6 Distributions

6a Dimension one

We know that every random variable $X : \Omega \to \mathbb{R}$ has its cumulative distribution function F_X. The function gives us probabilities of the form $\mathbb{P}\big(X \le x \big)$ directly; it also gives indirectly (recall (2a3)) probabilities of the form $\mathbb{P}\big(X \in B \big)$ for elementary sets B. However, we know (recall 5d6 and 5f6) that $\mathbb{P}\big(X \in B \big)$ is well-defined, and uniquely determined by F_X, for every Borel set B (elementary or not). We also know that a random variable may be defined (recall 5d7) without special attention to elementary sets. How to describe a distribution without special attention to elementary sets?

6a1 Theorem. Let $X : \Omega \to \mathbb{R}$ be a random variable. Then the function $P_X : \mathcal{B}_1 \to [0,1]$ defined by

$$P_X(B) = \mathbb{P}\big(X \in B \big)$$

is a Borel probability measure.[1]

Proof. We have $P_X(B) = P\big(X^{-1}(B)\big)$. Combining properties of the "inverse image" operation (recall the proof of 5d6) with properties of the probability measure P (recall 1a1) we get the needed properties of P_X:

$$\begin{aligned}
&P_X(\mathbb{R}) = P\big(X^{-1}(\mathbb{R})\big) = P(\Omega) = 1\,;\\
&P_X(A \uplus B) = P\big(X^{-1}(A \uplus B)\big) = P\big(X^{-1}(A) \uplus X^{-1}(B)\big) =\\
&\qquad = P\big(X^{-1}(A)\big) + P\big(X^{-1}(B)\big) = P_X(A) + P_X(B)\,;\\
&P_X(A_1 \uplus A_2 \uplus \dots) = P\big(X^{-1}(A_1 \uplus A_2 \uplus \dots)\big) = P\big(X^{-1}(A_1) \uplus X^{-1}(A_2) \uplus \dots\big) =\\
&\qquad = P\big(X^{-1}(A_1)\big) + P\big(X^{-1}(A_2)\big) + \dots = P_X(A_1) + P_X(A_2) + \dots
\end{aligned}$$

□

6a2 Definition. (a) A one-dimensional probability distribution[2] is a probability measure on the one-dimensional Borel σ-field $\mathcal{B}_1$.

(b) The distribution P_X of a random variable $X : \Omega \to \mathbb{R}$ is defined by $P_X(B) = P\big(\{\omega : X(\omega) \in B\}\big)$.

6a3 Note. If μ is a distribution then $(\mathbb{R}, \mathcal{B}, \mu)$ is a probability space.[3] The identical map $X_*^{(\mu)} : \mathbb{R} \to \mathbb{R}$, $X_*^{(\mu)}(x) = x$, may be treated as a random variable on that probability space, and the distribution of $X_*^{(\mu)}$ is μ. Indeed, $P_{X_*^{(\mu)}}(B) = \mathbb{P}\big(X_*^{(\mu)} \in B \big) = \mu\big((X_*^{(\mu)})^{-1}(B)\big) = \mu(B)$.

6a4 Proposition. The formula $F_\mu(x) = \mu\big((-\infty, x]\big)$ establishes a bijective correspondence between one-dimensional probability distributions μ and (one-dimensional, cumulative) distribution functions.[4]

---

[1] Recall 5f: a probability measure on the Borel σ-field is called a Borel probability measure.

[2] Or simply "a distribution".

[3] $\mathcal{B}$ stands for $\mathcal{B}_1$, as usual.

[4] Recall 2b15, 2b16.

Proof. (1) If μ is a probability distribution then F_μ is a distribution function. Indeed, consider the random variable $X = X_*^{(\mu)}$; by 6a3, μ is the distribution of X; thus, F_μ is the distribution function of X.

(2) If $F_\mu = F_\nu$ then $\mu = \nu$ by 5f6.

(3) If F is a distribution function then exists μ such that $F = F_\mu$. Indeed, by 2b15 exists a random variable X (on some probability space) such that $F = F_X$ (the distribution function of X). Let $\mu = P_X$ (the distribution of X), then $F_\mu = F$, since $F_\mu(x) = \mu\big((-\infty, x]\big) = P_X\big((-\infty, x]\big) = \mathbb{P}\big(X \le x \big) = F_X(x)$. $\square$

Given a probability distribution μ, we can construct two random variables $X_*^{(\mu)}$, $X^*_{(\mu)}$, both distributed μ, but defined on different probability spaces. Namely, $X_*^{(\mu)}$ is defined on $(\mathbb{R}, \mathcal{B}, \mu)$, while $X^*_{(\mu)}$ is defined on $\big((0,1), \mathcal{B}|_{(0,1)}, \operatorname{mes}|_{(0,1)}\big)$. The function $X_*^{(\mu)} : \mathbb{R} \to \mathbb{R}$ is just the identical function, while the function $X^*_{(\mu)} : (0,1) \to \mathbb{R}$ is the quantile function of μ (also of F_μ and of $X_*^{(\mu)}$).

Recall that random variables X, Y (possibly defined on different probability spaces) are called *identically distributed,* if $F_X = F_Y$; or equivalently, $P_X = P_Y$.

We see that $X_*^{(\mu)}$ and $X^*_{(\mu)}$ are identically distributed (both are distributed μ).

Given a random variable X, we may consider two other random variables, $X_* = X_*^{(\mu)}$ and $X^* = X^*_{(\mu)}$, where $\mu = P_X$ (the distribution of X). All the three random variables are identically distributed:

$$
\begin{aligned}
&(\Omega, \mathcal{F}, P) \xrightarrow{X} \mathbb{R}; && \text{(the original r.v.)} \\
&\big((0,1), \mathcal{B}|_{(0,1)}, \operatorname{mes}|_{(0,1)}\big) \xrightarrow{X^*} \mathbb{R}; && \text{(its quantile function)} \\
&(\mathbb{R}, \mathcal{B}, P_X) \xrightarrow{X_*} \mathbb{R}; && X_*(x) = x\,. \\
&X, X^*, X_* \text{ are identically distributed.}
\end{aligned} \tag{6a5}
$$

6a6 Lemma. Let X, Y be random variables (possibly on different probability spaces) and $\varphi : \mathbb{R} \to \mathbb{R}$ a Borel function. If X, Y are identically distributed then $\varphi(X), \varphi(Y)$ are identically distributed.

Proof. For every Borel set $B \subset \mathbb{R}$ we have $P_{\varphi(X)}(B) = \mathbb{P}\big(\varphi(X) \in B \big) = \mathbb{P}\big(X \in \varphi^{-1}(B) \big) = \mathbb{P}\big(Y \in \varphi^{-1}(B) \big) = \mathbb{P}\big(\varphi(Y) \in B \big) = P_{\varphi(Y)}(B)$. $\square$

It follows that $\mathbb{E}\,\varphi(X) = \mathbb{E}\,\varphi(X^*) = \mathbb{E}\,\varphi(X_*)$.[5] Writing each expectation as Lebesgue integral we get $\int \varphi(X)\,dP = \int \varphi(X^*)\,d(\operatorname{mes}) = \int \varphi\,dP_X$, that is,

$$
\underbrace{\int_\Omega \varphi\big(X(\omega)\big)\,dP(\omega)}_{\mathbb{E}\,(\varphi(X))} = \underbrace{\int_0^1 \varphi\big(X^*(p)\big)\,dp}_{\mathbb{E}\,(\varphi(X^*))} = \underbrace{\int_{-\infty}^{+\infty} \varphi(x)\,dF_X(x)}_{\mathbb{E}\,(\varphi(X_*))}\,. \tag{6a7}
$$

Here and henceforth dF_X means the same as dP_X (though F_X need not be differentiable, moreover, it may jump); that is a traditional notation.

[5] Either they all are integrable, or all give $+\infty$, or all give $-\infty$, or all are undefined $(\infty - \infty)$.

6a8 Note. Lebesgue integral over a Borel set $B \subset \mathbb{R}$ is defined by

$$\int_B f(x)\,dx = \int_{-\infty}^{+\infty} \mathbf{1}_B(x) f(x)\,dx\,.$$

6a9 Proposition. Let $f : \mathbb{R} \to \mathbb{R}$ be a Borel function such that $f(\cdot) \ge 0$ almost everywhere and $\int_{-\infty}^{+\infty} f(x)\,dx = 1$. Then the equality

$$\mu_f(B) = \int_B f(x)\,dx \qquad \text{for } B \in \mathcal{B}$$

defines a one-dimensional probability distribution μ_f.[6]

Proof. We have $\mathbf{1}_{A \uplus B} = \mathbf{1}_A + \mathbf{1}_B$, therefore $\mu_f(A \uplus B) = \int_{A\uplus B} f(x)\,dx = \int_{-\infty}^{+\infty} \mathbf{1}_{A\uplus B}(x) f(x)\,dx = \int_{-\infty}^{+\infty} \bigl(\mathbf{1}_A(x) + \mathbf{1}_B(x)\bigr) f(x)\,dx = \int_{-\infty}^{+\infty} \mathbf{1}_A(x) f(x)\,dx + \int_{-\infty}^{+\infty} \mathbf{1}_B(x) f(x)\,dx = \int_A f(x)\,dx + \int_B f(x)\,dx = \mu_f(A) + \mu_f(B)$. Clearly, $\mu_f(\mathbb{R}) = 1$, and $\mu_f : \mathcal{B} \to [0,1]$. If $B_n \uparrow B$ then $\mathbf{1}_{B_n}(x) \uparrow \mathbf{1}_B(x)$ for every x, therefore $\mathbf{1}_{B_n}(x) f(x) \uparrow \mathbf{1}_B(x) f(x)$; by Monotone convergence theorem, $\int \mathbf{1}_{B_n}(x) f(x)\,dx \uparrow \int \mathbf{1}_B(x) f(x)\,dx$, that is, $\mu_f(B_n) \uparrow \mu_f(B)$. Countable additivity of μ_f follows.

$\square$

If $f(\cdot) = g(\cdot)$ almost everywhere then $\mu_f = \mu_g$ (check it).

6a10 Lemma. Let $f, g : \mathbb{R} \to \mathbb{R}$ be integrable Borel functions such that $\int_B f(x)\,dx = \int_B g(x)\,dx$ for all Borel sets $B \subset \mathbb{R}$. Then $f(\cdot) = g(\cdot)$ almost everywhere.

Proof. The set $B = \{x : f(x) - g(x) > 0\}$ is Borel (since the function $f - g$ is Borel). We have $\mathbf{1}_B(x) \cdot \bigl(f(x) - g(x)\bigr) \ge 0$ for all x, and $\int \mathbf{1}_B(x) \cdot \bigl(f(x) - g(x)\bigr)\,dx = \int_B f(x)\,dx - \int_B g(x)\,dx = 0$; by (5g5), $\mathbf{1}_B(x) \cdot \bigl(f(x) - g(x)\bigr) = 0$ for almost all x, which means that $\mathbf{1}_B(\cdot) = 0$ almost everywhere, that is, $f(x) \le g(x)$ almost everywhere. Similarly, $f(x) \ge g(x)$ almost everywhere. $\square$

6a11 Corollary. Let $f, g : \mathbb{R} \to \mathbb{R}$ be integrable Borel functions such that $\int_{-\infty}^x f(u)\,du = \int_{-\infty}^x g(u)\,du$ for all $x \in \mathbb{R}$. Then $f(\cdot) = g(\cdot)$ almost everywhere.

Proof. We have $\mu_f\bigl((-\infty, x]\bigr) = \mu_g\bigl((-\infty, x]\bigr)$ for all x, therefore $\mu_f = \mu_g$ by 5f6, and 6a10 applies. $\square$

6a12 Note. We see that the function $F(x) = \int_{-\infty}^x f(u)\,du$ determines $f(\cdot)$ up to equivalence. In fact, the derivative $F'(x)$ exists and is equal to $f(x)$ for almost all x (which is more difficult to prove).

6a13 Definition. (a) A one-dimensional probability density[7] is a Borel function $f : \mathbb{R} \to \mathbb{R}$ such that $f(\cdot) \ge 0$ almost everywhere, and $\int_{-\infty}^{+\infty} f(x)\,dx = 1$.

(b) A distribution μ has a density, if $\mu = \mu_f$ for some density f. Each such f is a density of μ.

(c) A density of a random variable is, by definition, a density of its distribution.

---

[6] Integrals are treated as Lebesgue integrals, as usual. By definition, $\int_B f(x)\,dx = \int_{-\infty}^{+\infty} \mathbf{1}_B(x) f(x)\,dx$.

[7] Or simply “density”.

So,

$$\mathbb{P}\left(X \in B \right) = \int_B f_X(x)\,dx \tag{6a14}$$

for every Borel set $B \subset \mathbb{R}$, which is just the definition of the density f_X of X. Such f_X is determined by X up to equivalence (that is, equality almost everywhere). However, some distributions have no densities, as we'll see in Sect. 6c.

We may write (6a14) as $\mathbb{E}\left(\mathbf{1}_B(X)\right) = \int \mathbf{1}_B(x) f_X(x)\,dx$, which is $\mathbb{E}\,\varphi(X) = \int \varphi(x) f_X(x)\,dx$ for a Borel function $\varphi : \mathbb{R} \to \{0,1\}$.

6a15 Theorem. Let a random variable X have a density f_X, and $\varphi : \mathbb{R} \to \mathbb{R}$ be a bounded Borel function. Then

$$\mathbb{E}\,\varphi(X) = \int_{-\infty}^{+\infty} \varphi(x) f_X(x)\,dx\,.$$

Proof. We know that the statement holds when φ is an indicator. By linearity it holds when φ is a linear combination of indicators ($c_1\mathbf{1}_{B_1} + \dots + c_n\mathbf{1}_{B_n}$), that is, when φ takes on a finite number of values.

In general, similarly to 4a12, for any $\varepsilon > 0$ there are Borel functions[8] $\varphi_\ominus, \varphi_\oplus : \mathbb{R} \to \mathbb{R}$, each having a finite number of values, such that

$$\begin{gathered} \varphi_\ominus(x) \le \varphi(x) \le \varphi_\oplus(x)\,, \\ \varphi_\oplus(x) - \varphi_\ominus(x) \le \varepsilon \end{gathered}$$

for all $x \in \mathbb{R}$. The interval

$$\Big[\int \varphi_\ominus(x) f_X(x)\,dx, \int \varphi_\oplus(x) f_X(x)\,dx \Big] = \Big[\mathbb{E}\,\varphi_\ominus(X), \mathbb{E}\,\varphi_\oplus(X) \Big]$$

of length $\le \varepsilon$ contains both $\int \varphi(x) f_X(x)\,dx$ and $\mathbb{E}\,\varphi(X)$, therefore

$$\Big| \mathbb{E}\,\varphi(X) - \int \varphi(x) f_X(x)\,dx \Big| \le \varepsilon\,.$$

The inequality holds for all $\varepsilon > 0$, therefore $\mathbb{E}\,\varphi(X) = \int \varphi(x) f_X(x)\,dx$. □

6a16 Note. The formula $\mathbb{E}\,\varphi(X) = \int \varphi(x) f_X(x)\,dx$ holds also for unbounded φ,[9] which can be shown by applying Monotone convergence theorem.

6b Dimension two

Many statements of Sect. 6a have their 2-dim counterparts. A pair (X,Y) of random variables $X, Y : \Omega \to \mathbb{R}$ (defined on the same probability space) may be treated as a *two-dimensional random variable*, $(X,Y) : \Omega \to \mathbb{R}^2$.

[8] Do you understand why they are Borel functions? Could we use step functions instead?

[9] Once again: either both are integrable, or...

6b1 Theorem. Let $X, Y : \Omega \to \mathbb{R}$ be random variables. Then the function $P_{X,Y} : \mathcal{B}_2 \to [0,1]$ defined by

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

is a two-dimensional Borel probability measure.

The proof is similar to 6a1. Note that X, Y need not be independent.

6b2 Definition. (a) A two-dimensional probability distribution is a probability measure on the two-dimensional Borel σ-field $\mathcal{B}_2$.

(b) The *joint distribution* $P_{X,Y}$ of random variables $X, Y : \Omega \to \mathbb{R}$ is defined by $P_{X,Y}(B) = P\big(\{\omega : (X(\omega), Y(\omega)) \in B\}\big)$.

6b3 Note. If μ is a 2-dim distribution then $(\mathbb{R}^2, \mathcal{B}_2, \mu)$ is a probability space. 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 $X, Y : \mathbb{R}^2 \to \mathbb{R}$, $X(x,y) = x$, $Y(x,y) = y$. The joint distribution of X, Y is μ.

However, the quantile function $X^*(\cdot)$ has no straightforward 2-dim counterpart. We have no such things as $(X,Y)^*(p)$ or $(X,Y)^*(p,q)$. Of course, we still have $X^*(p)$ and $Y^*(p)$, but these give no information about dependence between X and Y. (A useful 2-dim generalization of quantile functions will be given in Sect. 7.)

Every 2-dim probability distribution μ has its *2-dim cumulative distribution function* $F_\mu(\cdot,\cdot)$:

$$F_\mu(x,y) = \mu\big((-\infty, x] \times (-\infty, y]\big) \,.$$

A 2-dim counterpart of 2a4, 2b15 (a complete list of properties of $F_\mu(\cdot,\cdot)$) is too complicated for our course.[10]

6b4 Proposition. If $F_\mu = F_\nu$ then $\mu = \nu$ (for 2-dim probability distributions μ, ν).

The proof is similar to 5f6 (and 5f5).

The 2-dim cumulative distribution function for a joint distribution,

$$\begin{aligned} F_{X,Y}(x,y) &= \mathbb{P}\,\big(X \le x \,\&\, Y \le y \big) = P\big(\{\omega : X(\omega) \le x \,\&\, Y(\omega) \le y\}\big) = \\ &= P\big(\{\omega : X(\omega) \le x\} \cap \{\omega : Y(\omega) \le y\}\big) = P\big(X^{-1}((-\infty, x]) \cap Y^{-1}((-\infty, y])\big) \end{aligned} \tag{6b5}$$

is useful, though less useful than the 1-dim cumulative distribution function.

6b6 Lemma. Let $X_1, Y_1 : \Omega_1 \to \mathbb{R}$ and $X_2, Y_2 : \Omega_2 \to \mathbb{R}$ be random variables and $\varphi : \mathbb{R}^2 \to \mathbb{R}$ a Borel function. If $P_{X_1,Y_1} = P_{X_2,Y_2}$ (that is, 2-dim random variables (X_1, Y_1) and (X_2, Y_2) are identically distributed) then $\varphi(X_1, Y_1)$ and $\varphi(X_2, Y_2)$ are identically distributed.

The proof is similar to 6a6.

6b7 Corollary. If $F_{X_1,Y_1} = F_{X_2,Y_2}$ then $\mathbb{E}\,(X_1 Y_1) = \mathbb{E}\,(X_2 Y_2)$.[11]

---

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

[11] Either both are integrable, or both give $+\infty$, or both give $-\infty$, or both are undefined $(\infty - \infty)$.

Proof. Combine 6b4 and 6b6 for $\varphi(x,y) = xy$:

$$F_{X_1,Y_1} = F_{X_2,Y_2} \quad\Longrightarrow\quad P_{X_1,Y_1} = P_{X_2,Y_2} \quad\Longrightarrow\quad P_{X_1Y_1} = P_{X_2Y_2} \quad\Longrightarrow\quad \mathbb{E}\,(X_1Y_1) = \mathbb{E}\,(X_2Y_2)\,.$$

$\square$

An explicit formula for $\mathbb{E}\,(XY)$ in terms of $F_{X,Y}$ (that is, a 2-dim generalization of (4a3)–(4a18)–4b5) is too complicated for our course.

Lebesgue integral defined by 5g1 for dimension one is easy to generalize for dimension two, provided that we have two-dimensional Lebesgue measure (mes_2). Similarly to 5f8, mes_2 is defined as a locally finite measure on $\mathcal{B}_2$ such that

$$\mathrm{mes}_2\big((a,b)\times(c,d)\big) = (b-a)(d-c)$$

whenever $-\infty < a < b < \infty$, $-\infty < c < d < \infty$. Uniqueness of mes_2 is ensured by the 2-dim counterpart of 5f7. Existence of mes_2 follows from measure theory (like 5f9), but we'll see in Sect. 6d that existence of mes_2 follows from existence of mes_1.

6b8 Definition. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a Borel function.

(a) f is *integrable,* if $\int_0^\infty \mathrm{mes}_2\{(x,y)\in\mathbb{R}^2 : |f(x,y)| > z\}\,dz < \infty$; here the integral is treated as improper Riemann integral;

(b) for an integrable f, its *Lebesgue integral* is

$$\iint_{\mathbb{R}^2} f(x,y)\,dxdy = \int_0^\infty \Big(\mathrm{mes}_2\{(x,y)\in\mathbb{R}^2 : f(x,y) > z\} - \mathrm{mes}_2\{(x,y)\in\mathbb{R}^2 : f(x,y) < -z\}\Big)\,dz\,;$$

the integral on the right-hand side is treated as improper Riemann integral.

Properties (5g3), (5g4), (5g5) and convergence theorems 5g6, 5g7 still hold in two dimensions. Lebesgue integral over a Borel set $B \subset \mathbb{R}^2$ is defined by

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

6b9 Proposition. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a Borel function such that $f(\cdot,\cdot) \ge 0$ almost everywhere and $\iint_{\mathbb{R}^2} f(x,y)\,dxdy = 1$. Then the equality

$$\mu_f(B) = \iint_B f(x,y)\,dxdy \qquad \text{for } B \in \mathcal{B}_2$$

defines a two-dimensional probability distribution μ_f.

The proof is similar to 6a9.

Still, if $f(\cdot,\cdot) = g(\cdot,\cdot)$ almost everywhere then $\mu_f = \mu_g$.

6b10 Lemma. Let $f, g : \mathbb{R}^2 \to \mathbb{R}$ be integrable Borel functions such that $\iint_B f(x,y)\,dxdy = \iint_B g(x,y)\,dxdy$ for all Borel sets $B \subset \mathbb{R}^2$. Then $f(\cdot,\cdot) = g(\cdot,\cdot)$ almost everywhere.

The proof is similar to 6a10.

6b11 Corollary. Let $f, g : \mathbb{R}^2 \to \mathbb{R}$ be integrable Borel functions such that $\iint_{(-\infty,x]\times(-\infty,y]} f(u,v)\,dudv = \iint_{(-\infty,x]\times(-\infty,y]} g(u,v)\,dudv$ for all $(x,y) \in \mathbb{R}^2$. Then $f(\cdot,\cdot) = g(\cdot,\cdot)$ almost everywhere.

The proof is similar to 6a11.

6b12 Note. We see that the function $F(x,y) = \iint_{(-\infty,x]\times(-\infty,y]} f(u,v)\,dudv$ determines $f(\cdot,\cdot)$ up to equivalence. In fact, $f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y)$ for almost all $(x,y) \in \mathbb{R}^2$, if $\frac{\partial^2}{\partial x \partial y}$ is defined appropriately (which is too complicated for our course).

6b13 Definition. (a) A two-dimensional probability density is a Borel function $f : \mathbb{R}^2 \to \mathbb{R}$ such that $f(\cdot,\cdot) \ge 0$ almost everywhere, and $\iint_{\mathbb{R}^2} f(x,y)\,dxdy = 1$.

(b) A two-dimensional distribution μ has a density, if $\mu = \mu_f$ for some two-dimensional density f. Each such f is a density of μ.

(c) A density of a two-dimensional random variable (X,Y) is, by definition, a density of its two-dimensional distribution.

So,

$$\mathbb{P}\big((X,Y) \in B \big) = \iint_B f_{X,Y}(x,y)\,dxdy \tag{6b14}$$

for every Borel set $B \subset \mathbb{R}^2$, which is just the definition of the joint density $f_{X,Y}$ of X, Y. Such $f_{X,Y}$ is determined by X, Y up to equivalence (that is, equality almost everywhere). However, some 2-dim distributions have no densities, as we'll see in Sect. 6e.

6b15 Theorem. Let random variables X, Y have a joint density $f_{X,Y}$, and $\varphi : \mathbb{R}^2 \to \mathbb{R}$ be a bounded Borel function. Then

$$\mathbb{E}\,\varphi(X,Y) = \iint_{\mathbb{R}^2} \varphi(x,y) f_{X,Y}(x,y)\,dxdy\,.$$

The proof is similar to 6a15.

6b16 Note. The formula of 6b15 holds also for unbounded φ (either both are integrable, or...).

6c Relations between dimensions one and two

Two random variables $X, Y : \Omega \to \mathbb{R}$ give rise to three probability distributions: P_X, P_Y and $P_{X,Y}$. We have

$$P_X(B) = P_{X,Y}(B \times \mathbb{R})\,, \qquad P_Y(B) = P_{X,Y}(\mathbb{R} \times B) \tag{6c1}$$

for every Borel set $B \subset \mathbb{R}$. Indeed, $P_{X,Y}(B \times \mathbb{R}) = \mathbb{P}\big((X,Y) \in B \times \mathbb{R} \big) = \mathbb{P}\big(X \in B \,\&\, Y \in \mathbb{R} \big) = \mathbb{P}\big(X \in B \big) = P_X(B)$. One-dimensional distributions P_X, P_Y are called *marginal distributions* (or simply "marginals") of the 2-dim distribution $P_{X,Y}$. Note that $P_{X,Y}$ determines P_X, P_Y uniquely, but the converse is wrong; P_X, P_Y (taken together) do not determine $P_{X,Y}$, since they give no information about dependence between X and Y. In terms of cumulative distribution functions, (6c1) becomes[12]

(6c2) $$F_X(x) = F_{X,Y}(x, +\infty)\,, \qquad F_Y(y) = F_{X,Y}(+\infty, y)\,.$$

Once again, $F_{X,Y}$ determines F_X, F_Y uniquely, but the converse is wrong.

What about marginal densities? Here we need another important result of measure theory: Fubini theorem.

6c3 Exercise. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a Borel function. Prove that for every x the section $f_x : \mathbb{R} \to \mathbb{R}$ defined by $f_x(y) = f(x, y)$ is a Borel function.

Hint. f_x is the composition of the (continuous) embedding $\mathbb{R} \ni y \mapsto (x, y) \in \mathbb{R}^2$ and the Borel function f.

A traditional notation for the integral of a section:

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

the integral is a function of x only. Note that

(6c4) $$\int_{-\infty}^{+\infty} f(x, y)g(x)\,dy = g(x)\int_{-\infty}^{+\infty} f(x, y)\,dy\,;$$

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

6c5 Theorem. (Fubini) For every Borel function $f : \mathbb{R}^2 \to \mathbb{R}$,

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

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

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

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

I give no proof.

6c6 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.

[12] What do you think about $F_{X,Y}(x, -\infty)$?

Proof. Functions f_X, f_Y defined by these integrals are densities (due to Fubini theorem). We have to prove that $P_X(B) = \int_B f_X(x)\,dx$, and the same for Y. We use (6c1), Fubini theorem and (6c4):

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

and the same for Y. □

6d Independence

6d1 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 example, binary digits α_1, α_2 of $\omega = (0.\alpha_1\alpha_2\dots)_2 \in (0,1)$ are independent (on the probability space $\bigl((0,1), \mathcal{B}|_{(0,1)}, \operatorname{mes}|_{(0,1)}\bigr)$). What about, say, the uniform distribution $U(0,1)$? Can we find functions $X, Y : (0,1) \to \mathbb{R}$ such that $X, Y \sim U(0,1)$ and X, Y are independent? Smooth functions cannot do so, since a smooth curve cannot fill the square $(0,1) \times (0,1)$.

Turn to binary digits: $X = (0.\beta_1\beta_2\dots)_2$, $Y = (0.\gamma_1\gamma_2\dots)_2$. For X to be uniform on $(0,1)$, its digits β_k must be "coin tossing". The same for Y. For X, Y to be independent, the whole collection $(\beta_1, \gamma_1, \beta_2, \gamma_2, \dots)$ of binary digits must be "coin tossing". Why not? We can do it as follows:

$$\omega = (0.\alpha_1\alpha_2\alpha_3\alpha_4\dots)_2 = (0.\beta_1\gamma_1\beta_2\gamma_2\dots)_2\,; \qquad \beta_k = \alpha_{2k-1},\ \gamma_k = \alpha_{2k}\,.$$

That is, we define functions $X, Y : (0,1) \to \mathbb{R}$ by

$$X\biggl(\sum_{k=1}^\infty \frac{\alpha_k}{2^k} \biggr) = \sum_{k=1}^\infty \frac{\alpha_{2k-1}}{2^k}\,, \qquad Y\biggl(\sum_{k=1}^\infty \frac{\alpha_k}{2^k} \biggr) = \sum_{k=1}^\infty \frac{\alpha_{2k}}{2^k}\,. \tag{6d2}$$

These functions are far from being smooth! They are discontinuous and non-monotone on *every* interval. Still, they are Borel functions,[13] satisfying 6d1 (as we'll see later). It means that a probability space (unlike a topological space) has, in some sense, no dimension! It

[13]Indeed, let $X\bigl(\sum_{k=1}^\infty \frac{\alpha_k}{2^k}\bigr) = \sum_{k=1}^n \frac{\alpha_{2k-1}}{2^k}$, then X_n is a step function, and $X_n(\omega) \xrightarrow[n\to\infty]{} X(\omega)$. (In fact, X and Y are *Riemann* integrable.)

shows also that existence of mes_2 can be deduced from existence of mes_1. Indeed, the joint distribution of these X, Y is $\mathrm{mes}_2 \big|_{(0,1)\times(0,1)}$.

However, one-dimensional models for two-dimensional distributions are rather unnatural. We turn to two-dimensional models.

6d3 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. Apply Fubini theorem and (6c4):

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

□

6d4 Corollary. $\mathrm{mes}_2(A \times B) = (\mathrm{mes}\,A)(\mathrm{mes}\,B)$ for all Borel sets $A, B \subset \mathbb{R}$.

Proof. It suffices to consider the case $\mathrm{mes}\,A < \infty$, $\mathrm{mes}\,B < \infty$. Apply 6d3 to $f = \mathbf{1}_A$ and $g = \mathbf{1}_B$:

$$\mathrm{mes}_2(A \times B) = \iint_{\mathbb{R}^2} \mathbf{1}_{A\times B}(x,y)\,dxdy = \iint_{\mathbb{R}^2} \mathbf{1}_A(x)\mathbf{1}_B(y)\,dxdy = \\ = \Big(\int_{\mathbb{R}} \mathbf{1}_A(x)\,dx \Big)\Big(\int_{\mathbb{R}} \mathbf{1}_B(y)\,dy \Big) = (\mathrm{mes}\,A)(\mathrm{mes}\,B) .$$

□

6d5 Proposition. Let μ, ν be one-dimensional probability distributions. Then there exists a probability space $(\Omega, \mathcal{F}, P)$ and random variables $X, Y : \Omega \to \mathbb{R}$ such that

$$P_X = \mu\,; \quad P_Y = \nu\,; \qquad X \text{ and } Y \text{ are independent.}$$

Proof. We know that there are increasing functions $X^*, Y^* : (0,1) \to \mathbb{R}$ such that X^* is distributed μ and Y^* is distributed ν. We take $\Omega = (0,1) \times (0,1)$, $\mathcal{F} = \mathcal{B}_2\big|_{\Omega}$, $P = \mathrm{mes}_2\big|_{\Omega}$, and define $X, Y : \Omega \to \mathbb{R}$ as follows: for every $\omega = (\omega_1, \omega_2) \in \Omega$,

$$X(\omega) = X(\omega_1, \omega_2) = X^*(\omega_1) , \\ Y(\omega) = Y(\omega_1, \omega_2) = Y^*(\omega_2) .$$

For any Borel sets $A, B \subset \mathbb{R}$,

$$\mathbb{P}\big(X \in A \,\&\, Y \in B \big) = P\big(\{\omega \in \Omega : X(\omega) \in A \,\&\, Y(\omega) \in B\}\big) = \\ = \mathrm{mes}_2\big(\{(\omega_1,\omega_2) : X^*(\omega_1) \in A_1 \,\&\, Y^*(\omega_2) \in B\}\big) = \mathrm{mes}_2\big((X^*)^{-1}(A) \times (Y^*)^{-1}(B)\big) = \\ = \mathrm{mes}\big((X^*)^{-1}(A)\big) \cdot \mathrm{mes}\big((Y^*)^{-1}(B)\big) = \mathbb{P}\big(X^* \in A \big) \cdot \mathbb{P}\big(Y^* \in B \big) = \mu(A)\nu(B) .$$

In particular, when $B = \mathbb{R}$ we get $\mathbb{P}\big(X \in A \big) = \mu(A)$; similarly $\mathbb{P}\big(Y \in B \big) = \nu(B)$. So, X is distributed μ, Y is distributed ν, and X, Y satisfy 6d1, that is, they are independent. □

So, the joint distribution $P_{X,Y}$ satisfies the condition $P_{X,Y}(A \times B) = \mu(A)\nu(B)$ for all $A, B \in \mathcal{B}$. The condition determines $P_{X,Y}$ uniquely, since it determines $P_{X,Y}$ on all 2-dim intervals.

6d6 Definition. The product $\mu \otimes \nu$ of one-dimensional probability distributions μ, ν is the two-dimensional distribution satisfying

$$(\mu \otimes \nu)(A \times B) = \mu(A)\nu(B)$$

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

The following definition is equivalent to 6d1.

6d7 Definition. Random variables $X, Y : \Omega \to \mathbb{R}$ are independent, if

$$P_{X,Y} = P_X \otimes P_Y \,.$$

Taking $A = (-\infty, x]$, $B = (-\infty, y]$ we get

$$F_{\mu\otimes\nu}(x, y) = F_\mu(x)F_\nu(y) \qquad \text{for all } x, y \in \mathbb{R}\,, \tag{6d8}$$

which leads to the third definition, equivalent to 6d1, 6d7 (due to 6b4).

6d9 Definition. Random variables $X, Y : \Omega \to \mathbb{R}$ are independent, if

$$F_{X,Y} = F_X(x)F_Y(y) \quad \text{for all } x, y \in R\,.$$

6d10 Exercise. Prove that X, Y defined by (6d2) are independent.

Hint. Show that $P_{X,Y}(A \times B) = P_X(A)P_Y(B)$ for $A = [\frac{k}{2^n}, \frac{k+1}{2^n})$, $B = [\frac{l}{2^n}, \frac{l+1}{2^n})$. Deduce that $F_{X,Y}(x, y) = F_X(x)F_Y(y) = xy$ for a dense set of points $(x, y) \in (0, 1) \times (0, 1)$.

6d11 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.

Proof. (a) The function $f(x, y) = f_X(x)f_Y(y)$ is a 2-dim density, since $\int f(x, y)\,dxdy = \bigl(\int f_X(x)\,dx\bigr)\bigl(\int f_Y(y)\,dy\bigr) = 1$ (recall 6d3). The corresponding 2-dim distribution μ_f satisfies

$$\begin{aligned}\mu_f(A \times B) = \iint_{A\times B} f(x, y)\,dxdy = \iint_{\mathbb{R}^2} \mathbf{1}_A(x)\mathbf{1}_B(y)f_X(x)f_Y(y)\,dxdy = \\ = \biggl(\int_{\mathbb{R}} \mathbf{1}_A(x)f_X(x)\,dx\biggr)\biggl(\int_{\mathbb{R}} \mathbf{1}_B(y)f_Y(y)\,dy\biggr) = P_X(A)P_Y(B)\,,\end{aligned}$$

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

(b) By 6c6, X and Y have densities f_X, f_Y, and[14]

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

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. □

6d12 Theorem. $\mathbb{E}\,(XY) = (\mathbb{E}\,X)(\mathbb{E}\,Y)$ for any independent integrable random variables X,Y.

Proof. Due to 6b7 we may restrict ourselves to the model of 6d5:

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

Apply Fubini theorem and 6d3:

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

□

6d13 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.

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

□

6d14 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.

6d15 Note. If they are not integrable, the formula still holds under appropriate conventions: $(+\infty)\cdot(+\infty) = +\infty$, $(-\infty)\cdot(+\infty) = -\infty$, (a positive number) $\cdot\,(+\infty) = +\infty$, (a positive number) $\cdot\,(\infty-\infty) = \infty-\infty$, etc.

[14] Equalities for densities hold almost everywhere, as usual.

6e Discrete, singular, and absolutely continuous distributions

6e1 Definition. A probability distribution μ (one-dimensional or two-dimensional) is called *absolutely continuous,* if $\mu(S) = 0$ for all Borel sets S of zero Lebesgue measure (mes_1 or mes_2, respectively).

In other words, absolute continuity means that

$$\begin{aligned} \mathrm{mes}(S) = 0 \quad &\Longrightarrow \quad \mu(S) = 0 \qquad \text{(1-dim case)}, \\ \mathrm{mes}_2(S) = 0 \quad &\Longrightarrow \quad \mu(S) = 0 \qquad \text{(2-dim case)}. \end{aligned}$$

6e2 Lemma. If a distribution (1-dim or 2-dim) has a densiity then it is absolutely continuous.

Proof. If $\mathrm{mes}(S) = 0$ then $\mu(S) = \int_S f(x)\,dx = \int \mathbf{1}_S(x) f(x)\,dx = 0$, since the integrand vanishes almost everywhere. □

It follows that discrete distributions cannot have densities. And not only discrete! A 2-dim distribution concentrated on a straight line (or a smooth curve, etc) cannot have a density.

The notion "continuous distribution" is simpler (than "absolutely continuous distribution"). Let us formulate the former in parallel to the latter.

6e3 Definition. A probability distribution μ (one-dimensional or two-dimensional) is called *continuous,* if $\mu(A) = 0$ for all at most countable[15] sets A ($A \subset \mathbb{R}$ or $A \subset \mathbb{R}^2$, respectively).

Due to countable additivity of μ we may restrict ourselves to singletons $A = \{x\}$. That is, μ is continuous if and only if μ is non-atomic. Recall that x is called an atom of μ if (and only if) $\mu(\{x\}) > 0$. (In dimension two, of course, an atom is a point of $\mathbb{R}^2$.) "Non-atomic" means: having no atoms.

A one-dimensional distribution μ is continuous ($=$ non-atomic) if and only if its cumulative distribution function F_μ is continuous. In dimension two, (F_μ is continuous) $\Longrightarrow$ (μ is continuous), but the converse is wrong (think, why). The following two definitions are applicable in both dimensions, 1 and 2.

6e4 Definition. A probability distribution μ is called *discrete,*[16] if $\mu(A) = 1$ for some at most countable set A.

6e5 Definition. A continuous probability distribution μ is called *singular,* if $\mu(S) = 1$ for some Borel set S of zero Lebesgue measure.

An example. Let X have a continuous distribution. Then the 2-dim distribution $P_{X,X}$ is singular.[17]

---

[15] That is, finite or countable.

[16] Or "purely atomic".

[17] Think, why. Further, what about, say, F_{X,X^2}?

Let μ be a probability distribution (1-dim or 2-dim), and A_μ be the set of all its atoms (maybe empty). The set A_μ is at most countable, since there are at most n atoms of probability $\ge 1/n$. There are three possible cases:

(6e6)
$$\begin{aligned} &\mu(A_\mu) = 0\,; && \mu \text{ is continuous } (=\text{non-atomic}); \qquad A_\mu = \emptyset\,;\\ &\mu(A_\mu) \in (0,1)\,; && \text{a mix of discrete and continuous;}\\ &\mu(A_\mu) = 1\,; && \mu \text{ is discrete.} \end{aligned}$$

Consider the case $\mu(A_\mu) \in (0,1)$. Introduce

(6e7)
$$\begin{gathered} p_{\mathrm d} = \mu(A_\mu)\,, \qquad p_{\mathrm c} = \mu(\mathbb{R}\setminus A_\mu)\,,\\ \mu_{\mathrm d}(B) = \frac{\mu(B\cap A_\mu)}{\mu(A_\mu)}\,, \qquad \mu_{\mathrm c}(B) = \frac{\mu(B\setminus A_\mu)}{\mu(\mathbb{R}\setminus A_\mu)} \quad \text{for Borel sets } B\,; \end{gathered}$$

then $\mu_{\mathrm d}, \mu_{\mathrm c}$ are probability distributions,[18] and

(6e8)
$$\begin{gathered} \mu = p_{\mathrm d}\mu_{\mathrm d} + p_{\mathrm c}\mu_{\mathrm c}\,, \qquad p_{\mathrm d} \ge 0\,, \quad p_{\mathrm c} \ge 0\,, \quad p_{\mathrm d} + p_{\mathrm c} = 1\,,\\ \mu_{\mathrm d} \text{ is discrete}, \qquad \mu_{\mathrm c} \text{ is continuous.} \end{gathered}$$

Pure cases (continuous and discrete) can be included into (6e8), since $p_{\mathrm d}$ and $p_{\mathrm c}$ are allowed to vanish.[19]

Given a random variable X, we apply (6e8) to its distribution:

(6e9)
$$\underbrace{\mathbb{P}\big(X\in B\big)}_{P_X(B)} = \underbrace{\mathbb{P}\big(\,X\in B \bigm| X\in A_X\,\big)}_{P_{X,\mathrm d}(B)} \cdot \underbrace{\mathbb{P}\big(X\in A_X\big)}_{p_{\mathrm d}} + \underbrace{\mathbb{P}\big(\,X\in B \bigm| X\notin A_X\,\big)}_{P_{X,\mathrm c}(B)} \cdot \underbrace{\mathbb{P}\big(X\notin A_X\big)}_{p_{\mathrm c}}\,;$$
$$P_X = p_{\mathrm d} P_{X,\mathrm d} + p_{\mathrm c} P_{X,\mathrm c}\,.$$

It follows from (6e8) that

(6e10)
$$\int \varphi\,d\mu = p_{\mathrm d}\int \varphi\,d\mu_{\mathrm d} + p_{\mathrm c}\int\varphi\,d\mu_{\mathrm c}$$

for any bounded Borel function $\varphi:\mathbb{R}\to\mathbb{R}$ (or $\mathbb{R}^2\to\mathbb{R}$, if μ is 2-dim).[20] Apply it to P_X:

(6e11)
$$\begin{gathered} \int\varphi\,dP_X = p_{\mathrm d}\int\varphi\,dP_{X,\mathrm d} + p_{\mathrm c}\int\varphi\,dP_{X,\mathrm c}\,, \quad \text{that is,}\\ \mathbb{E}\,\varphi(X) = \mathbb{E}\,\big(\,\varphi(X)\bigm| X\in A_X\,\big)\;\mathbb{P}\big(X\in A_X\big) + \mathbb{E}\,\big(\,\varphi(X)\bigm| X\notin A_X\,\big)\;\mathbb{P}\big(X\notin A_X\big)\,. \end{gathered}$$

(Conditional distributions and conditional expectations will be studied systematically in Sect. 7.) For unbounded φ, as usual, the formula holds under appropriate conventions about

[18] Check it!

[19] Though, if $p_{\mathrm d}=0$ then $\mu_{\mathrm d}$ is arbitrary. Similarly, if $p_{\mathrm c}=0$ then $\mu_{\mathrm c}$ is arbitrary.

[20] This is another linearity property of Lebesgue integral; not only $\int(a\varphi+b\psi)\,d\mu = a\int\varphi\,d\mu + b\int\psi\,d\mu$, but also $\int\varphi\,d(a\mu+b\nu) = a\int\varphi\,d\mu + b\int\varphi\,d\nu$. The proof is simple: integrate in y the trivial equality $(a\mu+b\nu)\big(\{x:\varphi(x)>y\}\big) = a\mu\big(\{x:\varphi(x)>y\}\big) + b\nu\big(\{x:\varphi(x)>y\}\big)$.

$\pm\infty$. For the discrete part, probabilities and expectations are sums:

(6e12)
$$
\begin{gathered}
p_{\mathrm{d}}\mu_{\mathrm{d}}(B) = \sum_{x\in B\cap A_\mu} \mu\big(\{x\}\big)\,,\\
\mathbb{P}\,\big(\, X\in B \,\big|\, X\in A_X \,\big)\; \mathbb{P}\,\big(\, X\in A_X \,\big) = \sum_{x\in B\cap A_\mu} \mathbb{P}\,\big(\, X=x \,\big)\,,\\
p_{\mathrm{d}}\int \varphi\, d\mu_{\mathrm{d}} = \sum_{x\in A_\mu} \varphi(x)\mu(\{x\})\,,\\
\mathbb{E}\,\big(\, \varphi(X) \,\big|\, X\in A_X \,\big)\; \mathbb{P}\,\big(\, X\in A_X \,\big) = \sum_{x\in A_\mu} \varphi(x)\mathbb{P}\,\big(\, X=x \,\big)\,;
\end{gathered}
$$

the sum contains finitely many terms if A_X is finite; otherwise it is a series. The order of terms does not matter, if $\varphi(X)$ is integrable, since in that case $\sum_{x\in A_\mu} |\varphi(x)|\mathbb{P}\,\big(\, X=x \,\big) < \infty$. All that holds in both dimensions, 1 and 2.

Assume now that μ is a continuous distribution. We have no such notion as "a singular point", thus we cannot say simply, "let S_μ be the set of all singular points of μ".[21] Instead, we may maximize $\mu(S)$ over all sets S of zero Lebesgue measure; the maximum is reached on some set;[22] denote this set by S_μ (though, it is not uniquely determined by μ).

Similarly to (6e6), there are three possible cases:

(6e13)
$$
\begin{aligned}
&\mu(S_\mu) = 0\,; &&\mu \text{ is absolutely continuous}\,;\\
&\mu(S_\mu) \in (0,1)\,; &&\text{a mix of singular and absolutely continuous}\,;\\
&\mu(S_\mu) = 1\,; &&\mu \text{ is singular}\,.
\end{aligned}
$$

Consider the case $\mu(S_\mu)\in(0,1)$. Similarly to (6e7), introduce

(6e14)
$$
\begin{gathered}
p_{\mathrm{s}} = \mu(S_\mu)\,, \qquad p_{\mathrm{ac}} = \mu(\mathbb{R}\setminus S_\mu)\,,\\
\mu_{\mathrm{s}}(B) = \frac{\mu(B\cap S_\mu)}{\mu(S_\mu)}\,, \qquad \mu_{\mathrm{ac}}(B) = \frac{\mu(B\setminus S_\mu)}{\mu(\mathbb{R}\setminus S_\mu)} \quad \text{for Borel sets } B\,;
\end{gathered}
$$

then $\mu_{\mathrm{s}}, \mu_{\mathrm{ac}}$ are probability distributions, and (similarly to (6e8))

(6e15)
$$
\begin{gathered}
\mu = p_{\mathrm{s}}\mu_{\mathrm{s}} + p_{\mathrm{ac}}\mu_{\mathrm{ac}}\,, \qquad p_{\mathrm{s}}\ge 0\,, \quad p_{\mathrm{ac}}\ge 0\,, \quad p_{\mathrm{s}}+p_{\mathrm{ac}} = 1\,,\\
\mu_{\mathrm{s}} \text{ is singular,} \qquad \mu_{\mathrm{ac}} \text{ is absolutely continuous.}
\end{gathered}
$$

Combining (6e8) and (6e15) (the latter applied to the continuous part), we get finally, for an arbitrary probability distribution μ,

(6e16)
$$
\begin{gathered}
\mu = p_{\mathrm{d}}\mu_{\mathrm{d}} + p_{\mathrm{s}}\mu_{\mathrm{s}} + p_{\mathrm{ac}}\mu_{\mathrm{ac}}\,, \qquad p_{\mathrm{d}}\ge 0\,, \quad p_{\mathrm{s}}\ge 0\,, \quad p_{\mathrm{ac}}\ge 0\,, \quad p_{\mathrm{d}}+p_{\mathrm{s}}+p_{\mathrm{ac}} = 1\,,\\
\mu_{\mathrm{d}} \text{ is discrete,} \qquad \mu_{\mathrm{s}} \text{ is singular,} \qquad \mu_{\mathrm{ac}} \text{ is absolutely continuous.}
\end{gathered}
$$

---

[21] Though, it can be done; roughly speaking, x is a singular point of μ, if $F'_\mu(x) = +\infty$ (in dimension 1).

[22] Indeed, let $p_{\mathrm{s}} = \sup\{\mu(S) : \operatorname{mes}(S) = 0\}$; take S_n such that $\mu(S_n)\to p_{\mathrm{s}}$ and $\operatorname{mes}(S_n) = 0$; take $S = S_1\cup S_2\cup\dots$, then $\mu(S) = p_{\mathrm{s}}$ (think, why) and $\operatorname{mes}(S) = 0$.

Given an arbitrary random variable X, we apply (6e16) to its distribution:

$$\begin{aligned}\underbrace{\mathbb{P}\big(X \in B \big)}_{P_X(B)} &= \underbrace{\mathbb{P}\Big(X \in B \,\Big|\, X \in A_X \Big)}_{P_{X,\mathrm{d}}(B)} \cdot \underbrace{\mathbb{P}\big(X \in A_X \big)}_{p_\mathrm{d}} + \\ &\quad + \underbrace{\mathbb{P}\Big(X \in B \,\Big|\, X \in S_X \setminus A_X \Big)}_{P_{X,\mathrm{s}}(B)} \cdot \underbrace{\mathbb{P}\big(X \in S_X \setminus A_X \big)}_{p_\mathrm{s}} + \\ &\quad + \underbrace{\mathbb{P}\Big(X \in B \,\Big|\, X \notin A_X \cup S_X \Big)}_{P_{X,\mathrm{ac}}(B)} \cdot \underbrace{\mathbb{P}\big(X \notin A_X \cup S_X \big)}_{p_\mathrm{ac}}; \\ P_X &= p_\mathrm{d} P_{X,\mathrm{d}} + p_\mathrm{s} P_{X,\mathrm{s}} + p_\mathrm{ac} P_{X,\mathrm{ac}} \,. \end{aligned} \tag{6e17}$$

Accordingly,

$$\begin{aligned}\mathbb{E}\big(\varphi(X)\big) &= \mathbb{E}\Big(\varphi(X) \,\Big|\, X \in A_X \Big)\ \mathbb{P}\big(X \in A_X \big) + \\ &\quad + \mathbb{E}\Big(\varphi(X) \,\Big|\, X \in S_X \setminus A_X \Big)\ \mathbb{P}\big(X \in S_X \setminus A_X \big) + \\ &\quad + \mathbb{E}\Big(\varphi(X) \,\Big|\, X \notin A_X \cup S_X \Big)\ \mathbb{P}\big(X \notin A_X \cup S_X \big) \,. \end{aligned} \tag{6e18}$$

In two dimensions it means that

$$\begin{aligned}\mathbb{E}\big(\varphi(X,Y)\big) &= \mathbb{E}\Big(\varphi(X,Y) \,\Big|\, (X,Y) \in A_{X,Y} \Big)\ \mathbb{P}\big((X,Y) \in A_{X,Y} \big) + \\ &\quad + \mathbb{E}\Big(\varphi(X,Y) \,\Big|\, (X,Y) \in S_{X,Y} \setminus A_{X,Y} \Big)\ \mathbb{P}\big((X,Y) \in S_{X,Y} \setminus A_{X,Y} \big) + \\ &\quad + \mathbb{E}\Big(\varphi(X,Y) \,\Big|\, (X,Y) \notin A_{X,Y} \cup S_{X,Y} \Big)\ \mathbb{P}\big((X,Y) \notin A_{X,Y} \cup S_{X,Y} \big) \,. \end{aligned} \tag{6e19}$$

For the discrete part, probabilities and expectations are sums (recall (6e12) and think how does it look in two dimensions). For the absolutely continuous part, probabilities and expectations are integrals due to (6a14), (6b14), 6a15, 6b15 and the following theorem of Measure theory (it holds in both dimensions, 1 and 2).

6e20 Theorem. (Radon-Nikodym). Every absolutely continuous distribution has a density.

I give no proof. Note that 6e20 is the converse to 6e2.

For the singular part, there is no universal recipe. Typically, a singular distribution appears when a 2-dim distribution is concentrated on a curve. Often it is of the form $P_{X,\psi(X)}$, that is, $P_{X,Y}$ with $Y = \psi(X)$, where X is absolutely continuous. In that case

$$\mathbb{E}\,\varphi(X,Y) = \mathbb{E}\,\varphi\big(X,\psi(X)\big) = \int \varphi\big(x,\psi(x)\big) f_X(x)\,dx \,. \tag{6e21}$$

A more general case will be studied in Sect. 7. In full generality, however, singular distributions are too complicated for our course.

Here is an example of a one-dimensional singular distribution. Define $X : (0,1) \to \mathbb{R}$ via binary digits as follows:

$$X\bigg(\sum_{k=1}^\infty \frac{\alpha_k}{2^k} \bigg) = \sum_{k=1}^\infty \frac{3\alpha_k}{4^k} \,. \tag{6e22}$$

In other words,

$$X\Big(\sum_{k=1}^\infty \frac{\alpha_k}{2^k}\Big) = \sum_{k=1}^\infty \frac{\beta_k}{2^k}\,, \qquad \text{where } \beta_{2k-1} = \beta_{2k} = \alpha_k\,. \tag{6e23}$$

Then X is a strictly increasing function (think, why) and the set $S = X\big((0,1)\big)$ is of zero Lebesgue measure.[23] Therefore X is singular.

6e24 Exercise. Prove the following multiplication table:

	d	s	ac
d	d	s	s
s	s	s	s
ac	s	s	ac

That is, if μ is discrete and ν is absolutely continuous then $\mu \otimes \nu$ is singular, etc.

Hint. Show that (1) if (x, y) is an atom for $\mu \otimes \nu$, then x is an atom for μ and y is an atom for ν; (2) if $\operatorname{mes}(A) = 0$ then $\operatorname{mes}_2(A \times B) = 0$; (3) if μ and ν are absolutely continuous then $\mu \otimes \nu$ is absolutely continuous (combine 6e20 and 6d11a).

6f Regression and correlation

This topic was studied in "Introduction to probability" for discrete distributions; definitions, statements and proofs remain the same for arbitrary distributions. I only recall it in short.

A random variable X is called *square integrable,*[24] if $\mathbb{E}(X^2) < \infty$. It follows that $\mathbb{E}|X| < \infty$, that is, X is integrable.

A square integrable random variable X has a variance

$$\operatorname{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = \mathbb{E}\big(X - \mathbb{E}X\big)^2 = \min_{a\in\mathbb{R}} \mathbb{E}\big((X-a)^2\big)$$

and a mean square deviation[25]

$$\sigma_X = \sqrt{\operatorname{Var}(X)}\,.$$

If X is a constant (that is, $X = c$ with probability 1 for some $c \in \mathbb{R}$), then $\operatorname{Var}(X) = 0$, otherwise $\operatorname{Var}(X) > 0$. Concentration of X around $\mathbb{E}X$ can be estimated by Chebyshev inequality

$$\mathbb{P}\big(\,|X - \mathbb{E}X| \ge a\sigma_X\,\big) \le \frac{1}{a^2}\,; \qquad \mathbb{P}\big(\,|X - \mathbb{E}X| \ge \varepsilon\,\big) \le \frac{\sigma_X^2}{\varepsilon^2}\,.$$

Two square integrable random variables $X, Y : \Omega \to \mathbb{R}$ are called *uncorrelated,* if[26] $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$. If X, Y are independent then they are uncorrelated; the converse is wrong.

---

[23] Indeed, $S = S_1 \cap S_2 \cap \dots$, where $S_n = \{\sum_{k=1}^\infty \frac{\beta_k}{2^k} : \beta_1 = \beta_2, \dots, \beta_{2n-1} = \beta_{2n}\}$. Clearly, $\operatorname{mes} S_n = 2^{-n} \xrightarrow[n\to\infty]{} 0$.

[24] All square integrable random variables on a probability space $(\Omega, \mathcal{F}, P)$ form a Hilbert space $L_2(\Omega, \mathcal{F}, P)$.

[25] In $L_2(\Omega, \mathcal{F}, P)$, σ_X is the distance between X and the one-dimensional space of constants.

[26] If X, Y are square integrable then XY is integrable; indeed, $|XY| \le \frac12(X^2 + Y^2)$.

6f1 Exercise. Let random variables $X, Y : \Omega \to \mathbb{R}$ be square integrable. Prove that X, Y are independent if and only if $\varphi(X), \psi(Y)$ are uncorrelated for all bounded Borel functions $\varphi, \psi : \mathbb{R} \to \mathbb{R}$.

Hint. "$\Longrightarrow$": recall 6d14; "$\Longleftarrow$": take $\varphi = \mathbf{1}_A$, $\psi = \mathbf{1}_B$.

Two square integrable random variables $X, Y : \Omega \to \mathbb{R}$ have a covariance[27]

$$\operatorname{Cov}(X, Y) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y) = \mathbb{E}\big((X - \mathbb{E}X)(Y - \mathbb{E}Y)\big) ;$$

if neither X nor Y is constant, they have a correlation coefficient[28]

$$\rho_{X,Y} = \frac{\operatorname{Cov}(X, Y)}{\sigma_X \sigma_Y} .$$

If Y is a linear function of X (that is, $Y = aX + b$ with probability 1 for some $a, b \in \mathbb{R}$) then $\rho(X, Y) = \operatorname{sgn} a = \pm 1$; otherwise $\rho(X, Y) \in (-1, +1)$.

Linear regression of Y on X, or (optimal) linear predictor of Y by X, is the linear function $aX + b$ of X closest to Y in the sense that[29]

$$\mathbb{E}\big(Y - (aX + b)\big)^2 \quad \text{is minimal over all } a, b \in \mathbb{R};$$

the optimal $aX + b = \hat{Y}$ is given by the formula

$$\frac{\hat{Y} - \mathbb{E}Y}{\sigma_Y} = \rho_{X,Y} \frac{X - \mathbb{E}X}{\sigma_X} ;$$

thus, $\hat{Y} = \rho_{X,Y} \dfrac{\sigma_Y}{\sigma_X}(X - \mathbb{E}X) + \mathbb{E}Y$. The error of the optimal approximation is given by the formula

$$\mathbb{E}\left(\frac{Y - \mathbb{E}Y}{\sigma_Y} - \rho_{X,Y} \frac{X - \mathbb{E}X}{\sigma_X} \right)^2 = 1 - \rho_{X,Y}^2 ;$$

thus, $\mathbb{E}\big((Y - \hat{Y})^2\big) = (1 - \rho_{X,Y}^2)\sigma_Y^2$. The error vanishes if and only if $\rho_{X,Y} = \pm 1$. Concentration of Y around $\hat{Y} = aX + b$ can be estimated by Chebyshev inequality:

$$\mathbb{P}\big(|Y - \hat{Y}| \ge \varepsilon \big) \le \frac{(1 - \rho_{X,Y}^2)\sigma_Y^2}{\varepsilon^2} .$$

Note that $\rho_{X,Y} = \rho_{Y,X}$, however, the two linear regresssion functions (Y on X, and X on Y) are *not* mutually inverse functions (unless $\rho_{X,Y} = \pm 1$).

---

[27] Scalar product in the quotient space, $L_2(\Omega, \mathcal{F}, P)$ modulo the one-dimensional subspace of constants.

[28] The cosine of the angle in the quotient space.

[29] The orthogonal projection of Y onto the 2-dim subspace spanned by X and constants.

6g Transformations

Recall the simplest case, a linear 1-dim transformation $Y = aX + b$, studied in Sect. 3a. In the smooth case we have (see (3a7))

(6g1) $$|a| f_Y(y) = f_X(x) \qquad (y = ax + b,\ a \ne 0)\,.$$

The same holds in full generality. Namely, Y is absolutely continuous if and only if X is absolutely continuous, in which case (6g1) holds almost everywhere.

The coefficient $|a|$ appears because Lebesgue measure is not preserved by the transformation:

(6g2) $$\operatorname{mes}\bigl(T(B)\bigr) = |a| \operatorname{mes}(B)$$

for all Borel sets $B \subset \mathbb{R}$; here $T(x) = ax + b$ and, of course, $T(B) = \{T(x) : x \in B\}$. Formula (6g2) evidently holds for intervals, therefore, for elementary sets. Its generalization for Borel sets is easily achieved via Monotone class theorem 5d4 (try it).

Turn to a linear 2-dim transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$,

(6g3) $$T(x,y) = \bigl(ax + by,\ cx + dy\bigr)\,.$$

If $B = (0,1) \times (0,1)$ is the unit square, then $T(B)$ is a parallelepiped of the area $|ad - bc|$. In general,

(6g4) $$\operatorname{mes}_2\bigl(T(B)\bigr) = |J| \operatorname{mes}_2(B)$$

for all Borel sets $B \subset \mathbb{R}^2$; here T is given by (6g3), and

(6g5) $$J = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

is the so-called Jacobian of T; we assume that $J \ne 0$ (which is equivalent to existence of the inverse transformation T^{-1}). Formula (6g4) is a simple geometric fact for 2-dim intervals B; therefore it holds for elementary sets, and (again by Monotone class theorem) for all Borel sets. It can be deduced that[30]

(6g6) $$|J| \cdot f_{U,V}(u,v) = f_{X,Y}(x,y) \qquad \text{when } (u,v) = T(x,y)$$

almost everywhere. Here (X,Y) is a 2-dim random variable, (U,V) is another 2-dim random variable, and $(U,V) = T(X,Y)$ is assumed (with probability 1). Absolute continuity of (U,V) is equivalent to absolute continuity of (X,Y).

For nonlinear smooth 1-dim transformations we have (recall (3b1))

(6g7) $$\left|\frac{dy}{dx}\right| f_Y(y) = f_X(x) \qquad \text{when } y = \varphi(x)$$

[30] A sandwich argument, similar to the proof of 6a15, is used.

provided that the transformation is one-one (continuity of f_X, f_Y, stipulated in Section 3, may be discarded now). Comparing (6g1), (6g6) and (6g7) it is easy to guess that

(6g8) $$\left|\frac{\partial(u,v)}{\partial(x,y)}\right| f_{U,V}(u,v) = f_{X,Y}(x,y) \qquad \text{when } (u,v) = T(x,y)$$

for any (nonlinear) 2-dim one-one transformation T; here

(6g9) $$\frac{\partial(u,v)}{\partial(x,y)} = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix}$$

is a convenient notation for the Jacobian of T. We'll deduce (6g8) from some results of Analysis, formulated below (6g11, 6g12).

Let $D \subset \mathbb{R}^2$ be an open set, and $T : D \to \mathbb{R}^2$ a map, $T(x,y) = \bigl(u(x,y), v(x,y)\bigr)$. Assume that functions u, v have continuous first-order partial derivatives on D. Introduce the Jacobian

(6g10) $$J_T(x,y) = \frac{\partial(u,v)}{\partial(x,y)} = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} \, .$$

Call T *smooth,* if it is one-one (from D onto $T(D)$) and $J_T(x,y) \ne 0$ for all $(x,y) \in D$.

6g11 Lemma. If T is smooth then $T(D)$ is an open subset of $\mathbb{R}^2$, and the inverse map $T^{-1} : T(D) \to D$ is smooth, and

$$\frac{\partial(u,v)}{\partial(x,y)} \cdot \frac{\partial(x,y)}{\partial(u,v)} = 1 \, .$$

I give no proof.

6g12 Theorem. (Change of variables). Let $T : D \to \mathbb{R}^2$ be a smooth transformation, $D_1 = T(D)$, and $f : D_1 \to \mathbb{R}$ a Borel function. Then

$$\iint\limits_D f\bigl(T(x,y)\bigr) \, |J_T(x,y)| \, dxdy = \iint\limits_{D_1} f(u,v) \, dudv$$

provided that f is integrable on D_1 and the function $(x,y) \mapsto f\bigl(T(x,y)\bigr) \, |J_T(x,y)|$ is integrable on D. Otherwise, both functions are non-integrable.

I give no proof.

6g13 Exercise. Prove that $\mathrm{mes}_2(A) = 0 \iff \mathrm{mes}_2\bigl(T(A)\bigr) = 0$ for Borel sets $A \subset D$.

Hint. Apply 6g12 to $f = \mathbf{1}_{T(A)}$.

6g14 Theorem. Let $T : D \to \mathbb{R}^2$ be a smooth transformation, $D_1 = T(D)$. Let $X, Y, U, V : \Omega \to \mathbb{R}$ be random variables such that $(X,Y) \in D$ and $T(X,Y) = (U,V)$ almost sure.[31]

[31] Therefore $(U,V) \in D_1$ almost sure.

Then (X,Y) is absolutely continuous if and only if (U,V) is absolutely continuous, and if they are, then[32]

$$\left|\frac{\partial(u,v)}{\partial(x,y)}\right| f_{U,V}(u,v) = f_{X,Y}(x,y) \qquad \text{when } (u,v) = T(x,y)\,.$$

Proof. The first claim (about absolute continuity) follows from 6g13. Define $f_{U,V}$ on D_1 by $f_{U,V}(u,v) = J_{T^{-1}}(u,v) f\big(T^{-1}(u,v)\big)$, then $f_{U,V}$ satisfies the equality $\left|\frac{\partial(u,v)}{\partial(x,y)}\right| f_{U,V}(u,v) = f_{X,Y}(x,y)$ when $(u,v) = T(x,y)$. We have to prove that $f_{U,V}$ is a density for (U,V), which means

$$\iint_A f_{U,V}(u,v)\,dudv = P_{U,V}(A)$$

for every Borel set $A \subset D_1$. Apply 6g12 to $\mathbf{1}_A f_{U,V}$:

$$\iint_D \mathbf{1}_A\big(T(x,y)\big) \underbrace{f_{U,V}\big(T(x,y)\big)|J_T(x,y)|}_{f_{X,Y}(x,y)}\,dxdy = \iint_{D_1} \mathbf{1}_A(u,v) f_{U,V}(u,v)\,dudv\,;$$

$$\iint_{T^{-1}(A)} f_{X,Y}(x,y)\,dxdy = \iint_A f_{U,V}(u,v)\,dudv\,;$$

the left-hand side is equal to $\mathbb{P}\,\big(\,(X,Y) \in T^{-1}(A)\,\big) = \mathbb{P}\,\big(\,(U,V) \in A\,\big) = P_{U,V}(A)$. $\square$

Some examples:

$$\begin{aligned} U &= X+Y \\ V &= X \end{aligned} \quad\Longrightarrow\quad f_{U,V}(u,v) = f_{X,Y}(x,y)\,;$$

$$\begin{aligned} X &= R\cos\Phi \\ Y &= R\sin\Phi \end{aligned} \quad\Longrightarrow\quad r f_{X,Y}(x,y) = f_{R,\Phi}(r,\varphi)\,.$$

Implication:

$$\begin{aligned} f_{X+Y}(u) &= \int_{-\infty}^{+\infty} f_{X,Y}(x,u-x)\,dx\,; \\ f_{\sqrt{X^2+Y^2}}(r) &= r\int_0^{2\pi} f_{X,Y}(r\cos\varphi, r\sin\varphi)\,d\varphi\,; \\ f_\Phi(\varphi) &= \int_0^\infty r f_{X,Y}(r\cos\varphi, r\sin\varphi)\,dr\,. \end{aligned}$$

[32]Of course, $f_{X,Y}$ vanishes outside D, and $f_{U,V}$ vanishes outside D_1.