Recall

Let X be a continuous random variable with PDF f_X. Suppose $g\colon \mathbb{R} \to \mathbb{R}$ is $\begin{cases} \text{strictly monotone,} \\ \text{differentiable on } \mathbb{R}. \end{cases}$

Then $f_{g(X)}(y) = \begin{cases} f_X(g^{-1}(y)) \cdot \left|\frac{dg^{-1}(y)}{dy}\right| & \text{, if } y = g(x) \text{ for some } x \in \mathbb{R}, \\ 0 & \text{, otherwise.} \end{cases}$

Joint distributions

Let X, Y be two random variables. The *joint cumulative distribution function* (joint CDF) of X, Y is

$$F(a,b) \coloneqq P(X \le a, Y \le b) \quad , \forall\, a, b \in \mathbb{R}.$$

Then the *marginal distributions* (marginal CDFs) are

$$F_X(a) = \lim_{b\to\infty} F(a,b) =: F(a,\infty) \quad , \forall\, a \in \mathbb{R},$$
$$F_Y(b) = \lim_{a\to\infty} F(a,b) =: F(\infty,b) \quad , \forall\, b \in \mathbb{R}.$$

All the joint probability questions about X, Y can be answered in terms of joint CDF. In particular, $P(X > a, Y > b) = 1 - F(a,\infty) - F(\infty,b) + F(a,b)$.

- If X, Y are discrete, then the *joint probability mass function* (joint PMF) is

  $$p(x,y) \coloneqq P(X = x, Y = y) \quad , \forall\, x, y \in \mathbb{R}.$$

  Moreover, we have the *marginal PMFs* of X, Y

  $$p_X(x) = \sum_y p(x,y) \quad , \forall\, x \in \mathbb{R},$$
  $$p_Y(y) = \sum_x p(x,y) \quad , \forall\, y \in \mathbb{R}.$$

  and the joint CDF becomes $F(a,b) = \sum_{\substack{x \le a \\ y \le b}} p(x,y)$ for all $a, b \in \mathbb{R}$.

- We call two random variables X, Y *joint continuous* if there exists a *joint probability density function* (joint PDF) $f\colon \mathbb{R}^2 \to [0,\infty)$ such that

  $$P\{(X,Y) \in C\} = \iint_C f(x,y)dxdy$$

  for all 'measurable' sets $C \subset \mathbb{R}^2$. Fortunately, the countable unions of rectangles are 'measurable' on which we can do the computations. In particular, the joint CDF becomes

  $$F(a,b) = \int_{-\infty}^{b} \int_{-\infty}^{a} f(x,y)dxdy \quad , \forall\, a, b \in \mathbb{R}.$$

  On the other hand, if f is continuous at (a,b), then $f(a,b) = \frac{\partial^2}{\partial a \partial b} F(a,b)$.

And X, Y are continuous random variables with *marginal PDFs* obtained by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy \quad , \forall\, x \in \mathbb{R},$$
$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx \quad , \forall\, y \in \mathbb{R}.$$

Independent random variables

Two random variables X and Y are *independent* if

$$\begin{tikzcd}
 & P(X \in A, Y \in B) = P(X \in A)P(Y \in B), \forall\, A, B \subset \mathbb{R} \arrow[d, Leftrightarrow] & \\
 & F(a,b) = F_X(a)F_Y(b), \forall\, a, b \in \mathbb{R} \arrow[dl, Leftrightarrow, "X{,}Y \text{ discrete}"'] \arrow[dr, Leftrightarrow, "X{,}Y \text{ joint continuous}"] & \\
p(x,y) = p_X(x)p_Y(y), \forall\, x, y \in \mathbb{R} & & f(x,y) = f_X(x)f_Y(y), \forall\, x, y \in \mathbb{R}.
\end{tikzcd}$$

Examples

Example 1. Let X, Y be random variables with joint PDF

$$f(x,y) = \begin{cases} ce^{-x}e^{-2y} & , x, y \in (0, +\infty) \\ 0 & , \text{otherwise.} \end{cases}$$

Find the value of c, $P(X > 1, Y < 1)$, $P(X < Y)$ and marginal PDFs f_X, f_Y. Are X and Y independent?

Solution. Since

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = \int_0^{\infty}\int_0^{\infty} ce^{-x}e^{-2y}dxdy = c\left(-e^{-x}\Big|_0^{\infty}\right)\left(-\frac{1}{2}e^{-2y}\Big|_0^{\infty}\right) = \frac{c}{2},$$

we have $c = 2$.

Then

$$\begin{aligned} P(X > 1, Y < 1) = \int_{-\infty}^{1}\int_{1}^{\infty} f(x,y)dxdy = \int_0^1\int_1^{\infty} 2e^{-x}e^{-2y}dxdy &= 2e^{-1}(-\frac{1}{2}e^{-2} + \frac{1}{2}) \\ &= e^{-1}(1 - e^{-2}), \end{aligned}$$

and

$$P(X < Y) = \int_0^{\infty}\int_0^{y} 2e^{-x}e^{-2y}dxdy = \int_0^{\infty} 2e^{-2y}(1 - e^{-y})dy = \frac{1}{3}.$$

By formula, if $x \le 0$, then $f_X(x) = 0$ and if $x > 0$, then

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^{\infty} 2e^{-x}e^{-2y}dy = e^{-x}.$$

Similarly, if $y \leq 0$, then $f_Y(y) = 0$ and if $y > 0$, then

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_0^{\infty} 2e^{-x}e^{-2y}dx = 2e^{-2y}.$$

Hence for any $x, y \in \mathbb{R}$, $f(x,y) = f_X(x)f_Y(y)$, thus X and Y are independent. □

Remark. There is an optional safe check that $\int_{-\infty}^{\infty} f_X(x)dx = 1$ to avoid computational mistakes.

Example 2. Let X, Y be random variables with joint PDF

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

Find $E[X]$ and $E[Y]$. Are X and Y independent?

Solution. By formula, if $x \notin (0,1)$, then $f_X(x) = 0$ and if $x \in (0,1)$, then

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^x \frac{1}{x}dy = \frac{1}{x} \times x = 1.$$

Similarly, if $y \notin (0,1)$, then $f_Y(y) = 0$ and if $y \in (0,1)$, then

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_y^1 \frac{1}{x}dx = -\ln x\Big|_y^1 = \ln y.$$

Hence $f(x,y) \neq f_X(x)f_Y(y)$, thus X and Y are not independent.

Then

$$E[X] = \int_{-\infty}^{\infty} xf_X(x)dx = \int_0^1 xdx = \frac{1}{2},$$

and

$$E[Y] = \int_{-\infty}^{\infty} yf_Y(y)dy = \int_0^1 -y\ln ydy = \left(\frac{y^2}{2}\ln y\Big|_0^1\right) + \int_0^1 \frac{1}{y}\frac{y^2}{2}dy = 0 + \int_0^1 \frac{y}{2}dy = \frac{1}{4}$$

where 0 follows from $\lim_{y\to 0} y^2 \ln y = 0$. Recall "exponential" $\geq$ "polynomial" $\geq$ "logarithmic" (proved by e.g., L'Hospital). □

Example 3. Let X, Y be random variables with joint PDF

$$f(x,y) = \begin{cases} 8xy & , 0 < y < x < 1 \\ 0 & , \text{otherwise} \end{cases}$$

Are X and Y independent? (Please take a quick answer in mind before computations)

Solution. By formula, if $x \notin (0,1)$, then $f_X(x) = 0$ and if $x \in (0,1)$, then

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^x 8xydy = 4x^3.$$

Similarly, if $y \notin (0,1)$, then $f_Y(y) = 0$ and if $y \in (0,1)$, then

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_y^1 8xydy = 4y\left(x^2\Big|_y^1\right) = 4y(1-y^2) = 4y - 4y^3.$$

Hence $f(x,y) \neq f_X(x)f_Y(y)$, thus X and Y are not independent. □

Remark. In Example 3, a **wrong** quick answer is easily obtained that X, Y are independent by viewing $4xy = (4x)(y)$. However, in that way we have overlook the dependence hiding in the region $0 < y < x < 1$. To be more precise, let χ denote indicator functions, then we can write $f(x,y) = 4xy\chi_{\{0<y<x<1\}}$ in which we can not split $\chi_{\{0<y<x<1\}}$.

Hence it is natural to arrive at the following example.

Example 4. Let A, B be two **fixed** subsets of $\mathbb{R}$. Suppose that random variables X, Y have joint PDF

$$f(x,y) = \begin{cases} g(x)h(y) & , x \in A, y \in B \\ 0 & , \text{otherwise.} \end{cases}$$

for some functions $g, h \colon \mathbb{R} \to \mathbb{R}$. Then X and Y are independent.

Proof. Let χ_C denote the indicator function for $C \subset \mathbb{R}^2$. Then

$$f(x,y) = g(x)h(y)\chi_{A\times B} = \Big(g(x)\chi_{A\times Y}\Big)\Big(h(y)\chi_{X\times B}\Big),$$

thus X and Y are independent.

Alternatively, if $x \notin A$, then $f_X(x) = 0$ and if $x \in A$

$$f_X(x) = \int_{\mathbb{R}} f(x,y)dy = \int_B g(x)h(y)dy = g(x)\int_B h(y)dy.$$

If $y \notin B$, then $f_Y(y) = 0$ and if $y \in B$

$$f_Y(y) = \int_{\mathbb{R}} f(x,y)dx = \int_A g(x)h(y)dx = h(y)\int_A g(x)dx.$$

It follows from the unit integral of f that $\Big(\int_A g(x)dx\Big)\Big(\int_B h(y)dy\Big) = 1$. Then for $x \in A$ and $y \in B$,

$$f_X(x)f_Y(y) = g(x)h(y)\Big(\int_A g(x)dx\Big)\Big(\int_B h(y)dy\Big) = g(x)h(y) = f(x,y),$$

and for $x \notin A$ or $y \notin B$, $f(x,y) = 0 = f_X(x)f_Y(y)$. Hence X and Y are independent. $\square$