2 Random variables

2a The definition

Discrete probability defines a random variable X as a function $X : \Omega \to \mathbb{R}$. The probability of a possible value $x \in \mathbb{R}$ is $\mathbb{P}\big(X = x \big) = P\big(\{\omega \in \Omega : X(\omega) = x\}\big)$. The probability of an interval, say, $\mathbb{P}\big(a < X < b \big)$ is the sum of probabilities of all possible values $x \in (a, b)$:

(2a1)
$$\mathbb{P}\big(a < X < b \big) = \sum_{x \in (a,b)} \mathbb{P}\big(X = x \big) \, .$$

Equivalently,

(2a2)
$$\mathbb{P}\big(a < X < b \big) = P\big(\{\omega \in \Omega : a < X(\omega) < b\}\big) \, .$$

Continuous probability cannot use (2a1), since $\mathbb{P}\big(X = x \big)$ it typically 0. Only (2a2) is used. The set $\{\omega \in \Omega : a < X(\omega) < b\}$ must be an event (otherwise its probability is not defined). The following definition uses intervals of the form $(-\infty, x]$ rather than (a, b), but it is the same, as we'll see. As usual, $(\Omega, \mathcal{F}, P)$ is a given probability space.

2a3 Definition. A *random variable* is a function $X : \Omega \to \mathbb{R}$ such that

$$\forall x \in \mathbb{R} \quad \{\omega \in \Omega : X(\omega) \le x\} \in \mathcal{F} \, .$$

2b Distribution function

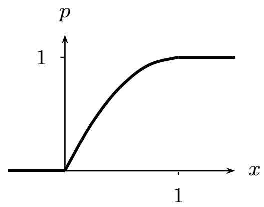
2b1 Example. Let $(\Omega, \mathcal{F}, P)$ be the square $(0, 1) \times (0, 1)$ with the Lebesgue measure (recall 1f13), and

$$X(s, t) = |s - t| \quad \text{for } s, t \in (0, 1) \, .$$

(Recall Sect. 1a; interpret s as the time of one friend, t — the other; X shows how long one of them has to wait for the other. Note that $\omega = (s, t)$.)

For $x = 1/3$ the set $\{\omega \in \Omega : X(\omega) \le x\}$ was considered in Sect. 1a (its probability is 5/9). For another $x \in (0, 1)$ it is similar;[12] the probability is[13] $2x(1 - x) + x^2 = 2x - x^2 = x(2 - x)$. For $x = 0$ the set is the diagonal ($s = t$) of probability 0; for $x < 0$ the set is empty (therefore, of probability 0). For $x \ge 1$ the set is the whole Ω (therefore, of probability 1).

(2b2)
$$\mathbb{P}\big(X \le x \big) = \begin{cases} 0 & \text{for } x \in (-\infty, 0], \\ x(2 - x) & \text{for } x \in [0, 1], \\ 1 & \text{for } x \in [1, \infty). \end{cases}$$



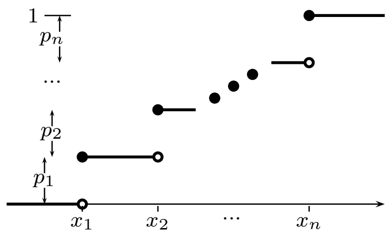
[12] The set $\{\omega \in \Omega : X(\omega) \le x\} = \{(s, t) \in (0, 1) \times (0, 1) : |s - t| \le x\}$ is a Borel set, since it is the intersection of the *open* set $(0, 1) \times (0, 1)$ and the *closed* set $\{(s, t) \in \mathbb{R}^2 : |s - t| \le x\}$. The set is also Jordan measurable (just a polygon), therefore its Lebesgue measure is equal to its area.

[13] Or, even simpler, $1 - 2 \cdot \frac{1}{2}(1 - x)^2 = x(2 - x)$.

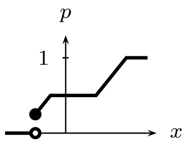
2b3 Definition. A (cumulative) *distribution function* of a random variable X is the function $F_X : \mathbb{R} \to [0,1]$ defined by

$$F_X(x) = \mathbb{P}\big(X \le x \big) \quad \text{for } x \in \mathbb{R}\,.$$

So, (2b2) is an example of a distribution function. Note that it is continuous. In contrast, a discrete distribution has a discontinuous distribution function:



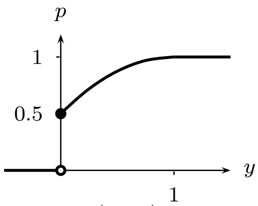
Any combination of discrete and continuous is also possible:



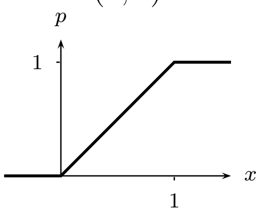
2b4 Example. Let $(\Omega, \mathcal{F}, P)$ be as in 2b1, and

$$Y(s,t) = (t-s)^+ = \begin{cases} t-s & \text{when } s \le t, \\ 0 & \text{when } s \ge t. \end{cases}$$

(Friend A waits for friend B during Y.) Here $\mathbb{P}\big(Y = 0\big) = 1/2$, but the rest of the distribution is continuous:



2b5 Example. The uniform distribution $\mathrm{U}(0,1)$ has a very simple distribution function



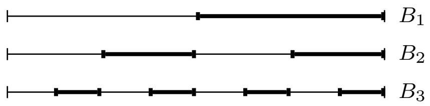
Turn to decimal digits,

$$X = \big(0.\alpha_1\alpha_2\dots\big)_{10} = \sum_{k=1}^{\infty} \frac{\alpha_k}{10^k}\,, \quad \alpha_k \in \{0,1,\dots,9\}\,;$$

$X \sim \mathrm{U}(0,1)$ means that $\alpha_1, \alpha_2, \dots$ are independent discrete random variables, each one distributed uniformly on $\{0,1,\dots,9\}$ (recall 1f4). Binary digits β_k,

$$X = \big(0.\beta_1\beta_2\dots\big)_2 = \sum_{k=1}^{\infty} \frac{\beta_k}{2^k}\,, \quad \beta_k \in \{0,1\}\,,$$

are independent random variables, namely, indicators of corresponding independent events $B_k = \{\beta_k = 1\}$, $\mathbb{P}\left(B_k \right) = 1/2$:



(Just infinite coin tossing.)

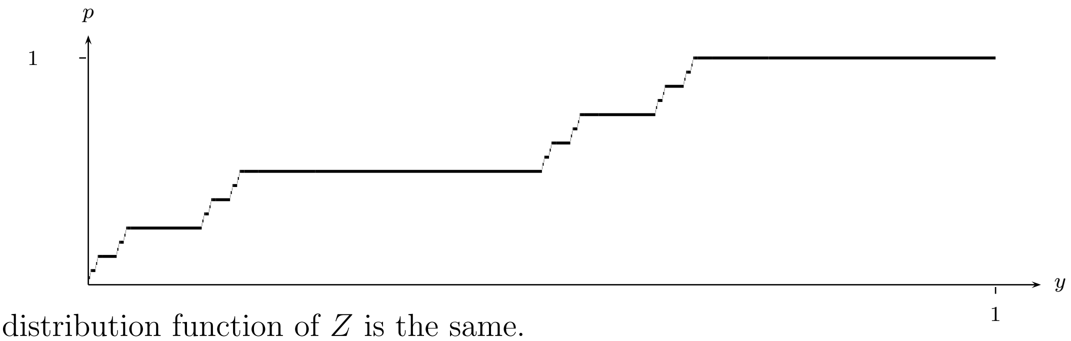
2b6 Example. Still $X = \left(0.\beta_1\beta_2\dots\right)_2$. We have

$$X = Y + \frac{1}{2}Z\,, \qquad Y \text{ and } Z \text{ are independent,}$$

$$Y = \left(0.\beta_1 0 \beta_3 0 \beta_5 0 \dots\right)_2 = \sum_{k=1}^{\infty} \frac{\beta_{2k-1}}{2^{2k-1}}\,,$$

$$\frac{1}{2}Z = \left(0.0\beta_2 0 \beta_4 0 \beta_6 \dots\right)_2 = \sum_{k=1}^{\infty} \frac{\beta_{2k}}{2^{2k}}\,.$$

The distribution function of Y is continuous but bizarre:



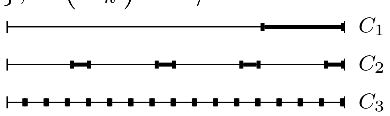
The distribution function of Z is the same.

2b7 Example. Still $X = \left(0.\beta_1\beta_2\dots\right)_2$. Introduce

$$\gamma_1 = \beta_1\beta_2,\ \gamma_2 = \beta_3\beta_4,\ \dots\ ;$$

$$Y = \left(0.\gamma_1\gamma_2\dots\right)_2 = \sum_{k=1}^{\infty} \frac{\beta_{2k-1}\beta_{2k}}{2^k}\,.$$

Binary digits γ_k of Y are independent random variables, namely, indicators of corresponding independent events $C_k = \{\gamma_k = 1\}$, $\mathbb{P}\left(C_k \right) = 1/4$:



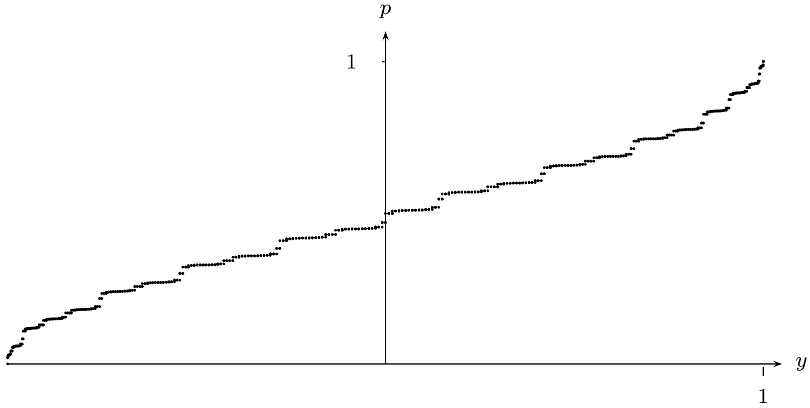
The distribution function of Y is continuous but bizarre:

You see, random digits often lead to bizarre distributions. However, discrete probability also can lead to bizarre distributions.

2b8 Example. Let X be a discrete random variable distributed geometrically:

$$\begin{array}{ccccc} x & 0 & 1 & 2 & \dots \\ \mathbb{P}\big(X=x\big) & p & pq & pq^2 & \dots \end{array}$$

where $p \in (0,1)$ is a parameter. Let $Y = \sin X$ (in radians, of course). The distribution function of Y is bizarre, and discontinuous on *every* interval $(a,b) \subset (0,1)$:

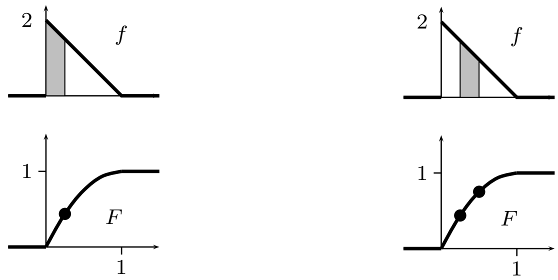


(The case $p = 0.03$ is shown.)

2c Density

Usually we deal with smooth distribution functions, like (2b2). Such a function F has a piecewise continuous derivative f, $f(x) = F'(x)$, and is the integral of f:

$$F(x) = \int_{-\infty}^{x} f(x_1)\,dx_1 \qquad F(b) - F(a) = \int_a^b f(x)\,dx$$



At some points f may be discontinuous. No need to define a value of f at such points, since these values do not influence the integral.

2c1. A function $f : \mathbb{R} \to \mathbb{R}$ is called a *density* of a random variable X, if

$$\mathbb{P}\big(a < X < b\big) = \int_a^b f(x)\,dx \tag{2c2}$$

whenever $-\infty < a < b < \infty$.

Unfortunately, 2c1 is not a definition, since the integration is not specified. Usually, f is Riemann integrable on every bounded interval (a,b), and we may use Riemann integration in 2c2. However, the integral $\int_{-\infty}^{x} f(x_1)\,dx_1$ is improper (rather than Riemann):

$$\int_{-\infty}^{x} f(x_1)\,dx_1 = \lim_{a\to-\infty} \int_a^x f(x_1)\,dx_1 \,.$$

Similarly,

$$\int_{-\infty}^{+\infty} f(x)\,dx = \lim_{\substack{a\to-\infty\\ b\to+\infty}} \int_a^b f(x)\,dx \,.$$

As you probably guess,

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

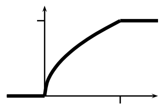
whenever f is a density of a random variable.[14]

Sometimes f is not Riemann integrable even on bounded intervals.

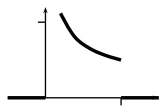
2c3 Example. Let $X \sim \mathrm{U}(0,1)$ (that is, X is a random variable distributed uniformly on $(0,1)$), and $Y = X^2$. (Think, say, about the area of a random square.) Then

$$F_Y(y) = \mathbb{P}\big(Y \le y \big) = \mathbb{P}\big(X^2 \le y \big) = \mathbb{P}\big(X \le \sqrt{y} \big) = \sqrt{y} \quad \text{for } y \in [0,1]\,;$$

$$F_Y(y) = \begin{cases} 0 & \text{for } y \in (-\infty,0],\\ \sqrt{y} & \text{for } y \in [0,1],\\ 1 & \text{for } y \in [1,+\infty); \end{cases}$$



$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} & \text{for } y \in (0,1),\\ 0 & \text{otherwise.} \end{cases}$$



Now f_Y is not Riemann integrable on $(0,1)$, since it is not bounded. Improper integral is used here:

$$\int_0^b f_Y(y)\,dy = \lim_{a\to+0} \int_a^b f_Y(y)\,dy \,.$$

More generally, if f has singularities, say, at 1/3 and 2/3, we write

$$\int_0^1 f(x)\,dx = \lim_{\varepsilon\to0+} \bigg(\int_0^{1/3-\varepsilon} f(x)\,dx + \int_{1/3+\varepsilon}^{2/3-\varepsilon} f(x)\,dx + \int_{2/3+\varepsilon}^{1} f(x)\,dx \bigg)$$

etc.

In principle, a density f may be such a bizarre function that Riemann integration is utterly inapplicable. Then so-called Lebesgue integration must be used in 2c2.[15]

---

[14] We'll return to the point later.

[15] We'll return to Lebesgue integral later. If you are curious to see a bizarre density, here is the simplest example known to me:

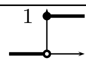
$$f(x) = 1 + \frac{2}{\pi} \arctan \bigg(\sum_{k=1}^{\infty} \frac{1}{k} \sin(\cdot 2^k x\pi) \bigg) \qquad \text{for } x \in (0,1)\,.$$

Sorry, I am unable to draw its graph; it is dense in the rectangle $[0,1] \times [0,2]$.

If f is Riemann integrable on (a,b), then the two-dimensional region $\{(x,y) \in \mathbb{R}^2 : x \in (a,b), y \in (0, f(x))\}$ is Jordan measurable, and its area is equal to $\int_a^b f(x)\,dx$. In general, Lebesgue measure of the area is equal to Lebesgue integral of the function.

Anyway, A DENSITY IS DEFINED IN TERMS OF INTEGRATION rather than differentiation. Consequently, a density may be changed at will (or left undefined) at any point, or finite set of points.[16]

IS THERE A DENSITY OF A DISCRETE DISTRIBUTION? A discussion follows.

Y: Consider for instance the function $F =$  and differentiate it. Clearly, $F'(x) = 0$ for $x \ne 0$. For $x = 0$ we have

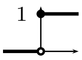
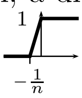
$$F'(0) = \lim_{\varepsilon \to 0} \frac{F(0+\varepsilon) - F(0)}{\varepsilon} = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} = \infty\,.$$

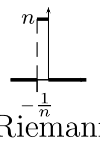
So, the function

$$f(x) = \begin{cases} \infty & \text{for } x = 0, \\ 0 & \text{otherwise} \end{cases}$$

is the derivative of F. Thus, the density of a discrete distribution exists.

N: First, you treat $\lim_{\varepsilon \to 0}$ as $\lim_{\varepsilon \to 0-}$; your derivative is in fact one-sided. Second, it is illegal for f to take on the value ∞.

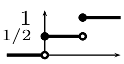
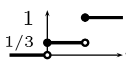
Y: For $f : \mathbb{R} \to \mathbb{R}$ is is illegal, but for $f : \mathbb{R} \to [0, +\infty]$ it is legal. One-sided limit... so what? I mean, a discontinuous function is a limit of a sequence of continuous functions,  $= \lim_{n\to\infty}$  and I take the limit of derivatives,

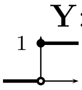
$$f = \lim_{n \to \infty} f_n\,, \qquad f_n = F_n' =$$ 

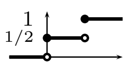
It is as legal as, say, using improper integral instead of Riemann integral. Call it improper derivative, if you want.

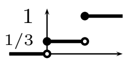
N: Anyway, your 'improper density' is useless. Consider for example

$$f(x) = \begin{cases} \infty & \text{for } x = 0, \\ \infty & \text{for } x = 1, \\ 0 & \text{otherwise.} \end{cases}$$

What is the corresponding distribution function F? Is it  or maybe , etc? You see, $2 \cdot \infty = \infty$.

Y: That is right, '∞' does not calibrate a singularity. Let us denote the derivative of  by δ, then

$F =$  $\implies \quad f(x) = \frac{1}{2}\delta(x) + \frac{1}{2}\delta(x-1)\,,$

$F =$  $\implies \quad f(x) = \frac{1}{3}\delta(x) + \frac{2}{3}\delta(x-1)\,.$

[16]In fact, on any set of zero Lebesgue measure.

N: If you put $\delta(x) = \begin{cases} \infty & \text{for } x = 0, \\ 0 & \text{otherwise,} \end{cases}$ you get $2\delta = \delta$. Your new notation '$\delta(0)$' is not better than our old '∞'.

Y: The 'new' function δ (invented long ago by a physicist Paul Dirac, not by me just now) is specified not by $\delta(x) = \begin{cases} \infty & \text{for } x = 0, \\ 0 & \text{otherwise,} \end{cases}$ but by

$$\int_a^b \delta(x)\, dx = \begin{cases} 1 & \text{if } 0 \in (a,b), \\ 0 & \text{if } 0 \notin [a,b]. \end{cases}$$

N: A function consists of its values, not integrals. Could you agree if I introduce a 'new' number Δ 'defined' by $\Delta + 1 = \Delta$, $\Delta - 1 = -\Delta$?

Y: However, physicists use Dirac's delta-function! It is useful, and does not lead to paradoxes (unless you insist that it must belong to 'old-fashioned' functions).

N: Not just physicists. Also mathematicians use Dirac's 'delta-function'. It exists, but not among functions. It exists among so-called Schwartz distributions[17] (known also as 'generalized functions'). Use them, *if* you are acquainted with their theory, otherwise you do not know what is legal and what is not.

According to the conventional terminology, A DENSITY IS A FUNCTION (rather than, say, a Schwartz distribution), and the integral in (2c2) is treated (most generally) as Lebesgue integral.[18]

If X has a density f_X, then its distribution function F_X is continuous.[19] That is, a discontinuous F_X has no density. In particular, DISCRETE DISTRIBUTIONS HAVE NO DENSITIES.

If F_X is continuous, it does not mean that X has a density. Bizarre distribution functions of examples 2b6, 2b7 are continuous, but nevertheless, have no densities.[20]

2d Distributions

2d1 Proposition. Let $X : \Omega \to \mathbb{R}$ be a random variable, and $B \in \mathcal{B}$ (that is, $B \subset \mathbb{R}$ is a Borel set). Then the set $\{\omega \in \Omega : X(\omega) \in B\}$ is an event.[21]

The probability $\mathbb{P}\big(X \in B \big) = P\big(\{\omega \in \Omega : X(\omega) \in B\}\big)$ is therefore well-defined for *every* Borel set $B \subset \mathbb{R}$ (not just interval).

2d2 Proposition. Let $X : \Omega \to \mathbb{R}$ be a random variable. Then the function $P_X : \mathcal{B} \to [0,1]$ defined by

(2d3) $$P_X(B) = \mathbb{P}\big(X \in B \big)$$

---

[17] Schwartz distributions are in general *not* probability distributions; the two ideas of 'distribution' are related but different.

[18] It conforms with proper and improper Riemann integration when the latter is applicable.

[19] Which follows from the theory of Lebesgue integration.

[20] There is a necessary and sufficient condition for existence of a density, the so-called *absolute continuity.* These bizarre functions are continuous but not absolutely continuous.

[21] That is, belongs to $\mathcal{F}$. As usual, $(\Omega, \mathcal{F}, P)$ is a given probability space.

is a probability measure on $(\mathbb{R}, \mathcal{B})$.[22]

2d4 Definition. The probability measure P_X defined by (2d3) is called the *distribution* of a random variable X.

Note that $(\mathbb{R}, \mathcal{B}, P_X)$ is another probability space.

Clearly,

$$F_X(x) = P_X\big((-\infty, x]\big) . \tag{2d5}$$

Thus, $P_X = P_Y$ implies $F_X = F_Y$. The converse is also true.

2d6 Proposition. If $F_X = F_Y$ then $P_X = P_Y$.

You can easily deduce 2d6 from 1f11.

So, distributions are in a one-one correspondence with distribution functions. A good luck; we cannot draw a distribution itself, but we can draw (the graph of) its distribution function.

2d7 Definition. Random variables X, Y are called *identically distributed,* if $P_X = P_Y$.

The latter definition is applicable also to the case of different probability spaces $(\Omega_1, \mathcal{F}_1, P_1)$, $(\Omega_2, \mathcal{F}_2, P_2)$, when $X : \Omega_1 \to \mathbb{R}$, $Y : \Omega_2 \to \mathbb{R}$. (Usually, speaking about two random variables, we mean 'on the same probability space'.)

Every probability measure on $(\mathbb{R}, \mathcal{B})$ corresponds to some random variable[23] on *some* probability space. The proof is immediate: given a probability measure $P : \mathcal{B} \to [0, 1]$, consider a probability space $(\mathbb{R}, \mathcal{B}, P)$ and a random variable $X : \mathbb{R} \to \mathbb{R}$, $X(\omega) = \omega$ for all $\omega \in \mathbb{R}$.[24]

2d8 Proposition. Let P be a probability measure on $(\mathbb{R}, \mathcal{B})$. Then the corresponding distribution function

$$F(x) = P\big((-\infty, x]\big)$$

satisfies the following conditions.

(a) $\forall x \in \mathbb{R} \quad 0 \le F(x) \le 1$.

(b) F increases, that is, $x_1 \le x_2 \quad \Longrightarrow \quad F(x_1) \le F(x_2)$.

(c) $F(-\infty) = 0$, that is, $\lim_{x \to -\infty} F(x) = 0$.

(d) $F(+\infty) = 1$, that is, $\lim_{x \to +\infty} F(x) = 1$.

(e) F is right continuous, that is, $F(x+) = F(x)$.

You can prove the proposition easily, using a simple but important consequence of sigma-additivity given below.

---

[22] A probability measure on $(\Omega, \mathcal{F})$ was defined in Sect. 1e for any σ-field $\mathcal{F}$ on any set Ω. In particular, it is well-defined for the case $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$.

[23] Of course, there are many such random variables.

[24] In the next section we'll see that, moreover, every probability measure on $(\mathbb{R}, \mathcal{B})$ corresponds to some random variable defined on the standard probability space, $(0, 1)$ with Lebesgue measure.

2d9 Exercise. Prove that

$$A_n \uparrow A \quad \Longrightarrow \quad \mathbb{P}\big(A_n\big) \uparrow \mathbb{P}\big(A\big),$$
$$A_n \downarrow A \quad \Longrightarrow \quad \mathbb{P}\big(A_n\big) \downarrow \mathbb{P}\big(A\big)$$

for any events $A_1, A_2, \dots$

Here '$A_n \uparrow A$' means that $A_1 \subset A_2 \subset \dots$ and $A = A_1 \cup A_2 \cup \dots$ Similarly, '$A_n \downarrow A$' means that $A_1 \supset A_2 \supset \dots$ and $A = A_1 \cap A_2 \cap \dots$.[25] And of course, $P(A_n) \uparrow P(A)$ means that $P(A_1) \le P(A_2) \le \dots$ and $P(A) = \lim_{k\to\infty} P(A_k)$.

Did you understand why F must be right continuous but not left continuous? Since

$$x_n \downarrow x \quad \Longrightarrow \quad (-\infty, x_n] \downarrow (-\infty, x],$$

however,

$$x_n \uparrow x \quad \not\Longrightarrow \quad (-\infty, x_n] \uparrow (-\infty, x];$$

in fact, $(-\infty, x_n] \uparrow (-\infty, x)$ except for a degenerate case.

Note that all our examples of distribution functions (including bizarre examples) satisfy 2d8(a–e).

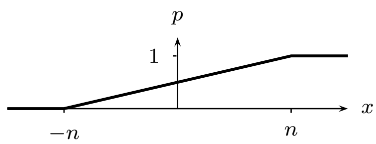
2d10 Exercise. Using (2c2) and 2d8(c,d) prove that $F(x) = \int_{-\infty}^{x} f(x_1)\,dx_1$ and $\int_{-\infty}^{+\infty} f(x)\,dx = 1$ whenever a density exists.

Is there a uniform distribution on the whole $\mathbb{R}$? A discussion follows.

Y: For every n there is a uniform distribution $\mathrm{U}(-n, n)$ on the interval $(-n, n)$. Its limit for $n \to \infty$ is the uniform distribution on $\mathbb{R}$.

N: The distribution function for $\mathrm{U}(-n, n)$ is

$$F_n(x) = \begin{cases} 0 & \text{for } x \in (-\infty, -n], \\ \frac{x+n}{2n} & \text{for } x \in [-n, n], \\ 1 & \text{for } x \in [n, +\infty). \end{cases}$$



Its limit for $n \to \infty$ is $F(x) = \lim_{n\to\infty} \frac{x+n}{2n} = \frac{1}{2}$. However, F is not a distribution function, it violates 2d8(c,d).

Y: I feel, something is wrong with 2d8(c,d).

N: I can say it in other words. Imagine the uniform distribution P on $\mathbb{R}$. What is $P\big([-1,1]\big)$?

Y: It tends to 0, since $[-1,1]$ is an infinitesimal part of the whole $\mathbb{R}$.

N: You must return to Sect. 1c; especially, see page 4. You say, $P\big([-1,1]\big)$ *tends* to 0, and you must add something like 'when $n \to \infty$', but you have no n here, unless you use an infinite sequence of models (these are $\mathrm{U}(-n,n)$) instead of a single model. You must say:

[25] Generally, $A_n \to A \quad \Longrightarrow \quad \mathbb{P}\big(A_n\big) \to \mathbb{P}\big(A\big)$ also for non-monotone sequences $A_1, A_2, \dots$ if $\lim A_n$ is defined appropriately. However, we do not need it now.

$P\big([-1,1]\big) = 0$. Similarly, $P\big([-2,2]\big) = 0$ and so on. However, $[-n,n] \uparrow \mathbb{R}$ and you get $P(\mathbb{R}) = 0$ instead of $P(\mathbb{R}) = 1$.

Y: Recall, you agree to define $\int_{-\infty}^{+\infty} f(x)\,dx$ as $\lim_{n\to\infty} \int_{-n}^{n} f(x)\,dx$. Similarly, I may define $P(B) = \lim_{n\to\infty} P_n(B)$ for any Borel set $B \subset \mathbb{R}$; here P_n is $\mathrm{U}(-n,n)$. Why not?

N: First, the limit need not exist. For example, it does not exist for $B = [1,2] \cup [4,8] \cup [16,32] \cup \dots$ Second, your definition gives $P\big([n,n+1)\big) = 0$ for every n, but $P(\mathbb{R}) = 1$, in contradiction to sigma-additivity.

Y: I feel, something is wrong with sigma-additivity. It is too restrictive.

N: I do not agree. Anyway, let me ask you another question. How could I choose a real number $x \in \mathbb{R}$ at random, uniformly on the whole $\mathbb{R}$?

Y: What is the problem? Just choose it at once.

N: No, that is an illusion. Try to do it gradually, like choosing $X \sim \mathrm{U}(0,1)$ by tossing a coin for its binary digits (though you may prefer decimal digits). What are digits of your number (uniform on the whole $\mathbb{R}$)? They must be independent and uniform, all digits, both of the fractional part and of the integral part.

Y: Nice, that is a way to choose x gradually. Just choose digits by tossing a coin.

N: And I get a *two-sided* infinite sequence $(\dots, \beta_{-2}, \beta_{-1}, \beta_0, \beta_1, \beta_2, \dots)$. Should I write $X = \pm \sum_{k=-\infty}^{+\infty} \beta_k 2^k$? Almost surely, the sum does not converge, since infinitely many of $\beta_1, \beta_2, \dots$ are non-zero. Your two-sided digital monster is not a real number.

Such a set function as

$$P(B) = \lim_{n\to\infty} \frac{1}{2n} \operatorname{mes}\big(B \cap [-n,n]\big)$$

is legal[26] and sometimes useful. However, P is *not a probability measure.* Speaking about 'uniform distribution on the whole $\mathbb{R}$' one escapes the usual probability theory. In principle, you may try it *if* you know what are you doing. However, in the framework of the usual probability theory, THERE IS NO UNIFORM DISTRIBUTION ON THE WHOLE $\mathbb{R}$ (or another set of infinite Lebesgue measure). By the way, discrete probability says the same: there is no uniform distribution on $\{1,2,\dots\}$; there is no countably infinite symmetric probability space.

Every distribution P corresponds to a distribution function F satisfying 2d8(a–e).

2d11 Exercise. Prove that

$$\begin{aligned}
&\mathbb{P}\big(X \in (-\infty,x] \big) = F_X(x)\,; &&\mathbb{P}\big(X \in (-\infty,x) \big) = F_X(x-)\,;\\
&\mathbb{P}\big(X \in [x,+\infty) \big) = 1 - F_X(x-)\,; &&\mathbb{P}\big(X \in (x,+\infty) \big) = 1 - F_X(x)\,;\\
&\mathbb{P}\big(X \in [a,b] \big) = F_X(b) - F_X(a-)\,; &&\mathbb{P}\big(X \in (a,b) \big) = F_X(b-) - F_X(a)\,;\\
&\mathbb{P}\big(X \in [a,b) \big) = F_X(b-) - F_X(a-)\,; &&\mathbb{P}\big(X \in (a,b] \big) = F_X(b) - F_X(a)\,;
\end{aligned}$$

$$\mathbb{P}\big(X \in [a,b] \cup (c,d) \big) = F_X(d-) - F_X(c) + F_X(b) - F_X(a-) \qquad (a<b<c<d)\,.$$

2d12 Exercise. Prove that

$$\mathbb{P}\big(X = x \big) = F_X(x) - F_X(x-)$$

[26] However, one must bother about existence of the limit.

and

$$\mathbb{P}\big(X \in B\big) = \sum_{x \in B} \big(F_X(x) - F_X(x-)\big)$$

for any *finite or countable* set $B \subset \mathbb{R}$. Can you generalize the formula for uncountable sets?

2d13 Exercise. For the random variable X of Example 2b8 prove that

$$\mathbb{P}\big(X \in B\big) = \sum_{A \cap B} \big(F_X(x) - F_X(x-)\big) = \sum_{k : \sin k \in B} pq^k$$

for any Borel set $B \subset \mathbb{R}$; here $A = \{\sin k : k = 0, 1, 2, \dots\}$ is a countable set dense in $[-1, 1]$. Can you generalize the formula for non-Borel sets?

2d14 Definition. A number $x \in \mathbb{R}$ is called an *atom* of (the distribution of) a random variable X, if

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

The random variable X (as well as its distribution) is called *nonatomic,* if

$$\forall x \in \mathbb{R} \quad \mathbb{P}\big(X = x\big) = 0\,.$$

Clearly, X is nonatomic if and only if F_X is continuous. If X has a density then it is nonatomic. The converse is false (think, why).[27]

2d15 Definition. The *support* of (the distribution of) a random variable X is the set of all $x \in \mathbb{R}$ such that

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

The support is always a closed set S of probability 1; I mean, $\mathbb{P}\big(X \in S\big) = 1$. In fact, the support is the least closed set of probability 1.

2d16 Exercise. Find atoms and the support of the bizarre distribution of Example 2b8.

2e Quantile function

The notion of a median appears even in newspapers; it was proposed to replace 'mean salary' with 'median salary', that is, a number higher than a half of the salaries and lower than the other half.

2e1 Definition. Let X be a random variable, $x \in \mathbb{R}$, $p \in (0, 1)$. The number x is called a *p-quantile* of X, if

$$\mathbb{P}\big(X < x\big) \le p \le \mathbb{P}\big(X \le x\big)\,. \tag{2e2}$$

A *median* is an $\frac{1}{2}$-quantile.

[27] I do not like the term 'continuous distribution' since it is somewhat ambiguous; some people interpret it as 'nonatomic distribution', others as 'distribution that has a density'.

In terms of F_X we may rewrite (2e2) as

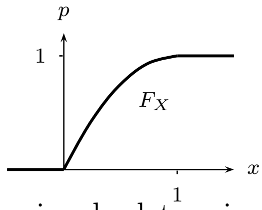
$$F_X(x-) \le p \le F_X(x)\,. \tag{2e3}$$

If F_X is continuous, it means simply

$$F_X(x) = p\,. \tag{2e4}$$

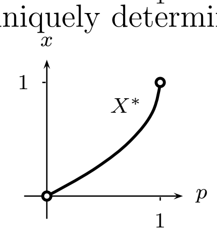
2e5 Example. Recall Example 2b1:

$$F_X(x) = \begin{cases} 0 & \text{for } x \in (-\infty, 0], \\ x(2-x) & \text{for } x \in [0,1], \\ 1 & \text{for } x \in [1, \infty). \end{cases}$$



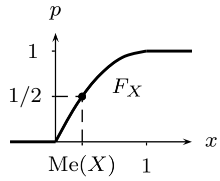
Here F_X is continuous, thus (2e2) becomes (2e4), and x is uniquely determined by p:

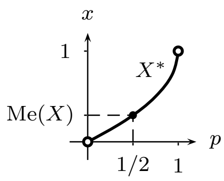
$$x(2-x) = p; \quad 1-(1-x)^2 = p;$$
$$(1-x)^2 = 1-p; \quad x = 1 - \sqrt{1-p}\,.$$



It is the function inverse to F_X, or rather, to the restriction $F_X|_{(0,1)}$. In particular, the median is

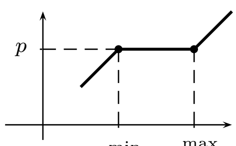
$$\mathrm{Me}(X) = 1 - \sqrt{1-\frac{1}{2}} = 1 - \frac{1}{\sqrt{2}} \approx 0.293$$





Usually, F_X is continuous and strictly increasing on some $[a,b]$ such that $F_X(a) = 0$, $F_X(b) = 1$. Then, denoting by F_X^{-1} the function inverse to $F_X|_{(a,b)}$, we have $F_X^{-1} : (0,1) \to (a,b)$, and $F_X^{-1}(p)$ is a p-quantile for every $p \in (0,1)$. The case $a = -\infty$, $b = +\infty$ is also usual; here, F_X is continuous and strictly increasing on the whole $\mathbb{R}$, and $F_X^{-1} : (0,1) \to (-\infty, +\infty)$. Of course, it can happen that $a = -\infty$ but $b < +\infty$ (or the opposite).

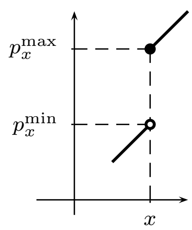
However, a p-quantile need not be unique, since a distribution may have a gap:



$$\{x : F_X(x) = p\} = [x_p^{\min}, x_p^{\max}]\,.$$

Here, every $x \in [x_p^{\min}, x_p^{\max}]$ is a p-quantile. See Example 2b6 for a lot of gaps. In fact, the support is the complement of the union of all gaps (treated as open intervals).

On the other hand, a single number may be a p-quantile for many values of p, since a distribution may have an atom:



$$F_X(x-) = p_x^{\min} < p_x^{\max} = F_X(x)\,.$$

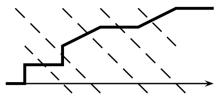
Here, x is a p-quantile for every $p \in [p_x^{\min}, p_x^{\max}]$.

We define the *quantile line* of X as the set of all $(x,p) \in \mathbb{R}^2$ such that

$$\begin{aligned} &p \in (0,1) \text{ and } x \text{ is a } p\text{-quantile,} \\ &\text{or } p = 0 \text{ and } F_X(x-) = 0, \\ &\text{or } p = 1 \text{ and } F_X(x) = 1. \end{aligned}$$

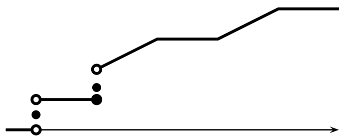
For example:

In general, the quantile line is not a graph of a function $p = f(x)$, nor $x = g(p)$, since its intersection with a vertical or horizontal line may be a segment rather than a single point. However, for every u the line $x + p = u$ intersects the quantile line at one and only one point.[28]



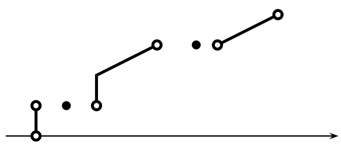
The quantile line divides R^2 into two regions. One region, above the line and to the left, is $\{(x,p) \in \mathbb{R}^2 : F_X(x-) \le p\}$; the other region, below and to the right, is $\{(x,p) \in \mathbb{R}^2 : p \le F_X(x)\}$; I mean closed regions; their intersection is the quantile line $\{(x,p) \in \mathbb{R}^2 : F_X(x-) \le p \le F_X(x)\}$.

The graph $\{(x,p) \in \mathbb{R}^2 : p = F_X(x)\}$ is a subset of the quantile line. Their relation is easy to describe: replace every vertical segment of the quantile line with its highest point, and you get the graph. Using the lowest point instead, you get the graph of the left-continuous function $x \mapsto F_X(x-) = \mathbb{P}\big(X < x \big)$. Choosing an arbitrary point, you get a function f such that $F_X(x-) \le f(x) \le F_X(x+)$ for all x. Then $f(x-) = F_X(x-)$ and $f(x+) = F_X(x+)$ despite the arbitrary choice.



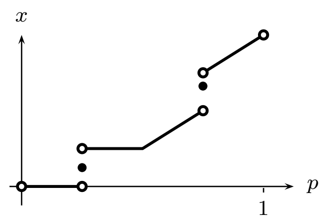
(An elementary case is shown on the picture, but the statements hold in full generality, including bizarre functions of 2b6, 2b7, 2b8.)

Similarly, we may replace every *horizontal* segment of the quantile line with a single point chosen arbitrarily on the segment. (Horizontal rays at $p = 0$ and $p = 1$ are just removed.) We get the graph $\{(x,p) \in \mathbb{R} \times (0,1) : x = g(p)\}$ of a function $g : (0,1) \to \mathbb{R}$ such that $g(p)$ is a p-quantile.



[28] The quantile line can be described by *continuous* functions $u \mapsto x_u$, $u \mapsto p_u$ of the new variable $u = x+p$. Moreover, $0 \le x_{u+\Delta u} - x_u \le \Delta u$, $0 \le p_{u+\Delta u} - p_u \le \Delta u$. Of course, $x + p$ is only a convenient trick; $x + 2p$ works equally well.

2e6 Definition. A function $X^* : (0,1) \to \mathbb{R}$ is called a *quantile function* of a random variable X, if $X^*(p)$ is a p-quantile of X whenever $p \in (0,1)$.



In the invertible case, $X^* = F_X^{-1}$, and so, a quantile function is unique. In general, $X^*(p-)$ and $X^*(p+)$ are uniquely determined, but $X^*(p) \in [X^*(p-), X^*(p+)]$ is arbitrary if X^* is discontinuous at p. Anyway, X^* is an increasing function.[29] Also,[30]

$$
\begin{aligned}
X^* \text{ is continuous} &\iff F_X \text{ is strictly increasing}\,; \\
X^* \text{ is strictly increasing} &\iff F_X \text{ is continuous}\,.
\end{aligned}
\tag{2e7}
$$

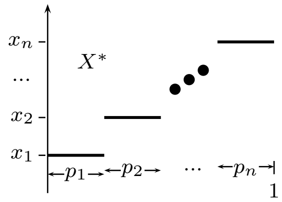
Try to apply it to Examples 2b6, 2b7, 2b8.

The two regions can be described in terms of X^* as well as F_X:

$$
\begin{aligned}
F_X(x-) \le p &\iff x \le X^*(p+)\,, \\
p \le F_X(x+) &\iff X^*(p-) \le x\,, \\
F_X(x-) \le p \le F_X(x+) &\iff X^*(p-) \le x \le X^*(p+)\,;
\end{aligned}
\tag{2e8}
$$

the latter describes the quantile line; of course, $F_X(x+) = F_X(x)$.

Looking at the discrete case,



we see that X^* is distributed like X. Indeed, $X^*(\cdot) = x_1$ on an interval of length p_1; $X^*(\cdot) = x_2$ on an interval of length p_2; and so on.[31]

2e9 Theorem. Let $(\Omega, \mathcal{F}, P)$ be a probability space, $X : \Omega \to \mathbb{R}$ a random variable, and $X^* : (0,1) \to \mathbb{R}$ a quantile function of X. Consider X^* as a random variable on the probability space $(0,1)$ (equipped with Lebesgue measure). Then random variables X and X^* are identically distributed.

It means simply that the set $\{p \in (0,1) : X^*(p) \le x\}$ is either $\big(0, F_X(x)\big)$ or $\big(0, F_X(x)\big]$. (Both cases are possible; think, why.)

Theorem 2e9 is useful for simulating random variables.[32] Having a random numbers generator that gives p distributed uniformly on $(0,1)$, we get $x = X^*(p)$ distributed like X.

[29] Not strictly increasing, in general.

[30] Strict increase of F_X does not relate to x such that $F_X(x) = 0$ or $F_X(x) = 1$.

[31] Arbitrary values at jumping points do not matter.

[32] Other ways may be more effective for special distributions, but this way is quite universal.

Moreover, the quantile function may be used for constructing a random variable from scratch. Assume that F is a function satisfying 2d8(a–e). For now we do not know, whether $F = F_X$ for some X, or not.[33] Nevertheless we may define the quantile line of F as $\{(x,p) \in \mathbb{R}^2 : F(x-) \le p \le F(x)\}$. It appears that all needed properties of the quantile line follow from 2d8(a–e). As we know, a quantile line leads to a quantile function $X^* : (0,1) \to \mathbb{R}$. Though, there is no X for now. However, we may take $X = X^*$; the very X^* is a random variable! It appears that $F_X = F$. So, Conditions 2d8(a–e) are not only necessary but also sufficient.

2e10 Theorem. For any function $F : \mathbb{R} \to \mathbb{R}$, Conditions 2d8(a–e) are necessary and sufficient for existence of a probability measure P on $(\mathbb{R}, \mathcal{B})$ such that

$$\forall x \in \mathbb{R} \quad P\big((-\infty, x]\big) = F(x) \,.$$

If such P exists, it is unique (recall 2d6), and we get the following fact.

2e11 Corollary. The formula

$$\forall x \in \mathbb{R} \quad P\big((-\infty, x]\big) = F(x)$$

establishes a one-one correspondence between probability distributions P on $\mathbb{R}$ and functions F satisfying 2d8(a–e).

2e12 Corollary. For every function $f : \mathbb{R} \to \mathbb{R}$ satisfying[34]

$$\begin{gathered} \forall x \quad f(x) \ge 0 \,, \\ \int_{-\infty}^{+\infty} f(x)\, dx = 1 \end{gathered}$$

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

Indeed, the function $F(x) = \int_{-\infty}^{x} f(x_1)\, dx_1$ satisfies 2d8(a–e).

[33] In other words, we do not know, whether or not $F(x) = P\big((-\infty, x]\big)$ for some probabilty measure P on $(\mathbb{R}, \mathcal{B})$; recall the paragraph before 2d8.

[34] The function must be good enough for its integral to exist; recall Sect. 2c.