5 Joint distributions and independence

5a Distributions in dimension 2 (and more)

We know from Introduction to Probability, that a random variable has its distribution, but two random variables have not just two distributions; rather, they have a two-dimensional distribution.

A pair (X, Y) of random variables[56] $X, Y : \Omega \to \mathbb{R}$ may be treated as a *two-dimensional random variable*, $(X, Y) : \Omega \to \mathbb{R}^2$. The condition $\forall x \in \mathbb{R}\ \{\omega \in \Omega : X(\omega) \le x\} \in \mathcal{F}$ stipulated for one-dimensional random variables (recall 2a3) implies its two-dimensional counterpart,

$$\forall (x, y) \in \mathbb{R}^2 \quad \{\omega \in \Omega : X(\omega) \le x,\, Y(\omega) \le y\} \in \mathcal{F} \tag{5a1}$$

(think, why). Similarly to 2b3, the *joint* (cumulative) *distribution function* of X, Y is the function $F_{X,Y} : \mathbb{R}^2 \to [0, 1]$ defined by

$$F_{X,Y}(x, y) = \mathbb{P}\big(X \le x, Y \le y \big) . \tag{5a2}$$

5a3 Exercise. Let $X \sim \mathrm{U}(0, 1)$ and $Y = X$. Calculate $F_{X,Y}$ (that is, $F_{X,X}$). Is it continuous? Is $F_{X,Y}(x, y)$ equal to $F_X(x)F_Y(y)$?

Unfortunately, distribution functions are less illuminating in dim 2 than these in dim 1. Say, continuity of $F_{X,Y}$ is not a natural property of X, Y. Also, $F_{X,Y}$ does not lead to something like a 2-dim quantile function.

Two-dim counterparts of Conditions 2d8(a–e) and Theorem 2e10, being well-known, are of little use.[57] Distributions (and densities) are more useful.

Propositions 2d1, 2d2 have 2-dim counterparts.

5a4 Proposition. Let $X, Y : \Omega \to \mathbb{R}$ be random variables and $B \in \mathcal{B}_2$ (that is, $B \subset \mathbb{R}^2$ is a Borel set). Then the set $\{\omega \in \Omega : \big(X(\omega), Y(\omega)\big) \in B\}$ is an event.

(About 2-dim Borel sets, recall 1f9, 1f10.)

5a5 Proposition. Let $X, Y : \Omega \to \mathbb{R}$ be random variables. Then the function $P_{X,Y} : \mathcal{B}_2 \to [0, 1]$ defined by

$$P_{X,Y}(B) = \mathbb{P}\big((X, Y) \in B \big) \tag{5a6}$$

is a probability measure on $(\mathbb{R}^2, \mathcal{B}_2)$.

5a7 Definition. The probability measure $P_{X,Y}$ defined by (5a6) is called the *joint distribution* of random variables X, Y (or the distribution of the two-dimensional random variable (X, Y)).

[56] Both are defined on *the same* probability space $(\Omega, \mathcal{F}, P)$.

[57] You may think about $F_{X,Y}(x, y) = xy$ and $F_{X,Y}(x, y) = 1 - (1 - x)(1 - y)$ for $x, y \in (0, 1)$. The former is possible, but the latter is not. Do you understand, why?

That is similar to 2d4. Note that $(\mathbb{R}^2, \mathcal{B}_2, P_{X,Y})$ is another probability space. Similarly to (2d5), 2d6 we have

$$F_{X,Y}(x,y) = P_{X,Y}\big((-\infty,x] \times (-\infty,y]\big)\,, \tag{5a8}$$

$$F_{X,Y} = F_{U,V} \quad\Longleftrightarrow\quad P_{X,Y} = P_{U,V}\,. \tag{5a9}$$

5a10 Exercise. Do two equalities $P_X = P_U$ and $P_Y = P_V$ imply $P_{X,Y} = P_{U,V}$? Hint. Return to 5a3 and consider also $F_{X,-X}$.

Any probability measure P on $(\mathbb{R}^2, \mathcal{B}_2)$ corresponds to some random variables X, Y on *some* probability space. Namely, consider the probability space $(\mathbb{R}^2, \mathcal{B}_2, P_{X,Y})$. The identical map $\mathbb{R}^2 \to \mathbb{R}^2$ may be treated as a 2-dim random variable on that probability space, or as a pair (X,Y) of (1-dim) random variables $X, Y : \mathbb{R}^2 \to \mathbb{R}$, $X(x,y) = x$, $Y(x,y) = y$. Then $P_{X,Y} = P$ (think, why).

Here are 2-dim counterparts of 2d14 and 2d15.

5a11 Definition. A point $(x,y) \in \mathbb{R}^2$ is called an *atom* of (the distribution of) a two-dimensional random variable (X,Y), if

$$\mathbb{P}\big(\,(X,Y) = (x,y)\,\big) > 0\,.$$

5a12 Definition. The support of (the distribution of) a two-dimensional random variable (X,Y) is the set of all points $(x,y) \in \mathbb{R}^2$ such that

$$\forall \varepsilon > 0 \quad \mathbb{P}\big(\,x - \varepsilon < X < x + \varepsilon,\, y - \varepsilon < Y < y + \varepsilon\,\big) > 0\,.$$

Still, the support is the least *closed* set of probability 1.

By the way (in contrast to dimension one), if $F_{X,Y}$ is strictly increasing in a neighborhood of a given point, it does not mean that the point belongs to the support.

All said (in Sect. 5a) about dimension 2 holds for all dimensions $d = 1, 2, 3, \dots$ You can easily formulate such generalizations.

Do we need two-dimensional probability spaces? A discussion follows.

Y: If we restrict ourselves to a 1-dim probability space, say $(0,1)$, then any 2-dim random variable $(X,Y) : (0,1) \to \mathbb{R}^2$ is concentrated on a line. Its distribution is not really 2-dim.

N: Recall Example 2b8 ($Y = \sin X$, $X = 0, 1, 2, \dots$). There, X is discrete, it may be defined on a discrete (0-dim) probability space. Nevertheless, the support of Y is $[-1,+1]$, a 1-dim set. Similarly, consider $Y = \sin \alpha X$, $Z = \sin \beta X$, $X = 0, 1, 2, \dots$ for 'generic' α, β (I mean, α/π, β/π and α/β are irrational). The support of (Y,Z) is the whole square $[-1,+1] \times [-1,+1]$. This way we produce dim 2 out of dim 0. Of course, we may also produce dim 2 out of dim 1. Say, $Y = \sin \alpha X$, $Z = \sin \beta X$, where the support of X is the whole $\mathbb{R}$ (and again, α/β is irrational).

Y: If a countable set is dense in a square, it does not mean that it is really 2-dim. It is still 0-dim. Similarly, a line dense in the square is still 1-dim.

N: What is really 2-dim?

Y: The uniform distribution on the square. You cannot reach it by X, Y on $(0,1)$.

N: However, I can! Recall 2b6; I use the same idea:

$$(0,1) \ni \omega = \big(0.\beta_1\beta_2\beta_3\dots\big)_2\,; \qquad \begin{aligned} X(\omega) &= \big(0.\beta_1\beta_3\beta_5\dots\big)_2\,, \\ Y(\omega) &= \big(0.\beta_2\beta_4\beta_6\dots\big)_2\,. \end{aligned} \tag{5a13}$$

The distribution of (X,Y) is uniform on the square $(0,1)\times(0,1)$.

Y: Why?

N: Since the sequence $\big(\beta_1,\beta_2,\dots\big)$ is just infinite coin tossing; therefore $\big(\beta_1,\beta_3,\dots\big)$ is also infinite coin tossing, as well as $\big(\beta_2,\beta_4,\dots\big)$, and the two sequences are independent.

Y: Does it mean that the evident distinction between dimensions is an illusion?

N: The interval and the square are different (non-isomorphic) as *topological* spaces, but identical (isomorphic) as *probability* spaces.

Here are 2-dim counterparts of 3d1, 3d3.

5a14 Definition. A function $\varphi:\mathbb{R}^2\to\mathbb{R}$ is called a Borel function, if

$$\forall z\in\mathbb{R}\quad \{(x,y)\in\mathbb{R}^2 : \varphi(x,y)\le z\}\in\mathcal{B}_2$$

or equivalently, if for every one-dimensional Borel set $B\subset\mathbb{R}$ its inverse image $\varphi^{-1}(B) = \{(x,y)\in\mathbb{R}^2 : \varphi(x,y)\in B\}$ is a two-dimensional Borel set.[58]

5a15 Exercise. If $X,Y:\Omega\to\mathbb{R}$ are random variables and $\varphi:\mathbb{R}^2\to\mathbb{R}$ a Borel function then the function $Z:\Omega\to\mathbb{R}$ defined by $\forall\omega\ Z(\omega)=\varphi\big(X(\omega),Y(\omega)\big)$ is also a random variable. Prove it. (Hint: recall 3d4; use 5a4 instead of 2d1.)

5a16 Exercise. Every continuous function $\varphi:\mathbb{R}^2\to\mathbb{R}$ is a Borel function. Prove it. (Hint: similar to 3d5.)[59] Apply it to functions

$$\begin{aligned} \varphi(x,y)&=x \qquad & \varphi(x,y)&=x+y \\ \varphi(x,y)&=y \qquad & \varphi(x,y)&=xy \end{aligned}$$

and others. Reconsider Proposition 3d9.

5a17 Exercise. Generalize 3d9, 3d10 and 3d11 for dimension 2.

Here are 2-dim counterparts of 3d14 and (4e5).

5a18 Exercise. Let $P_{X,Y}=P_{U,V}$, and $\varphi:\mathbb{R}^2\to\mathbb{R}$ is a Borel function. Then random variables $\varphi(X,Y)$ and $\varphi(U,V)$ are identically distributed. Prove it. (Hint: find $P_{\varphi(X,Y)}$ in terms of $P_{X,Y}$.)

5a19 Exercise. For any Borel function $\varphi:\mathbb{R}^2\to\mathbb{R}$,

$$\mathbb{E}\,\varphi(X,Y)=\iint_{\mathbb{R}^2}\varphi\,dP_{X,Y}$$

(a number, or $-\infty$, or $+\infty$, or $\infty-\infty$). Prove it. (Hint: similar to (4e5).)

---

[58]Thus, 5a14 is not only a definition but also a statement (generalizing 3d2).

[59]It was stated in 1f10 without proof, that all open sets (and all closed sets) in $\mathbb{R}^2$ (and $\mathbb{R}^d$) are Borel sets. Here is a hint toward a proof. Let $U\subset\mathbb{R}^2$ be open. Consider *all* rectangles $(a,b)\times(c,d)$ such that a,b,c,d are rational numbers, and $(a,b)\times(c,d)\subset U$. The set of such rectangles is countable, and their union is equal to U.

5b Densities in dimension 2 (and more)

As was said in Sect. 2c, a density is defined in terms of integration. You surely guess that a 2-dim density will be defined in terms of 2-dim integration. Fortunately, we are now acquainted with Lebesgue integration; being fairly general, it works also on the plane $\mathbb{R}^2$. Namely, according to (4d5),

(5b1) $$\iint_B f(x,y)\,dxdy = \int_0^\infty \operatorname{mes}_2\{(x,y)\in B : f(x,y)>z\}\,dz \in [0,+\infty]$$

for any Borel set $B \subset \mathbb{R}^2$ and any Borel function $f : B \to [0,\infty)$.

5b2 Exercise.

$$\iint_B f(x,y)\,dxdy = \iint_{\mathbb{R}^2} f(x,y)\mathbf{1}_B(x,y)\,dxdy\,,$$

where $\mathbf{1}_B$ is the indicator of B. Prove it.

5b3 Definition. A Borel function $f : \mathbb{R}^2 \to [0,\infty)$ is called a density of (the distribution of) a two-dimensional random variable (X,Y), or a *joint density* of X,Y, if

$$P_{X,Y}(B) = \iint_B f(x,y)\,dxdy$$

for all Borel sets $B \subset \mathbb{R}^2$.

Here are 2-dim counterparts of 4d14, 4d15 and (4c4).

5b4 Proposition. If a 2-dim distribution P has a density f then

$$\iint_{\mathbb{R}^2} \varphi\,dP = \iint_{\mathbb{R}^2} \varphi(x,y)f(x,y)\,dxdy$$

for any Borel function $\varphi : \mathbb{R}^2 \to \mathbb{R}$. Both integrals are Lebesgue integrals. The four cases (a number, $-\infty$, $+\infty$, $\infty-\infty$) for the former integral correspond to the four cases for the latter integral.[60]

5b5 Corollary. If X,Y have a joint density $f_{X,Y}$ then

$$\mathbb{E}\,(XY) = \iint_{\mathbb{R}^2} xyf_{X,Y}(x,y)\,dxdy$$

(the four cases correspond...)

5b6 Proposition. Let $f : \mathbb{R}^2 \to [0,\infty)$ be a Borel function such that $\iint_{\mathbb{R}^2} f(x,y)\,dxdy = 1$. Then the function $P_f : \mathcal{B}_2 \to [0,1]$ defined by

$$P_f(B) = \iint_B f(x,y)\,dxdy$$

is a probability measure on $(\mathbb{R}^2, \mathcal{B}_2)$.

[60] A hint toward a proof (if you are curious). First, check the equality for 'simple' φ (that is, taking on only a finite number of values). Second, a 'sandwich argument' extends the equality to bounded φ. Last, use a limiting procedure for unbounded φ.

5b7 Corollary. For every Borel function $f : \mathbb{R}^2 \to \mathbb{R}$ satisfying

$$\forall (x, y) \in \mathbb{R}^2 \quad f(x, y) \ge 0\,,$$
$$\iint_{\mathbb{R}^2} f(x, y)\, dxdy = 1$$

there is one and only one two-dimensional distribution P such that f is a density of P.

5b8 Exercise. (a) Explain, why 5b7, being formulated as a 2-dim counterpart of 2e12, results from quite different arguments. (You see, 5b6 is not at all parallel to 2e11 or 2e10.)

(b) Reconsider 2c1 and 2e12 in the light of 5b3 and 5b7. (Be informed that Proposition 5b6 holds not only for the dimension $d = 2$, but also for $d = 1$, and in fact for all $d = 1, 2, 3, \dots$)

5b9 Exercise. P_f (defined in 5b6) determines f uniquely up to (a change on) a set of measure 0. That is, if $\iint_B f_1(x, y)\, dxdy = \iint_B f_2(x, y)\, dxdy$ for all $B \in \mathcal{B}_2$ then $f_1(x, y) = f_2(x, y)$ for all $(x, y) \in \mathbb{R}^2$ except (maybe) for a set of measure 0. Prove it. (Hint: consider $B = \{(x, y) : f_1(x, y) > f_2(x, y)\}$ and show that $\operatorname{mes}_2 B = 0$; the same for $f_1(x, y) < f_2(x, y)$.)

5b10 Exercise. Let $f_{X,Y}$ be a density for (X, Y), $B \subset \mathbb{R}^2$ a Borel set, $0 \le a \le b < \infty$ and

$$\forall (x, y) \in B \quad a \le f_{X,Y}(x, y) \le b\,.$$

Then

$$a \operatorname{mes}_2 B \le \mathbb{P}\big((X, Y) \in B \big) \le b \operatorname{mes}_2 B\,.$$

Prove it. (Hint: use monotonicity of Lebesgue integral, stated in Sect. 4d.)

5b11 Exercise. Let $f_{X,Y}$ be a density of (X, Y). Assume that $f_{X,Y}$ is continuous at a point (x, y). Then

$$f_{X,Y}(x, y) = \lim_{\varepsilon \to 0} \frac{1}{4\varepsilon^2} \mathbb{P}\big(x - \varepsilon < X < x + \varepsilon,\ y - \varepsilon < Y < y + \varepsilon \big)\,.$$

Prove it. (Hint: use 5b10.) What about disks instead of squares?

5b12 Exercise. Formulate and prove one-dimensional counterparts to 5b10 and 5b11.

All said (in Sect. 5b) about dimension 2 holds for all dimensions $d = 1, 2, 3, \dots$ You can easily formulate such generalizations.

5c Relations between dimensions 1 and 2 (and more)

These relations will be considered for Borel sets and functions, for distributions, Lebesgue integrals, and densities.

5c1 Exercise. If $B_1, B_2 \subset \mathbb{R}$ are (1-dim) Borel sets then their product

$$B = B_1 \times B_2 = \{(x, y) \in \mathbb{R}^2 : x \in B_1,\ y \in B_2\}$$

is a (2-dim) Borel set. Prove it. (Hint: note that $B_1 \times B_2 = (B_1 \times \mathbb{R}) \cap (\mathbb{R} \times B_2)$; recall 5a16 for $\varphi(x, y) = x$ and $\varphi(x, y) = y$.)

5c2 Exercise. If $B \subset \mathbb{R}^2$ is a (2-dim) Borel set then for every $x \in \mathbb{R}$ the *section*

$$B_x = \{y \in \mathbb{R} : (x,y) \in B\}$$

is a (1-dim) Borel set.[61] Prove it. (Hint: define Borel functions $\mathbb{R} \to \mathbb{R}^2$; show that the (continuous) embedding $\mathbb{R} \ni y \mapsto (x,y) \in \mathbb{R}^2$ is a Borel function.)

5c3 Exercise. If $\varphi_1, \varphi_2 : \mathbb{R} \to \mathbb{R}$ are Borel functions, then functions

$$\begin{aligned} &(x,y) \mapsto \varphi_1(x), \qquad &(x,y) \mapsto \varphi_1(x) + \varphi_2(y), \\ &(x,y) \mapsto \varphi_2(y), \qquad &(x,y) \mapsto \varphi_1(x)\varphi_2(y) \end{aligned}$$

are Borel functions $\mathbb{R}^2 \to \mathbb{R}$. Prove it.

5c4 Exercise. If $\varphi : \mathbb{R}^2 \to \mathbb{R}$ is a Borel function then for every $x \in \mathbb{R}$ the *section*

$$\varphi_x : \mathbb{R} \to \mathbb{R}\,, \qquad \varphi_x(y) = \varphi(x,y)$$

is a Borel function.

We turn to distributions. A 2-dim distribution $P_{X,Y}$ determines uniquely 1-dim distributions P_X, P_Y by

$$P_X(B) = P_{X,Y}(B \times \mathbb{R})\,, \qquad P_Y(B) = P_{X,Y}(\mathbb{R} \times B) \tag{5c5}$$

(think, why). Thus, P_X, P_Y are marginal distributions, as defined below.

5c6 Definition. Given a two-dimensional distribution P, its *marginal* distributions P_1, P_2 are one-dimensional distributions defined by

$$\forall B \in \mathcal{B} \quad P_1(B) = P(B \times \mathbb{R})\,, \quad P_2(B) = P(\mathbb{R} \times B)\,.$$

5c7 Exercise. These P_1, P_2 are indeed distributions (that is, probability measures on $(\mathbb{R}, \mathcal{B})$). Prove it. (Hint: $(B_1 \cup B_2) \times \mathbb{R} = (B_1 \times \mathbb{R}) \cup (B_2 \times \mathbb{R})$, etc.)

5c8 Exercise. Marginal distribution functions F_X, F_Y are determined by the joint distribution function $F_{X,Y}$ as follows:

$$\begin{aligned} F_X(x) &= F_{X,Y}(x, +\infty) = \lim_{y \to \infty} F_{X,Y}(x,y)\,, \\ F_Y(y) &= F_{X,Y}(+\infty, y) = \lim_{x \to \infty} F_{X,Y}(x,y)\,. \end{aligned}$$

Prove it.

As was noted, a 2-dim distribution *is not* uniquely determined by its marginal distributions (see 5a10).

[61]The converse is wrong. Say, for an arbitrary one-to-one function $\mathbb{R} \to \mathbb{R}$ its graph has single-point sections (in both variables) but need not be a Borel set.

5c9 Exercise. If a marginal distribution F_X is non-atomic, then a joint distribution $F_{X,Y}$ is also non-atomic. Prove it. The converse is false. Find a counterexample. What about supports?

We turn now to integrals. Formula (5b1) is not a practical way of calculating 2-dim integrals; the integrand $\operatorname{mes}_2\{\dots\}$, usually the area of a domain, is not easy to calculate. Fortunately, 2-dim integration can be reduced to 1-dim integration (applied twice) by an important theorem (due to Fubini).

First, note a traditional notation for the integral of a section $f_x : \mathbb{R} \to \mathbb{R}$ of a function $f : \mathbb{R}^2 \to \mathbb{R}$ (you know, $f_x(y) = f(x, y)$):

$$\int_{-\infty}^{+\infty} f_x(y)\, dy = \int_{-\infty}^{+\infty} f(x, y)\, dy\,;$$

the integral is a function of x only. Note that

$$\int_{-\infty}^{+\infty} f(x, y) g(x)\, dy = g(x) \int_{-\infty}^{+\infty} f(x, y)\, dy\,; \tag{5c10}$$

indeed, the section $f(x, \cdot)g(x)$ is the number $g(x)$ times the function $f(x, \cdot)$.

5c11 Theorem. (Fubini) For every Borel function $f : \mathbb{R}^2 \to \mathbb{R}$,

$$\text{(a)} \qquad \iint\limits_{\mathbb{R}^2} |f(x, y)|\, dxdy = \int_{-\infty}^{+\infty} \Big(\int_{-\infty}^{+\infty} |f(x, y)|\, dy \Big) dx \in [0, \infty]\,,$$

the internal integral being a Borel function $\mathbb{R} \to [0, +\infty]$ (of x); if the integral (a) is finite, then

$$\text{(b)} \qquad \iint\limits_{\mathbb{R}^2} f(x, y)\, dxdy = \int_{-\infty}^{+\infty} \Big(\int_{-\infty}^{+\infty} f(x, y)\, dy \Big) dx \in \mathbb{R}\,,$$

the internal integral coinciding almost everywhere with a Borel function $\mathbb{R} \to \mathbb{R}$ (of x).

Think about the area under a graph. . .

5c12 Exercise. Consider the disk $B = \{(x, y) : x^2 + y^2 \le 1\} \subset \mathbb{R}^2$. For any Borel function $f : \mathbb{R}^2 \to [0, \infty)$,

$$\iint_B f(x, y)\, dxdy = \int_{-1}^{+1} \bigg(\int_{-\sqrt{1-x^2}}^{+\sqrt{1-x^2}} f(x, y)\, dy \bigg) dx\,.$$

Prove it. (Hint: apply Fubini theorem to the function $f\mathbf{1}_B$, where $\mathbf{1}_B$ is the indicator of B.) What about $f : \mathbb{R}^2 \to \mathbb{R}$ (rather than $[0, \infty)$)?

5c13 Exercise. Let $-\infty \le a < b \le \infty$. For every Borel function $f : (a, b) \to [0, \infty)$,

$$\int_a^b f(x)\, dx = \operatorname{mes}_2\{(x, y) : x \in (a, b),\ 0 \le y \le f(x)\}\,.$$

Prove it. (Hint: apply Fubini theorem to the indicator of the two-dimensional set.)

5c14 Theorem. Let random variables X, Y have a joint density $f_{X,Y}$. Then X and Y have densities f_X, f_Y, and

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

almost everywhere.

Proof. We have to prove that $P_X(B) = \int_B f_X(x)\,dx$, and the same for Y. We use (5c5), 5b2, Fubini theorem and (5c10):

$$\begin{aligned}
P_X(B) = P_{X,Y}(B\times\mathbb{R}) &= \iint_{B\times\mathbb{R}} f_{X,Y}(x,y)\,dxdy = \\
&= \iint_{\mathbb{R}^2} \mathbf{1}_{B\times\mathbb{R}}(x,y) f_{X,Y}(x,y)\,dxdy = \iint_{\mathbb{R}^2} \mathbf{1}_B(x) f_{X,Y}(x,y)\,dxdy = \\
= \int_{-\infty}^{+\infty} \biggl(\int_{-\infty}^{+\infty} & \mathbf{1}_B(x) f_{X,Y}(x,y)\,dy \biggr) dx = \int_{-\infty}^{+\infty} \mathbf{1}_B(x) \biggl(\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dy \biggr) dx = \\
&= \int_B f_X(x)\,dx\,,
\end{aligned}$$

and the same for Y. $\square$

All said (in Sect. 5c) about dimensions 1 and 2 holds also for other dimensions. You can easily formulate such generalizations.

5d Independence

5d1 Definition. Random variables $X, Y : \Omega \to \mathbb{R}$ are *independent,* if

$$\mathbb{P}\bigl(X \in A, \, Y \in B \bigr) = \mathbb{P}\bigl(X \in A \bigr) \mathbb{P}\bigl(Y \in B \bigr)$$

for all Borel sets $A, B \subset \mathbb{R}$.

For discrete X, Y independence is evidently possible. For continuous X, Y, say, $X, Y \sim \mathrm{U}(0,1)$, independence can be reached on the 'one-dimensional' probability space $(0,1)$ (recall the end of Sect. 5a). However, the 'two-dimensional' probability space $(0,1)\times(0,1)$ is much more natural here. (And do not think that X^*, Y^* are independent!)

Given any two increasing functions $X^*, Y^* : (0,1) \to \mathbb{R}$, we may construct

$$\begin{aligned} X(\omega) &= X(\omega_1,\omega_2) = X^*(\omega_1)\,, \\ Y(\omega) &= Y(\omega_1,\omega_2) = Y^*(\omega_2)\,, \end{aligned} \tag{5d2}$$

then X, Y are random variables on $\Omega = (0,1)\times(0,1)$ (with 2-dim Lebesgue measure, of course).

5d3 Exercise. Prove that X^* is a quantile function for X defined by (5d2). The same for Y. Prove that $F_{X,Y}(x,y) = F_X(x) F_Y(y)$. (Hint: $\{\omega : X(\omega) \le x\}$ is of the form $(0,p_1)\times(0,1)$ or $(0,p_1]\times(0,1)$.)

We see that X, Y of (5d2) satisfy 5d1 for the special case of $A = (-\infty, x]$, $B = (-\infty, y]$. The general case needs a lemma. (Note that in general, the product of two integrable functions need not be integrable.)

5d4 Lemma. Let $f, g : \mathbb{R} \to \mathbb{R}$ be integrable Borel functions. Then

$$\iint_{\mathbb{R}^2} f(x)g(y)\,dxdy = \Big(\int_{\mathbb{R}} f(x)\,dx \Big) \Big(\int_{\mathbb{R}} g(y)\,dy \Big) .$$

Proof. We apply Fubini theorem and (5c10):[62]

$$\iint_{\mathbb{R}^2} f(x)g(y)\,dxdy = \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} f(x)g(y)\,dy \Big) dx = \\ = \int_{\mathbb{R}} \Big(f(x) \underbrace{\int_{\mathbb{R}} g(y)\,dy}_{=\text{const}} \Big) dx = \Big(\int_{\mathbb{R}} g(y)\,dy \Big) \int_{\mathbb{R}} f(x)\,dx .$$

□

5d5 Exercise. Prove that $\operatorname{mes}_2(A \times B) = (\operatorname{mes} A)(\operatorname{mes} B)$ for all Borel sets $A, B \subset \mathbb{R}$. (Hint: use 5d4 and 4d6.)

5d6 Exercise. Prove that X, Y of (5d2) are independent. (Hint: use 5d5.)

5d7 Definition. The *tensor product* $P = P_1 \otimes P_2$ of one-dimensional probability distributions P_1, P_2 is the two-dimensional distribution P satisfying

$$P(A \times B) = P_1(A) P_2(B)$$

for all Borel sets $A, B \subset \mathbb{R}$.

5d8 Exercise. Prove that the definition is correct, that is, such P exists and is unique. (Hint: for the existence use (5d2)–5d6; for the uniqueness use (5a9).)

5d9. So, the following conditions are equivalent:

- X, Y are independent;
- $P_{X,Y} = P_X \otimes P_Y$;
- $\forall x, y \;\; F_{X,Y}(x, y) = F_X(x) F_Y(y)$.

5d10 Exercise. Prove that random variables X, Y of (5a13) are independent. (Hint: $F_{X,Y}(x, y) = xy$ for all $x, y \in [0, 1]$ of the form $k/2^n$, therefore for all $x, y \in [0, 1]$.)

5d11 Theorem. (a) Let random variables $X, Y : \Omega \to \mathbb{R}$ have densities f_X, f_Y. If X, Y are independent, then the joint density $f_{X,Y}$ exists, and

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) \quad \text{almost everywhere.}$$

(b) Let random variables $X, Y : \Omega \to \mathbb{R}$ have a joint density $f_{X,Y}$. If $f_{X,Y}(x, y)$ can be written in the form $g(x)h(y)$ for some $g, h : \mathbb{R} \to \mathbb{R}$ then X, Y are independent.

[62]First, we apply Fubini theorem to $|f(x)|$ and $|g(y)|$, in order to check integrability. After that, we apply it again, to $f(x)$ and $g(y)$.

Proof. (a) The function $f(x,y) = f_X(x)f_Y(y)$ is a 2-dim density, since $\int f(x,y)\,dxdy = \big(\int f_X(x)\,dx\big)\big(\int f_Y(y)\,dy\big) = 1$ (recall 5d4). The corresponding 2-dim distribution P_f satisfies

$$P_f(A \times B) = \iint_{A\times B} f(x,y)\,dxdy = \iint_{\mathbb{R}^2} \mathbf{1}_A(x)\mathbf{1}_B(y)f_X(x)f_Y(y)\,dxdy = \\ = \bigg(\int_{\mathbb{R}} \mathbf{1}_A(x)f_X(x)\,dx\bigg)\bigg(\int_{\mathbb{R}} \mathbf{1}_B(y)f_Y(y)\,dy\bigg) = P_X(A)P_Y(B)\,,$$

which means that $P_f = P_X \otimes P_Y$. However, $P_X \otimes P_Y = P_{X,Y}$ due to independence. So, $P_{X,Y} = P_f$, which means that f is a joint density of X, Y.

(b) By 5c14, X and Y have densities f_X, f_Y, and[63]

$$f_X(x) = \int f_{X,Y}(x,y)\,dy = \int g(x)h(y)\,dy = g(x)\int h(y)\,dy\,;$$

similarly, $f_Y(y) = h(y)\int g(x)\,dx$. It follows that $\big(\int g(x)\,dx\big)\big(\int h(y)\,dy\big) = 1$. Therefore $f_X(x)f_Y(y) = g(x)h(y) = f_{X,Y}(x,y)$. It follows (as was seen in the proof of (a)) that

$$\underbrace{\iint_{A\times B} f_{X,Y}(x,y)\,dxdy}_{P_{X,Y}(A\times B)} = P_X(A)P_Y(B)\,,$$

which means that X, Y are independent. □

5d12 Theorem. $\mathbb{E}\,(XY) = (\mathbb{E}\,X)(\mathbb{E}\,Y)$ for any independent integrable random variables X, Y.

Proof. Due to 5a18 we may restrict ourselves to the model of (5d2):

$$(\Omega, \mathcal{F}, P) = \big((0,1)\times(0,1), \mathcal{B}_2|_{(0,1)\times(0,1)}, \operatorname{mes}_2|_{(0,1)\times(0,1)}\big)\,;\\ X(\omega) = X(\omega_1,\omega_2) = X^*(\omega_1)\,,\\ Y(\omega) = Y(\omega_1,\omega_2) = Y^*(\omega_2)\,.$$

We apply 5d4:

$$\mathbb{E}\,(XY) = \int_\Omega XY\,dP = \iint_{(0,1)\times(0,1)} X(\omega_1,\omega_2)Y(\omega_1,\omega_2)\,d\omega_1 d\omega_2 = \\ \iint_{(0,1)\times(0,1)} X^*(\omega_1)Y^*(\omega_2)\,d\omega_1 d\omega_2 = \bigg(\int_0^1 X^*(\omega_1)\,d\omega_1\bigg)\bigg(\int_0^1 Y^*(\omega_2)\,d\omega_2\bigg) = (\mathbb{E}\,X)(\mathbb{E}\,Y)\,.$$

□

In terms of distributions,

$$\iint xy\,d(P_X \otimes P_Y) = \bigg(\int x\,dP_X\bigg)\bigg(\int y\,dP_Y\bigg)\,.$$

[63] Equalities for densities hold almost everywhere, as usual. Note that g, h are Borel functions, by 5c4.

In terms of densities, if they exist,

$$\iint xy \underbrace{f_{X,Y}(x,y)}_{=f_X(x)f_Y(y)} \, dxdy = \Big(\int x f_X(x) \, dx \Big) \Big(\int y f_Y(y) \, dy \Big) ;$$

compare it with 5d4.

5d13 Lemma. Let X, Y be random variables, and $\varphi, \psi : \mathbb{R} \to \mathbb{R}$ Borel functions. If X and Y are independent then $\varphi(X)$ and $\psi(Y)$ are independent.

Proof.

$$\begin{aligned} \mathbb{P}\big(\varphi(X) \in A, \, \psi(Y) \in B \big) &= \mathbb{P}\big(X \in \varphi^{-1}(A), \, Y \in \psi^{-1}(B) \big) = \\ &= \mathbb{P}\big(X \in \varphi^{-1}(A) \big) \, \mathbb{P}\big(Y \in \psi^{-1}(B) \big) = \mathbb{P}\big(\varphi(X) \in A \big) \, \mathbb{P}\big(\psi(Y) \in B \big) . \end{aligned}$$

$\square$

5d14 Corollary. $\mathbb{E}\big(\varphi(X)\psi(Y) \big) = \big(\mathbb{E}\, \varphi(X) \big) \big(\mathbb{E}\, \psi(Y) \big)$ for any independent random variables X, Y and Borel functions φ, ψ such that $\varphi(X)$ and $\psi(Y)$ are integrable.

5d15 Note. If they are not integrable, the formula still holds under appropriate conventions: $(+\infty) \cdot (+\infty) = +\infty$, $(-\infty) \cdot (+\infty) = -\infty$, $(\text{a positive number}) \cdot (+\infty) = +\infty$, $(\text{a positive number}) \cdot (\infty - \infty) = \infty - \infty$, etc. (Think, however: what about $0 \cdot \infty$?)

All said (in Sect. 5d) about dimension 2 holds for all dimensions $d = 2, 3, \dots$ You can easily formulate such generalizations. (Do not confuse independence of $X_1, \dots, X_n$ with their *pairwise* independence.)

5e Independence of infinite sequences

5e1 Definition. (a) Random variables $X_1, X_2, \dots : \Omega \to \mathbb{R}$ are *independent*, if $X_1, \dots, X_n$ are independent for each n.

(b) Events $A_1, A_2, \dots \subset \Omega$ are *independent*, if their indicators $\mathbf{1}_{A_1}, \mathbf{1}_{A_2}, \dots$ are independent random variables.

That is, (a) means

$$\mathbb{P}\big(X_1 \in B_1, \dots, X_n \in B_n \big) = \mathbb{P}\big(X_1 \in B_1 \big) \dots \mathbb{P}\big(X_n \in B_n \big)$$

for all n and all Borel sets $B_1, \dots, B_n \subset \mathbb{R}$.

Do not think that (b) requires just $\mathbb{P}\big(A_1 \cap \dots \cap A_n \big) = \mathbb{P}\big(A_1 \big) \dots \mathbb{P}\big(A_n \big)$ for all n. It requires also, say, $\mathbb{P}\big(A_1 \cap \overline{A_2} \cap A_3 \big) = \mathbb{P}\big(A_1 \big) \mathbb{P}\big(\overline{A_2} \big) \mathbb{P}\big(A_3 \big)$, etc. (Of course, $\overline{A} = \Omega \setminus A$ and $\mathbb{P}\big(\overline{A} \big) = 1 - \mathbb{P}\big(A \big)$.)

Infinite random sequences often are defined on infinite-dimensional probability spaces, but they can be defined on $(0, 1)$ by a trick that generalizes (5a13):

$$(0,1) \ni \omega = \big(0.\beta_1\beta_2\beta_3 \dots \big)_2 ; \qquad \begin{aligned} X_1(\omega) &= \big(0.\beta_1\beta_3\beta_5 \dots \big)_2 , \\ X_2(\omega) &= \big(0.\beta_2\beta_6\beta_{10} \dots \big)_2 , \\ X_3(\omega) &= \big(0.\beta_4\beta_{12}\beta_{20} \dots \big)_2 , \\ &\dots \end{aligned} \tag{5e2}$$