

Continuous random variables

Many experiments have outcomes that take values on a continuous scale. For example, in Chapter 2 we encountered the load at which a model of a bridge collapses. These experiments have *continuous* random variables naturally associated with them.

5.1 Probability density functions

One way to look at continuous random variables is that they arise by a (never-ending) process of refinement from discrete random variables. Suppose, for example, that a discrete random variable associated with some experiment takes on the value 6.283 with probability p . If we refine, in the sense that we also get to know the fourth decimal, then the probability p is spread over the outcomes 6.2830, 6.2831, \dots , 6.2839. Usually this will mean that each of these new values is taken on with a probability that is much smaller than p —the sum of the ten probabilities is p . Continuing the refinement process to more and more decimals, the probabilities of the possible values of the outcomes become smaller and smaller, approaching zero. However, the probability that the possible values lie in some fixed interval $[a, b]$ will settle down. This is closely related to the way sums converge to an integral in the definition of the integral and motivates the following definition.

DEFINITION. A random variable X is *continuous* if for some function $f : \mathbb{R} \rightarrow \mathbb{R}$ and for any numbers a and b with $a \leq b$,

$$P(a \leq X \leq b) = \int_a^b f(x) \, dx.$$

The function f has to satisfy $f(x) \geq 0$ for all x and $\int_{-\infty}^{\infty} f(x) \, dx = 1$. We call f the *probability density function* (or *probability density*) of X .

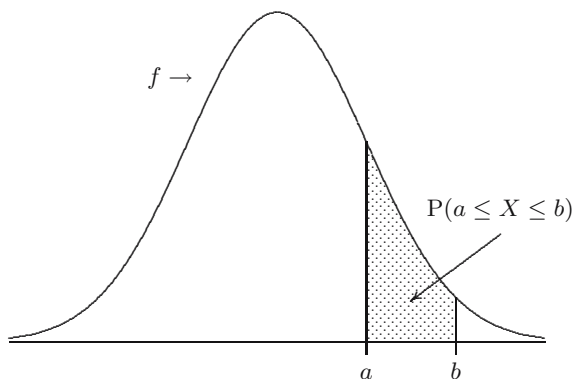


Fig. 5.1. Area under a probability density function f on the interval $[a, b]$.

Note that the probability that X lies in an interval $[a, b]$ is equal to the area under the probability density function f of X over the interval $[a, b]$; this is illustrated in Figure 5.1. So if the interval gets smaller and smaller, the probability will go to zero: for any positive ε

$$P(a - \varepsilon \leq X \leq a + \varepsilon) = \int_{a-\varepsilon}^{a+\varepsilon} f(x) dx,$$

and sending ε to 0, it follows that for any a

$$P(X = a) = 0.$$

This implies that for continuous random variables you may be careless about the precise form of the intervals:

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a < X < b) = P(a \leq X < b).$$

What does $f(a)$ represent? Note (see also Figure 5.2) that

$$P(a - \varepsilon \leq X \leq a + \varepsilon) = \int_{a-\varepsilon}^{a+\varepsilon} f(x) dx \approx 2\varepsilon f(a) \quad (5.1)$$

for small positive ε . Hence $f(a)$ can be interpreted as a (relative) measure of how likely it is that X will be near a . However, do not think of $f(a)$ as a probability: $f(a)$ can be arbitrarily large. An example of such an f is given in the following exercise.

QUICK EXERCISE 5.1 Let the function f be defined by $f(x) = 0$ if $x \leq 0$ or $x \geq 1$, and $f(x) = 1/(2\sqrt{x})$ for $0 < x < 1$. You can check quickly that f satisfies the two properties of a probability density function. Let X be a random variable with f as its probability density function. Compute the probability that X lies between 10^{-4} and 10^{-2} .

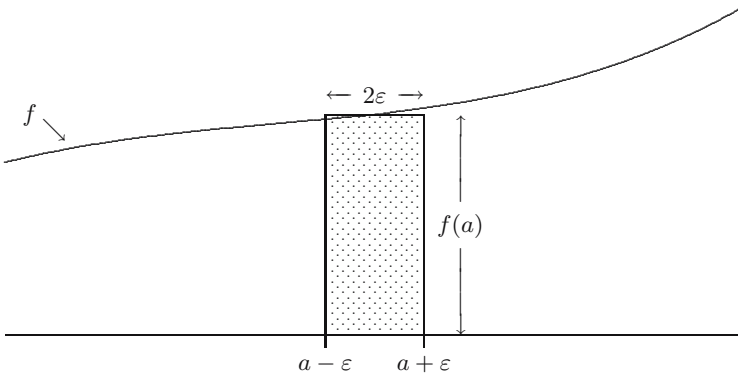


Fig. 5.2. Approximating the probability that X lies ε -close to a .

You should realize that discrete random variables do not have a probability density function f and continuous random variables do not have a probability mass function p , but that *both* have a distribution function $F(a) = P(X \leq a)$. Using the fact that for $a < b$ the event $\{X \leq b\}$ is a disjoint union of the events $\{X \leq a\}$ and $\{a < X \leq b\}$, we can express the probability that X lies in an interval $(a, b]$ directly in terms of F for *both* cases:

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a).$$

There is a simple relation between the distribution function F and the probability density function f of a continuous random variable. It follows from integral calculus that

$$F(b) = \int_{-\infty}^b f(x) dx \quad \text{and}^1 \quad f(x) = \frac{d}{dx} F(x).$$

Both the probability density function and the distribution function of a continuous random variable X contain all the probabilistic information about X ; the *probability distribution* of X is described by either of them.

We illustrate all this with an example. Suppose we want to make a probability model for an experiment that can be described as “an object hits a disc of radius r in a completely arbitrary way” (of course, this is not *you* playing darts—nevertheless we will refer to this example as the darts example). We are interested in the distance X between the hitting point and the center of the disc. Since distances cannot be negative, we have $F(b) = P(X \leq b) = 0$ when $b < 0$. Since the object hits the disc, we have $F(b) = 1$ when $b > r$. That the dart hits the disk in a completely arbitrary way we interpret as that the probability of hitting any region is proportional to the area of that region. In particular, because the disc has area πr^2 and the disc with radius b has area πb^2 , we should put

¹ This holds for all x where f is continuous.

$$F(b) = P(X \leq b) = \frac{\pi b^2}{\pi r^2} = \frac{b^2}{r^2} \quad \text{for } 0 \leq b \leq r.$$

Then the probability density function f of X is equal to 0 outside the interval $[0, r]$ and

$$f(x) = \frac{d}{dx}F(x) = \frac{1}{r^2} \frac{d}{dx}x^2 = \frac{2x}{r^2} \quad \text{for } 0 \leq x \leq r.$$

QUICK EXERCISE 5.2 Compute for the darts example the probability that $0 < X \leq r/2$, and the probability that $r/2 < X \leq r$.

5.2 The uniform distribution

In this section we encounter a continuous random variable that describes an experiment where the outcome is completely arbitrary, except that we know that it lies between certain bounds. Many experiments of physical origin have this kind of behavior. For instance, suppose we measure for a long time the emission of radioactive particles of some material. Suppose that the experiment consists of recording in each hour at what times the particles are emitted. Then the outcomes will lie in the interval $[0, 60]$ minutes. If the measurements would concentrate in any way, there is either something wrong with your Geiger counter or you are about to discover some new physical law. Not concentrating in any way means that subintervals of the same length should have the same probability. It is then clear (cf. equation (5.1)) that the probability density function associated with this experiment should be constant on $[0, 60]$. This motivates the following definition.

DEFINITION. A continuous random variable has a *uniform distribution* on the interval $[\alpha, \beta]$ if its probability density function f is given by $f(x) = 0$ if x is not in $[\alpha, \beta]$ and

$$f(x) = \frac{1}{\beta - \alpha} \quad \text{for } \alpha \leq x \leq \beta.$$

We denote this distribution by $U(\alpha, \beta)$.

QUICK EXERCISE 5.3 Argue that the distribution function F of a random variable that has a $U(\alpha, \beta)$ distribution is given by $F(x) = 0$ if $x < \alpha$, $F(x) = 1$ if $x > \beta$, and $F(x) = (x - \alpha)/(\beta - \alpha)$ for $\alpha \leq x \leq \beta$.

In Figure 5.3 the probability density function and the distribution function of a $U(0, \frac{1}{3})$ distribution are depicted.

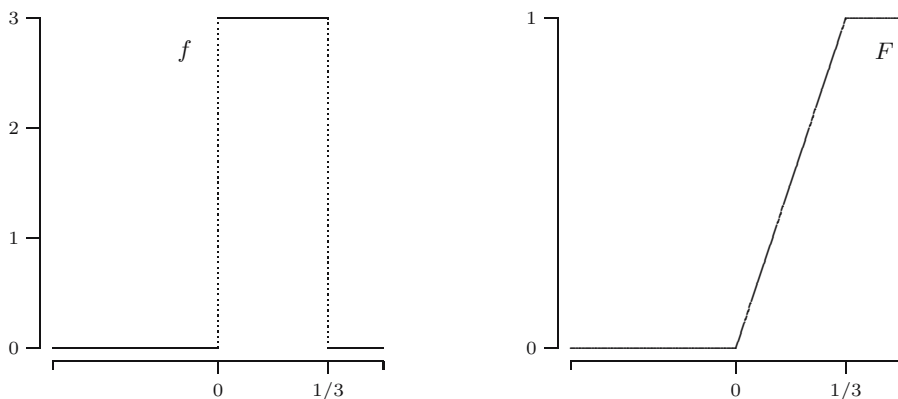


Fig. 5.3. The probability density function and the distribution function of the $U(0, \frac{1}{3})$ distribution.

5.3 The exponential distribution

We already encountered the exponential distribution in the chemical reactor example of Chapter 3. We will give an argument why it appears in that example. Let v be the effluent volumetric flow rate, i.e., the volume that leaves the reactor over a time interval $[0, t]$ is vt (and an equal volume enters the vessel at the other end). Let V be the volume of the reactor vessel. Then in total a fraction $(v/V) \cdot t$ will have left the vessel during $[0, t]$, when t is not too large. Let the random variable T be the residence time of a particle in the vessel. To compute the distribution of T , we divide the interval $[0, t]$ in n small intervals of equal length t/n . Assuming perfect mixing, so that the particle's position is uniformly distributed over the volume, the particle has probability $p = (v/V) \cdot t/n$ to have left the vessel during any of the n intervals of length t/n . If we assume that the behavior of the particle in different time intervals of length t/n is independent, we have, if we call “leaving the vessel” a success, that T has a geometric distribution with success probability p . It follows (see also Quick exercise 4.6) that the probability $P(T > t)$ that the particle is still in the vessel at time t is, for large n , well approximated by

$$(1 - p)^n = \left(1 - \frac{vt}{Vn}\right)^n.$$

But then, letting $n \rightarrow \infty$, we obtain (recall a well-known limit from your calculus course)

$$P(T > t) = \lim_{n \rightarrow \infty} \left(1 - \frac{vt}{V} \cdot \frac{1}{n}\right)^n = e^{-\frac{v}{V}t}.$$

It follows that the distribution function of T equals $1 - e^{-\frac{v}{V}t}$, and differentiating we obtain that the probability density function f_T of T is equal to

$$f_T(t) = \frac{d}{dt}(1 - e^{-\frac{v}{V}t}) = \frac{v}{V}e^{-\frac{v}{V}t} \quad \text{for } t \geq 0.$$

This is an example of an exponential distribution, with parameter v/V .

DEFINITION. A continuous random variable has an *exponential distribution* with parameter λ if its probability density function f is given by $f(x) = 0$ if $x < 0$ and

$$f(x) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0.$$

We denote this distribution by $\text{Exp}(\lambda)$.

The distribution function F of an $\text{Exp}(\lambda)$ distribution is given by

$$F(a) = 1 - e^{-\lambda a} \quad \text{for } a \geq 0.$$

In Figure 5.4 we show the probability density function and the distribution function of the $\text{Exp}(0.25)$ distribution.

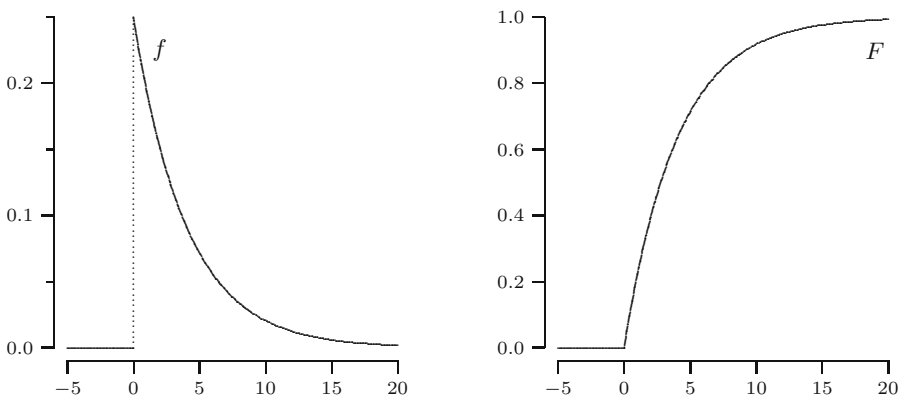


Fig. 5.4. The probability density and the distribution function of the $\text{Exp}(0.25)$ distribution.

Since we obtained the exponential distribution directly from the geometric distribution it should not come as a surprise that the exponential distribution *also* satisfies the memoryless property, i.e., if X has an exponential distribution, then for all $s, t > 0$,

$$P(X > s + t \mid X > s) = P(X > t).$$

Actually, this follows directly from

$$P(X > s + t \mid X > s) = \frac{P(X > s + t)}{P(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t).$$

QUICK EXERCISE 5.4 A study of the response time of a certain computer system yields that the response time in seconds has an exponentially distributed time with parameter 0.25. What is the probability that the response time exceeds 5 seconds?

5.4 The Pareto distribution

More than a century ago the economist Vilfredo Pareto ([20]) noticed that the number of people whose income exceeded level x was well approximated by C/x^α , for some constants C and $\alpha > 0$ (it appears that for all countries α is around 1.5). A similar phenomenon occurs with city sizes, earthquake rupture areas, insurance claims, and sizes of commercial companies. When these quantities are modeled as realizations of random variables X , then their distribution functions are of the type $F(x) = 1 - 1/x^\alpha$ for $x \geq 1$. (Here 1 is a more or less arbitrarily chosen starting point—what matters is the behavior for large x .) Differentiating, we obtain probability densities of the form $f(x) = \alpha/x^{\alpha+1}$. This motivates the following definition.

DEFINITION. A continuous random variable has a *Pareto distribution* with parameter $\alpha > 0$ if its probability density function f is given by $f(x) = 0$ if $x < 1$ and

$$f(x) = \frac{\alpha}{x^{\alpha+1}} \quad \text{for } x \geq 1.$$

We denote this distribution by $Par(\alpha)$.

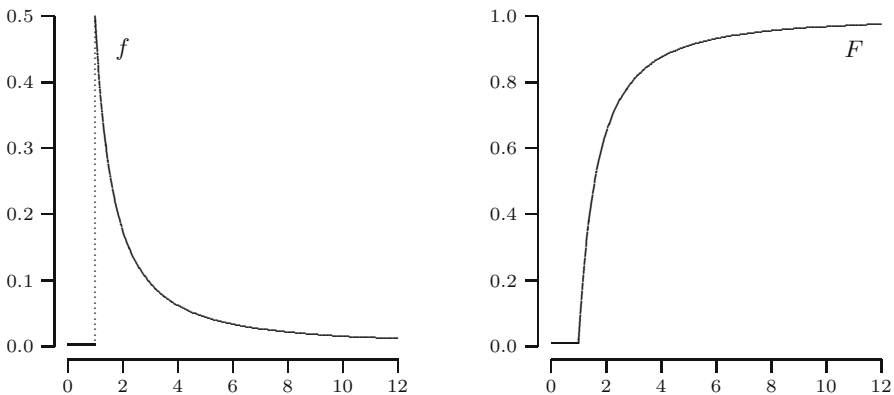


Fig. 5.5. The probability density and the distribution function of the $Par(0.5)$ distribution.

In Figure 5.5 we depicted the probability density f and the distribution function F of the $Par(0.5)$ distribution.

5.5 The normal distribution

The normal distribution plays a central role in probability theory and statistics. One of its first applications was due to C.F. Gauss, who used it in 1809 to model observational errors in astronomy; see [13]. We will see in Chapter 14 that the normal distribution is an important tool to approximate the probability distribution of the average of independent random variables.

DEFINITION. A continuous random variable has a *normal distribution* with parameters μ and $\sigma^2 > 0$ if its probability density function f is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{for } -\infty < x < \infty.$$

We denote this distribution by $N(\mu, \sigma^2)$.

In Figure 5.6 the graphs of the probability density function f and distribution function F of the normal distribution with $\mu = 3$ and $\sigma^2 = 6.25$ are displayed.

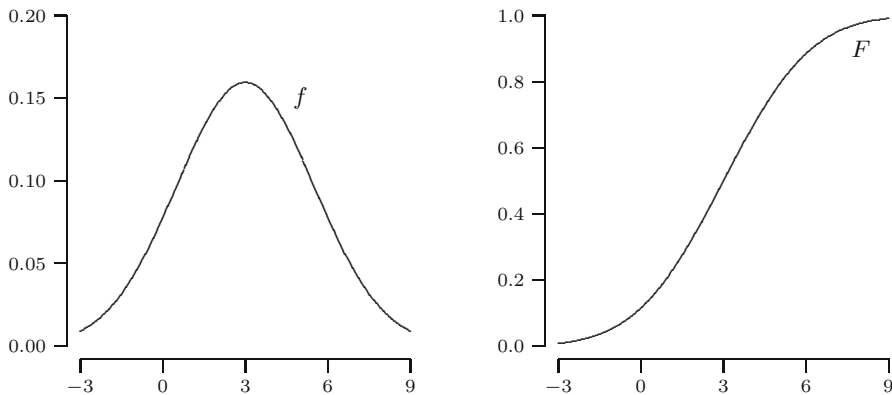


Fig. 5.6. The probability density and the distribution function of the $N(3, 6.25)$ distribution.

If X has an $N(\mu, \sigma^2)$ distribution, then its distribution function is given by

$$F(a) = \int_{-\infty}^a \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad \text{for } -\infty < a < \infty.$$

Unfortunately there is no explicit expression for F ; f has no antiderivative. However, as we shall see in Chapter 8, any $N(\mu, \sigma^2)$ distributed random variable can be turned into an $N(0, 1)$ distributed random variable by a simple transformation. As a consequence, a table of the $N(0, 1)$ distribution suffices. The latter is called the *standard* normal distribution, and because of its special role the letter ϕ has been reserved for its probability density function:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad \text{for } -\infty < x < \infty.$$

Note that ϕ is symmetric around zero: $\phi(-x) = \phi(x)$ for each x . The corresponding distribution function is denoted by Φ . The table for the standard normal distribution (see Table B.1) does not contain the values of $\Phi(a)$, but rather the so-called *right tail probabilities* $1 - \Phi(a)$. If, for instance, we want to know the probability that a standard normal random variable Z is smaller than or equal to 1, we use that $P(Z \leq 1) = 1 - P(Z \geq 1)$. In the table we find that $P(Z \geq 1) = 1 - \Phi(1)$ is equal to 0.1587. Hence $P(Z \leq 1) = 1 - 0.1587 = 0.8413$. With the table you can handle tail probabilities with numbers a given to two decimals. To find, for instance, $P(Z > 1.07)$, we stay in the same row in the table but move to the seventh column to find that $P(Z > 1.07) = 0.1423$.

QUICK EXERCISE 5.5 Let the random variable Z have a standard normal distribution. Use Table B.1 to find $P(Z \leq 0.75)$. How do you know—without doing any calculations—that the answer should be larger than 0.5?

5.6 Quantiles

Recall the chemical reactor example, where the residence time T , measured in minutes, has an exponential distribution with parameter $\lambda = v/V = 0.25$. As we shall see in the next chapters, a consequence of this choice of λ is that the *mean* time the particle stays in the vessel is 4 minutes. However, from the viewpoint of process control this is not the quantity of interest. Often, there will be some minimal amount of time the particle has to stay in the vessel to participate in the chemical reaction, and we would want that at least 90% of the particles stay in the vessel this minimal amount of time. In other words, we are interested in the number q with the property that $P(T > q) = 0.9$, or equivalently,

$$P(T \leq q) = 0.1.$$

The number q is called the 0.1th quantile or 10th percentile of the distribution. In the case at hand it is easy to determine. We should have

$$P(T \leq q) = 1 - e^{-0.25q} = 0.1.$$

This holds exactly when $e^{-0.25q} = 0.9$ or when $-0.25q = \ln(0.9) = -0.105$. So $q = 0.42$. Hence, although the mean residence time is 4 minutes, 10% of

the particles stays less than 0.42 minute in the vessel, which is just slightly more than 25 seconds! We use the following general definition.

DEFINITION. Let X be a continuous random variable and let p be a number between 0 and 1. The p th *quantile* or 100th *percentile* of the distribution of X is the smallest number q_p such that

$$F(q_p) = P(X \leq q_p) = p.$$

The *median* of a distribution is its 50th percentile.

QUICK EXERCISE 5.6 What is the median of the $U(2, 7)$ distribution?

For continuous random variables q_p is often easy to determine. Indeed, if F is *strictly* increasing from 0 to 1 on some interval (which may be infinite to one or both sides), then

$$q_p = F^{\text{inv}}(p),$$

where F^{inv} is the inverse of F . This is illustrated in Figure 5.7 for the $\text{Exp}(0.25)$ distribution.

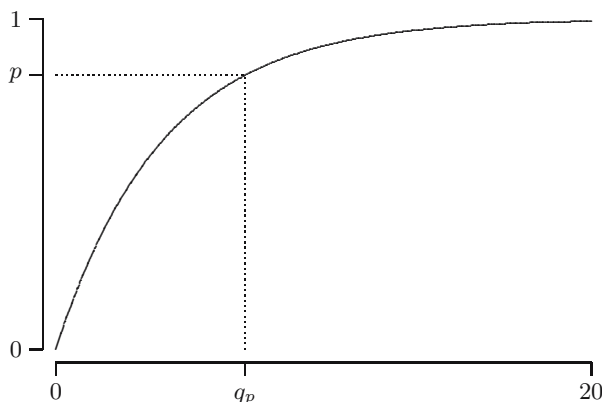


Fig. 5.7. The p th quantile q_p of the $\text{Exp}(0.25)$ distribution.

For an exponential distribution it is easy to *compute* quantiles. This is different for the standard normal distribution, where we have to use a table (like Table B.1). For example, the 90th percentile of a standard normal is the number $q_{0.9}$ such that $\Phi(q_{0.9}) = 0.9$, which is the same as $1 - \Phi(q_{0.9}) = 0.1$, and the table gives us $q_{0.9} = 1.28$. This is illustrated in Figure 5.8, with both the probability density function and the distribution function of the standard normal distribution.

QUICK EXERCISE 5.7 Find the 0.95th quantile $q_{0.95}$ of a standard normal distribution, accurate to two decimals.

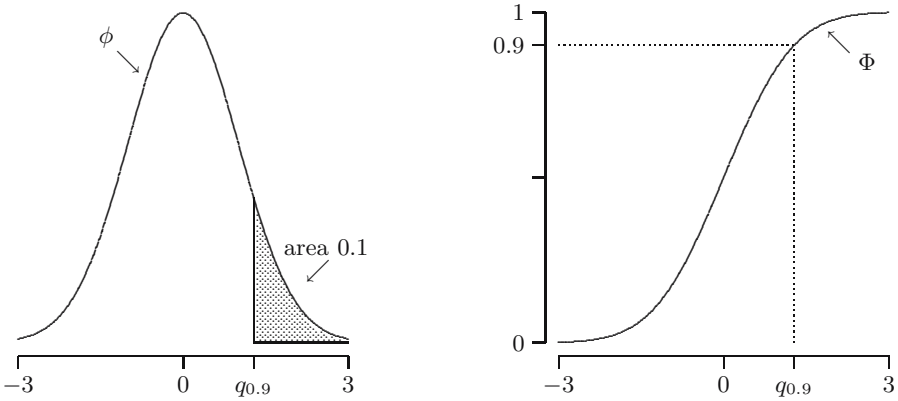


Fig. 5.8. The 90th percentile of the $N(0,1)$ distribution.

5.7 Solutions to the quick exercises

5.1 We know from integral calculus that for $0 \leq a \leq b \leq 1$

$$\int_a^b f(x) \, dx = \int_a^b \frac{1}{2\sqrt{x}} \, dx = \sqrt{b} - \sqrt{a}.$$

Hence $\int_{-\infty}^{\infty} f(x) \, dx = \int_0^1 1/(2\sqrt{x}) \, dx = 1$ (so f is a probability density function—nonnegativity being obvious), and

$$\begin{aligned} P(10^{-4} \leq X \leq 10^{-2}) &= \int_{10^{-4}}^{10^{-2}} \frac{1}{2\sqrt{x}} \, dx \\ &= \sqrt{10^{-2}} - \sqrt{10^{-4}} = 10^{-1} - 10^{-2} = 0.09. \end{aligned}$$

Actually, the random variable X arises in a natural way; see equation (7.1).

5.2 We have $P(0 < X \leq r/2) = F(r/2) - F(0) = (1/2)^2 - 0^2 = 1/4$, and $P(r/2 < X \leq r) = F(r) - F(r/2) = 1 - 1/4 = 3/4$, no matter what the radius of the disc is!

5.3 Since $f(x) = 0$ for $x < \alpha$, we have $F(x) = 0$ if $x < \alpha$. Also, since $f(x) = 0$ for all $x > \beta$, $F(x) = 1$ if $x > \beta$. In between

$$F(x) = \int_{-\infty}^x f(y) \, dy = \int_{\alpha}^x \frac{1}{\beta - \alpha} \, dy = \left[\frac{y}{\beta - \alpha} \right]_{\alpha}^x = \frac{x - \alpha}{\beta - \alpha}.$$

In other words; the distribution function increases linearly from the value 0 in α to the value 1 in β .

5.4 If X is the response time, we ask for $P(X > 5)$. This equals

$$P(X > 5) = e^{-0.25 \cdot 5} = e^{-1.25} = 0.2865 \dots$$

5.5 In the eighth row and sixth column of the table, we find that $1 - \Phi(0.75) = 0.2266$. Hence the answer is $1 - 0.2266 = 0.7734$. Because of the symmetry of the probability density ϕ , half of the mass of a standard normal distribution lies on the negative axis. Hence for any number $a > 0$, it should be true that $P(Z \leq a) > P(Z \leq 0) = 0.5$.

5.6 The median is the number $q_{0.5} = F^{\text{inv}}(0.5)$. You either see directly that you have got half of the mass to both sides of the middle of the interval, hence $q_{0.5} = (2 + 7)/2 = 4.5$, or you solve with the distribution function:

$$\frac{1}{2} = F(q) = \frac{q-2}{7-2}, \quad \text{and so} \quad q = 4.5.$$

5.7 Since $\Phi(q_{0.95}) = 0.95$ is the same as $1 - \Phi(q_{0.95}) = 0.05$, the table gives us $q_{0.95} = 1.64$, or more precisely, if we interpolate between the fourth and the fifth column; 1.645.

5.8 Exercises

5.1 Let X be a continuous random variable with probability density function

$$f(x) = \begin{cases} \frac{3}{4} & \text{for } 0 \leq x \leq 1 \\ \frac{1}{4} & \text{for } 2 \leq x \leq 3 \\ 0 & \text{elsewhere.} \end{cases}$$

- a. Draw the graph of f .
- b. Determine the distribution function F of X , and draw its graph.

5.2 \square Let X be a random variable that takes values in $[0, 1]$, and is further given by

$$F(x) = x^2 \quad \text{for } 0 \leq x \leq 1.$$

Compute $P(\frac{1}{2} < X \leq \frac{3}{4})$.

5.3 Let a continuous random variable X be given that takes values in $[0, 1]$, and whose distribution function F satisfies

$$F(x) = 2x^2 - x^4 \quad \text{for } 0 \leq x \leq 1.$$

- a. Compute $P(\frac{1}{4} \leq X \leq \frac{3}{4})$.
- b. What is the probability density function of X ?

5.4 \boxplus Jensen, arriving at a bus stop, just misses the bus. Suppose that he decides to walk if the (next) bus takes longer than 5 minutes to arrive. Suppose also that the time in minutes between the arrivals of buses at the bus stop is a continuous random variable with a $U(4, 6)$ distribution. Let X be the time that Jensen will wait.

- a. What is the probability that X is less than $4\frac{1}{2}$ (minutes)?
- b. What is the probability that X equals 5 (minutes)?
- c. Is X a discrete random variable or a continuous random variable?

5.5 \square The probability density function f of a continuous random variable X is given by:

$$f(x) = \begin{cases} cx + 3 & \text{for } -3 \leq x \leq -2 \\ 3 - cx & \text{for } 2 \leq x \leq 3 \\ 0 & \text{elsewhere.} \end{cases}$$

- a. Compute c .
- b. Compute the distribution function of X .

5.6 Let X have an $Exp(0.2)$ distribution. Compute $P(X > 5)$.

5.7 The score of a student on a certain exam is represented by a number between 0 and 1. Suppose that the student passes the exam if this number is at least 0.55. Suppose we model this experiment by a continuous random variable S , the score, whose probability density function is given by

$$f(x) = \begin{cases} 4x & \text{for } 0 \leq x \leq \frac{1}{2} \\ 4 - 4x & \text{for } \frac{1}{2} \leq x \leq 1 \\ 0 & \text{elsewhere.} \end{cases}$$

- a. What is the probability that the student fails the exam?
- b. What is the score that he will obtain with a 50% chance, in other words, what is the 50th percentile of the score distribution?

5.8 \boxplus Consider Quick exercise 5.2. For another dart thrower it is given that *his* distance to the center of the disc Y is described by the following distribution function:

$$G(b) = \sqrt{\frac{b}{r}} \quad \text{for } 0 < b < r$$

and $G(b) = 0$ for $b \leq 0$, $G(b) = 1$ for $b \geq r$.

- a. Sketch the probability density function $g(y) = \frac{d}{dy}G(y)$.
- b. Is this person “better” than the person in Quick exercise 5.2?
- c. Sketch a distribution function associated to a person who in 90% of his throws hits the disc no further than $0.1 \cdot r$ of the center.

5.9 \square Suppose we choose arbitrarily a point from the square with corners at (2,1), (3,1), (2,2), and (3,2). The random variable A is the area of the triangle with its corners at (2,1), (3,1) and the chosen point (see Figure 5.9).

- a. What is the largest area A that can occur, and what is the set of points for which $A \leq 1/4$?

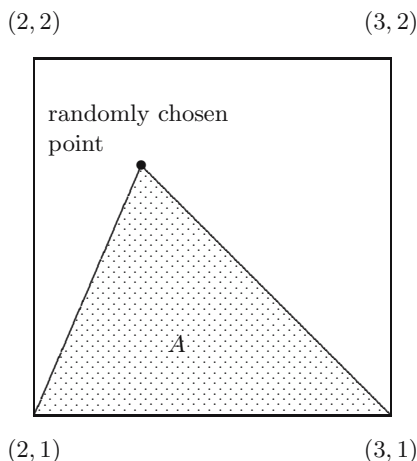


Fig. 5.9. A triangle in a square.

- b. Determine the distribution function F of A .
- c. Determine the probability density function f of A .

5.10 Consider again the chemical reactor example with parameter $\lambda = 0.5$. We saw in Section 5.6 that 10% of the particles stay in the vessel no longer than about 12 seconds—while the mean residence time is 2 minutes. Which percentage of the particles stay no longer than 2 minutes in the vessel?

5.11 Compute the median of an $Exp(\lambda)$ distribution.

5.12 \square Compute the median of a $Par(1)$ distribution.

5.13 \boxplus We consider a random variable Z with a standard normal distribution.

- a. Show why the symmetry of the probability density function ϕ of Z implies that for any a one has $\Phi(-a) = 1 - \Phi(a)$.
- b. Use this to compute $P(Z \leq -2)$.

5.14 Determine the 10th percentile of a standard normal distribution.