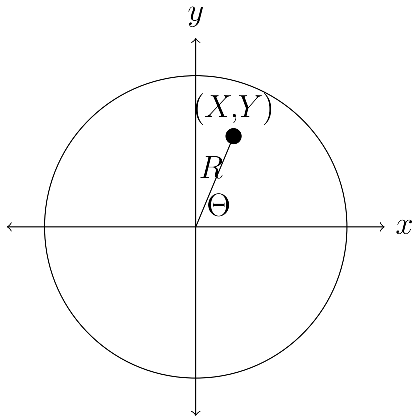
$$\frac{24}{\pi}x^2y^2\neq\frac{16}{\pi}x^2(1-x^2)^{3/2}\frac{16}{\pi}y^2(1-y^2)^{3/2},$$

even more obviously, outside the circle, but still with $-1\leq y\leq 1$ and $-1\leq x\leq 1$ the joint pdf is 0, but the marginal pdfs are not and we have

$$0\neq\frac{16}{\pi}x^2(1-x^2)^{3/2}\frac{16}{\pi}y^2(1-y^2)^{3/2}.$$

So X and Y are not independent. Intuitively, if X is near 1, then Y must be small, because of the constraint that the random variables take values inside the circle.

EXAMPLE 1.9. *Choose a point on the unit disk with pdf given by the previous example. Let R denote the distance from the origin of this point and let Θ be the angle that the line from this point to the origin makes with the positive x-axis. See the picture.*



Are R and Θ independent?

Solution: Using Polar Coordinates we have

$$x = r\cos(\theta) \text{ and } y = r\sin(\theta) \text{ and } dxdy = rdrd\theta$$

So the joint pdf of R, Θ is

$$f_{R,\Theta}(r,\theta) = \begin{cases} r^5\cos^2(\theta)\sin^2(\theta), & \text{if } 0 \le r \le 1,\ 0 \le \theta \le 2\pi \\ 0, & \text{otherwise.} \end{cases}$$

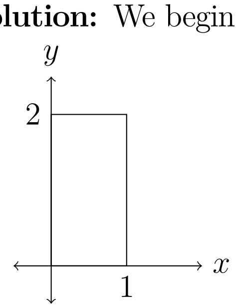
Since the domain is of the form $0 \le r \le 1$, $0 \le \theta \le 2\pi$ (the bounds are just constant numbers) and we can write $f_{R,\Theta}(r,\theta) = g(r)h(\theta)$ with $g(r) = r^5$ and $h(\theta) = \cos^2(\theta)\sin^2(\theta)$. We see that R and Θ are independent.

EXAMPLE 1.10. *Let the X_1 and X_2 be random variables with joint pdf*

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{3}(x+y), & \text{if } 0 \le x \le 1,\ 0 \le y \le 2 \\ 0, & \text{otherwise.} \end{cases}$$

Are X and Y independent?

Solution: We begin by drawing the domain



then we compute the marginal distributions: for $0 \leq x \leq 1$:

$$f_X(x) = \int_0^2 \frac{1}{3}(x+y)dy = \frac{1}{3}\left(xy + \frac{y^2}{2}\right)_0^2 = \frac{2}{3}(x+1)$$

and 0 otherwise. An exercise for you to check that:

$$f_Y(y) = \begin{cases} \frac{1}{6}(1+2y), & \text{if } 0 \leq y \leq 2 \\ 0, & \text{otherwise.} \end{cases}$$

Now we multiply the conditional probabilities to get:

$$f_Y(y)f_X(x) = \frac{1}{6}(1+2y) * \frac{2}{3}(x+1) \neq \frac{1}{3}(x+y).$$

So the random variables are not independent.

2 Conditional distributions of independent random variables

If X and Y are independent continuous random variables then

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

where the first equality is just the definition, the second uses the independence assumption and the final is just canceling.

Which very similar to our formula for conditional probabilities of independent events.