Chapter 5

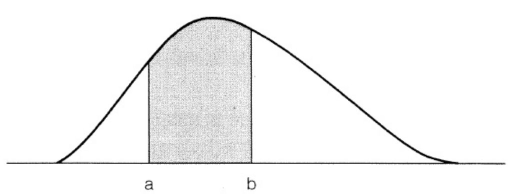
Continuous Distributions

5.1 Density Functions

In many situations random variables can take any value on the real line or in a subset of the real line such as the nonnegative numbers or the interval $[0, 1]$. For concrete examples, consider the height or weight of a person chosen at random or the time it takes a person to drive from Los Angeles to San Francisco. A random variable X is said to have a **continuous distribution** with **density function** f if for all $a \leq b$ we have

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx \tag{5.1}$$

Geometrically, $P(a \leq X \leq b)$ is the area under the curve f between a and b.



For the purposes of understanding and remembering formulas, it is useful to think of $f(x)$ as $P(X = x)$ even though the last event has probability zero. To explain the last remark and to prove $P(X = x) = 0$, note that taking $a = x$ and $b = x + \Delta x$ in (2.1) we have

$$P(x \leq X \leq x + \Delta x) = \int_x^{x+\Delta x} f(y)\, dy \approx f(x)\Delta x$$

when Δx is small. Letting $\Delta x \to 0$, we see that $P(X = x) = 0$, but $f(x)$ tells us how likely it is for X to be near x. That is,

$$\frac{P(x \le X \le x + \Delta x)}{\Delta x} \approx f(x)$$

In order for $P(a \le X \le b)$ to be nonnegative for all a and b and for $P(-\infty < X < \infty) = 1$ we must have

$$f(x) \ge 0 \quad \text{and} \quad \int f(x)\,dx = 1 \tag{5.2}$$

Here, and in what follows, if the limits of integration are not specified, the integration is over all values of x from $-\infty$ to ∞. Any function f that satisfies (5.2) is said to be a **density function**. Some important density functions are:

Example 5.1. Uniform distribution. Given $a < b$ we define

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

The idea here is that we are picking a value "at random" from (a, b). That is, values outside the interval are impossible, and all those inside have the same probability (density).

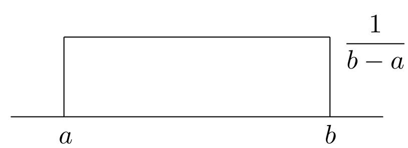


Figure 5.1: Density function for the uniform on $[0, 1]$.

If we set $f(x) = c$ when $a < x < b$ and 0 otherwise then

$$\int f(x)\,dx = \int_a^b c\,dx = c(b-a)$$

So we have to pick $c = 1/(b-a)$ to make the integral 1. The most important special case occurs when $a = 0$ and $b = 1$. Random numbers generated by a computer are typically uniformly distributed on $(0, 1)$. Another case that comes up in applications is $a = -1/2$ and $b = 1/2$. If we take a measurement and round it off to the nearest integer then it is reasonable to assume that the "round-off error" is uniformly distributed on $(-1/2, 1/2)$.

Example 5.2. Exponential distribution. Given $\lambda > 0$ we define

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

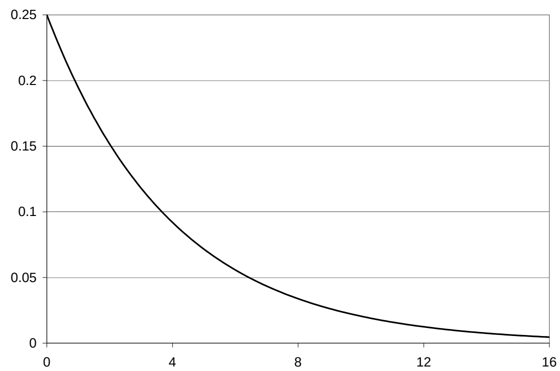


Figure 5.2: Exponential density with $\lambda = 0.25$.

To check that this is a density function, we note that

$$\int_0^\infty \lambda e^{-\lambda x}\, dx = -e^{-\lambda x}\Big|_0^\infty = 0 - (-1) = 1$$

Exponentially distributed random variables often come up as waiting times between events; for example, the arrival times of customers at a bank or ice cream shop. Sometimes we will indicate that X has an exponential distribution with parameter λ by writing $X = \text{exponential}(\lambda)$.

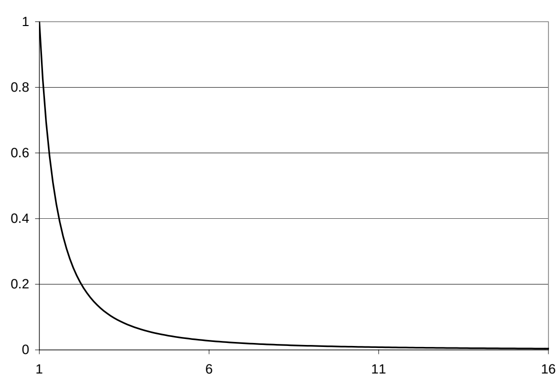


Figure 5.3: Power law density with $\rho = 2$.

Example 5.3. Power laws.

$$f(x) = \begin{cases} (\rho - 1)x^{-\rho} & x \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

Here $\rho > 1$ is a parameter that governs how fast the probabilities go to 0 at ∞.

To check that this is a density function, we note that

$$\int_1^\infty (\rho-1)x^{-\rho}\,dx = -x^{-(\rho-1)}\Big|_1^\infty = 0-(-1) = 1$$

These distributions are often used in situations where $P(X > x)$ does not go to 0 very fast as $x \to \infty$. For example, the Italian economist Pareto used them to describe the distribution of family incomes.

A fourth example, and perhaps the most important distribution of all, is the normal distribution. However, because some treatments might skip this chapter, we delay its consideration until Section 6.4.

Expected value

Given a discrete random variable X and a function $r(x)$, the expected value of $r(X)$ is defined by

$$Er(X) = \sum_x r(x)P(X = x).$$

To define the expected value for a continuous random variables, we replace the probability function by the density function and the sum by an integral.

$$Er(X) = \int r(x)f(x)\,dx \tag{5.3}$$

As in section 1.6, if $r(x) = x^k$, EX^k is called the kth moment of X and the variance m is defined by

$$\text{var}\,(X) = E(X-EX)^2 = E(X^2)-(EX)^2$$

Example 5.4. Uniform distribution. Suppose X has density function $f(x) = 1/(b-a)$ for $a < x < b$ and 0 otherwise. Then

$$EX = \frac{a+b}{2} \qquad \text{var}\,(X) = \frac{(b-a)^2}{12} \tag{5.4}$$

Notice that $(a+b)/2$ is the midpoint of the interval and hence is the natural choice for the average value of X. The variance only depends on the length of the interval not its location.

We begin with the case $a = 0$, $b = 1$.

$$EX = \int_0^1 x\,dx = \frac{x^2}{2}\Big|_0^1 = 1/2$$
$$E(X^2) = \int_0^1 x^2\,dx = \frac{x^3}{3}\Big|_0^1 = 1/3$$
$$\text{var}\,(X) = E(X^2)-(EX)^2 = (1/3)-(1/2)^2 = 1/12$$

To extend to the general case we recall from Section 1.6 that if $Y = c + dX$ then

$$E(Y) = c + dEX \qquad \text{var}\,(Y) = d^2\,\text{var}\,(X) \tag{5.5}$$

Taking $c = a$ and $d = b - a$,

$$EY = a + \frac{b-a}{2} = \frac{a+b}{2}$$
$$\text{var}\,(Y) = \frac{(b-a)^2}{12}$$

Example 5.5. Exponential distribution. Suppose X has density function $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$ and 0 otherwise.

$$EX = 1/\lambda \qquad \text{var}\,(X) = 1/\lambda^2 \tag{5.6}$$

To explain the form of the answers, we note that if Y is exponential(1) then $X = Y/\lambda$ is exponential(λ), and then use (5.5) to conclude $EX = EY/\lambda$ and $\text{var}\,(X) = \text{var}\,(Y)/\lambda^2$. Because the mean is inversely proportional to λ, λ is sometimes called the rate.

To compute the moments we need the integration by parts formula:

$$\int_a^b g(x)h'(x)\,dx = g(x)h(x)|_a^b - \int_a^b g'(x)h(x)\,dx \tag{5.7}$$

Integrating by parts with $g(x) = x$, $h'(x) = \lambda e^{-\lambda x}$, so $g'(x) = 1$ and $h(x) = -e^{-\lambda x}$.

$$\begin{aligned} EX &= \int_0^\infty x\,\lambda e^{-\lambda x}\,dx \\ &= -xe^{-\lambda x}|_0^\infty + \int_0^\infty e^{-\lambda x}\,dx \\ &= 0 + 1/\lambda \end{aligned}$$

To check the last step note that $-xe^{-\lambda x} = 0$ when $x = 0$ and when $x \to \infty$ $-xe^{-\lambda x} \to 0$ since the exponential decreases to 0 much faster than x grows.

To compute $E(X^2)$ we integrate by parts with $g(x) = x^2$, $h'(x) = \lambda e^{-\lambda x}$, so $g'(x) = 2x$ and $h(x) = -e^{-\lambda x}$.

$$\begin{aligned} E(X^2) &= \int_0^\infty x^2\,\lambda e^{-\lambda x}\,dx \\ &= -x^2 e^{-\lambda x}|_0^\infty + \int_0^\infty 2xe^{-\lambda x}\,dx \\ &= 0 + \frac{2}{\lambda}\int_0^\infty x\lambda e^{-\lambda x}\,dx = \frac{2}{\lambda^2} \end{aligned}$$

by the result for EX. Combining the last two results:

$$\text{var}\,(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

Example 5.6. Power laws. Let $\rho > 1$ and $f(x) = (\rho - 1)x^{-\rho}$ for $x \geq 1$. with $f(x) = 0$ otherwise. To compute EX

$$EX = \int_1^\infty x(\rho - 1)x^{-\rho}\, dx = \frac{\rho - 1}{\rho - 2}$$

if $\rho > 2$. $EX = \infty$ if $1 < \rho \leq 2$. The second moment

$$E(X^2) = \int_1^\infty x^2(\rho - 1)x^{-\rho}\, dx = \frac{\rho - 1}{\rho - 3}$$

if $\rho > 3$. $E(X^2) = \infty$ if $1 < \rho \leq 2$. If $\rho > 3$

$$\operatorname{var}(X) = \frac{\rho - 1}{\rho - 3} - \left(\frac{\rho - 1}{\rho - 2}\right)^2 = \frac{\rho - 1}{(\rho - 2)(\rho - 3)}$$

If, for example, $\rho = 4$, $EX = 3/2$, $E(X^2) = 3$ and $\operatorname{var}(X) = 3 - (3/2)^2 = 3/4$.

5.2 Distribution Functions

Any random variable (discrete, continuous, or in between) has a *distribution function* defined by $F(x) = P(X \le x)$. If X has a density function $f(x)$ then

$$F(x) = P(-\infty < X \le x) = \int_{-\infty}^{x} f(y)\,dy$$

That is, F is an antiderivative of f, and a special one $F(x) \to 0$ as $x \to -\infty$ and $F(x) \to 1$ as $x \to \infty$.

One of the reasons for computing the distribution function is explained by the next formula. If $a < b$ then $\{X \le b\} = \{X \le a\} \cup \{a < X \le b\}$ with the two sets on the right-hand side disjoint so

$$P(X \le b) = P(X \le a) + P(a < X \le b)$$

or, rearranging,

$$P(a < X \le b) = P(X \le b) - P(X \le a) = F(b) - F(a) \tag{5.8}$$

The last formula is valid for any random variable. When X has density function f, it says that

$$\int_a^b f(x)\,dx = F(b) - F(a)$$

i.e., the integral can be evaluated by taking the difference of any antiderivative at the two endpoints.

To see what distribution functions look like, and to explain the use of (5.8), we return to our examples.

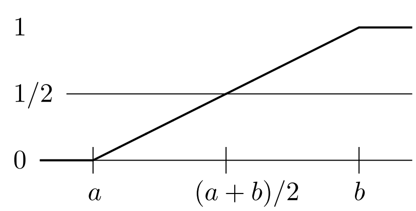


Figure 5.4: Distribution function for the uniform on $[a, b]$.

Example 5.7. Uniform distribution. $f(x) = 1/(b-a)$ for $a < x < b$.

$$F(x) = \begin{cases} 0 & x \le a \\ (x-a)/(b-a) & a \le x \le b \\ 1 & x \ge b \end{cases}$$

To check this, note that $P(a < X < b) = 1$ so $P(X \le x) = 1$ when $x \ge b$ and $P(X \le x) = 0$ when $x \le a$. For $a \le x \le b$ we compute

$$P(X \le x) = \int_{-\infty}^{x} f(y)\,dy = \int_a^x \frac{1}{b-a}\,dy = \frac{x-a}{b-a}$$

In the most important special case $a = 0$, $b = 1$ we have $F(x) = x$ for $0 \le x \le 1$.

Example 5.8. Exponential distribution. $f(x) = \lambda e^{-\lambda x}$ for $x \ge 0$.

$$F(x) = \begin{cases} 0 & x \le 0 \\ 1 - e^{-\lambda x} & x \ge 0 \end{cases}$$

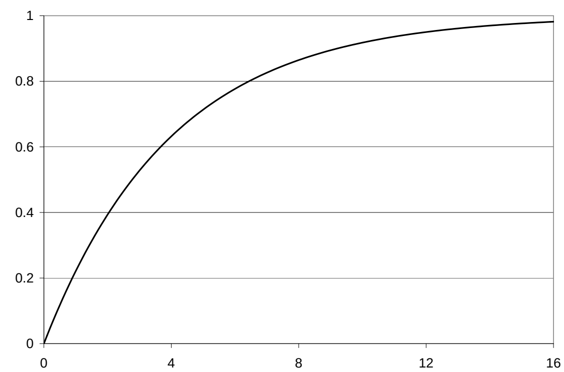


Figure 5.5: Exponential distribution function, $\lambda = 0.25$.

The first line of the answer is easy to see. Since $P(X > 0) = 1$ we have $P(X \le x) = 0$ for $x \le 0$. For $x \ge 0$ we compute

$$\begin{aligned} P(X \le x) = \int_{-\infty}^{x} f(y)\,dy &= \int_0^x \lambda e^{-\lambda y}\,dy \\ &= -e^{-\lambda y}\Big|_0^x = -e^{-\lambda x} - (-1) \end{aligned}$$

Suppose X has an exponential distribution with parameter λ. If $t \ge 0$ then $P(X > t) = 1 - P(X \le t) = 1 - F(t) = e^{-\lambda t}$, so if $s \ge 0$ then

$$P(T > t + s | T > t) = \frac{P(T > t + s)}{P(T > t)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = P(T > s)$$

This is the **lack of memory property** of the exponential distribution. Given that you have been waiting t units of time, the probability you must wait an additional s units of time is the same as if you had not been waiting at all.

Example 5.9. Power laws. $f(x) = (\rho - 1)x^{-\rho}$ for $x \ge 1$ where $\rho > 1$.

$$F(x) = \begin{cases} 0 & x \le 1 \\ 1 - x^{-(\rho-1)} & x \ge 1 \end{cases}$$

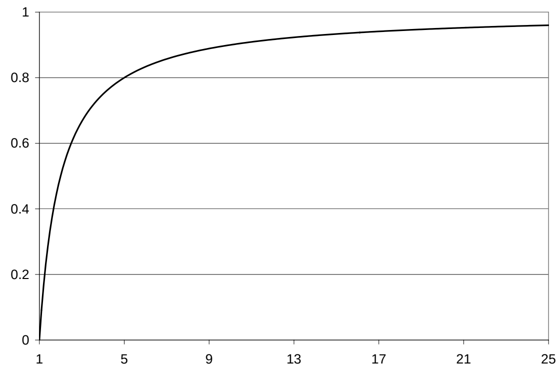


Figure 5.6: Power law distribution function, $\rho = 2$.

The first line of the answer is easy to see. Since $P(X > 1) = 1$, we have $P(X \le x) = 0$ for $x \le 1$. For $x \ge 1$ we compute

$$\begin{aligned} P(X \le x) = \int_{-\infty}^{x} f(y)\,dy &= \int_1^x (\rho-1)y^{-\rho}\,dy \\ &= -y^{-(\rho-1)}\Big|_1^x = 1 - x^{-(\rho-1)} \end{aligned}$$

To illustrate the use of (5.8) we note that if $\rho = 3$ then

$$P(2 < X \le 4) = (1 - 4^{-2}) - (1 - 2^{-2}) = \frac{1}{4} - \frac{1}{16} = \frac{3}{16}$$

Distribution functions are somewhat messier in the discrete case.

Example 5.10. Binomial(3,1/2). Flip three coins and let X be the number of heads that we see. The probability function is given by

x	0	1	2	3
$P(X = x)$	1/8	3/8	3/8	1/8

In this case the distribution function is:

$$F(x) = \begin{cases} 0 & x < 0 \\ 1/8 & 0 \le x < 1 \\ 4/8 & 1 \le x < 2 \\ 7/8 & 2 \le x < 3 \\ 1 & 3 \le x \end{cases}$$

To check this, note for example that for $1 \le x < 2$, $P(X \le x) = P(X \in \{0, 1\}) = 1/8 + 3/8$. The reader should note that F is discontinuous at each possible value of X and the height of the jump there is $P(X = x)$. The little black dots in Figure 5.7 are there to indicate that at 0 the value is 1/8, at 1 it is 1/2, etc.

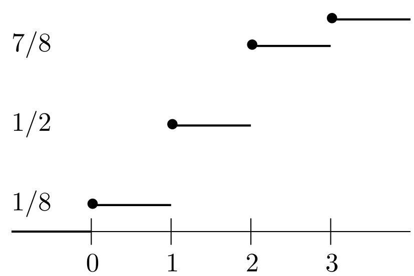


Figure 5.7: Distribution function for Binomial(3,1/2).

Theorem 5.1. *All distribution functions have the following properties*

(i) If $x_1 < x_2$ *then* $F(x_1) \le F(x_2)$ *i.e.,* F *is nondecreasing.*

(ii) $\lim_{x \to -\infty} F(x) = 0$

(iii) $\lim_{x \to \infty} F(x) = 1$

(iv) $\lim_{y \downarrow x} F(y) = F(x)$*, i.e.,* F *is continuous from the right.*

(v) $\lim_{y \uparrow x} F(y) = P(X < x)$

(vi) $\lim_{y \downarrow x} F(y) - \lim_{y \uparrow x} F(y) = P(X = x)$*,*
i.e., the jump in F *at* x *is equal to* $P(X = x)$*.*

Proof. To prove (i) we note that $\{X \le x_1\} \subset \{X \le x_2\}$, so (1.4) implies $F(x_1) = P(X \le x_1) \le P(X \le x_2) = F(x_2)$.

For (ii), we note that $\{X \le x\} \downarrow \emptyset$ as $x \downarrow -\infty$ (here $\downarrow$ is short for "decreases and converges to"), so (1.5) implies that $P(X \le x) \downarrow P(\emptyset) = 0$.

The argument for (iii) is similar $\{X \le x\} \uparrow \Omega$ as $x \uparrow \infty$ (here $\uparrow$ is short for "increases and converges to"), so (1.5) implies that $P(X \le x) \uparrow P(\Omega) = 1$.

For (iv), we note that if $y \downarrow x$ then $\{X \le y\} \downarrow \{X \le x\}$, so (1.5) implies that $P(X \le y) \downarrow P(X \le x)$.

The argument for (v) is similar. If $y \uparrow x$ then $\{X \le y\} \uparrow \{X < x\}$ since $\{X = x\} \not\subset \{X \le y\}$ when $y < x$. Using (1.5) now, (v) follows.

Subtracting (v) from (iv) gives (vi). □

Two useful transformations

The first result can often be used to reduce a general continuous distribution to the special case of a uniform.

Theorem 5.2. *Suppose X has a continuous distribution. Then $Y = F(X)$ is uniform on $(0,1)$.*

Proof. Even though F may not be strictly increasing, we can define an inverse of F by

$$F^{-1}(y) = \min\{x : F(x) \geq y\}$$

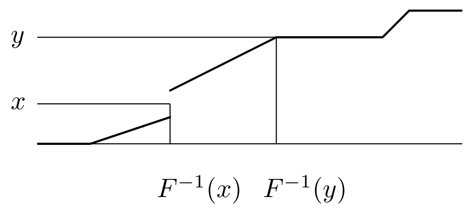


Figure 5.8: Inverse of a distribution function.

Using this definition of F^{-1}, we have

$$P(Y \leq y) = P(X \leq F^{-1}(y)) = F(F^{-1}(y)) = y$$

the last equality holding since F is continuous. □

This is the key to many results in nonparametric statistics. For example, suppose we have a sample of 10 men's heights and 10 women's heights. To test the hypothesis that men and women have the same height distribution, we can look at the ranks of the men's heights in the overall sample of size 20. For example, these might be 1, 2, 3, 4, 6, 8, 9, 11, 13, and 14. Since applying the distribution function to the data points does not change the ranks, Theorem 5.2 implies that the distribution of the rank sum does not depend on the underlying distribution.

Reversing the ideas in the proof of Theorem 5.2, we get a result that is useful to construct random variables with a specified distribution.

Theorem 5.3. *Suppose U has a uniform distribution on (0,1). Then $Y = F^{-1}(U)$ has distribution function F.*

Proof. The definition of F^{-1} was chosen so that if $0 < x < 1$ then

$$F^{-1}(y) \leq x \text{ if and only if } F(x) \leq y$$

and this holds for any distribution function F. Taking $y = U$, it follows that

$$P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$$

since $P(U \leq u) = u$. □

For a concrete example, suppose we want to construct an exponential distribution with parameter λ. Setting $1-e^{-\lambda x} = y$ and solving gives $-\ln(1-y)/\lambda = x$. So if U is uniform on $(0,1)$ then $-\ln(1-U)/\lambda$ has the desired exponential distribution. Of course since $1-U$ is uniform on $(0,1)$ we could also use $-\ln(U)/\lambda$. In the case of a power law, setting $1-x^{-(\rho-1)} = y$ and solving gives $(1-y)^{-1/(\rho-1)} = x$. So if U is uniform on $(0,1)$ then $U^{-1/(\rho-1)}$ has the desired power law distribution.

Implement in Matlab

Medians

Intuitively, the median is the place where $F(x)$ crosses $1/2$. The precise definition we are about to give is complicated by the fact that $\{x : F(x) = 1/2\}$ may be empty or contain more than one point.

m is a **median** for F if $P(X \le m) \ge 1/2$ and $P(X \ge m) \ge 1/2$.

We begin with our three favorite examples.

Example 5.11. Uniform distribution. Suppose X has density $1/(b-a)$ for $a \le x \le b$. As we computed in Example 5.7 the distribution function is $(x-a)/(b-a)$ for $a \le x \le b$. The computation of the median is illustrated in Figure 5.7. To find the median,we set $(x-a)/(b-a) = 1/2$, i.e., $2x - 2a = b - a$ or solving we have $x = (b+a)/2$. To see that this is the only median, we observe that if $m < (a+b)/2$ then $P(X \le m) < 1/2$ while if $m > (a+b)/2$ then $P(X \ge m) < 1/2$. In this case the median is equal to the mean, but this is a rare occurrence.

Example 5.12. Exponential distribution. Suppose X has density $\lambda e^{-\lambda x}$ for $x \ge 0$. As we computed in Example 5.8, the distribution function is $F(x) = 1 - e^{-\lambda x}$. To find the median we set $P(X \le m) = 1/2$, i.e., $1 - e^{-\lambda m} = 1/2$, and solve to find $m = (\ln 2)/\lambda$, compared to the mean $1/\lambda$.

In the context of radioactive decay, which is commonly modeled with an exponential distribution, the median is sometimes called the **half-life**, since half of the particles will have broken down by that time. One reason for interest in the half-life is that

$$P(X > k \ln 2/\lambda) = e^{-k \ln 2} = 2^{-k}$$

or in words, after k half-lives only $1/2^k$ particles remain radioactive.

Example 5.13. Power laws. Suppose X has density $(\rho-1)x^{-\rho}$ for $x \ge 1$, where $\rho > 1$. As we computed in Example 5.9 the distribution function is $1 - x^{-\rho}$. To find the median we set $1 - m^{-\rho} = 1/2$, i.e., $1/2 = m^{-\rho}$ and solving gives $m = 2^{1/\rho}$. This contrasts to the mean $(\rho-1)/(\rho-2)$ which is finite for $\rho > 2$. For a concrete example note that when $\rho = 4$ the mean is $3/2$ while the median is $2^{1/4} = 1.189$.

We now turn to unusual cases where there may be no solution to $P(X \le x) = 1/2$ or more than one.

Example 5.14. Multiple solutions: Binomial(3,1/2). The distribution function was computed in Example 5.10.

$$F(x) = \begin{cases} 0 & x < 0 \\ 1/8 & 0 \le x < 1 \\ 1/2 & 1 \le x < 2 \\ 7/8 & 2 \le x < 3 \\ 1 & 3 \le x \end{cases}$$

If $1 \le m \le 2$ then $P(X \le m) \ge 1/2$ and $P(X \ge m) \ge 1/2$ so the set of medians is $[1,2]$. For a picture see Figure 5.7.

Example 5.15. No solution: Uniform on 1,2,3. Suppose X takes values 1, 2, 3 with probability 1/3 each. The distribution function is

$$F(x) = \begin{cases} 0 & x < 1 \\ 1/3 & 1 \le x < 2 \\ 2/3 & 2 \le x < 3 \\ 1 & 3 \le x \end{cases}$$

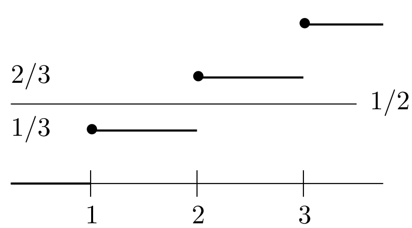


Figure 5.9: Distribution function for Uniform on $\{1,2,3\}$.

To check that 2 is a median, we note that

$$\begin{aligned} P(X \le 2) &= P(X \in \{1,2\}) = 2/3 \\ P(X \ge 2) &= P(X \in \{2,3\}) = 2/3 \end{aligned}$$

This is the only median, since if $x < 2$ then $P(X \le x) \le P(X < 2) \le 1/3$ and if $x > 2$ then $P(X \ge x) \le P(X > 2) = 1/3$.

5.3 Functions of Random Variables

In this section we will answer the question: If X has density function f and $Y = r(X)$, then what is the density function for Y? Before proving a general result, we will consider an example:

Example 5.16. Suppose X has an exponential distribution with parameter λ. What is the distribution of $Y = X^2$?

To solve this problem we will use the distribution function. First we recall from Example 5.8 that $P(X \le x) = 1 - e^{-\lambda x}$ so if $y \ge 0$ then

$$P(Y \le y) = P(X^2 \le y) = P(X \le \sqrt{y}) = 1 - e^{-\lambda y^{1/2}}$$

Differentiating, we see that the density function of Y is given by

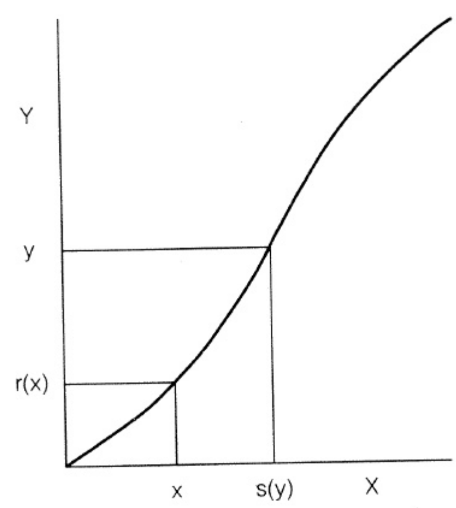
$$f_Y(y) = \frac{d}{dy} P(Y \le y) = \frac{\lambda y^{-1/2}}{2} e^{-\lambda y^{1/2}} \qquad \text{for } y \ge 0$$

and 0 otherwise.

Generalizing from the last example, we get

Theorem 5.4. *Suppose X has density f and $P(a < X < b) = 1$. Let $Y = r(X)$. Suppose $r : (a,b) \to (\alpha, \beta)$ is continuous and strictly increasing, and let $s : (\alpha,\beta) \to (a,b)$ be the inverse of r. Then Y has density*

$$g(y) = f(s(y))s'(y) \qquad \textit{for } y \in (\alpha, \beta) \tag{5.9}$$



Before proving this, let's see how it applies to the last example. There X has density $f(x) = \lambda e^{-\lambda x}$ for $x \ge 0$ so we can take $a = 0$ and $b = \infty$. The function $r(x) = x^2$ is indeed continuous and strictly increasing on $(0, \infty)$. To find the inverse function we set $y = x^2$ and solve to get $x = y^{1/2}$ so $s(y) = y^{1/2}$. Differentiating, we have $s'(y) = y^{-1/2}/2$ and plugging into the formula, we have

$$g(y) = \lambda e^{-\lambda y^{1/2}} \cdot y^{-1/2}/2 \qquad \text{for } y > 0$$

Proof. If $y \in (\alpha, \beta)$ then

$$P(Y \le y) = P(r(X) \le y) = P(X \le s(y))$$

since r is increasing and s is its inverse.

Writing $F(x)$ for $P(X \le x)$ and differentiating with respect to y now gives

$$g(y) = \frac{d}{dy} P(Y \le y) = \frac{d}{dy} F(s(y)) = F'(s(y))s'(y) = f(s(y))s'(y)$$

by the chain rule. □

For our next example, we will consider a special case of Theorem **??**.

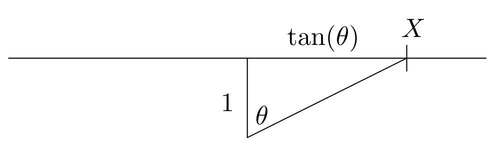
Example 5.17. Suppose X has an exponential distribution with parameter 3. That is, X has density function $3e^{-3x}$ for $x \ge 0$. Find the distribution function of $Y = 1 - e^{-3X}$.

Here, $r(x) = 1 - e^{-3x}$ is increasing on $(0, \infty)$, $\alpha = r(0) = 0$, and $\beta = r(\infty) = 1$. To find the inverse function we set $y = 1 - e^{-3x}$ and solve to get $s(y) = (-1/3)\ln(1 - y)$. Differentiating, we have $s'(y) = -(-1/3)/(1 - y)$. So plugging into (3.1), the density function of Y is

$$f(s(y))s'(y) = 3e^{\ln(1-y)} \cdot \frac{1/3}{(1-y)} = 1$$

for $0 < y < 1$. That is, Y is uniform on $(0, 1)$.

Example 5.18. Cauchy distribution. A drunk standing one foot from a wall shines a flashlight at a random angle that is uniformly distributed between $-\pi/2$ and $\pi/2$. Find the density function of the place X where the light hits the wall.



The angle Θ is uniformly distributed on $[-\pi/2, \pi/2]$ and has density $1/\pi$. As you can see from the picture $r(\theta) = \tan(\theta)$. The inverse function $s(x) = \tan^{-1}(x)$, has $s'(x) = 1/(1 + x)$ so using 5.9 X has density function

$$\frac{1}{\pi} \cdot \frac{1}{1 + x^2}$$

This is the Cauchy density. Its median is 0 but its mean does not exist since

$$E|X| = \int \frac{|x|}{\pi(1+x^2)}\,dx = \infty$$

To check the last conclusion note that the integrand is $\approx 1/|x|$ when $|x|$ is large.

Example 5.19. How not to water your lawn. The head of a lawn sprinkler, which is a metal rod with a line of small holes in it, revolves back and forth so that drops of water shoot out at angles between 0 and $\pi/2$ radians (i.e., between 0 and 90 degrees). If we use x to denote the distance from the sprinkler and y the height off the ground, then a drop of water released at angle θ with velocity v_0 will follow a trajectory

$$x(t) = (v_0 \cos\theta)t \qquad y(t) = (v_0 \sin\theta)t - gt^2/2$$

where g is the gravitational constant, 32 ft/sec^2. The drop lands when $y(t_0) = 0$ that is, at time $t_0 = (2v_0 \sin\theta)/g$. At this time

$$x(t_0) = \frac{2v_0^2}{g}\sin\theta\cos\theta = \frac{v_0^2}{g}\sin(2\theta)$$

If we assume that the sprinkler moves evenly back and forth between 0 and $\pi/2$, it will spend an equal amount of time at each angle. Letting $K = v_0^2/g$, this leads us to the following question:

If Θ is uniform on $[0, \pi/2]$ then what is the distribution of $Z = K\sin(2\Theta)$?

The first difficulty we must confront when solving this problem is that $\sin(2x)$ is increasing on $[0, \pi/4]$ and decreasing on $[\pi/4, \pi/2]$. The solution to this problem is simple, however. The function $\sin(2x)$ is symmetric about $\pi/4$, so if we let X be uniform on $[0, \pi/4]$ then $Z = K\sin(2\Theta)$ and $Y = K\sin(2X)$ have the same distribution. To apply (5.9) we let $r(x) = K\sin(2x)$ and solve $y = K\sin(2x)$ to get $s(y) = (1/2)\sin^{-1}(y/K)$. Plugging into (5.9) and recalling

$$\frac{d}{dx}\sin^{-1}(x) = \frac{1}{\sqrt{1-x^2}}$$

we see that Y has density function

$$f(s(y))s'(y) = \frac{4}{\pi}\cdot\frac{1}{2\sqrt{1-y^2/K^2}}\cdot\frac{1}{K} = \frac{2}{\pi\sqrt{K^2-y^2}}$$

when $0 < y < K$ and 0 otherwise. The title of this example comes from the fact that the density function goes to ∞ as $y \to K$ so the lawn gets very soggy at the edge of the sprinkler's range. This is due to the fact that $s'(K) = \infty$, which in turn is caused by $r'(\pi/4) = 0$.

5.4 Joint Distributions

Two random variables are said to have **joint density function** f if for any $A \subset \mathbf{R}^2$

$$P((X,Y) \in A) = \iint_A f(x,y)\,dx\,dy \tag{5.10}$$

where $f(x,y) \geq 0$ and $\iint f(x,y)\,dx\,dy = 1$.

In words, we find the probability that (X,Y) lies in A by integrating f over A. As we will see a number of times below, it is useful to think of $f(x,y)$ as $P(X = x, Y = y)$ even though the last event has probability 0. As in Section 5.1, the precise interpretation of $f(x,y)$ is

$$\begin{aligned} P(x \leq X \leq x + \Delta x,\, y \leq Y \leq y + \Delta y) &= \int_x^{x+\Delta x} \int_y^{y+\Delta y} f(u,v)\,dv\,du \\ &\approx f(x,y)\Delta x \Delta y \end{aligned}$$

when Δx and Δy are small, so $f(x,y)$ indicates how likely it is for (X,Y) to be near (x,y).

For a concrete example of a joint density function, consider

Example 5.20.

$$f(x,y) = \begin{cases} e^{-y} & 0 < x < y < \infty \\ 0 & \textit{otherwise} \end{cases}$$

The story behind this example will be told later in Example 5.28. To check that f is a density function, we observe that

$$\int_0^\infty \int_0^y e^{-y}\,dx\,dy = \int_0^\infty ye^{-y}\,dy$$

and integrating by parts with $g(y) = y$, $h'(y) = e^{-y}$ (so $g'(y) = 1$, $h(y) = -e^{-y}$).

$$\int_0^\infty ye^{-y}\,dy = -ye^{-y}\Big|_0^\infty + \int_0^\infty e^{-y}dy = 0 + (-e^{-y})\Big|_0^\infty = 1$$

To illustrate the use of (5.10) we will now compute $P(X \leq 1)$, which can be written as $P((X,Y) \in A)$ where $A = \{(x,y) : x \leq 1\}$. The formula in (5.10) tells us that we find $P((X,Y) \in A)$ by integrating the joint density over A. However, the joint density is only positive on $B = \{(x,y) : 0 < x < y < \infty\}$ so we only need to integrate over $A \cap B = \{(x,y) : 0 < x \leq 1, x < y\}$, and doing this we find

$$P(X \leq 1) = \int_0^1 \int_x^\infty e^{-y}\,dy\,dx$$

To evaluate the double integral we begin by observing that

$$\int_x^\infty e^{-y}\,dy = (-e^{-y})\Big|_x^\infty = 0 - (-e^{-x}) = e^{-x}$$

so $P(X < 1) = \int_0^1 e^{-x}\,dx = (-e^{-x})\big|_0^1 = 1 - e^{-1}$.

Example 5.21. Uniform on a ball. Pick a point at random from the ball $B = \{(x, y) : x^2 + y^2 \leq 1\}$. By "at random from B" we mean that a choice outside of B is impossible and that all the points in B should be equally likely. In terms of the joint density this means that $f(x, y) = 0$ when $(x, y) \notin B$ and there is a constant $c > 0$ so that $f(x, y) = c$ when $(x, y) \in B$.

Our $f(x, y) \geq 0$. To make the integral of f equal to 1, we have to choose c appropriately. Now,

$$\iint f(x, y)\, dx\, dy = \iint_B c\, dx\, dy = c \text{ (area of } B) = c\pi$$

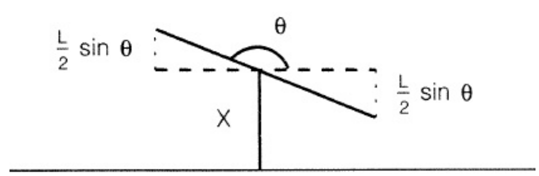
So we choose $c = 1/\pi$ to make the integral 1 and define

$$f(x, y) = \begin{cases} 1/\pi & x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

The arguments that led to the last conclusion generalize easily to show that if we pick a point "at random" from a set S with area a then

$$f(x, y) = \begin{cases} 1/a & (x, y) \in S \\ 0 & \text{otherwise} \end{cases} \tag{5.11}$$

Example 5.22. Buffon's needle. A floor consists of boards of width 1. If we drop a needle of length $L \leq 1$ on the floor, what is the probability it will touch one of the cracks (i.e., the small spaces between the boards)? To make the question simpler to answer, we assume that the needle and the cracks have width zero.



Let X be the distance from the center of the needle to the nearest crack and Θ be the angle $\in [0, \pi)$ that the top half of the needle makes with the crack. (We make this choice to have $\sin \Theta > 0$.) We assume that all the ways the needle can land are equally likely, that is, the joint distribution of (X, Θ) is

$$f(x, \theta) = \begin{cases} 2/\pi & \text{if } x \in [0, 1/2),\ \theta \in [0, \pi) \\ 0 & \text{otherwise} \end{cases}$$

The formula for the joint density follows from (5.11). We are picking a point "at random" from a set S with area $\pi/2$, so the joint density is $2/\pi$ on S.

By drawing a picture (like the one above), one sees that the needle touches the crack if and only if $(L/2)\sin\Theta \geq X$. (5.10) tells us that the probability of this event is obtained by integrating the joint density over

$$A = \{(x,\theta) \in [0,1/2) \times [0,\pi) : x \leq (L/2)\sin\theta\}$$

so the probability we seek is

$$\begin{aligned}\iint_A f(x,\theta)\,dx\,d\theta &= \int_0^\pi \int_0^{(L/2)\sin\theta} \frac{2}{\pi}\,dx\,d\theta \\ &= \frac{2}{\pi}\int_0^\pi \frac{L}{2}\sin\theta\,d\theta = \frac{L}{\pi}(-\cos\theta)\Big|_0^\pi = 2L/\pi\end{aligned}$$

Buffon wanted to use this as a method of estimating π. Taking $L = 1/2$ and performing the experiment 10,000 times on a computer, we found that 1 over the fraction of times the needle hit the crack was 3.2310, 3.1368, and 3.0893 in the three times we tried this. We will see in Chapter 5 that these numbers are typical outcomes and that to compute π to 4 decimal places would require about 10^8 (or 100 million) tosses.

Remark. Before leaving the subject of joint densities, we would like to make one remark that will be useful later. If X and Y have joint density $f(x,y)$ then $P(X = Y) = 0$. To see this, we observe that $\iint_A f(x,y)\,dx\,dy$ is the volume of the region over A underneath the graph of f, but this volume is 0 if A is the line $x = y$.

Joint distribution function

The joint distribution of two random variables is occasionally described by giving the **joint distribution function**:

$$F(x,y) = P(X \leq x, Y \leq y)$$

The next example illustrates this notion but also shows that sometimes the density function is easier to write down.

Example 5.23. Suppose (X,Y) is uniformly distributed over the square $\{(x,y) : 0 < x < 1, 0 < y < 1\}$. That is,

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

Here, we are picking a point "at random" from a set with area 1, so the formula follows from (5.11).

By patiently considering the possible cases, one finds that

$$F(x,y) = \begin{cases} 0 & \text{if } x < 0 \text{ or } y < 0 \\ xy & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ x & \text{if } 0 \leq x \leq 1 \text{ and } y > 1 \\ y & \text{if } x > 1 \text{ and } 0 \leq y \leq 1 \\ 1 & \text{if } x > 1 \text{ and } y > 1 \end{cases}$$

The answer is probably easier to understand in a picture:

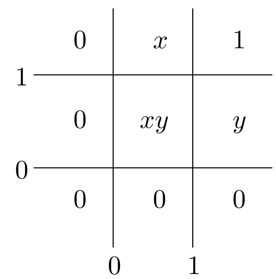


Figure 5.10: Distribution function for the uniform on $[0,1]^2$.

The first case should be clear: If $x < 0$ or $y < 0$ then $\{X \le x, Y \le y\}$ is impossible since X and Y always lie between 0 and 1. For the second case we note that when $0 \le x \le 1$ and $0 \le y \le 1$,

$$P(X \le x, Y \le y) = \int_0^x \int_0^y 1\,dv\,du = xy$$

In the third case, since values of $Y > 1$ are impossible,

$$P(X \le x, Y \le y) = P(X \le x, Y \le 1) = x$$

by the formula for the second case. The fourth case is similar to the third, and the fifth is trivial. X and Y are always smaller than 1 so if $x > 1$ and $y > 1$ then $\{X \le x, Y \le y\}$ has probability 1.

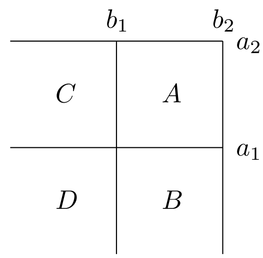
We will not use the joint distribution function in what follows. For completeness, however, we want to mention two of its important properties. The first formula is the two-dimensional generalization of $P(a < X \le b) = F(b) - F(a)$.

$$\begin{aligned} &P(a_1 < X \le b_1, a_2 < Y \le b_2) \\ &\qquad = F(b_1,b_2) - F(a_1,b_2) - F(b_1,a_2) + F(a_1,a_2) \end{aligned} \tag{5.12}$$

Proof. The reasoning we use here is much like that employed in studying the probabilities of unions in Section 1.6. By adding and subtracting the probabilities on the right, we end up with the desired area counted exactly once.

Using A as shorthand for $P((X,Y) \in A)$, etc., and consulting the picture.

$$\begin{array}{rclllll} F(b_1,b_2) & = & A & +B & +C & +D \\ -F(a_1,b_2) & = & & -B & & -D \\ -F(b_1,a_2) & = & & & -C & -D \\ F(a_1,a_2) & = & & & & +D \end{array}$$



Adding the last four equations gives the one in (5.12). □

The next formula tells us how to recover the joint density function from the joint distribution function. To motivate the formula, we recall that in one dimension $F' = f$ since $F(x) = \int_\infty^x f(u)\,du$.

$$\frac{\partial^2 F}{\partial x \partial y} = f \tag{5.13}$$

To explain why this formula is true, we note that

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\,dv\,du$$

and differentiating twice kills the two integrals. To check that (5.13) works in Example 5.23, $F(x,y) = xy$ when $0 < x < 1$ and $0 < y < 1$, so $\frac{\partial^2 F}{\partial x \partial y} = 1$ there and it is 0 otherwise.

5.5 Marginal and Conditional Distributions

In the discrete case the marginal distributions are obtained from the joint distribution by summing

$$P(X = x) = \sum_y P(X = x, Y = y) \qquad P(Y = y) = \sum_x P(X = x, Y = y)$$

In the continuous case if X and Y have joint density $f(x, y)$, then the **marginal densities** of X and Y are given by

$$f_X(x) = \int f(x, y)\, dy \qquad f_Y(y) = \int f(x, y)\, dx \tag{5.14}$$

The verbal explanation of the first formula is similar to that of the discrete case: if $X = x$ then Y will take on some value y, so to find $P(X = x)$ we integrate the joint density $f(x, y)$ over all possible values of y.

To illustrate the use of these formulas we look at Example 5.20.

Example 5.24.

$$f(x, y) = \begin{cases} e^{-y} & 0 < x < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

In this case

$$f_X(x) = \int_x^\infty e^{-y}\, dy = (-e^{-y})\big|_x^\infty = e^{-x}$$

since (5.14) tells us to integrate $f(x, y)$ over all values of y but we only have $f > 0$ when $y > x$. Similarly,

$$f_Y(y) = \int_0^y e^{-y}\, dx = ye^{-y}$$

The next result is the continuous analogue of (??):

Theorem 5.5. *Two random variables with joint density f are independent if and only if*

$$f(x, y) = f_X(x) f_Y(y)$$

that is, if the joint density is the product of the marginal densities.

We will now consider three examples that parallel the ones used in the discrete case.

Example 5.25.

$$f(x, y) = \begin{cases} e^{-y} & 0 \le x \le y < \infty \\ 0 & \text{otherwise} \end{cases}$$

We calculated the joint distribution in the previous example but we can settle the question without computation. $f(3,2) = 0$ while $f_X(3)$ and $f_Y(2)$ are both positive so

$$f(3,2) = 0 < f_X(3) f_Y(2)$$

and Theorem 5.5 implies that X and Y are not independent. In general, if the set of values where $f > 0$ is not a rectangle then X and Y are not independent.

Example 5.26.

$$f(x,y) = \begin{cases} (1+x+y)/2 & 0 \le x, y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

In this case the set where $f > 0$ is a rectangle, so the joint distribution passes the first test and we have to compute the marginal densities

$$\begin{aligned} f_X(x) &= \int_0^1 (1+x+y)/2 \, dy = \left(\frac{1+x}{2}\right) y + \frac{y^2}{4} \Big|_0^1 = \frac{x}{2} + \frac{3}{4} \\ f_Y(y) &= \frac{y}{2} + \frac{3}{4} \qquad \text{by symmetry} \end{aligned}$$

These formulas are valid for $0 \le x \le 1$ and $0 \le y \le 1$ respectively. To check independence we have to see if

$$(\star) \qquad \frac{1+x+y}{2} = \left(\frac{x}{2} + \frac{3}{4}\right) \cdot \left(\frac{y}{2} + \frac{3}{4}\right)$$

A simple way to see that $(\star)$ is wrong is simply to note that when $x = y = 0$ it says that $1/2 = 9/16$. Some readers (or instructors) may note that, strictly speaking, it is not enough for the condition to fail at one point, but the formulas here are continuous so failing at one point means they also fail near the point.

Example 5.27.

$$f(x,y) = \begin{cases} \dfrac{y^{-3/2}\cos x}{(e-2+e^{-1})} e^{\sin x - (1/2y)} & 0 < x < \pi/2, \; y > 0 \\ 0 & \text{otherwise} \end{cases}$$

In this case, the integration does not look like much fun, so we adopt another approach.

Theorem 5.6. *If $f(x,y)$ can be written as $g(x)h(y)$ then there is a constant c so that $f_X(x) = cg(x)$ and $f_Y(y) = h(y)/c$. It follows that $f(x,y) = f_X(x) f_Y(y)$ and hence X and Y are independent.*

In words, if we can write $f(x,y)$ as a product of a function of x and a function of y then these functions must be constant multiples of the marginal densities. Theorem 5.6 takes care of our example since

$$f(x,y) = \left(\frac{\cos x \, e^{\sin x}}{(e-2+e^{-1})}\right) (y^{-3/2} e^{-(1/2y)})$$

Proof. We begin by observing

$$f_X(x) = \int f(x,y)\,dy = g(x)\int h(y)\,dy$$
$$f_Y(y) = \int f(x,y)\,dx = h(y)\int g(x)\,dx$$
$$1 = \int\int f(x,y)\,dx\,dy = \int g(x)\,dx \int h(y)\,dy$$

So if we let $c = \int h(y)\,dy$ then the last equation implies $\int g(x)\,dx = 1/c$, and the first two give us $f_X(x) = cg(x)$ and $f_Y(y) = h(y)/c$. □

Conditional distributions

Introducing $f_X(x|Y = y)$ as notation for the **conditional density of** X **given** $Y = y$ (which we think of as $P(X = x|Y = y)$), we have

$$f_X(x|Y = y) = \frac{f(x,y)}{f_Y(y)} = \frac{f(x,y)}{\int f(u,y)\,du} \tag{5.15}$$

In words, we fix y, consider the joint density function as a function of x, and then divide by the integral to make it a probability density. To see how formula (5.15) works, we return to Example 5.15.

Example 5.28.

$$f(x,y) = \begin{cases} e^{-y} & 0 < x < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

In this case we have computed $f_Y(y) = ye^{-y}$ (in Example 5.24) so

$$f_X(x|Y = y) = \frac{e^{-y}}{ye^{-y}} = \frac{1}{y} \quad \text{for } 0 < x < y$$

That is, the conditional distribution is uniform on $(0, y)$. This should not be surprising since the joint density does not depend on x.

To compute the other conditional distribution we recall $f_X(x) = e^{-x}$ so

$$f_Y(y|X = x) = \frac{e^{-y}}{e^{-x}} = e^{-(y-x)} \quad \text{for } y > x$$

That is, given $X = x$, $Y - x$ is exponential with parameter 1. From this it follows that if Z_1, Z_2 are independent exponential(1) then $X = Z_1$, $Y = Z_1 + Z_2$ has the joint distribution given above. If we condition on $X = x$ then $Z_1 = x$ and $Y = x + Z_2$.

The multiplication rule says

$$P(X = x, Y = y) = P(X = x)P(Y = y|X = x)$$

Substituting in the analogous continuous quantities, we have

$$f(x,y) = f_X(x) f_Y(y|X=x) \tag{5.16}$$

The next example demonstrates the use of (5.16) to compute a joint distribution.

Example 5.29. Suppose we pick a point uniformly distributed on $(0,1)$, call it X, and then pick a point Y uniformly distributed on $(0,X)$.

To find the joint density of (X,Y) we note that

$$\begin{aligned} f_X(x) &= 1 && \text{for } 0 < x < 1 \\ f_Y(y|X=x) &= 1/x && \text{for } 0 < y < x \end{aligned}$$

So using (5.16), we have

$$f(x,y) = f_X(x) f_Y(y|X=x) = 1/x \quad \text{for } 0 < y < x < 1$$

To complete the picture we compute

$$\begin{aligned} f_Y(y) &= \int f(x,y)\,dx = \int_y^1 \frac{1}{x}\,dx = -\ln y \\ f_X(x|Y=y) &= \frac{f(x,y)}{f_Y(y)} = \frac{1/x}{-\ln y} \qquad \text{for } y < x < 1 \end{aligned}$$

Again the conditional density of X given $Y=y$ is obtained by fixing y, regarding the joint density function as a function of x, and then normalizing so that the integral is 1. The reader should note that although X is uniform on $(0,1)$ and Y is uniform on $(0,X)$, X is not uniform on $(Y,1)$ but has a greater probability of being near Y.

5.6 Exercises

Density functions

1. Suppose X has density function $f(x) = c(3 - |x|)$ when $-3 < x < 3$. What value of c makes this a density function?

2. Consider $f(x) = c(1 - x^2)$ for $-1 < x < 1$, 0 otherwise. What value of c should we take to make f a density function?

3. Suppose X has density function $6x(1 - x)$ for $0 < x < 1$ and 0 otherwise. Find (a) EX, (b) $E(X^2)$, and (c) $\text{var}\,(X)$.

4. Suppose X has density function $x^2/9$ for $0 < x < 3$ and 0 otherwise. Find (a) EX, (b) $E(X^2)$, and (c) $\text{var}\,(X)$.

5. Suppose X has density function $x^{-2/3}/21$ for $1 < x < 8$ and 0 otherwise. Find (a) EX, (b) $E(X^2)$, and (c) $\text{var}\,(X)$.

Distribution functions

6. $F(x) = 3x^2 - 2x^3$ for $0 < x < 1$ (with $F(x) = 0$ if $x \le 0$ and $F(x) = 1$ if $x \ge 1$) defines a distribution function. Find the corresponding density function.

7. Let $F(x) = e^{-1/x}$ for $x \ge 0$, $F(x) = 0$ for $x \le 0$. Is F a distribution function? If so, find its density function.

8. Let $F(x) = 3x - 2x^2$ for $0 \le x \le 1$, $F(x) = 0$ for $x \le 0$, and $F(x) = 1$ for $x \ge 1$. Is F a distribution function? If so, find its density function.

9. Suppose X has density function $f(x) = x/2$ for $0 < x < 2$, 0 otherwise. Find (a) the distribution function, (b) $P(X < 1)$, (c) $P(X > 3/2)$, (d) the median.

10. Suppose X has density function $f(x) = 4x^3$ for $0 < x < 1$, 0 otherwise. Find (a) the distribution function, (b) $P(X < 1/2)$, (c) $P(1/3 < X < 2/3)$, (d) the median.

11. Suppose X has density function $x^{-1/2}/2$ for $0 < x < 1$, 0 otherwise. Find (a) the distribution function, (b) $P(X > 3/4)$, (c) $P(1/9 < X < 1/4)$, (d) the median.

12. Suppose $P(X = x) = x/21$ for $x = 1, 2, 3, 4, 5, 6$. Find all the medians of this distribution.

13. Suppose X has a Poisson distribution with $\lambda = \ln 2$. Find all the medians of X.

14. Suppose X has a geometric distribution with success probability 1/4, i.e., $P(X = k) = (3/4)^{k-1}(1/4)$. Find all the medians of X.

15. Suppose X has density function $3x^{-4}$ for $x \ge 1$. (a) Find a function g so that $g(X)$ is uniform on $(0, 1)$. (b) Find a function h so that if U is uniform on $(0, 1)$, $h(U)$ has density function $3x^{-4}$ for $x \ge 1$.

16. Suppose $X_1, \ldots, X_n$ are independent and have distribution function $F(x)$. Find the distribution functions of (a) $Y = \max\{X_1, \ldots, X_n\}$ and (b) $Z = \min\{X_1, \ldots, X_n\}$

17. Suppose $X_1, \ldots, X_n$ are independent exponential(λ). Show that

$$\min\{X_1, \ldots, X_n\} = \text{exponential}(n\lambda)$$

Functions of random variables

18. Suppose X has density function $f(x)$ for $a \le x \le b$ and $Y = cX + d$ where $c > 0$. Find the density function of Y.

19. Show that if $X = \text{exponential}(1)$ then $Y = X/\lambda$ is exponential(λ).

20. Suppose X is uniform on $(0,1)$. Find the density function of $Y = X^n$.

21. Suppose X has density x^{-2} for $x \ge 1$ and $Y = X^{-2}$. Find the density function of Y.

22. Suppose X has an exponential distribution with parameter λ and $Y = X^{1/\alpha}$. Find the density function of Y. This is the *Weibull distribution.*

23. Suppose X has an exponential distribution with parameter 1 and $Y = \ln(X)$. Find the distribution function of X. This is the *double exponential distribution.*

24. Suppose X is uniform on $(0, \pi/2)$ and $Y = \sin X$. Find the density function of Y. The answer is called the *arcsine law* because the distribution function contains the arcsine function.

25. Suppose X has density function $f(x)$ for $-1 \le x \le 1$, 0 otherwise. Find the density function of (a) $Y = |X|$, (b) $Z = X^2$.

26. Suppose X has density function $x/2$ for $0 < x < 2$, 0 otherwise. Find the density function of $Y = X(2 - X)$ by computing $P(Y \ge y)$ and then differentiating.

Joint distributions

27. Suppose X and Y have joint density $f(x,y) = c(x+y)$ for $0 < x, y < 1$. (a) What is c? (b) What is $P(X < 1/2)$?

28. Suppose X and Y have joint density $f(x,y) = 6xy^2$ for $0 < x, y < 1$. What is $P(X + Y < 1)$?

29. Suppose X and Y have joint density $f(x,y) = 2$ for $0 < y < x < 1$. (a) Find $P(X - Y > z)$.

30. Suppose X and Y have joint density $f(x,y) = 1$ for $0 < x, y < 1$. Find $P(XY \le z)$.

31. Two people agree to meet for a drink after work but they are impatient and each will only wait 15 minutes for the other person to show up. Suppose that

they each arrive at independent random times uniformly distributed between 5 p.m. and 6 p.m. What is the probability they will meet?

32. Suppose X and Y have joint density $f(x,y) = e^{-(x+y)}$ for $x, y > 0$. Find the distribution function.

33. Suppose X is uniform on $(0,1)$ and $Y = X$. Find the joint distribution function of X and Y.

34. A pair of random variables X and Y take values between 0 and 1 and have $P(X \le x, Y \le y) = x^3y^2$ when $0 \le x, y \le 1$. Find the joint density function.

35. Given the joint distribution function $F_{X,Y}(x,y) = P(X \le x, Y \le y)$, how do you recover the marginal distribution Function $F_X(x) = P(X \le x)$?

36. Suppose X and Y have joint density $f(x,y)$. Are X and Y independent if
(a) $f(x,y) = xe^{-x(1+y)}$ for $x, y \ge 0$?
(b) $f(x,y) = 6xy^2$ when $x, y \ge 0$ and $x + y \le 1$?
(c) $f(x,y) = 2xy + x$ when $0 < x < 1$ and $0 < y < 1$?
(d) $f(x,y) = (x+y)^2 - (x-y)^2$ when $0 < x < 1$ and $0 < y < 1$?
In each case $f(x,y) = 0$ otherwise.

37. Suppose a point (X,Y) is chosen at random from the disk $x^2+y^2 \le 1$. Find (a) the marginal density of X, (b) the conditional density of Y given $X = x$.

38. Suppose X and Y have joint density $f(x,y) = x + 2y^3$ when $0 < x < 1$ and $0 < y < 1$. (a) Find the marginal densities of X and Y. (b) Are X and Y independent?

39. Suppose X and Y have joint density $f(x,y) = 6y$ when $x > 0$, $y > 0$, and $x + y < 1$. (a) Find the marginal densities of X and Y, (b) the conditional density of X given $Y = y$.

40. Suppose X and Y have joint density $f(x,y) = 10x^2y$ when $0 < y < x < 1$. (a) Find the marginal densities of X and Y. (b) the conditional density of Y given $X = x$.