5 CONTINUOUS PROBABILITIES AND RANDOM VARIABLES

We have thus far restricted discussion to discrete spaces and discrete random variables – those consisting of at most a countably infinite number of outcomes. This is not because it is not possible, interesting, or useful to consider probabilities on an uncountably infinite set such as the real line or the interval $(0, 1)$. Instead, there are a few technicalities that arise when discussing such probabilities that are best avoided until they are needed. That time is now.

5.1 UNCOUNTABLE SAMPLE SPACES AND DENSITIES

Suppose we want to randomly select a number on the interval $(0, 1)$ in some uniform way. In the discrete setting we would have said that "uniform" meant that every outcome in our sample space $S = (0, 1)$ was equally likely. Suppose we took that same approach here and declared that there was some value p for which $P(\{x\}) = p$ for every $x \in (0, 1)$. Then if we let E be the event $E = \{\frac{1}{n} : n = 2, 3, 4, \dots\} \subset S$, we find that

$$\begin{aligned} P(E) &= P(\{\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots\}) \\ &= P(\frac{1}{2}) + P(\frac{1}{3}) + P(\frac{1}{4}) + \dots \\ &= p + p + p + \dots \end{aligned}$$

If $p > 0$ this sum diverges to infinity, which cannot be since it describes a probability. Therefore it must be that $p = 0$. If every individual outcome in $S = (0, 1)$ is equally likely, then each outcome must have a probability of zero. After several chapters considering only discrete probabilities many readers may suspect that this, in and of itself, is a contradiction. How is it possible for $P(S) = 1$ when every single element of S has probability zero? Could not one then show

$$\begin{aligned} P(S) &= P(\bigcup_{s \in S} \{s\}) \\ &= \sum_{s \in S} P(\{s\}) \\ &= \sum_{s \in S} 0 \\ &= 0 \end{aligned}$$

using the probability axioms? The answer to that question is "no". The probability space axiom that allows us to write the probability of a disjoint union as the sum of separate probabilities only applies to countable collections of events. But the events $\{s\}$ that combine to create $(0, 1)$ are an uncountable collection. If S is uncountable, we could still have $P(S) = 1$ even if every individual element of $s \in S$ has probability zero.

However, all of that does not yet explain how to define a uniform probability on $(0, 1)$. Knowing that each individual outcome has probability zero does not tell us how to calculate $P([\frac{1}{4}, \frac{3}{4}])$, for example, since we cannot simply add up the probabilities of each of the constituent outcomes individually. Instead we need to reinterpret what we mean by "uniform" in this situation. It would make sense to suggest that the event $[\frac{1}{4}, \frac{3}{4}]$ should have a probability of $\frac{1}{2}$ since its length is exactly half of the length of $(0, 1)$. Indeed it is tempting (and essentially correct) to declare that $P(A)$ should be the length of the set A. The complication with making such a statement is that, although length is easy to define if A is an interval or even a countable collection of disjoint intervals, it is not even possible to consistently define a length for every single subset of $(0, 1)$. Because of this unfortunate fact, we will need to reconsider which subsets of S are actually events which will be assigned a probability.

At a minimum we will want events to include any interval. The axioms and basic properties of probability spaces also require that for any collection of events we must be able to consider complements and countable unions of these events. Further, the entire sample space S should also be considered a legitimate event. Consequently we make the following definition.

Definition 5.1.1. **(σ-field)** *If S is a sample space, then a σ-field $\mathcal{F}$ is a collection of subsets of S such that*

(1) $S \in \mathcal{F}$

(2) If $A \in \mathcal{F}$ *then* $A^c \in \mathcal{F}$

(3) If $A_1, A_2, \ldots$ *is a countable collection of sets in* $\mathcal{F}$ *then* $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$

We shall refer to an element of the σ-field as an event.

If S happens to be the set of real numbers there is a smallest σ-field that contains all intervals, and this collection of subsets of $\mathbb{R}$ is known as the Borel sets. It happens that the concept of the "length" of a set can be consistantly described for such sets. Because of this we will modify our definition of probability space slightly at this point.

Definition 5.1.2. **(Probability Space Axioms)** *Let S be a sample space and let $\mathcal{F}$ be a σ-field of S. A "***probability***" is a function* $P : \mathcal{F} \to [0, 1]$ *such that*

(1) $P(S) = 1$;

(2) If $E_1, E_2, \ldots$ *are a countable collection of disjoint events in* $\mathcal{F}$*, then*

$$P(\bigcup_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} P(E_j).$$

The triplet $(S, \mathcal{F}, P)$ is referred to as a probability space.

Our old definition is simply a special case where the σ-field was the collection of all subset of S, so all results we have previously seen in the discrete setting are still legitimate in this new framework. There are many technicalities that arise due to the fact that not every set may be viewed as an event, but these issues would be distracting from the primary goal of this text. Thus we give the definitions above only to provide the modern definition of probability space.

Throughout the remainder of the sections on continuous probability spaces we will restrict our attention to the sample space being $\mathbb{R}$. Whenever we state or prove anything for an event A (a

Borel set) we shall restrict ourselves to the case the event A is a finite or countable unions of intervals. This will enable us to use standard results from calculus and thereby avoid technicalities. A thorough study of the Borel sets and the related theory of integration is beyond the scope of this text (the interested reader may see [AS09] in the bibliography for additional information).

5.1.1 *Probability Densities on* $\mathbb{R}$

The primary way we will define continuous probabilities on $\mathbb{R}$ is through a "density function". We begin by providing an example of what should be meant by a uniform distribution on $(0,1)$.

EXAMPLE 5.1.3. Let $f : \mathbb{R} \to \mathbb{R}$ be a function defined by

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

For an event A define $P(A) = \int_A f(x)\,dx$. Note that for an interval $A = [a,b] \subset (0,1)$ it happens that $P(A)$ is just the length of the interval.

$$\begin{aligned} P(A) &= \int_A f(x)\,dx \\ &= \int_a^b 1\,dx \\ &= b - a \end{aligned}$$

For disjoint unions of intervals, the lengths simply add. For instance if $A = [\frac{1}{5}, \frac{2}{5}] \cup [\frac{3}{5}, \frac{4}{5}]$, then

$$\begin{aligned} P(A) &= \int_{[\frac{1}{5},\frac{2}{5}]\cup[\frac{3}{5},\frac{4}{5}]} f(x)\,dx \\ &= \int_{\frac{1}{5}}^{\frac{2}{5}} 1\,dx + \int_{\frac{3}{5}}^{\frac{4}{5}} 1\,dx \\ &= \frac{1}{5} + \frac{1}{5} = \frac{2}{5} \end{aligned}$$

which is the sum of the lengths of the two component intervals. In particular note that $P((0,1)) = 1$ while $P(\{c\}) = 0$ for any c since a single point has no length. Similarly, if $A = [a,b]$ is an interval that is disjoint from $(0,1)$, then

$$\begin{aligned} P(A) &= \int_A f(x)\,dx \\ &= \int_a^b 0\,dx \\ &= 0 \end{aligned}$$

We will soon see that P defines a probability on $\mathbb{R}$. From the computation above this probability gives equal likelihood to all equal-width intervals within $(0,1)$ and assigns zero probability to any interval outside of $(0,1)$. Therefore it is consistant with the properties a uniform probability on $(0,1)$ should have. ■

The function f from the example above is known as a density. What properties must be required of such a function in order for it to define a probability? The fact that probabilities cannot be negative suggests we will need to require $f(x)$ to be non-negative for all real numbers x. The fact that $P(S) = 1$ means that $\int_{-\infty}^{\infty} f(x)\,dx$ has to be 1. It turns out that these two requirements are essentially all that are needed. The only other assumption we will make is that a density funciton be piecewise continuous. Though this final requirement is more restrictive than necessary, the assumption will help avoid technicalities and will include all densities of interest to us in the remainder of the text. We give a precise definition.

Definition 5.1.4. *Let $f : \mathbb{R} \to \mathbb{R}$ is called a density function if f satisfies the following:*

(i) $f(x) \geq 0$,

(ii) f *is piecewise-continuous, and*

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

We proceed to state and prove a result that will help us construct probabilities on $\mathbb{R}$ with the help of density functions. This will also ensure that in Example 5.1.3 we indeed constructed a probability on $\mathbb{R}$.

Theorem 5.1.5. *Let $f(x)$ be a density function. Define*

$$P(A) = \int_A f(x)\,dx, \tag{5.1.1}$$

for any event $A \subset \mathbb{R}$. Then P defines a probability on $\mathbb{R}$. The function f is called the "density function" for the probability P.

Proof - First note

$$\begin{aligned} P(\mathbb{R}) &= \int_{\mathbb{R}} f(x)\,dx \\ &= \int_{-\infty}^{\infty} f(x)\,dx = 1 \end{aligned}$$

by assumption, so the entire sample space has probability 1. Now let A be a Borel subset of $\mathbb{R}$. Since $f(x)$ is non-negative,

$$\begin{aligned} P(A) &= \int_A f(x)\,dx \geq 0, \qquad \text{and} \\ P(A) &= \int_A f(x)\,dx \leq \int_{\mathbb{R}} f(x)\,dx = 1, \end{aligned}$$

so $P(A) \in [0,1]$. Finally, if $E_1, E_2, \ldots$ are a countable collection of disjoint events, then

$$\begin{aligned} P(\bigcup_{n=1}^{\infty} E_n) &= \int_{\bigcup_{n=1}^{\infty} E_n} f(x)\,dx \\ &= \sum_{n=1}^{\infty} \int_{E_n} f(x)\,dx \\ &= \sum_{n=1}^{\infty} P(E_n). \end{aligned}$$

Therefore P satisfies the conditions of a probability space on $\mathbb{R}$. ■

EXAMPLE 5.1.6. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} 3x^2 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

f is piecewise continuous, $f(x)$ is non-negative for all x and

$$\int_{\mathbb{R}} f(x)\,dx = \int_0^1 3x^2\,dx = x^3\big|_0^1 = 1.$$

As it satisfies $(i) - (iii)$ in Definition 5.1.4, f is a density function. Let P be as defined in (5.1.1). As with the uniform example, $f(x)$ is zero outside of $(0, 1)$, so events lying outside this interval will have zero probability. However, note that

$$P([\frac{1}{5}, \frac{2}{5}]) = \int_{\frac{1}{5}}^{\frac{2}{5}} 3x^2\,dx = \frac{7}{125} \qquad \text{while}$$

$$P([\frac{3}{5}, \frac{4}{5}]) = \int_{\frac{3}{5}}^{\frac{4}{5}} 3x^2\,dx = \frac{37}{125}.$$

In other words, intervals of the same length do not have equal probabilities; this probability is not uniform on $(0, 1)$.

The probability of individual points is still zero, so $P(\{\frac{1}{5}\}) = P(\{\frac{2}{5}\}) = 0$, but in terms of the density function, $f(\frac{2}{5})$ is four times as large as $f(\frac{1}{5})$. What does this mean in practical terms?

Let ϵ be a small positive quantity (certianly less than $\frac{1}{5}$). Then

$$P([\frac{1}{5} - \epsilon, \frac{1}{5} + \epsilon]) = \frac{2}{25}\epsilon + 2\epsilon^3 \approx \frac{2}{25}\epsilon \qquad \text{while}$$

$$P([\frac{2}{5} - \epsilon, \frac{2}{5} + \epsilon]) = \frac{8}{25}\epsilon + 2\epsilon^3 \approx \frac{8}{25}\epsilon.$$

The fact that $f(\frac{2}{5})$ is four times as large as $f(\frac{1}{5})$ essentially means that a tiny interval around $\frac{2}{5}$ has approximately four times the probability of a similarly sized interval around $\frac{1}{5}$. ■

EXERCISES

Ex. 5.1.1. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} 2x & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Show that f is a probability density function.

(b) Use f to calculate $P((0, \frac{1}{2}))$.

Ex. 5.1.2. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} x & \text{if } 0 < x < 1 \\ 2 - x & \text{if } 1 \le x < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Sketch a graph of the function f.

(b) Show that f is a probability density function.

(c) Use f to calculate : $P((0, \frac{1}{4})$, $P((\frac{3}{2}, 2))$, $P((-3, -2))$ and $P((\frac{1}{2}, \frac{3}{2}))$.

Ex. 5.1.3. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} k & \text{if } 0 < x < \frac{1}{4} \\ 2k & \text{if } \frac{1}{4} \leq x < \frac{3}{4} \\ 3k & \text{if } \frac{3}{4} \leq x < 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find k that makes f a probability density function.

(b) Sketch a graph of the function f.

(c) Use f to calculate : $P((0, \frac{1}{4})$, $P((\frac{1}{4}, \frac{3}{4}))$, $P((\frac{3}{4}, 1))$.

Ex. 5.1.4. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} k \cdot \sin(x) & \text{if } 0 < x < \pi \\ 0 & \text{otherwise} \end{cases}$$

(a) Determine the value of k that makes f a probability density function.

(b) Calculate $P((0, \frac{1}{2}))$ and $P((\frac{1}{2}, 1))$.

(c) Which will be larger, $P((0, \frac{1}{4}))$ or $P((\frac{1}{4}, \frac{1}{2}))$? Explain how you can answer this question without actually calculating either probability.

(d) A game is played in the following way. A random variable X is selected with a density described by f above. You must select a number r and you win the game if the random variable results in an outcome in the interval $(r - 0.01, r + 0.01)$. Explain how you should choose r to maximize your chance of winning the game. (A formal proof requires only basic calculus, but it should take very little computation to determine the correct answer).

Ex. 5.1.5. Let $\lambda > 0$ and $f : \mathbb{R} \to \mathbb{R}$ be defined by

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

(a) Show that f is a probability density function.

(b) Let $a > 0$. Find $P((a, \infty))$.

Ex. 5.1.6. Let $a, b \in \mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$ be defined by

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

(a) Show that f is a probability density function.

(b) Show that if $I, J \subset [a, b]$ are two intervals that have the same length, then $P(I) = P(J)$.

Ex. 5.1.7. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} \frac{1}{x^2} & \text{if } 1 < x \\ 0 & \text{otherwise} \end{cases}$$

(a) Show that f is a probability density function.

(b) Let $a > 1$. Find $P((a, \infty))$.

Ex. 5.1.8. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} \frac{1}{6}x^2 e^{-x} & \text{if } 0 < x \\ 0 & \text{otherwise} \end{cases}$$

Show that f is a probability density function.

Ex. 5.1.9. For any $x \in \mathbb{R}$, the hyperbolic secant is defined as $\operatorname{sech} x = \frac{2}{(e^x + e^{-x})}$. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \frac{1}{2} \operatorname{sech}(\frac{\pi}{2} x), x \in \mathbb{R}$$

Show that f is a probability density function.

Ex. 5.1.10. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad x \in \mathbb{R}$$

Follow the steps below to show that the function f is a density function.

(a) Let $I = \int_{-\infty}^{\infty} e^{-x^2/2}\, dx$ and then explain why

$$I^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\, dx\, dy$$

(Hint: Write I^2 as a product of two integrals each over a different variable and explain why the resulting expression may be written as the double integral above).

(b) Explain why

$$I^2 = \int_0^{2\pi} \int_0^{\infty} r \cdot e^{-r^2/2}\, dr\, d\theta$$

after switching from rectangular to polar coordinates. (Hint: Use the fact from multivariate calculus that after the change of variables $(dx\, dy)$ becomes $(r\, dr\, d\theta)$ and explain the new limits of integration based on the region being described in the plane).

(c) Compute the integral from (b) to find the value of I.

(d) Use (c) to show that $\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = 1$. (Hint: Use a u-substitution $u = \frac{x-\mu}{\sigma}$).

5.2 CONTINUOUS RANDOM VARIABLES

Just as the move from discrete to continuous spaces required a slight change in the definition of probability space, so it also requires a slight change in the definition of random variable. In the discrete setting we frequently needed to consider the preimage $X^{-1}(A)$ of a set. Now we need to make sure that such a preimage is a legitimate event.

DEFINITION 5.2.1. *Let $(S, \mathcal{F}, P)$ be a probability space and let $X : S \to \mathbb{R}$ be a function. Then X is a random variable provided that whenever B is an event in $\mathbb{R}$ (i.e. a Borel set), $X^{-1}(B)$ is also an event in $\mathcal{F}$.*

Note that in the discrete setting this extra condition was met trivially as every subset of S was an event. Therefore the discrete setting is simply a special case of this new definition. As with the introduction of σ-fields, we include this definition for completeness. We will only consider functions which meet this criterion. In this section we shall consider only continuous random variables. These are defined next.

DEFINITION 5.2.2. *Let $(S, \mathcal{F}, P)$ be a probability space. A random variable $X : S \to \mathbb{R}$ is called a continuous random variable if there exists a density function $f_X : \mathbb{R} \to \mathbb{R}$ such that for any event A in $\mathbb{R}$,*

$$P(X \in A) = \int_A f_X(x)\,dx. \tag{5.2.1}$$

f_X is called the probability density function of X.

The following lemma demonstrates an elementary property of continous random variables that distinguishes them from discrete random variables.

LEMMA 5.2.3. *Let X be a continuous random variable. Then for any $a \in \mathbb{R}$,*

$$P(X = a) = 0 \tag{5.2.2}$$

Proof- Let $a \in \mathbb{R}$, then $P(X = a) = \int_a^a f(x)dx = 0$. ■

Random variables may also be described using a “distribution function” (also commonly known as a “cumulative distribution function”).

DEFINITION 5.2.4. *If X is a random variable then its distribution funciton $F : \mathbb{R} \to [0, 1]$ is defined by*

$$F(x) = P(X \leq x). \tag{5.2.3}$$

When it must be emphasized that a distribution function belongs to a particular random variable X the notation $F_X(x)$ will be used to indicate the random variable.

Though a distribution function is defined for any real-valued random variable, there is a special relationship between $f_X(x)$ and $F_X(x)$ when the random variable has a density.

THEOREM 5.2.5. *Let X be a random variable with a piecewise continuous density function $f(x)$. If $F(x)$ denotes the distriubtion function of X then*

$$F(x) = \int_{-\infty}^{x} f(x)\,dx. \tag{5.2.4}$$

Moreover, where $f(x)$ is continuous, $F(x)$ is differentiable and $F'(x) = f(x)$.

Proof - By definition $F(x) = P(X \leq x) = P(X \in (-\infty, x])$, but this probability is described in terms of an integral over the density of X, so $F(x) = \int_{-\infty}^{x} f(x)\, dx$.

The result that $F'(x) = f(x)$ then follows from the fundamental theorem of calculus after taking derivatives of both sides of the equation (when such a derivative exists). Note, in particular, that since densities are assumed to be piecewise continuous, their corresponding distribution functions are piecewise differentiable. ■

This theorem will be useful for computation, but it also shows that the distribution of a continuous random variable X is completely determined by its distribution function F_X. That is, if we know $F_X(x)$ and want to calculate $P(X \in A)$ for some set A we could do so by differentiating $F_X(x)$ to find $f_X(x)$ and then integrating this density over the set A. In fact $F_X(x)$ always completely determines the distribution of X (regardless of whether or not X is a continuous random variable), but a proof of that fact is beyond the scope of the course and will not be needed for subsequent results.

5.2.1 *Common Distributions*

In the literature random variables whose distributions satisfy (5.2.1) are called absolutely continuous random variables and those that satisfy (5.2.2)are referred to as continous random variables. Since we shall only consider continuous random variables that satisfy (5.2.1) we refer to them as continous random variables.

There are many continuous distributions that commonly arise. Some of these are continuous analogs of discrete random variables we have already studied. We will define these in the context of continuous random variables having the corresponding distributions. We begin with the already discussed uniform distribution but on an arbitrary interval.

DEFINITION 5.2.6. $X \sim$ **Uniform**(a, b): *Let (a, b) be an open interval. If X is a random variable with its probabilty density function given by*

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

then X is said to be uniformly distributed on (a, b). Note that this is consistant with the example at the beginning of the section since the density of a Uniform$(0, 1)$ is one on the interval $(0, 1)$ and zero elsewhere. Further, recall that in Exercise 5.1.6 we have shown that f is indeed a probability density function.

Since X only takes values on (a, b) if $x < a$ then $P(X \leq x) = 0$ while if $x > b$ then $P(X \leq x) = 1$. So let $a \leq x \leq b$. Then,

$$P(X \leq x) = \int_{-\infty}^{x} f_X(y)\, dy = \int_{-\infty}^{a} 0\, dy + \int_{a}^{x} \frac{1}{b-a}\, dy = \frac{x-a}{b-a}.$$

Therefore the distribution function for X is

$$F_X(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \leq x \leq b \\ 1 & \text{if } x > b \end{cases}$$

Exponential Random Variable

The next continuous distribution we introduce is called the exponential distribution. It is well known from physical experiments that the radioactive isotopes decay to its stable form. Suppose there were $N(0)$ atoms of a certain radioctive material at time 0 then one is interested in the fraction of radioactive material that have not decayed at time $t > 0$. It is observed from experiments that if $N(t)$ is the number of atoms of radioactive material that has not decayed by time t then the fraction

$$\frac{N(t)}{N(0)} \approx e^{-\lambda t},$$

for some $\lambda > 0$. One can introduce a probability model for the above experiment in the following manner. Suppose X represented the time taken by a randomly chosen radioactive atom to decay to its stable form. The distribution of the random variable X needs to satisfy

$$P(X \geq t) = e^{-\lambda t}, \tag{5.2.5}$$

for $t > 0$. It is possible to define such a random variable.

Definition 5.2.7. $X \sim \mathbf{Exp}(\lambda)$*: Suppose* $\lambda > 0$*. If* X *is a random variable with its probabilty density function given by*

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

it is said to be distributed exponentially with parameter λ*. Recall that in Exercise 5.1.5 we have shown that* f *is indeed a probability density function. Since* X *only takes values on* $(0, \infty)$ *if* $x < 0$ *then* $P(X \leq x) = 0$*. So let* $x \geq 0$*. Then,*

$$P(X \leq x) = \int_{-\infty}^{x} f_X(x)\, dx = \int_{-\infty}^{0} 0\, dx + \int_0^x \lambda e^{-\lambda y} dy \; = -e^{-\lambda y} \mid_0^x = 1 - e^{-\lambda x}.$$

Therefore the distribution function for X *is*

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\lambda x} & \text{if } 0 \leq x \end{cases}$$

We have previously seen that geometric random variables have the memoryless property (See (3.2.2)). It turns out that the exponential random variable also possess the *memoryless property* in continuous time. Clearly if $X \sim \text{Exp}(\lambda)$ then $P(X \geq 0) = 1$ and

$$P(X \geq t) = P(X \in [t, \infty)) = \int_t^{\infty} \lambda e^{-\lambda x}\, dx = -e^{-\lambda x} \mid_t^{\infty} = e^{-\lambda t},$$

for $t > 0$. Further if $s, t > 0$, $X > s + t$ imples $X > s$. So

$$\begin{aligned} P(X > s + t | X > s) &= \frac{P((X > s + t) \cap (X > s))}{P(X > s)} \\ &= \frac{P(X > s + t)}{P(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t}. \end{aligned}$$

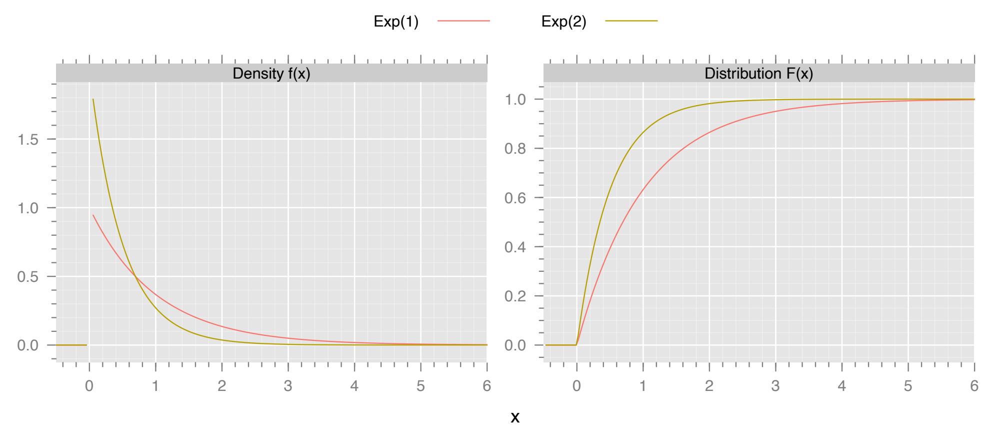

Figure 5.1: The shape of typical Exponential density and cumulative distribution functions.

Therefore for all $s, t > 0$

$$P(X > s + t | X > s) = P(X > t) \tag{5.2.6}$$

Thinking of the variables s and t in terms of time, this says that if an exponential random variable has not yet occurred by time s, then its distribution from that time onward continues to be distributed like an exponential random variable with the same parameter. Situations that involve waiting times such as the lifetime of a light bulb or the time spent in a queue at a service counter are often modelled with the exponential distribution. It is a fact (see Exercise 5.2.12) that if a positive continuous random variable has the memoryless property then it necessarily is an exponential random variable.

EXAMPLE 5.2.8. Let $X \sim \text{Exp}(2)$. Calculate the probability that X produces a value larger than 4.

The density of X is

$$f(x) = \begin{cases} 2e^{-2x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

So, $P(X > 4)$ may be calculated via an integral.

$$\begin{aligned} P(X > 4) &= \int_4^\infty 2e^{-2x}\, dx \\ &= -e^{-2x} \mid_4^\infty = 0 - (-e^{-8}) = e^{-8} \approx 0.000335 \end{aligned}$$

So there is only about a 0.0335% chance of such a result. ■

Normal Random Variable

Of all continuous distributions, The normal distribution (also sometimes called a "Gaussian distribution") is the most fundamental for applictions of statistics as it frequently arises as a limiting distribution of sampling procedures.

DEFINITION 5.2.9. $X \sim$ **Normal**(μ, σ^2)*: Let $\mu \in \mathbb{R}$ and let $\sigma > 0$. Then X is said to be normally distributed with parameters μ and σ^2 if it has the density*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{5.2.7}$$

for all $x \in \mathbb{R}$. We will prove that μ and σ are, respectively, the mean and standard deviation of such a random variable (See Definiton 6.1.1, Definition 6.1.9, Example 6.1.11). Recall that in Exercise 5.1.10 we have seen that f is a probability density function.

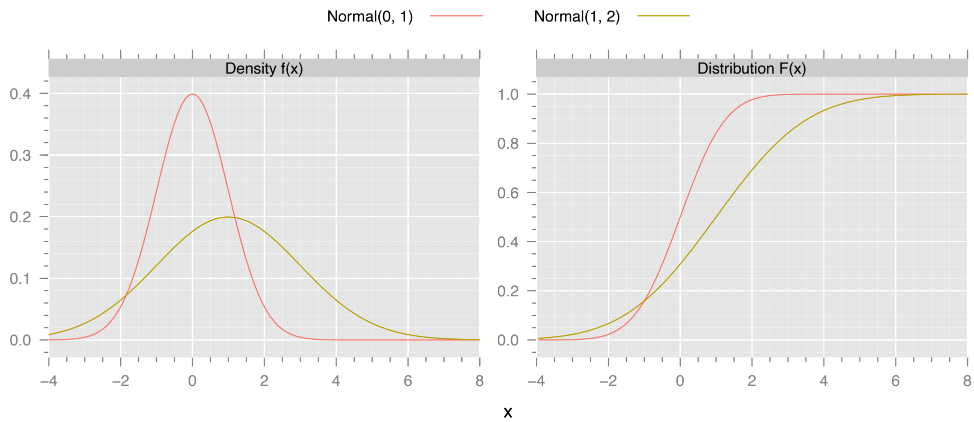

Figure 5.2: The shape of typical Normal density and cumulative distribution functions.

It is observed during statistical experiments that if X were to denote the number of leaves in an apple tree or the height of adult men in a population then X would be close to a normal random variable with approrpriate parameters μ and σ^2. It also arises as a limiting distribution. We shall discuss this aspect in general in Chapter 8, but here we will briefly mention one such limit that appears as an approximation for Binomial probabilities.

Suppose we have $X_1, X_2, \ldots, X_n$ are i.i.d Bernoulli (p) random variables. Then we know that $S_n = \sum_{i=1}^n X_i$ is a Binomial (n, p) random variable. In Theorem 2.2.2 we saw that for $\lambda > 0$, $k \geq 1$, $0 \leq p = \frac{\lambda}{n} < 1$,

$$\lim_{n\to\infty} P(S_n = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

Such an approximation was useful when p was decreasing to zero while n grew to infinity with np remaining constant. The De Moivre-Laplace Central Limit Theorem allows us to consider another form of limit where p remains fixed, but n increases.

THEOREM 5.2.10. **(De Moivre-Laplace Central Limit Theorem)** *Suppose $S_n \sim$ Binomial (n, p), where $0 < p < 1$. Then for any $a < b$*

$$\lim_{n\to\infty} P(a < \frac{S_n - np}{\sqrt{np(1-p)}} \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}\, dx \tag{5.2.8}$$

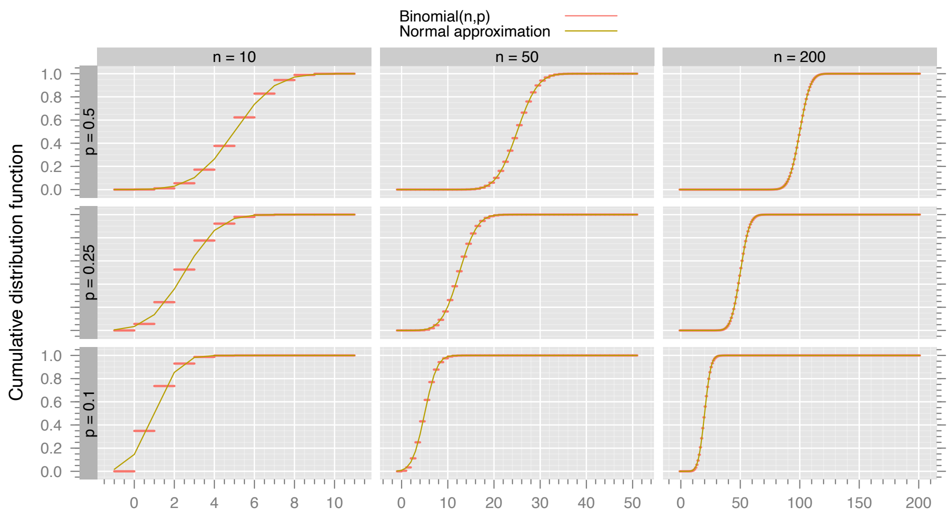

Figure 5.3: The normal approximation to binomial.

We shall omit the proof of the above Theorem for now. We prove it in a more general setting in Chapter 8. For the students well versed with Real Analysis the proof is sketched in Exercise 5.2.16. We refer the reader to [Ram97] for a detailed discussion of the Theorem 5.2.10.

Calculating Normal Probabilities and Necessity of Normal Tables

In a standard introduction to integral calculus one learns many different techniques for calculating integrals. But there are some functions whose indefinite integral has no closed-form solution in terms of simple functions. The density of a normal random variable is one such function. Because of this if $X \sim \text{Normal}(0,1)$ the probability

$$P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

cannot be expressed exactly in terms of standard functions. Many scientific calculators will have a feature that allows this expression to be evaluated. For example, in R, the command `pnorm(x)` evaluates the integral above. Another common solution in statistical texts is to provide a table of values.

Table 5.1 gives values only for positive values of z because for negative z, $P(X \leq z)$ can be easily computed using the symmetry of the $\text{Normal}(0,1)$ distribution as (see Figure 5.4)

$$P(X \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \int_{-z}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 1 - P(X \leq -z). \tag{5.2.9}$$

A more complete version of this table is given in the Appendix. A similar computation can be made for other normally distributed random variables by normalizing them. Suppose $Y \sim \text{Normal}$ (μ, σ^2) and we were interested in finding the distribution function of Y. Observe that

$$P(Y \leq y) \;=\; \int_{-\infty}^{y} \frac{1}{\sigma\sqrt{2\pi}} e^{-(z-\mu)^2/2\sigma^2}\, dz.$$

	0.00	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18
0.0	0.500	0.508	0.516	0.524	0.532	0.540	0.548	0.556	0.564	0.571
0.2	0.579	0.587	0.595	0.603	0.610	0.618	0.626	0.633	0.641	0.648
0.4	0.655	0.663	0.670	0.677	0.684	0.691	0.698	0.705	0.712	0.719
0.6	0.726	0.732	0.739	0.745	0.752	0.758	0.764	0.770	0.776	0.782
0.8	0.788	0.794	0.800	0.805	0.811	0.816	0.821	0.826	0.831	0.836
1.0	0.841	0.846	0.851	0.855	0.860	0.864	0.869	0.873	0.877	0.881
1.2	0.885	0.889	0.893	0.896	0.900	0.903	0.907	0.910	0.913	0.916
1.4	0.919	0.922	0.925	0.928	0.931	0.933	0.936	0.938	0.941	0.943
1.6	0.945	0.947	0.949	0.952	0.954	0.955	0.957	0.959	0.961	0.962
1.8	0.964	0.966	0.967	0.969	0.970	0.971	0.973	0.974	0.975	0.976
2.0	0.977	0.978	0.979	0.980	0.981	0.982	0.983	0.984	0.985	0.985

Table 5.1: Table of Normal$(0,1)$ probabilities. For $X \sim$ Normal$(0,1)$, the table gives values of $P(X \leq z)$ for various values of z between 0 and 2.18 upto three digits. The value of z for each entry is obtained by adding the corresponding row and column labels.

Now perform a change of variables $u = \frac{z-\mu}{\sigma}$ so that $du = \frac{1}{\sigma}\, dz$. This integral then becomes

$$P(Y \leq y) = \int_{-\infty}^{\frac{y-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du = P(X \leq \frac{y-\mu}{\sigma}), \tag{5.2.10}$$

where $X \sim$ Normal $(0,1)$. Now we may use Table 5.1 to compute the distribution function of Y. We conclude this section with two examples.

EXAMPLE 5.2.11. If $X \sim$ Normal$(0,1)$, how likely is it that X will be within one standard deviation of its expected value?

In this case the expected value of the random variable is zero and the standard deviation is one. Therefore the answer is given by

$$\begin{aligned} P(-1 \leq X \leq 1) &= \int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \\ &= \int_{-\infty}^{1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx - \int_{-\infty}^{-1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \\ &= P(X \leq 1) - P(X \leq -1) \end{aligned}$$

R tells us that

```
> pnorm(1)
[1] 0.8413447
> pnorm(-1)
[1] 0.1586553
> pnorm(1) - pnorm(-1)
[1] 0.6826895
```

Alternatively, using Table 5.1, we see that $P(X \leq 1) = 0.841$ (upto three decimal places), and by symmetry $P(X \leq -1) = P(X \geq 1) = 1 - P(X \leq 1) = 1 - 0.841 = 0.159$. Therefore, $P(-1 \leq X \leq 1) \approx 0.841 - 0.159 = 0.682$. In other words, there is roughly a 68% chance that a standardized normal random variable will produce a value within one standard deviation of expected value. ■

Figure 5.4: Computation of Normal$(0,1)$ probabilities as area under the normal density curve. For Normal$(0,1)$ and in fact for any symmetric distribution in general, it is enough to know the distribution function for positive values (see Exercise 5.2.8).

EXAMPLE 5.2.12. A machine fills bags with cashews. The intended weight of cashews in the bag is 200 grams. Assume the machine has a tolerance such that the actual weight of the cashews is a normally distributed random variables with an expected value of 200 grams and a standard deviation of 4 grams. How likely is it that a bag filled by this machine will have fewer than 195 grams of cashews in it?

We know $Y \sim$ Normal$(200, 4^2)$ and we want the probability $P(Y < 195)$. By above computation, (5.2.10),

$$P(Y < 195) = P(X < \frac{195 - 200}{4}) = P(X < -\frac{5}{4})$$

where $X \sim$ Normal$(0,1)$. If we were to use Table 5.1, we would first obtain

$$P(X < -\frac{5}{4}) = 1 - P(X < \frac{5}{4}) = 1 - P(X < 1.25) = 1 - 0.896 = 0.104$$

By the R command "pnorm(-5/4)" we obtain 0.1056498. That is, there is slightly more than a 10% chance of a bag this light being produced by the machine. ■

5.2.2 *A word about individual outcomes*

We began this section by noting that continuous random variables must necessarily give probability zero to any single outcome. It is an awkward consequence of this that two different densities may give rise to exactly the same probabilities. For instance, the functions

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

and

$$g(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

are different because they assign different values to the points $x = 0$ and $x = 1$. However, these individual points cannot affect the computation of probabilities so both $f(x)$ and $g(x)$ give rise to the same probability distribution. The same thing would occur even if $f(x)$ and $g(x)$ differed in a countably infinite number of points, since these will still have probability zero when taken collectively.

Because of this we will describe $f(x)$ and $g(x)$ as the same density (and sometimes even write $f(x) = g(x)$) when the two densities produce the same probabilities. We do this even when f and g may technically be different functions. Though it is a more restirctive assumption than is necessary, we have required densities to be piecewise continuous. As a consquence of the explanation above, altering the values of the function at the endpoints of intervals of continuity will not change the resulting probabilities and will result in the same density.

EXERCISES

Ex. 5.2.1. Suppose X was continuous random variable with distribution function F. Express the following probabilities in terms of F:

(a) $P(a < X \leq b)$, where $-\infty < a < b < \infty$

(b) $P(a < X < \infty)$ where $a \in \mathbb{R}$.

(c) $P(\mid X - a \mid \geq b)$ where $a, b \in \mathbb{R}$ and $b > 0$.

Ex. 5.2.2. Let $R > 0$ and $X \sim$ Uniform $[0, R]$. Let $Y = \min(X, \frac{R}{10})$. Find the distribution function of Y.

Ex. 5.2.3. Let X be a random variable with distribution function given by

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 < x < \frac{1}{4} \\ \frac{x}{2} + \frac{1}{8} & \text{if } \frac{1}{4} \leq x < \frac{3}{4} \\ 2x - 1 & \text{if } \frac{3}{4} \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

(a) Sketch a graph of the function F.

(b) Use F to calculate : $P([0, \frac{1}{4}))$, $P([\frac{1}{8}, \frac{3}{2}])$, $P((\frac{3}{4}, \frac{7}{8}])$.

(c) Find the probabilty density function of X.

Ex. 5.2.4. Let X be a continuous random variable with distribution function $F : \mathbb{R} \to [0, 1]$. Then $G : \mathbb{R} \to [0, 1]$ given by

$$G(x) = 1 - F(x)$$

is called the reliability function of X or the right tail distribution function of X. Suppose $T \sim$ Exponential(λ) for some $\lambda > 0$, then find the reliability function of T.

Ex. 5.2.5. Let X be a random variable whose probability density function $f : \mathbb{R} \to [0, 1]$ is given by

$$f(x) = \begin{cases} kx^{k-1}e^{-x^k} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the distribution function of X for $k = 2$.

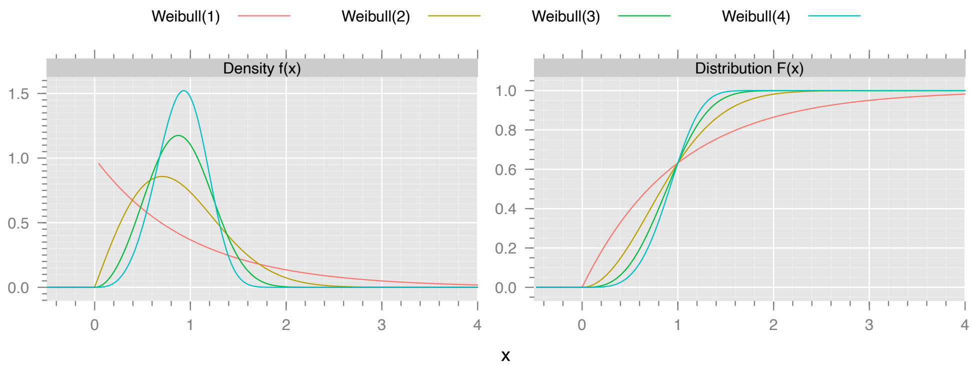

Figure 5.5: The shape of typical Weibull density and cumulative distribution functions.

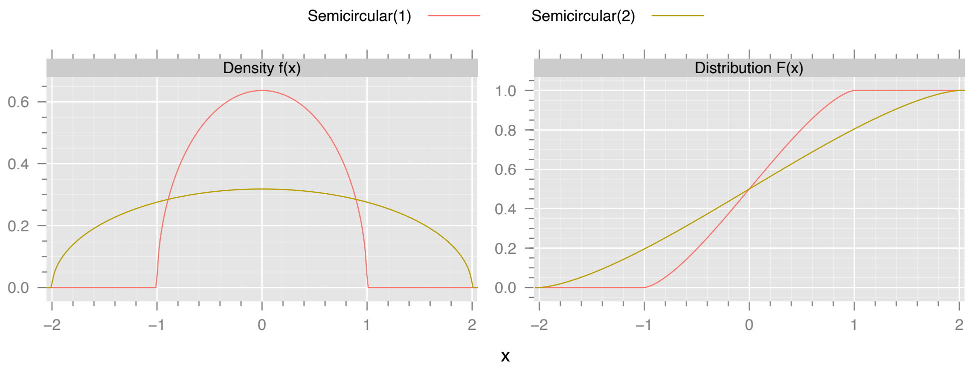

Figure 5.6: The shape of the semicircular density and cumulative distribution functions.

(b) Find the distribution function of X for general k.

The distribution of X is called the Weibull distribution. Figure 5.5 plots the Weibull distribution for selected values of k.

Ex. 5.2.6. Let X be a random variable whose probability density function $f : \mathbb{R} \to [0,1]$ is given by

$$f(x) = \begin{cases} \frac{2}{\pi R^2}\sqrt{R^2 - x^2} & \text{if } -R < x < R \\ 0 & \text{otherwise} \end{cases}$$

Find the distribution function of X. The distribution of X is called the semicircular distribution (see Figure 5.6).

Ex. 5.2.7. Let X be a random variable whose distribution function $F : \mathbb{R} \to [0,1]$ is given by

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \frac{2}{\pi}\arcsin(\sqrt{x}) & \text{if } 0 < x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

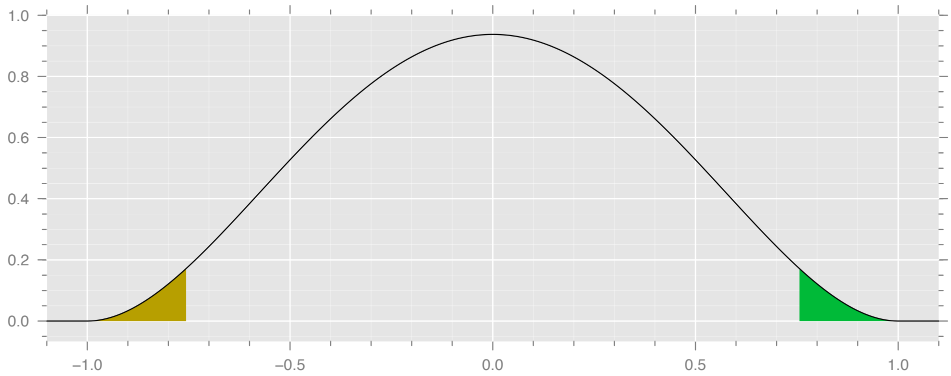

Figure 5.7: Computation of probabilities as area under the density curve. For symmetric distributions, it is enough to know the (cumulative) distribution function for positive values.

Find the probability density function of X. The distribution of X is called the standard arcsine law.

Ex. 5.2.8. Let X be a continuous random variable with probability density function f and distribution function F. Suppose f is a symmetric function, i.e. $f(x) = f(-x)$ for all $x \in \mathbb{R}$. Then show that

(a) $P(X \le 0) = P(X \ge 0) = \frac{1}{2}$,

(b) for $x \ge 0$, $F(x) = \frac{1}{2} + P(0 \le X \le x)$,

(c) for $x \le 0$, $F(x) = P(X \ge -x) = \frac{1}{2} + P(0 \le X \le -x)$.

We have observed this fact for the normal distribution earlier (see Figure 5.7).

Ex. 5.2.9. Let $X \sim \text{Exp}(\lambda)$. The "90th percentile" is a value a such that X is larger than a 90% of the time. Find the 90th percentile of X by determining the value of a for which $P(X < a) = 0.9$.

Ex. 5.2.10. The "median" of a continuous random variable X is the value of x for which $P(X > x) = P(X < x) = \frac{1}{2}$.

(a) If $X \sim \text{Uniform}(a, b)$ calculate the median of X.

(b) If $Y \sim \text{Exp}(\lambda)$ calcluate the median of Y.

(c) Let $Z \sim \text{Normal}(\mu, \sigma^2)$. Show that the median of Z is μ.

Ex. 5.2.11. Let $X \sim \text{Normal}(\mu, \sigma^2)$. Show that $P(|X - \mu| < k\sigma)$ does not depend on the values of μ or σ. (Hint: Use a change of variables for the appropriate integral).

Ex. 5.2.12. Above we saw that exponential random variables satisfied the memoryless property, (5.2.6). It can be shown that any positive, continuous random variable with the memoryless property must be exponential. Follow the steps below to prove a slightly weakened version of this result. For all parts, suppose X is a positive, continuous random variable with the memoryless property for which the distribution function $F_X(t)$ has a continuous derivative for $t > 0$. Suppose further that $lim_{t \to 0^+} F'(t)$ exists and call this quantity α. Let $G(t) = 1 - F_X(t) = P(X > t)$ and do the following.

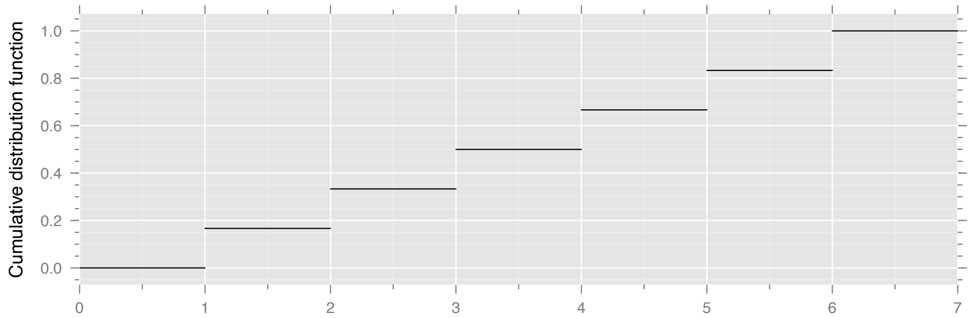

Figure 5.8: The cumulative distribution function for Exercise 5.2.14.

(a) Use the memoryless property to show that $G(s+t) = G(s) \cdot G(t)$ for all postiive s and t.

(b) Use part (a) to conclude that $G'(t) = -\alpha G(t)$. (Hint: Take a derivative with respect to s and then take an appropriate limit).

(c) It is a fact (which you may take as granted) that the differential equation from (b) has solutions of the form $G(t) = Ce^{-\alpha t}$. Use the fact that X is positive to explain why it must be that $C = 1$.

(d) Use part (c) to calculate $F_X(t)$ and then differentiate to find $f_X(t)$.

(e) Conclude that X must be exponentially distributed and determine the associated parameter in terms of α.

Ex. 5.2.13. Let X be a random variable with density $f(x) = 2x$ for $0 < x < 1$ (and $f(x) = 0$ otherwise). Calculate the distribution function of X.

Ex. 5.2.14. Let $X \sim \text{Uniform}(\{1, 2, 3, 4, 5, 6\})$. Despite the fact this is a discrete random variable without a density, the distribution function $F_X(x)$ is still defined. Find a piecewise defined expression for $F_X(x)$ (see Figure 5.8 for a plot).

Ex. 5.2.15. Suppose $F : \mathbb{R} \to [0, 1]$ is given by (5.2.3). Then show that

1. F is a monotonically increasing function.
2. $\lim_{x\to\infty} F(x) = 1$.
3. $\lim_{x\to-\infty} F(x) = 0$.
4. if, in addition, F is given by (5.2.4) then F is continuous.

Ex. 5.2.16. We use the notation as in Theorem 5.2.10.

(a) Let

$$A_n = \left\{k : 0 \le k \le n, np + a\sqrt{np(1-p)} \le k \le np + a\sqrt{np(1-p)}\right\}.$$

Show that

$$P(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b) = \sum_{k \in A_n} P(S_n = k).$$

(b) Let

$$\xi_{k,n} = \frac{k - np}{\sqrt{np(1-p)}}.$$

Using the definition of the Riemann integral show that

$$\lim_{n\to\infty} \sum_{k\in A_n} \frac{e^{-\frac{\xi_{k,n}^2}{2}}}{\sqrt{2\pi np(1-p)}} = \int_a^b \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

(c) Using Stirling's approximation show that

$$\lim_{n\to\infty} \sup_{k\in A_n} \frac{\binom{n}{k}p^k(1-p)^{n-k}}{\sqrt{2\pi np(1-p)}e^{-\frac{\xi_{k,n}^2}{2}}} = 1$$

(d) Prove Theorem 5.2.10 by observing

$$P(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b) =$$

$$\sum_{k\in A_n} \frac{e^{-\frac{\xi_{k,n}^2}{2}}}{\sqrt{2\pi np(1-p)}} + \sum_{k\in A_n} \frac{e^{-\frac{\xi_{k,n}^2}{2}}}{\sqrt{2\pi np(1-p)}} \left(\frac{\binom{n}{k}p^k(1-p)^{n-k}}{\frac{1}{\sqrt{2\pi np(1-p)}}e^{-\frac{\xi_{k,n}^2}{2}}} - 1 \right)$$

5.3 TRANSFORMATION OF CONTINUOUS RANDOM VARIABLES

In Section 3.3 we have discussed functions of discrete random variables and how to find their distributions. Suppose $g : \mathbb{R} \to \mathbb{R}$ and $Y = g(X)$, to find the distribution of Y we converted events associated with Y with events of X by inverting the function g. In the setting of continuous random variables distribution functions are used for calculating probabilities associated with functions of a known random variable. We next present a simple example for which $g(x) = x^2$ followed by a result that covers situations when $g(x)$ is any linear function.

EXAMPLE 5.3.1. Let $X \sim \text{Uniform}(0,1)$ and let $Y = X^2$. What is the density for Y?

Since X takes values on $(0,1)$ and since $Y = X^2$, it will also be the case that Y takes values on $(0,1)$. However, though X is uniform on the interval, there should be no expectation that Y will also be uniform. In fact, since squaring a positive number less than one results in a smaller number than the original, it should seem intuitive that results of Y will be more likely to be near to zero than they are to be near to one.

It is not easy to see how to calculate the density of Y directly from the density of X. However, it is a much easier task to compute the distribution of Y from the distribution of X. Therefore we will use the following plan in the calculation below – integrate $f_X(x)$ to find $F_X(x)$; use $F_X(x)$ to determine $F_Y(y)$; then differentiate $F_Y(y)$ to calculate $f_Y(y)$.

For the first step, note

$$F_X(x) = \int_{-\infty}^{x} f_X(x)\,dx = \begin{cases} 0 & \text{if } 0 < x \\ x & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x > 1 \end{cases}$$

Next, since Y takes values in $(0,1)$, if $y \leq 0$ then $F_Y(y) = P(Y \leq y) = 0$. But if $y > 0$ then

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}).$$

Since X is always positive, the event $(X < -\sqrt{y})$ has zero probability we may connect this to the distribution of X by writing

$$\begin{aligned} F_Y(y) &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= P(X < -\sqrt{y}) + P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= P((X < -\sqrt{y}) \cup (-\sqrt{y} \leq X \leq \sqrt{y})) \\ &= P(X \leq \sqrt{y}) = F_X(\sqrt{y}). \end{aligned}$$

Therefore,

$$F_Y(y) = \begin{cases} 0 & \text{if } 0 \leq y \\ \sqrt{y} & \text{if } 0 < y < 1 \\ 1 & \text{if } y \geq 1 \end{cases}$$

and finally by using the fact that $F'(y) = f(y)$ we can determine that

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} & \text{if } 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

As noted in the beginning of this example, this distribution is far from uniform and gives much more weight to intervals close to zero than it does intervals close to one. ■

LEMMA 5.3.2. *Let $a \neq 0$ and $b \in \mathbb{R}$. Suppose X is a continuous random variable with probability density function f_X. Let $g(x) = ax + b$ be any non-constant linear function (so $a \neq 0$) and let $Y = g(X)$ then Y is also a continuous random variable whose density function f_Y is given by*

$$f_Y(y) = \frac{1}{|a|} f_X(\frac{y-b}{a}), \tag{5.3.1}$$

for all $y \in \mathbb{R}$.

Proof- Let $y \in \mathbb{R}$. Assume first that $a > 0$. Then

$$P(Y \leq y) = P(aX + b \leq y) = P(X \leq \frac{y-b}{a}) = \int_{-\infty}^{\frac{y-b}{a}} f_X(z)dz$$

By a simple change of variable $z = \frac{u-b}{a}$ we obtain that

$$P(Y \leq y) = \int_{-\infty}^{y} \frac{1}{a} f_X(\frac{u-b}{a})du. \tag{5.3.2}$$

If $a < 0$ then

$$P(Y \leq y) = P(aX + b \leq y) = P(X \geq \frac{y-b}{a}) = \int_{\frac{y-b}{a}}^{\infty} f_X(z)dz$$

Again a simple change of variable $z = \frac{u-b}{a}$, with $a < 0$, we obtain that

$$P(Y \leq y) = \int_{-\infty}^{y} \frac{1}{-a} f_X(\frac{u-b}{a})du. \tag{5.3.3}$$

Using (5.3.2) and (5.3.3) we have that Y is a continuous random varable with density as in (5.3.1). ■

Lemma 5.3.2 provides a method to standardise the normal random variable.

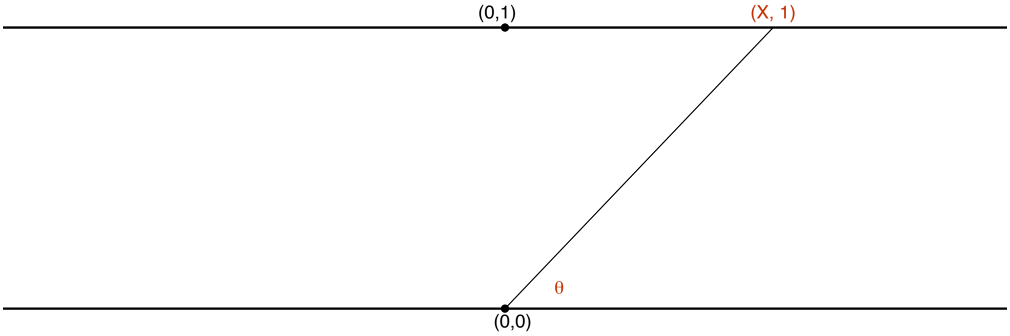

Figure 5.9: Illustration of Example 5.3.4.

COROLLARY 5.3.3. *(a) Let $X \sim Normal(0,1)$ and let $Y = aX + b$ with $a, b \in \mathbb{R}, a \neq 0$. Then, $Y \sim Normal(b, a^2)$.*

(b) Let $X \sim Normal(\mu, \sigma^2)$ and let $Z = \frac{X-\mu}{\sigma}$. Then $Z \sim Normal(0,1)$.

Proof - X has a probability density function given by (5.2.7).
(a)By Lemma 5.3.2, we have that the density of Y is given by

$$f_Y(y) = \frac{1}{\mid a \mid} f_X\left(\frac{y-b}{a}\right) = \frac{1}{\sqrt{2\pi} \mid a \mid} e^{-\frac{(z-b)^2}{2a^2}},$$

for all $y \in \mathbb{R}$. Hence $Y \sim$ Normal (b, a^2).

(b) By Lemma 5.3.2, with $a = \frac{1}{\sigma}$ and $b = -\frac{\mu}{\sigma}$ we have that the density of Z is given by

$$f_Z(z) = \sigma f_X\left(\sigma(z + \frac{\mu}{\sigma})\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}},$$

for all $z \in \mathbb{R}$. Hence $Z \sim$ Normal $(0,1)$. ■

EXAMPLE 5.3.4. Consider the two parallel lines in $\mathbb{R}^2$, given by $y = 0$ and $y = 1$. Piku is standing at the origin in the plane. She chooses an angle θ uniformly in $(0, \pi)$ and she draws a line segment between the lines $y = 0$ and $y = 1$ at an angle θ from the origin in $\mathbb{R}^2$. Suppose the line segment meets the line $y = 1$ at the point $(X, 1)$. Find the probability density function of X.

First observe that $X = \tan(\frac{\pi}{2} - \theta)$. We shall first find the distribution function of X. Let $x \in \mathbb{R}$. Observe that $\tan(x)$ is a strictly increasing function in the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$ and has an inverse denoted by $\arctan(x)$. So

$$\begin{aligned} P(X \leq x) &= P(\tan(\frac{\pi}{2} - \theta) \leq x) \\ &= P((\frac{\pi}{2} - \theta) \leq \arctan(x)) \\ &= P(\theta \geq \frac{\pi}{2} - \arctan(x)) \\ &= 1 - P(\theta \leq \frac{\pi}{2} - \arctan(x)) \end{aligned}$$

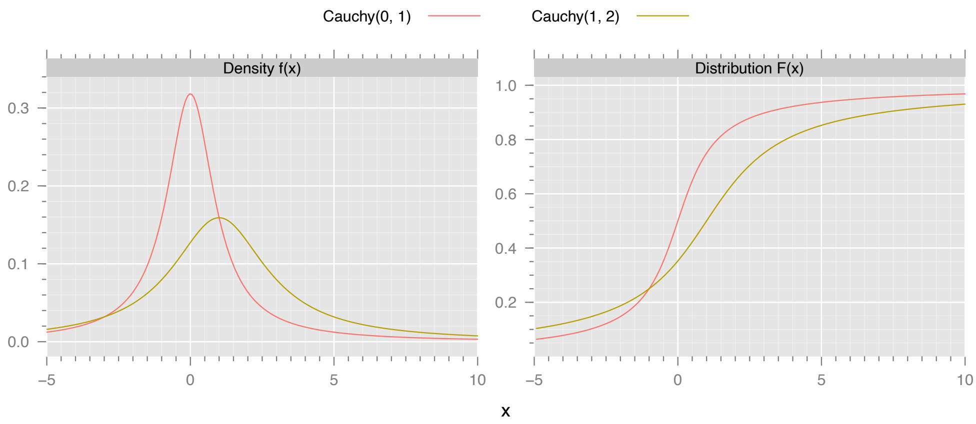

Figure 5.10: The shape of Cauchy density and cumulative distribution functions for selected parameter values.

For any $x \in \mathbb{R}$, $\frac{\pi}{2} - \arctan(x) \in (0, \pi)$. As θ has Uniform $(0, \pi)$ distribution, the above is

$$\begin{aligned} &= 1 - \frac{1}{\pi}\left(\frac{\pi}{2} - \arctan(x)\right) \\ &= \frac{1}{2} + \frac{1}{\pi}\arctan(x) \end{aligned}$$

Hence the distribution function of X is differentiable and therefore the probability density function of X is given by

$$f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2},$$

for all $x \in \mathbb{R}$. Such a random variable is an example of a Cauchy distribution which we define more generally next. ■

DEFINITION 5.3.5. $X \sim$ **Cauchy**(θ, α^2)*: Let $\theta \in \mathbb{R}$ and let $\alpha > 0$. Then X is said to have a Cauchy distribution with parameters θ and α^2 if it has the density*

$$f(x) = \frac{1}{\pi}\frac{\alpha}{\alpha^2 + (x-\theta)^2} \tag{5.3.4}$$

for all $x \in \mathbb{R}$. Here θ is referred to as the location parameter and α is referred to as the scale parameter. The distribution function of X is given by

$$F(x) = \frac{1}{\pi}\arctan\left(\frac{x-\theta}{\alpha}\right) \tag{5.3.5}$$

Figure 5.10 gives plots of the Cauchy density and distribution functions.

Similar computations as above are useful for simulations. Most computer progamming languages and spreadsheets have a "Random" function designed to approximate a Uniform$(0, 1)$ random variable. How could one use such a feature to simulate random variables with other densities? We start with an example.

Example 5.3.6. If $X \sim$ Uniform$(0, 1)$, our goal is to find a function $g : (0, 1) \to \mathbb{R}$ for which $Y = g(X) \sim$ Exponential (λ). We will try to find such a $g : (0, 1) \to \mathbb{R}$ which is strictly increasing so that it has an inverse. This will be important when it comes to relating the distributions of X and Y.

We require Y to Exponential(λ). So the distribution function of Y is

$$F_Y(y) = \begin{cases} 0 & \text{if } y \leq 0 \\ 1 - e^{-\lambda y} & \text{if } y > 0 \end{cases}$$

But

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

where the final equality comes from our decree that the function g should be strictly increasing. Therefore,

$$F_Y(y) = F_X(g^{-1}(y)).$$

But the distribution function of a uniform random variable has previously been computed. Hence,

$$F_X(g^{-1}(y)) = \begin{cases} 0 & \text{if } g^{-1}(y) \leq 0 \\ g^{-1}(y) & \text{if } 0 < g^{-1}(y) < 1 \\ 1 & \text{if } g^{-1}(y) \geq 1 \end{cases}$$

Thus we are forced to have

$$g^{-1}(y) = 1 - e^{-\lambda y}$$

for $y > 0$. So inverting the above formula, we get $g : (0, 1) \to (0, \infty)$ is given by

$$g(x) = -\frac{1}{\lambda} \ln(1 - x),$$

for $x \in (0, 1)$. Hence,

$$X \sim \text{ Uniform}(0, 1) \implies -\frac{1}{\lambda} \ln(1 - X) \sim \text{ Exponential}(\lambda).$$

In conclusion one could view g as the inverse of F_Y, on $(0, \infty)$. It turns out that this is a general result. We state a special case of this in the lemma below. ■

Lemma 5.3.7. *Let $U \sim$ Uniform $(0, 1)$ random variable. Let X be a continuous random variable such that its distribution function, F_X, is a strictly increasing continous function. Then*

(a) $Y = F_X^{-1}(U)$ has the same distribution as X.

(b) $Z = F_X(X)$ has the same distribution as U.

Proof- We observe that as F is strictly increasing continuous distribution function $F : \mathbb{R} \to (0, 1)$ and Range $(F) = (0, 1)$.

(a) We shall verify that Y and X have the same distribution function. Let $y \in \mathbb{R}$, then

$$F_Y(y) = P(Y \leq y) = P(F_X^{-1}(U) \leq y) = P(U \leq F_X(y)) = F_X(y)$$

Hence X and Y have the same distribution.

(b) We shall verify that Z and U have the same distribution function. Let $z \in \mathbb{R}$. If $z \leq 0$ then

$$P(Z \leq z) \quad = \quad P(F(X) \leq z) = 0$$

as $F : \mathbb{R} \to (0,1)$. If $z \geq 1$ then

$$P(Z \leq z) \quad = \quad P(F(X) \leq z) = 1$$

as $F : \mathbb{R} \to (0,1)$. If $0 < z < 1$ then $F^{-1}(z)$ is well defined as Range $(F) = (0,1)$ and

$$P(Z \leq z) \quad = \quad P(F(X) \leq z) = P(X \leq F^{-1}(z)) = F(F^{-1}(z)) = z.$$

Hence Z and U have the same distribution. ■

The previous lemma may be generalized even to the case when F is not strictly increasing. It requires a concept called the generalized inverse. The interested reader will find it discussed in Exercise 5.3.12.

EXERCISES

Ex. 5.3.1. Let $X \sim \text{Uniform}(0,1)$ and let $Y = \sqrt{X}$. Determine the density of Y.

Ex. 5.3.2. Let $X \sim \text{Uniform}(0,1)$ and let $Z = \frac{1}{X}$. Determine the density of Z.

Ex. 5.3.3. Let $X \sim \text{Uniform}(0,1)$. Let $r > 0$ and define $Y = rX$. Show that Y is uniformly distributed on $(0,r)$.

Ex. 5.3.4. Let $X \sim \text{Uniform}(0,1)$. Let $Y = 1 - X$. Show that $Y \sim \text{Uniform}(0,1)$ as well.

Ex. 5.3.5. Let $X \sim \text{Uniform}(0,1)$. Let a and b be real numbers with $a < b$ and let $Y = (b-a)X + a$. Show that $Y \sim \text{Uniform}(a,b)$.

Ex. 5.3.6. Let $X \sim \text{Uniform}(0,1)$. Find a function $g(x)$ (which is strictly increasing) such that the random variable $Y = g(X)$ has density $f_Y(y) = 3y^2$ for $0 < y < 1$ (and $f_Y(y) = 0$ otherwise).

Ex. 5.3.7. Let $X \sim Normal(\mu, \sigma^2)$. Let $g : (-\infty, \infty) \to \mathbb{R}$ be given by $g(x) = x^2$. Find the probability density function of $Y = g(X)$.

Ex. 5.3.8. Let $\alpha > 0$ and X be a random variable with the p.d.f given by

$$f(x) = \begin{cases} \frac{\alpha}{x^{\alpha+1}} & 1 \leq x < \infty \\ 0 & \text{otherwise} \end{cases}$$

The random variable X is said to have Pareto (α) distribution (see Figure 5.11).

(a) Find the distribution of $X_1 = X^2$

(b) Find the distribution of $X_2 = \frac{1}{X}$

(c) Find the distribution of $X_3 = \ln(X)$

In the above exercises we assume that the transformation function is defined as above when the p.d.f of X is positive and zero otherwise.

Ex. 5.3.9. Let X be a continuous random variable with probability density function $f_X : \mathbb{R} \to \mathbb{R}$. Let $a > 0$, $b \in \mathbb{R}$ $Y = \frac{1}{a}(X-b)^2$. Show that Y is also a continuous random variable with probability density function $f_Y : \mathbb{R} \to \mathbb{R}$ given by

$$f_Y(y) = \frac{\sqrt{a}}{2\sqrt{y}} \left[f_X(\sqrt{ay} + b) + f_X(-\sqrt{ay} + b) \right]$$

for $y > 0$.

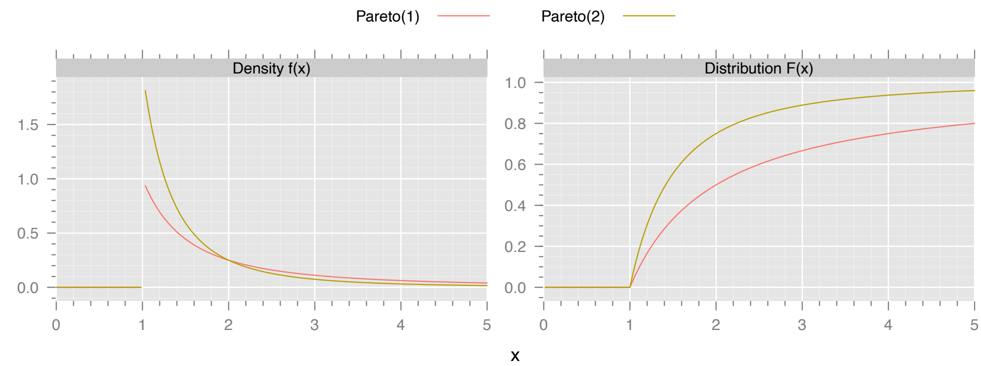

Figure 5.11: The shape of the pareto density and cumulative distribution functions.

Ex. 5.3.10. Let $-\infty \leq a < b \leq \infty$ and $I = (a, b)$ and $g : I \to \mathbb{R}$. Let X be a continuous random variable whose density f_X is zero on the complement of I. Set $Y = g(X)$.

(a) Let g be a differentiable strictly increasing function.

(i) Show that inverse of g exists and g^{-1} is strictly increasing on $g(I)$.

(ii) For any $y \in \mathbb{R}$, show that $P(Y \leq y) = P(X \leq g^{-1}(y))$

(iii) Show that Y has a density $f_Y(\cdot)$ given by

$$f_Y(y) = f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y).$$

(b) Let g be a differentiable strictly decreasing function.

(i) Show that inverse of g exists and g^{-1} is strictly decreasing on $g(I)$.

(ii) For any $y \in \mathbb{R}$, show that $P(Y \leq y) = 1 - P(X \leq g^{-1}(y))$

(iii) Show that Y has a density $f_Y(\cdot)$ given by

$$f_Y(y) = f_X(g^{-1}(y))\left(-\frac{d}{dy}g^{-1}(y)\right).$$

Ex. 5.3.11. Let X be a random variable having an exponential density. Let $g : [0, \infty) \to \mathbb{R}$ be given by $g(x) = x^{\frac{1}{\beta}}$, for some $\beta \neq 0$. Find the probability density function of $Y = g(X)$.

Ex. 5.3.12. Let $U \sim$ Uniform $(0, 1)$. Let X be a continuous random variable with a distribution function F. Extend $F : \mathbb{R} \to \mathbb{R}$ to $F : \mathbb{R} \cup \{-\infty\} \cup \{\infty\} \to \mathbb{R}$ by setting $F(\infty) = 1$ and $F(-\infty) = 0$. Define the generalised inverse of F, $G : [0, 1] \to \mathbb{R} \cup \{-\infty\} \cup \{\infty\}$ by

$$G(y) = \inf\{x \in \mathbb{R} : F(x) \geq y\}.$$

Show that

(a) Show that for all $y \in [0, 1]$, $F(G(y)) = y$.

(b) Show that for all $x \in \mathbb{R}$ and $y \in [0, 1]$

$$F(x) \geq y \iff x \geq G(y).$$

(c) $Y = G(U)$ has the same distribution as X.

(d) $Z = F(X)$ has the same distribution as U.

5.4 MULTIPLE CONTINUOUS RANDOM VARIABLES

When analyzing multiple random variables at once, one may consider a "joint density" analogous to the joint distribution of the discrete variable case. In this section we will restrict considerations to only two random variables, but we shall see in Chapter 8 that the definitions and results all generalize to any finite collection of variables.

THEOREM 5.4.1. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a non-negative function, piecewise-continuous in each variable for which*

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

For a Borel set $A \subset \mathbb{R}^2$ define

$$P(A) = \int_A f(x, y)\, dx\, dy.$$

Then P is a probability on $\mathbb{R}^2$ and f is called the density for P.

Proof- The proof of the theorem is essentially the same as in the one-variable version of Theorem 5.1.5. We will not reproduce it here. As in the discrete case we will typically associate such densities with random variables. ■

DEFINITION 5.4.2. *A pair of random variables (X, Y) is said to have a joint density $f(x, y)$ if for every Borel set $A \subset \mathbb{R}^2$*

$$P((X, Y) \in A) = \int_A f(x, y)\, dx\, dy.$$

As in the one-variable case we describe this in terms of "Borel sets" to be precise, but in practice we will only consider sets A which are simple regions in the plane. In fact regions such as $(-\infty, a] \times (-\infty, b]$, for all real numbers a, b are enough to characterise the joint distribution. As in the one variable case we can define a "joint distribution function" of (X, Y) as

$$F_{(X,Y)}(a, b) = P((X \le a) \cap (Y \le b)) = \int_{-\infty}^{a} \int_{-\infty}^{b} f(z, w) dw dz \tag{5.4.1}$$

for all $a, b \in \mathbb{R}$. We will usually denote the joint distribution function by F omiting the subscripts unless it is particularly needed. One can state and prove a similar type of result as Theorem 5.2.5 for $F(a, b)$ when (X, Y) have a joint density. In particular, we can conclude that since the joint densities are assumed to be piecewise continuous, the corresponding distribution functions are piecewise differentiable. Further, the joint distribution of two continuous random variables (X, Y) are completely determined by their joint distribution function F. That is, if we know the value of $F(a, b)$ for all $a, b \in \mathbb{R}$, we could use multivariable calculus to differentiate $F(a, b)$ to find $f(a, b)$. Then $P((X, Y) \in A)$ for any event A is obtained by integrating the joint density f over the event A. We illustrate this with a couple of examples.

EXAMPLE 5.4.3. Consider the open rectangle in $\mathbb{R}^2$ given by $R = (0,1) \times (3,5)$ and $\mid R \mid = 2$ denote its area. Let (X,Y) have a joint density $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x,y) = \begin{cases} \frac{1}{2} & \text{if } (x,y) \in R \\ 0 & \text{otherwise.} \end{cases}$$

The above is clearly a density function. So for any recntangle $A = (a,b) \times (c,d) \subset R$,

$$P((X,Y) \in A) = \int_c^d \int_a^b f(x,y) dx dy = \frac{(b-a)(d-c)}{2} = \frac{\mid A \mid}{\mid R \mid}.$$

■

In general one can use the following definition to define a uniform random variable on the plane.

DEFINITION 5.4.4. *Let $D \subset \mathbb{R}^2$ be non-empty and with positive area (assume D is a Borel set or in particular f or any simple region whose area is well defined). Then $(X,Y) \sim$Uniform (D) if it has a joint probability density function given by $f : \mathbb{R}^2 \to \mathbb{R}$ given by*

$$f(x,y) = \begin{cases} \frac{1}{|D|} & \textit{if } (x,y) \in D \\ 0 & \textit{otherwise,} \end{cases}$$

where $\mid D \mid$ denotes the area of D.

When $(X,Y) \sim$ Uniform (D) then the probability that (X,Y) lies in a region $A \subset D$ is proportional to the area of A.

EXAMPLE 5.4.5. Let (X,Y) have a joint density $f : \mathbb{R}^2 \to \mathbb{R}$ given by

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

We note that this really does describe a density. The function $f(x,y)$ is non-negative and

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\, dx\, dy &= \int_0^1 \int_0^1 x + y\, dx\, dy \\ &= \int_0^1 (\frac{1}{2}x^2 + xy) \mid_{x=0}^{x=1} dy \\ &= \int_0^1 \frac{1}{2} + y\, dy \\ &= \frac{1}{2}y + \frac{1}{2}y^2 \mid_{y=0}^{y=1} = 1. \end{aligned}$$

Calculating a probability such as $P((X < \frac{1}{2}) \cap (Y < \frac{1}{2}))$ requires integrating over the appropriate region.

$$\begin{aligned} P((X < \frac{1}{2}) \cap (Y < \frac{1}{2})) &= \int_{-\infty}^{1/2} \int_{-\infty}^{1/2} f(x,y)\,dx\,dy \\ &= \int_{0}^{1/2} \int_{0}^{1/2} x + y\,dx\,dy \\ &= \int_{0}^{1/2} \frac{1}{8} + \frac{1}{2} y\,dy \\ &= \frac{1}{8}. \end{aligned}$$

A probability only involving one variable may still be calculated from the joint density. For instance $P(X < \frac{1}{2})$ does not appear to involve Y, but this simply means that Y is unrestircted and the corresponding integral should range over all possible values of Y. Therefore,

$$\begin{aligned} P(X < \frac{1}{2}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{1/2} f(x,y)\,dx\,dy \\ &= \int_{0}^{1} \int_{0}^{1/2} x + y\,dx\,dy = \frac{3}{8}. \end{aligned}$$

It is just as easy to compute that $P(Y < \frac{1}{2}) = \frac{3}{8}$. Note that these computations also demonstrate that X and Y are not independent since

$$P(X < \frac{1}{2}) \cdot P(Y < \frac{1}{2}) = \frac{9}{64} \neq P((X < \frac{1}{2}) \cap (Y < \frac{1}{2})).$$

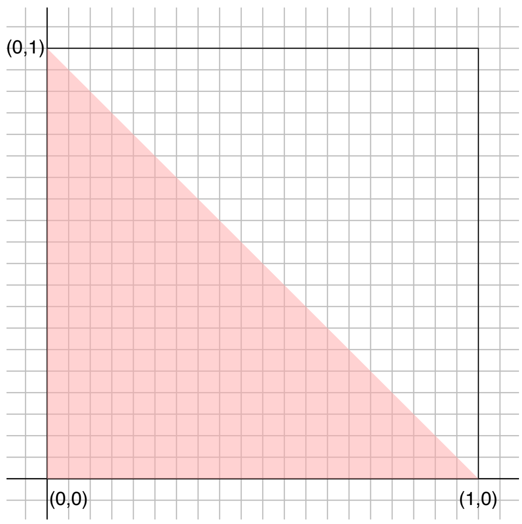

Figure 5.12: The subset A of the unit square that represents the region $x + y < 1$.

A probability such as $P(X+Y<1)$ can be found by integrating over a non-rectangular region in the plane, as shown in Figure 5.12. Let $A = \{(x,y) | x+y<1\}$. Then

$$\begin{aligned}
P(X+Y<1) &= \int_A f(x,y)\,dx\,dy \\
&= \int_0^1 \int_0^{1-y} x+y\,dx\,dy \\
&= \int_0^1 \frac{1}{2}x^2 + xy \,\Big|_0^{1-y}\; dy \\
&= \int_0^1 \frac{1}{2}(1-y)^2 + (1-y)y\,dy \\
&= \int_0^1 \frac{1}{2} - \frac{1}{2}y^2\,dy \\
&= \frac{1}{3}.
\end{aligned}$$

■

5.4.1 *Marginal Distributions*

As in the discrete case, when we begin with the joint density of many random variables, but want to speak of the distribution of an individual variable we will frequently refer to it as a "marginal distribution" .

Suppose (X,Y) are random variables and have a joint probability density function $f : \mathbb{R}^2 \to \mathbb{R}$. Then we obseve that

$$P(X \le x) = P(X \le x, -\infty < Y < \infty) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u,y)dydu.$$

If $g : \mathbb{R} \to \mathbb{R}$ is given by

$$g(u) = \int_{-\infty}^{\infty} f(u,y)dy$$

then

$$P(X \le x) \int_{-\infty}^{x} g(u)du.$$

Using Theorem 5.2.5, by the continuity assumptions on f, we find that the random variable X is also a continuous random variable with probability density function of X given by

$$f_X(x) = g(x) = \int_{-\infty}^{\infty} f(x,y)dy. \tag{5.4.2}$$

As it was derived from a joint probability density function, the density of X is referred to as the marginal density of X. Similarly one can show that Y is also a continuous random variable and its marginal density is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx. \tag{5.4.3}$$

Example 5.4.6. (Example 5.4.3 contd.) Going back to Example 5.4.3, we can compute the marginal density of X and Y. The marginal density of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \begin{cases} \int_3^5 \frac{1}{2} & \text{if } 0<x<1 \\ 0 & \text{otherwise.} \end{cases} = \begin{cases} 1 & \text{if } 0<x<1 \\ 0 & \text{otherwise.} \end{cases}$$

The marginal density of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \begin{cases} \int_0^1 \frac{1}{2} & \text{if } 3 < y < 5 \\ 0 & \text{otherwise.} \end{cases} = \begin{cases} \frac{1}{2} & \text{if } 3 < y < 5 \\ 0 & \text{otherwise.} \end{cases}$$

So we observe that $X \sim$ Uniform $(0,1)$ and $Y \sim$ Uniform $(3,5)$. ■

While it is routine to find the marginal densities from the joint density there is no standard way to get to the joint from the marginals. Part of the reason for this difficulty is that the marginal desnitieis offer no information about how the varaibles relate to each other, which is critical information for determining how they behave jointly. However, in the case that the random variables happen to be independent there is a convenient relationship between the joint and marginal densities.

5.4.2 *Independence*

THEOREM 5.4.7. *Let f be the joint density of random variables X and Y and let f_X and f_Y be the respective marginal densities. Then*

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

if and only if X and Y are independent.

Proof - First suppose X and Y are independent and consider the quantity $P((X \le x) \cap (Y \le y))$. On one hand independnece gives

$$P((X \le x) \cap (Y \le y)) = P(X \le x)P(Y \le y) = F_X(x)F_Y(y) \tag{5.4.4}$$

On the other hand, integrating the joint density yields

$$P((X \le x) \cap (Y \le y)) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x,y)\, dx\, dy. \tag{5.4.5}$$

Since equations 5.4.4 and 5.4.5 are equal we may differentiate both with respect to each of the variables x and y and they remain equal. However, differentiating the former gives $f_X(x)f_Y(y)$ because of the relationship between the distribution and the density, while differentiating the latter yields $f(x,y)$ by a two-fold application of the fundamental theorem of calculus.

To prove the opposite direction, suppose $f(x,y) = f_X(x)f_Y(y)$. Let A and B be Borel sets in $\mathbb{R}$. Then

$$\begin{aligned} P((X \in A) \cap (Y \in B)) &= \int_B \int_A f(x,y)\, dx\, dy \\ &= \int_B \int_A f_X(x) f_Y(y)\, dx\, dy \\ &= \left(\int_A f_X(x)\, dx \right) \left(\int_B f_Y(y)\, dy \right) \\ &= P(X \in A)P(Y \in B) \end{aligned}$$

Since this is true for all sets such sets A and B, the variables X and Y are independent. ■

EXAMPLE 5.4.8. (Example 5.4.3 contd.) We had observed that if $(X,Y) \sim$ Uniform (R) then $X \sim$ Uniform $(0,1)$ and $Y \sim$ Uniform $(3,5)$. Note further that

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

for all $x, y \in R$. Consequently X, Y are independent as well. ■

It is tempting to generalise and say that $(X, Y) \sim$ Uniform (D) for a region D with non-trivial area then X and Y would be independent. This is not the case, we illustrate in the example below.

Example 5.4.9. Consider the open disk in $\mathbb{R}^2$ given by $C = \{(x, y) : x^2 + y^2 < 25\}$ and $| C |= 25\pi$ denote its area. Let (X, Y) have a joint density $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x, y) = \begin{cases} \frac{1}{|C|} & \text{if } (x, y) \in C \\ 0 & \text{otherwise.} \end{cases}$$

As before for any Borel $A \subset C$,

$$P((X, Y) \in A) = \frac{| A |}{| C |},$$

and the probability that (X, Y) lies in A is proportional to the area of A. However the marginal density calculation is a little different. The marginal density of X is given by

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x, y)dy = \begin{cases} \displaystyle\int_{-\sqrt{25-x^2}}^{\sqrt{25-x^2}} \frac{1}{| C |}\, dy & \text{if } -5 < x < 5 \\ 0 & \text{otherwise.} \end{cases} \\ &= \begin{cases} \frac{2}{25\pi}\sqrt{25 - x^2} & \text{if } -5 < x < 5 \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

The distribution of X is the Semi-circular law described in Exercise 5.2.6. As the joint density f is symmetric in x and y (i.e $f(x, y) = f(y, x)$) the marginal density of Y is the same as that of X (why ?). It is easy to see

$$\frac{1}{25\pi} = f(0, 0) \neq f_X(0)f_Y(0) = \frac{4}{25\pi^2}$$

Consequently X, Y are not independent. This fact should make intuitive sense as well, for if X happens to take a value near 5 or -5 the range of possible values of Y is much more restricted than if X takes a value near 0. ■

We shall see the utility of independence when computing distributions of various functions of independent random variables (see Section 5.5). Independence of random variables also makes it easier to compute their joint density and hence probabilites. For instance, consider the following example.

Example 5.4.10. Suppose $X \sim$ Exponential(λ_1), $Y \sim$ Exponential(λ_2) are independent random variables. Find $P(X - Y < 0)$.

The joint density of (X, Y) is given by

$$f(x, y) = f_X(x)f_Y(y) = \begin{cases} \lambda_1\lambda_2 e^{-(\lambda_1 x + \lambda_2 y)} & \text{if } x > 0 \text{ and } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

Therefore

$$\begin{aligned}
P(X - Y < 0) &= \int_0^\infty \int_0^y \lambda_1\lambda_2 e^{-(\lambda_1 x + \lambda_2 y)} dx dy = \lambda_1\lambda_2 \int_0^\infty e^{-\lambda_2 y}[\int_0^y e^{-\lambda_1 x} dx] dy \\
&= \lambda_1\lambda_2 \int_0^\infty e^{-\lambda_2 y} \frac{1}{\lambda_1}[1 - e^{-\lambda_1 y}] dy \\
&= \lambda_2 \left[\int_0^\infty e^{-\lambda_2 y} - e^{-(\lambda_1+\lambda_2)y} dy\right] \\
&= \lambda_2 \left[\frac{-1}{\lambda_2}(e^{-\lambda_2 y} \mid_0^\infty) + \frac{1}{\lambda_1+\lambda_2}(e^{-(\lambda_1+\lambda_2)y} \mid_0^\infty)\right] \\
&= \lambda_2 \left[\frac{1}{\lambda_2} - \frac{1}{\lambda_1+\lambda_2}\right] \\
&= \frac{\lambda_1}{\lambda_1+\lambda_2}.
\end{aligned}$$

Similarly one can also compute $P(Y - X < 0) = \frac{\lambda_2}{\lambda_1+\lambda_2}$. This fact is quite useful when using exponential random variables to model waiting times, for $P(X - Y < 0) = P(X < Y)$, so we have determined the probability that one waiting time will be shorter than another. ■

5.4.3 *Conditional Density*

In Section 3.2.2 we have seen the notion of conditional distributions for discrete random variables and in Section 4.4 we have seen the notions of conditional expectation and variance for discrete random variables. Suppose X measures the parts per million of a particulate matter less than 10 microns in the air and Y is the incidence rate of asthma in the population. It is clear that X and Y ought to be related; for the distribution of one affects the other. Towards this, in this section we shall discuss conditional distributions for two continuous random variables having a joint probability density function. We recall from Definition 3.2.5 that if X is a random variable on a sample space S and $A \subset S$ be an event such that $P(A) > 0$, then the probability Q described by

$$Q(B) = P(X \in B|A)$$

is called the conditional distribution of X given the event A.

Suppose X and Y have a joint probability density function f. Given our discussion for discrete random variables it is natural to characterise the conditional distribution of X given some information on Y. In the discrete setting we typically considered an event $A = \{Y = b\}$ for some real number b in the range of Y. In the continuous setting such an event A would have zero probability, so the usual way of conditioning on an event would not be possible. However, there is a way to make such a conditioning meaningful and precise provided $f_Y(b) > 0$, where f_Y is the marginal density of Y.

Suppose we wish to find the following :

$$P(X \in [3, 4] \mid Y = b).$$

We shall argue heuristically and arrive at an expression for the above probability. Suppose the marginal density of X is $f_X(\cdot)$, and that of Y is $f_Y(\cdot)$. Assume first that f_Y is piecewise continuous and $f_Y(b) > 0$. Then it is a standard fact from real analysis to see that

$$P(Y \in [b, b + \frac{1}{n})) > 0,$$

for all $n \geq 1$. One can then view the conditional probability as before, that is

$$\begin{aligned} P(X \in [3,4] \mid X \in [b, b+\frac{1}{n})) &= \frac{P(X \in [3,4] \cap X \in [b, b+\frac{1}{n}))}{P(X \in [b, b+\frac{1}{n}))} \\ &= \frac{\int_3^4 \left(\int_b^{b+\frac{1}{n}} f(u,v)du \right) dv}{\int_b^{b+\frac{1}{n}} f_X(u)du} \\ &= \frac{\int_3^4 \left(n \int_b^{b+\frac{1}{n}} f(u,v)du \right) dv}{n \int_b^{b+\frac{1}{n}} f_X(u)du} \end{aligned}$$

From facts in real analysis (under some mild assumptions on f) the following can be established,

$$\lim_{n\to\infty} n \int_b^{b+\frac{1}{n}} f(u,v)du = f(b,v),$$

for all real numbers v and

$$\lim_{n\to\infty} n \int_b^{b+\frac{1}{n}} f_X(u)du = f_X(b).$$

We have seen earlier (see Exercise 1.1.13 (b))

$$\lim_{n\to\infty} P(Y \in [b, b+\frac{1}{n})) = P(Y = b).$$

Hence it would be reasonable to argue that $P(X \in [3,4] \mid Y = b)$ ought to be defined as

$$P(X \in [3,4] \mid Y = b) = \frac{\int_3^4 f(b,v)dv}{f_Y(b)}.$$

With the above motivation we are now ready to define conditional densities for two random variables.

Definition 5.4.11. *Let (X,Y) be random variables having joint density f. Let the marginal density of Y be $f_Y(\cdot)$. Suppose b is a real number such that $f_Y(b) > 0$ and is continuous at b then conditional density of X given $Y = b$ is given by*

$$f_{X|Y=b}(x) = \frac{f(x,b)}{f_Y(b)} \tag{5.4.6}$$

for all real numbers x. Similarly, let the marginal density of X be $f_X(\cdot)$. Suppose a is a real number such that $f_X(a) > 0$ and is continuous at a then conditional density of Y given $X = a$ is given by

$$f_{Y|X=a}(y) = \frac{f(a,y)}{f_X(a)}$$

for all real numbers y.

This definition genuinely defines a probability density function, for $f_{X|Y=b}(x) \geq 0$ since it is the ratio of a non-negative quantity and a positive quantity. Moreover,

$$\begin{aligned}\int_{-\infty}^{\infty} f_{X|Y=b}(x)dx &= \int_{-\infty}^{\infty} \frac{f(x,b)}{f_Y(b)}dx \\ &= \frac{1}{f_Y(b)}\int_{-\infty}^{\infty} f(x,b)dx = \frac{1}{f_Y(b)}f_Y(b) = 1\end{aligned}$$

Note that if X and Y are independent then

$$f_{X|Y=b}(x) = \frac{f(x,b)}{f_Y(b)} = \frac{f_X(x)f_Y(b)}{f_Y(b)} = f_X(x).$$

One can use the conditional density to compute the conditional probabilities, namely if (X,Y) are random variables having joint density f and b is a real number such that its marginal density has the property $f_Y(b) > 0$ then

$$P(X \in A \mid Y = b) = \int_A f_{X|Y=b}(x)dx = \int_A \frac{f(x,b)}{f_Y(b)}dx.$$

We conclude this section with two examples where we compute conditional densities. In both the examples the dependencies between the random variables imply that the conditional distributions are different from the marginal distributions.

Example 5.4.12. Let (X,Y) have joint probability density function f given by

$$f(x,y) = \frac{\sqrt{3}}{4\pi}e^{-\frac{1}{2}(x^2-xy+y^2)} \quad -\infty < x,y < \infty.$$

Let $x \in \mathbb{R}$, then the marginal density of X at x is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{-\infty}^{\infty} \frac{\sqrt{3}}{4\pi}e^{-\frac{1}{2}(x^2-xy+y^2)}dy$$

By a standard completing the square computation, $\frac{1}{2}(x^2-xy+y^2) = \frac{3x^2}{8} + \frac{1}{2}(y-\frac{x}{2})^2$. Therefore,

$$f_X(x) = \frac{\sqrt{3}}{4\pi}e^{-\frac{3x^2}{8}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(y-\frac{x}{2})^2}dy$$

Observing that $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(y-\frac{x}{2})^2}dy = 1$ (why ?), we have

$$f_X(x) = \frac{\sqrt{3}}{4\pi}e^{-\frac{3x^2}{8}}\sqrt{2\pi} = \sqrt{\frac{3}{4}}\frac{1}{\sqrt{2\pi}}e^{-\frac{3x^2}{8}}$$

Hence X is a Normal random variable with mean 0 and variance $\frac{4}{3}$. By symmetry (or calculating similarly as above) we can also show that Y is a Normal random variable with mean 0 and variance $\frac{4}{3}$. Also, we can easily see that

$$f_X(x)f_Y(y) = \frac{3}{8\pi}e^{-\frac{3}{8}(x^2+y^2)} \neq \frac{\sqrt{3}}{4\pi}e^{-\frac{1}{2}(x^2-xy+y^2)} = f(x,y)$$

for many $x,y \in \mathbb{R}$. Hence X and Y are not independent. Note that $f_X(x) \neq 0$ for all real numbers x and is continuous at all $x \in \mathbb{R}$. Fix $x \in \mathbb{R}$, the conditional density of Y given $X = x$ is given by

$$f_{Y|X=x}(y) = \frac{f(x,y)}{f_X(x)} = \frac{\frac{\sqrt{3}}{4\pi}e^{-\frac{1}{2}(x^2-xy+y^2)}}{\sqrt{\frac{3}{4}}\frac{1}{\sqrt{2\pi}}e^{-\frac{3x^2}{8}}} = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(y-\frac{x}{2})^2} \quad \forall\, y \in \mathbb{R}.$$

Hence though the marginal distribution of Y is Normal$(0, \frac{4}{3}$, the)the conditional distribution of Y given $X = x$ is Normal with mean $\frac{x}{2}$ and variance 1. Put another way, if we are given that $X = x$ the mean of Y changes from 0 to x and the variance reduces from $\frac{4}{3}$ to 1.

Such a pair (X, Y) is an example of a bivariate normal random variable and will be discussed in detail in Section 6.4. ■

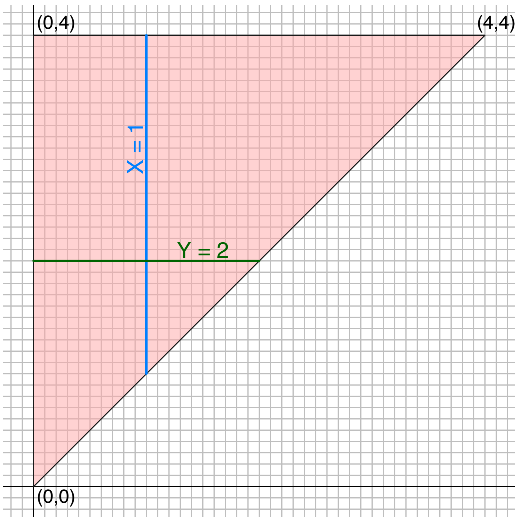

Figure 5.13: The region $T = \{(x, y) \mid 0 < x < y < 4\}$ from Example 5.4.13.

EXAMPLE 5.4.13. Suppose $T = \{(x, y) \mid 0 < x < y < 4\}$ and let $(X, Y) \sim$ Uniform (T). Therefore its joint density is given by (see Figure 5.13)

$$f(x, y) = \begin{cases} \frac{1}{8} & \text{if } (x, y) \in T \\ 0 & \text{otherwise.} \end{cases}$$

The marginal density of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy = \begin{cases} \int_x^4 \frac{1}{8}dy & \text{if } 0 < x < 4 \\ 0 & \text{otherwise.} \end{cases} = \begin{cases} \frac{4-x}{8} & \text{if } 0 < x < 4 \\ 0 & \text{otherwise.} \end{cases}$$

The marginal density of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx = \begin{cases} \int_0^y \frac{1}{8}dy & \text{if } 0 < y < 4 \\ 0 & \text{otherwise.} \end{cases} = \begin{cases} \frac{y}{8} & \text{if } 0 < y < 4 \\ 0 & \text{otherwise.} \end{cases}$$

Let us fix $0 < b < 4$. So $f_Y(\cdot)$ is non-zero at b and is continuous at b. The conditional density of $(X \mid Y = b)$ is given by

$$f_{X|Y=b}(x) = \frac{f(x, b)}{f_Y(b)} = \begin{cases} \frac{1/8}{b/8} & \text{if } 0 < x < b \\ 0 & \text{otherwise.} \end{cases} = \begin{cases} \frac{1}{b} & \text{if } 0 < x < b \\ 0 & \text{otherwise.} \end{cases}$$

Therefore $(X \mid Y = b) \sim$ Uniform $(0, b)$. Similarly if we fix $0 < a < 4$, we observe $f_X(\cdot)$ is non-zero at a and is continuous at a. The conditional density of $(Y \mid X = a)$ is given by

$$f_{Y|X=a}(y) = \frac{f(a, y)}{f_X(a)} = \begin{cases} \frac{1/8}{(4-a)/8} & \text{if } a < y < 4 \\ 0 & \text{otherwise.} \end{cases} = \begin{cases} \frac{1}{4-a} & \text{if } a < y < 4 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore $(Y \mid X = a) \sim$ Uniform $(a, 4)$.

Clearly X and Y are continuous random variables with distributions that are not uniform, but the conditional distributions turn out to be uniform. ■

EXERCISES

Ex. 5.4.1. Let (X, Y) be random variables whose probability density function is given by $f : \mathbb{R}^2 \to \mathbb{R}$. Find the probability density function of X and probability density function of Y in each of the following cases:-

(a) $f(x, y) = (x + y)$ if $0 \leq x \leq 1, 0 \leq y \leq 1$ and 0 otherwise

(b) $f(x, y) = 2(x + y)$ if $0 \leq x \leq y \leq 1$ and 0 otherwise

(c) $f(x, y) = 6x^2 y$ if $0 \leq x \leq 1, 0 \leq y \leq 1$ and 0 otherwise

(d) $f(x, y) = 15x^2 y$ if $0 \leq x \leq y \leq 1$ and 0 otherwise

Ex. 5.4.2. Let $c > 0$. Suppose that X and Y are random variables with joint probability density

$$f(x, y) = \begin{cases} c(xy + 1) & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find c.

(b) Compute the marginal densities $f_X(\cdot)$ and $f_Y(\cdot)$ and the conditional density $f_{X|Y=b}(\cdot)$

Ex. 5.4.3. Let $A = \{(x, y) \in \mathbb{R}^2 : x > 0, y > 0, x + y < 1\}$ and let X and Y be random variables defined by the joint density $f(x, y) = 24xy$ if $(x, y) \in A$ (and $f(x, y) = 0$ otherwise).

(a) Verify the claim that $f(x, y)$ is a density.

(b) Show that X and Y are dependent random variables.

(c) Explain why (b) doesn't violate Theorem 5.4.7 despite the fact that $24xy$ is a product of a function of x with a function of y.

Ex. 5.4.4. Consider the set $D = [-1, 1] \times [-1, 1]$. Let

$$L = \{(x, y) \in D : x = 0 \text{ or } \text{ or } x = -1 \text{ or } x = 1 \text{ or } y = 0 \text{ or } y = 1 \text{ or } y = -1\}$$

be the lines that create a tiling of D. Suppose we drop a coin of radius R at a uniformly chosen point in D what is the probability that it will intersect the set L ?

Ex. 5.4.5. Let X and Y be two independent uniform $(0, 1)$ random variables. Let $U = \max(X, Y)$ and $V = \min(X, Y)$.

(a) Find the joint distribution of U, V.

(b) Find the conditional distribution of $(V \mid U = 0.5)$

Ex. 5.4.6. Suppose X is a random variable with density

$$f(x) = \begin{cases} cx^2(1 - x) & \text{for } 0 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find:

(a) the value of c.

(b) the distribution function of X.

(c) the conditional probability $P(X > 0.2 \mid X < 0.5)$.

Ex. 5.4.7. Suppose $g : \mathbb{R} \to \mathbb{R}$ be a continuous probability density function, such that $g(x) = 0$ when $x \notin [0, 1]$. Let $D \subset \mathbb{R}^2$ be given by

$$D = \{(x, y) : x \in \mathbb{R} \text{ and } 0 \leq y \leq g(x)\}$$

Let (X, Y) be uniformly distributed on D. Find the probability density function of X.

Ex. 5.4.8. Continuous random variables X and Y have a joint density

$$f(x,y) = \begin{cases} \frac{1}{24}, & \text{for } 0 < x < 6, 0 < y < 4 \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find $P(2Y > X)$.

(b) Are X and Y independent?

Ex. 5.4.9. Let

$$f(x,y) = \begin{cases} \eta(y-x)^{\gamma} & \text{if } 0 \leq x < y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) For what values of γ can η be chosen so that f be a joint probability density function of X, Y.

(b) Given a γ from part (a), what is the value of η ?

(c) Given a γ and η from parts (a) and (b), find the marginal densities of X and Y.

Ex. 5.4.10. Let $D = \{(x, y) : x^3 \leq y \leq x\}$. A point (X, Y) is chosen uniformly from D. Find the joint probability density function of X and Y.

Ex. 5.4.11. Let X and Y be two random variables with the joint p.d.f given by

$$f(x,y) = \begin{cases} ae^{-by} & 0 \leq x \leq y \\ \\ 0 & \text{otherwise} \end{cases}$$

Find a conditions on a and b that make this a joint probability density function.

Ex. 5.4.12. Suppandi and Meera plan to meet at Gopalan Arcade between 7pm and 8pm. Each will arrive at a time (independent of each other) uniformly between 7pm and 8pm and will wait for 15 minutes for the other person before leaving. Find the probability that they will meet ?

5.5 FUNCTIONS OF INDEPENDENT RANDOM VARIABLES

In Section 5.3 we have seen how to compute the distribution of $Y = g(X)$ from the distribution of X for various $g : \mathbb{R}^2 \to \mathbb{R}$. Suppose (X, Y) are random variables having a joint probability density function $f : \mathbb{R}^2 \to \mathbb{R}$. Let $h : \mathbb{R}^2 \to \mathbb{R}$. A natural follow up objective is then to determine the distribution of

$$Z = h(X, Y).$$

In Section 3.3 we discussed an approach to this question when the random variables where discrete.

One could prove a result as attained in Exercise 5.3.10 for functions of two variables but this will require knowledge of Linear Algebra and multivariable calculus. Here we limit our objective and shall focus on two specific functions namely the sum and the product.

5.5.1 *Distributions of Sums of Independent Random variables*

Let X and Y be two independent continous random variables with densities f_X and f_Y. In this section we shall see how to compute the distribution of $Z = X + Y$. We first prove a proposition that describes the probability density function of Z.

PROPOSITION 5.5.1. *(Sum of two independent random variables) Let X and Y be two independent random variables with marginal densities given by $f_X : \mathbb{R} \to \mathbb{R}$ and $f_Y : \mathbb{R} \to \mathbb{R}$. Then $Z = X + Y$ has a probability density function $f_Z : \mathbb{R} \to \mathbb{R}$ given by*

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx. \tag{5.5.1}$$

Proof- Let us first find an expression for the distribution function of Z.

$$\begin{aligned}
F(z) &= P(Z \le z) \\
&= P(X + Y \le z) \\
&= \int\int_{\{(x,y):x+y\le z\}} f_X(x) f_Y(y) dy dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_X(x) f_Y(y) dy dx \\
&= \int_{-\infty}^{z} [\int_{-\infty}^{\infty} f_X(x) f_Y(u - x) dx] du.
\end{aligned}$$

As $f_X(\cdot)$ and $f_Y(\cdot)$ are densities, it can be shown that the integrand is a piecewise continuous function. Hence F is of the form (5.2.4) and Theorem 5.2.5 implies that the probability density function of Z is given by (5.5.1). ■

The integral expression on the right hand side of (5.5.1) is referred to as the convolution of f_X and f_Y and is denoted by $f_X \star f_Y(z)$. It is a property of convolutions that $f_X \star f_Y(z) = f_Y \star f_X(z)$ for all $z \in \mathbb{R}$. Thus if we view the sum of X and Y as $Z = X + Y$ or $Z = Y + X$ the distribution will be the same (See Exercise 5.5.8).

EXAMPLE 5.5.2. (Sum of Uniforms) Let X and Y be two independent Uniform $(0, 1)$ random variables. Let $Z = X + Y$. From the above proposition that Z has a density given by (5.5.1). Note that

$$f_X(x) f_Y(z - x) = \begin{cases} 1 & \text{if } 0 < x < 1, 0 < z - x < 1 \text{ and } 0 < z < 2 \\ 0 & \text{otherwise} \end{cases}$$

Therefore $f_X(x) f_Y(z - x)$ is non-zero if and only if $\max\{0, z - 1\} < x < \min\{1, z\}, 0 < z < 2$. So for $0 < z < 2$,

$$f_Z(z) = \int_{\max\{0,z-1\}}^{\min\{1,z\}} f_X(x) f_Y(z - x) dx = \int_{\max\{0,z-1\}}^{\min\{1,z\}} 1 dx = \min\{1, z\} - \max\{0, z - 1\}.$$

Therefore,

$$f_Z(z) = \begin{cases} \min\{1, z\} - \max\{0, z - 1\} & \text{if } 0 < z < 2 \\ 0 & \text{otherwise} \end{cases} = \begin{cases} z & \text{if } 0 < z \le 1 \\ 2 - z & \text{if } 1 < z < 2 \\ 0 & \text{otherwise} \end{cases}$$

A graph of this density is displayed in Figure 5.14. ■

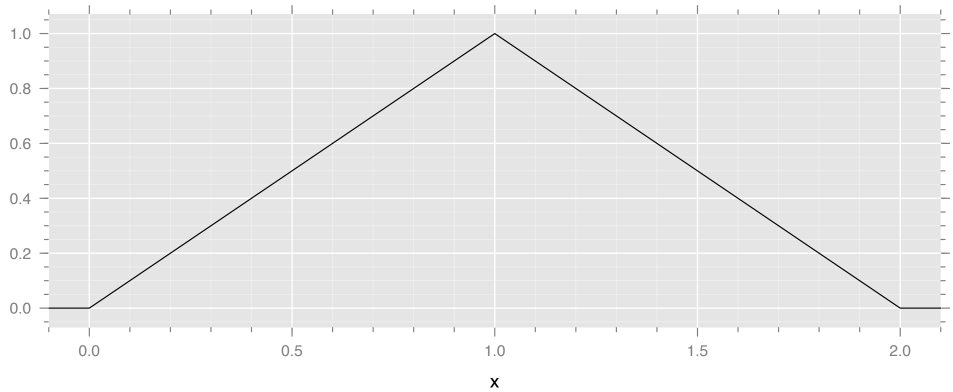

Figure 5.14: The region $T = \{(x, y) \mid 0 < x < y < 4\}$ from Example 5.5.2.

Our next example will deal with sum of two independent exponential random variables. This will lead us to the Gamma distribution which is of significant interest in statistics.

EXAMPLE 5.5.3. (Sum of Exponentials) Let $\lambda > 0$, X and Y be two independent Exponential (λ) random variables. Let $Z = X + Y$. Then we know and Z has a density given by (5.5.1). Further,

$$f_X(x)f_Y(z-x) = \begin{cases} \lambda^2 e^{-\lambda x} e^{-\lambda(z-x)} & \text{if } x \geq 0, z - x \geq 0 \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \lambda^2 e^{-\lambda z} & \text{if } x \geq 0, x \leq z, z \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Hence $f_X(x)f_Y(z-x)$ is non-zero if and only if $0 \leq x \leq z$. So

$$f_Z(z) \quad = \quad \int_0^z f_X(x) f_Y(z-x)dx = \lambda^2 e^{-\lambda z} \int_0^z 1dx = \lambda^2 z e^{-\lambda z},$$

for $z \geq 0$ and $f_Z(z) = 0$ otherwise. This is known as Gamma $(2, \lambda)$ distribution. ■

Before we define the Gamma distribution more generally we prove a lemma in real analysis, the proof of which can be skipped upon first reading.

LEMMA 5.5.4. *For $n \geq 1$, and $\lambda > 0$,*

$$\int_0^\infty x^{n-1} e^{-\lambda x} = \frac{(n-1)!}{\lambda^n} \tag{5.5.2}$$

Proof- For all $n \geq 1, \lambda > 0, a > 0$ define $u : [0, a] \to \mathbb{R}$ and $v : [0, a] \to \mathbb{R}$ by

$$u(x) = x^{n-1} \text{ and } v(x) = e^{-\lambda x}.$$

As u, v are continuous functions, clearly $I^a_{n,\lambda}$ given by

$$I^a_{n,\lambda} = \int_0^a x^{n-1} e^{-\lambda x}.$$

is well defined finite positive number. As $x^\alpha e^{-\beta x} \to 0$ as $x \to \infty$ for any $\alpha, \beta > 0$ there is a $K > 0$ such that

$$0 \leq x^{n-1} e^{-\lambda x} < e^{-\frac{\lambda x}{2}},$$

for all $K > 0$. Therefore $b > a > k$ we have

$$| I^a_{n,\lambda} - I^b_{n,\lambda} |= \int_a^b x^{n-1}e^{-\lambda x} \leq \int_a^b e^{-\frac{\lambda x}{2}} dx = 2(e^{-\frac{\lambda b}{2}} - e^{-\frac{\lambda a}{2}}).$$

From this it is standard to note that

$$I_{n,\lambda} := \int_0^\infty x^{n-1}e^{-\lambda x} = \lim_{a\to\infty} I^a_{n,\lambda}$$

is a well defined finite positive number. Now, as u, v are differentiable we have by the integration by parts formula

$$\int_0^a u(x)v'(x)dx = u(a)v(a) - u(0)v(0) - \int_0^a u'(x)v(x)dx.$$

Substituting for u, v above we get

$$-\lambda I^a_{n,\lambda} = a^{n-1}e^{-\lambda a} - (n-1)I^a_{n-1,\lambda}.$$

Taking limits as $a \to \infty$ we have

$$\lambda I_{n,\lambda} = (n-1)I_{n-1,\lambda}.$$

Applying the above inductively we have

$$I_{n,\lambda} = \prod_{i=1}^{n-1} \frac{(n-i)}{\lambda} I_{1,\lambda} = \frac{(n-1)!}{\lambda^{n-1}} I_{1,\lambda}.$$

Using the fact that $I_1 = \frac{1}{\lambda}$ we have the result. ■

DEFINITION 5.5.5. $X \sim$ **Gamma**(n, λ): *Let $\lambda > 0$ and $n \in \mathbb{N}$. Then X is said to be Gamma distributed with parameters n and λ if it has the density*

$$f(x) = \frac{\lambda^n}{(n-1)!}x^{n-1}e^{-\lambda x}, \tag{5.5.3}$$

where $x \geq 0$. The parameter n is referred to as the shape parameter and λ as the rate parameter. By (5.5.2) we know that f given by (5.5.3) is a density function.

We saw in Example 5.5.3 that sum of two exponential distributions resulted in a gamma distribution. If $X \sim$ Exponential (λ) then it can also be viewed as a Gamma$(1, \lambda)$ distribution. The result in Example 5.5.3 could be rephrased as follows: the sum of two gamma random variables with shape parameter 1 and rate parameter λ is distributed as a gamma random variable with shape parameter 2 and rate parameter λ. This holds more generally as we show in the next example.

EXAMPLE 5.5.6. (Sum of Gammas) Let $n \in \mathbb{N}, m \in \mathbb{N}, \lambda > 0$, X and Y be two independent Gamma(n, λ) and Gamma(m, λ) random variables respectively. Let $Z = X + Y$. Then we know that Z has a density given by (5.5.1). Further,

$$\begin{aligned} f_X(x)f_Y(z-x) &= \begin{cases} \frac{\lambda^n}{(n-1)!}x^{n-1}e^{-\lambda x}\frac{\lambda^m}{(m-1)!}(z-x)^{m-1}e^{-\lambda(z-x)} & \text{if } x \geq 0, z-x \geq 0 \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} \frac{e^{-\lambda z}\lambda^{n+m}}{(n-1)!(m-1)!}x^{n-1}(z-x)^{m-1} & \text{if } x \geq 0, x \leq z, z \geq 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

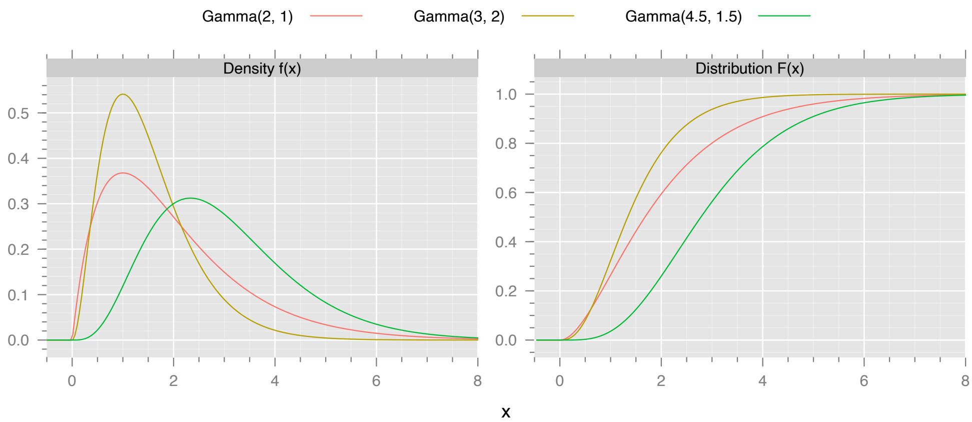

Figure 5.15: The Gamma density and cumulative distribution functions for various shape and rate parameters.

For $z \geq 0$, we have

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(z-x)dx = \int_0^z f_{X_1}(x) f_{X_2}(z-x)dx \\ &= \frac{e^{-\lambda z}\lambda^{n+m}}{(n-1)!(m-1)!}\int_0^z x^{n-1}(z-x)^{m-1}dx \end{aligned}$$

We now make a change of variable $x = zu$ so that $dx = zdu$ to obtain

$$f_Z(z) = \frac{z^{n+m-1}e^{-\lambda z}\lambda^{n+m}}{(n-1)!(m-1)!}\int_0^1 u^{n-1}(1-u)^{m-1}du$$

Define

$$c(n,m) = \frac{\int_0^1 u^{n-1}(1-u)^{m-1}du}{(n-1)!(m-1)!}.$$

Thus we have the probability density of Z is given by,

$$f_Z(z) = \begin{cases} c(n,m)\cdot\lambda^{n+m}z^{n+m-1}e^{-\lambda z} & \text{if } z \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

To evaluate $c(n,m)$ we use the following fact. From Proposition 5.5.1 $f_Z(\cdot)$ (given by (5.5.1)) is a Probability density function. Therefore,

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f_Z(z)dz \\ &= c(n,m)\lambda^{n+m}\int_0^{\infty} z^{n+m-1}e^{-\lambda z}dz \\ &= c(n,m)[(n+m-1)!], \end{aligned}$$

where in the last line we have used (5.5.2) with n replaced by $n+m$. So $c(n,m) = \frac{1}{(n+m-1)!}$. Hence Z has Gamma $(n+m,\lambda)$ distribution. From the definition of $c(n,m)$ we also have

$$\int_0^1 u^{n-1}(1-u)^{m-1}du = \frac{(n+m-1)!}{(n-1)!(m-1)!}.$$

The above calculation is easily extended by an induction argument to obtain the fact that if $\lambda > 0$, X_i, $1 \leq i \leq n$ are independent Gamma(n_i, λ) distributed random variables (respectively). Then $Z = \sum_{i=1}^n X_i$ has Gamma $(\sum_{i=1}^n n_i, \lambda)$ distribution.

As Exponential (λ) is the same as Gamma$(1, \lambda)$ random variable, the above implies that the sum of n independent Exponential (λ) random variables is a Gamma(n, λ) random variable. ■

It is possible to define the Gamma distribution when the shape parameter is not necessarily an integer.

Definition 5.5.7. *$X \sim$ **Gamma**(α, λ): Let $\lambda > 0$ and $\alpha > 0$. Then X is said to be Gamma distributed with shape parameter α and rate parameter λ if it has the density*

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \tag{5.5.4}$$

where $x \geq 0$ and for $\alpha > 0$

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \tag{5.5.5}$$

One can imitate the calculation done in Example 5.5.6 as well for such a Gamma distribution.

The distribution function of a gamma random variable involves an indefinite form of the integral in (5.5.5). Such integrals are known as incomplete gamma functions, and have no closed-form solution in terms of simple functions. In R, $F(x)$ for the gamma distribution

$$F(x) = P(X \leq x) = \int_0^x \frac{\lambda^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\lambda z} dz\,, x > 0$$

can be evaluated numerically with a function call of the form `pgamma(x, alpha, lambda)`. For example,

```
> pgamma(1, 2, 1)
[1] 0.2642411
> pgamma(3, 4.5, 1.5)
[1] 0.5627258
```

Similarly, the density function $f(x)$ in (5.5.4) involves the normalising constant $\Gamma(\alpha)$ (also known as the gamma function) which usually cannot be computed explicitly when α is not an integer. Using R, one can evaluate $f(x)$ numerically using the `dgamma()` function as

```
> dgamma(1, 2, 1)
[1] 0.3678794
> dgamma(3, 4.5, 1.5)
[1] 0.2769272
```

5.5.2 *Distributions of Quotients of Independent Random varibles.*

Let X and Y be two independent continous random variables with densities f_X and f_Y. In this section we shall find out the probability density function of $Z = \frac{X}{Y}$. As $P(Y = 0) = 0$, Z is well defined random variable.

Proposition 5.5.8. *(Quotient of two independent random variables) Let X and Y be two independent random variables with marginal densities given by $f_X : \mathbb{R} \to \mathbb{R}$ and $f_Y : \mathbb{R} \to \mathbb{R}$. Then $Z = \frac{X}{Y}$ has a probability density function $f_Z : \mathbb{R} \to \mathbb{R}$ given by*

$$f_Z(z) = \int_{-\infty}^{\infty} \mid y \mid f_X(zy) f_Y(y) dy. \tag{5.5.6}$$

Proof- Let us find an expression for the distribution function of Z.

$$\begin{aligned}
F(z) &= P(Z \leq z) \\
&= P(\frac{X}{Y} \leq z) \\
&= \int\int_{\{(x,y): y \neq 0, \frac{x}{y} \leq z\}} f_X(x) f_Y(y) dy dx \\
&= \int\int_{\{(x,y): y < 0, \frac{x}{y} \leq z\}} f_X(x) f_Y(y) dy dx + \int\int_{\{(x,y): y > 0, \frac{x}{y} \leq z\}} f_X(x) f_Y(y) dy dx \\
&= \int\int_{\{(x,y): y < 0, x \geq yz\}} f_X(x) f_Y(y) dy dx + \int\int_{\{(x,y): y > 0, x \leq yz\}} f_X(x) f_Y(y) dy dx \\
&= \int_{-\infty}^{0}\int_{yz}^{\infty} f_X(x) f_Y(y) dx dy + \int_{0}^{\infty}\int_{-\infty}^{yz} f_X(x) f_Y(y) dx dy \\
&= I + II
\end{aligned}$$

Let us make a u-substituion $x = yu$ in both I and II. For I, $y < 0$, so we will obtain,

$$\begin{aligned}
I &= \int_{-\infty}^{0}\int_{z}^{-\infty} y f_X(yu) f_Y(y) du dy \\
&= \int_{-\infty}^{0}\int_{-\infty}^{z} (-y) f_X(yu) f_Y(y) du dy \\
&= \int_{-\infty}^{z}\int_{-\infty}^{0} (-y) f_X(yu) f_Y(y) dy du,
\end{aligned}$$

where in the last line we have changed the order of integration[1]. For II, $y > 0$ so we will obtain (similarly as in I),

$$\begin{aligned}
II &= \int_{0}^{\infty}(\int_{-\infty}^{z} y f_X(yu) f_Y(y) du dy \\
&= \int_{-\infty}^{z}\int_{0}^{\infty} y f_X(yu) f_Y(y) dy du,
\end{aligned}$$

Therefore

$$\begin{aligned}
F(z) &= I + II \\
&= \int_{-\infty}^{z}\int_{-\infty}^{0} (-y) f_X(yu) f_Y(y) dy du + \int_{-\infty}^{z}\int_{0}^{\infty} y f_X(yu) f_Y(y) dy du \\
&= \int_{-\infty}^{z}\int_{-\infty}^{\infty} \mid y \mid f_X(yu) f_Y(y) dy du
\end{aligned}$$

[1] The change of order of integration is justifiable under certain hypothesis for the intgegrand. We shall assume these are satisfied, as it is not possible to state and verify them within the scope of this book

As $f_X(\cdot)$ and $f_Y(\cdot)$ are densities, it can be shown that the integrand is a piecewise continuous function. Hence the F is of the form (5.2.4) and Theorem 5.2.5 implies that the probability density function of Z is given by (5.5.6). ■

Using the above method for finding the distribution of quotient of two random variables, we shall present three examples that will lead us to standard continuous distributions that are useful in applications. We begin with an example that constructs the Cauchy distribution.

EXAMPLE 5.5.9. Let X and Y be two independent Normal random variables with mean 0 and variance $\sigma^2 \neq 0$. Let $Z = \frac{X}{Y}$. We know that the probability density function of Z is given by (5.5.6). Further, for any $y, z \in \mathbb{R}$

$$f_X(zy)f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{z^2y^2}{2\sigma^2}}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{y^2}{2\sigma^2}} = \frac{1}{2\pi\sigma^2}\exp\left(-\left(\frac{1+z^2}{2\sigma^2}\right)y^2\right)$$

Fix $z \in \mathbb{R}$.

$$\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} |\, y \,| \frac{1}{2\pi\sigma^2}\exp\left(-\left(\frac{1+z^2}{2\sigma^2}\right)y^2\right)dy \\
&= \frac{1}{2\pi\sigma^2}\left[\int_{-\infty}^{0} |\, y \,| \exp\left(-\left(\frac{1+z^2}{2\sigma^2}\right)y^2\right)dy + \int_{0}^{\infty} |\, y \,| \exp\left(-\left(\frac{1+z^2}{2\sigma^2}\right)y^2\right)dy\right] \\
&= \frac{1}{2\pi\sigma^2}\left[\int_{-\infty}^{0} (-y) \exp\left(-\left(\frac{1+z^2}{2\sigma^2}\right)y^2\right)dy + \int_{0}^{\infty} y \exp\left(-\left(\frac{1+z^2}{2\sigma^2}\right)y^2\right)dy\right]
\end{aligned}$$

It is easy to see that two integrals are the same (perform a substitution of $u = -y$ in the first integral). So the above is

$$= \frac{1}{\pi\sigma^2}\int_{0}^{\infty} y \exp\left(-\left(\frac{1+z^2}{2\sigma^2}\right)y^2\right)dy.$$

Now perform a substitution $\left(\frac{1+z^2}{2\sigma^2}\right)y^2 = t$, so $\frac{1+z^2}{\sigma^2}ydy = dt$.

$$\begin{aligned}
f_Z(z) &= \frac{1}{\pi\sigma^2}\frac{\sigma^2}{1+z^2}\int_{0}^{\infty}\exp(-t)dt. \\
&= \frac{1}{\pi(1+z^2)}(-e^{-t}\,|_0^{\infty}) = \frac{1}{\pi(1+z^2)}.
\end{aligned}$$

Therefore Z has the Cauchy distribution, which we first saw in the context of Example 5.3.4. ■

The next example considers the ratio of two gamma random variables. This motivates a standard distribution called the F-distribution, also very important for statistics.

EXAMPLE 5.5.10. Let $m \in \mathbb{N}, n \in \mathbb{N}, \lambda > 0$, X and Y be two independent Gamma (m, λ) and Gamma (n, λ) random variables respectively. Let $Z = \frac{X}{Y}$. We know that the probability density function of Z is given by (5.5.6). Further,

$$\begin{aligned}
f_X(zy)f_Y(y) &= \begin{cases} \frac{\lambda^m}{(m-1)!}(zy)^{m-1}e^{-\lambda(zy)}\frac{\lambda^n}{(n-1)!}y^{n-1}e^{-\lambda y} & \text{if } y \geq 0, zy \geq 0 \\ 0 & \text{otherwise} \end{cases} \\
&= \begin{cases} \frac{\lambda^{n+m}}{(n-1)!(m-1)!}y^{n+m-2}z^{m-1}e^{-\lambda(1+z)y} & \text{if } y \geq 0, z \geq 0 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

Fix $z > 0$,

$$
\begin{aligned}
f_Z(z) &= \int_0^\infty y \frac{\lambda^{n+m}}{(n-1)!(m-1)!} y^{n+m-2} z^{m-1} e^{-\lambda(1+z)y} dy \\
&= \frac{z^{m-1}\lambda^{n+m}}{(n-1)!(m-1)!} \int_0^\infty y^{n+m-1} e^{-\lambda(1+z)y} dy
\end{aligned}
$$

Now perform a substition $(1+z)y = t$, so $(1+z)dy = dt$ and the above is

$$
= \frac{z^{m-1}}{(1+z)^{m+n}} \frac{\lambda^{m+n}}{(m-1)!(n-1)!} \int_0^\infty t^{m+n-1} e^{-\lambda t} dt
$$

Using (5.5.2) we have that

$$
f_Z(z) = \begin{cases} \frac{(m+n-1)!}{(m-1)!(n-1)!} z^{m-1}(1+z)^{-(m+n)} & \text{if } z \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.5.7}
$$

■

We shall define the F-distribution in Chapter 8 (See Example 8.1.7) and see further applications of it in Chapter **??**. Out next example is a construction of the Beta-distribution.

Example 5.5.11. Let $m \in \mathbb{N}, n \in \mathbb{N}, \lambda > 0$. Let X and Y be two independent Gamma (m, λ) and Gamma (n, λ) random variables respectively. Let $Z = \frac{X}{X+Y}$.

Let $W = \frac{Y}{X}$. Note that $Z = \frac{1}{1+W}$. In Example 5.5.10 we found the probability density function of W. We shall use this to find the distribution funciton of Z. As $P(W \geq 0) = 1$,

$$
P(Z \leq z) = \begin{cases} 0 & \text{if } z < 0 \\ 1 & \text{if } z > 1. \end{cases}
$$

For $0 < z < 1$,

$$
\begin{aligned}
P(Z \leq z) &= P(\frac{1}{1+W} \leq z) = P(W \geq \frac{1-z}{z}) \\
&= 1 - P(W \leq \frac{1-z}{z})
\end{aligned}
$$

Using (5.5.7) we obtain that the above is

$$
\begin{aligned}
&= 1 - \int_0^{\frac{1-z}{z}} \frac{(m+n-1)!}{(m-1)!(n-1)!} u^{m-1}(1+u)^{-(m+n)} du \\
&= 1 - \frac{(m+n-1)!}{(m-1)!(n-1)!} \int_0^{\frac{1-z}{z}} u^{m-1}(1+u)^{-(m+n)} du
\end{aligned}
$$

For $0 < z < 1$, by the fundamental theorem of calculus, differentiating in z

$$
\begin{aligned}
f_Z(z) &= \frac{1}{z^2} \cdot \frac{(m+n-1)!}{(m-1)!(n-1)!} \left(\frac{1-z}{z}\right)^{m-1} \left(1 + \frac{1-z}{z}\right)^{-(m+n)} \\
&= \frac{(m+n-1)!}{(m-1)!(n-1)!} z^{n-1}(1-z)^{m-1}
\end{aligned}
$$

Z is said to have the Beta(m, n) distribution. ■

We define the distribution in general next.

DEFINITION 5.5.12. $X \sim \textbf{Beta}(\alpha, \beta)$*: Let* $\alpha > 0$ *and* $\beta > 0$*. Then* X *is said to be Beta distributed with parameters* α *and* β *if it has the density*

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & 0 < x < 1 \\ 0 & \textit{otherwise.} \end{cases} \tag{5.5.8}$$

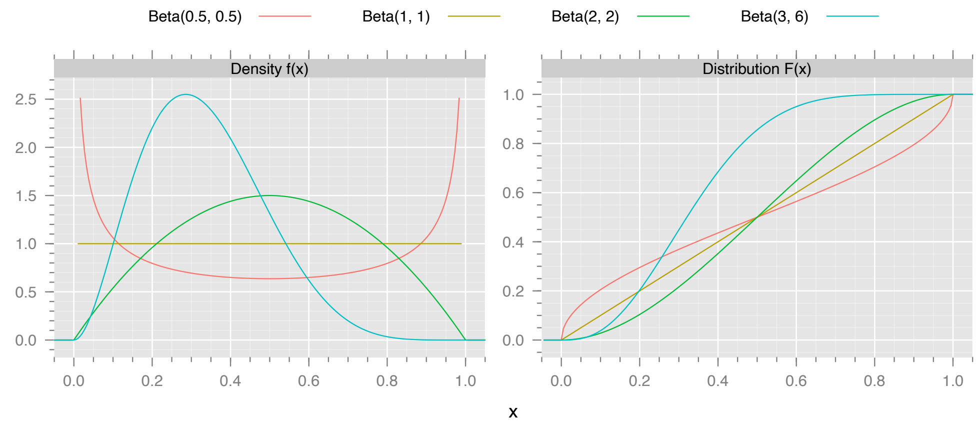

Figure 5.16: The Beta density and cumulative distribution functions for selected shape parameters.

The distribution function of a beta random variable is given by an indefinite integral which in general has no closed-form solution in terms of simple functions. In R, $F(x)$ for the beta distribution

$$F(x) = P(X \leq x) = \int_0^x \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} u^{\alpha-1}(1-u)^{\beta-1} du \,, 0 < x < 1$$

can be evaluated numerically with a function call of the form `pbeta(x, alpha, beta)`. For example,

```
> pbeta(0.5, 0.5, 0.5)
[1] 0.5
> pbeta(0.5, 3, 6)
[1] 0.8554688
> pbeta(0.2, 6, 1)
[1] 0.2030822
> pbeta(0.2, 1, 6)
[1] 0.737856
```

In the special case where either α or β equals 1, the distribution function of X can be computed explicitly. Another special case is the standard arcsine law we previously encountered in Exercise 5.2.7 in terms of its explicit distribution function; it is easy to see that this is the same as the Beta$(\frac{1}{2}, \frac{1}{2})$ distribution. The semicircular distribution encountered in Exercise 5.2.6 is also related, in the sense that it can be viewed as a location and scale transformed beta random variable.

EXERCISES

Ex. 5.5.1. Suppose that X and Y are random variables with joint probability density

$$f(x,y) = \begin{cases} \frac{4}{5}(xy+1) & \text{if } 0 \le x \le 1 \text{ and } 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Compute the marginal densities of X and Y ?

(b) Compute the conditional density $(X|Y = y)$ (for appropriate y).

(c) Are X and Y independent?

Ex. 5.5.2. Let X and Y be two random variables with the joint p.d.f given by

$$f(x,y) = \begin{cases} \lambda^2 e^{-\lambda y} & 0 \le x \le y \\ \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the marginal distribution of X and Y.

(b) Find the conditional distribution of $(Y \mid X = x)$ for some $x > 0$

Ex. 5.5.3. Let $a, b > 0$. Let $X \sim$ Gamma (a, b) and $Y \sim$ Exponential (X).

(a) Find the joint density of X and Y.

(b) Find the marginal density of Y.

(c) Find the conditional density of $(X \mid Y = y)$.

Ex. 5.5.4. Let X_1, X_2, X_3 be independent and identically distributed Uniform $(0,1)$ random variables. Let $A = X_1X_3$ and $B = X_2^2$. Find the $P(A < B)$.

Ex. 5.5.5. Let X and Y be two independent exponential random variables each with mean 1.

(a) Find the density of $U_1 = X^{\frac{1}{2}}$.

(b) Find the density of $U_2 = X + Y + 1$.

(c) Find $P(\max\{X, Y\} > 1)$.

Ex. 5.5.6. Suppose X is a uniform random variable in the interval $(0,1)$ and Y is an independent exponential(2) random variable. Find the distribution of $Z = X + Y$.

Ex. 5.5.7. Let $\alpha > 0, \beta > 0, \lambda > 0$, X and Y be two independent Gamma(α, λ) and Gamma(β, λ) random variables respectively. Then $Z = X + Y$ is distributed as a Gamma $(\alpha + \beta, \lambda)$.

Ex. 5.5.8. Let X and Y be two independent random variables with probability density function $f_X(\cdot)$ and $f_Y(\cdot)$. Show that $X + Y$ and $Y + X$ have the same distribution by showing that the integral expression defining $f_X \star f_Y(\cdot)$ is equal to the integral expression defining $f_Y \star f_X(\cdot)$).

Ex. 5.5.9. Let $\alpha > 0$ and $\Gamma(\alpha)$ as in (5.5.5).

(a) Using the same technique as in Lemma 5.5.4, show that $0 < \Gamma(\alpha) < \infty$.

(b) Show that $\Gamma(\frac{1}{2}) = \int_0^\infty x^{-0.5} e^{-x} dx = \sqrt{\pi}$.

Ex. 5.5.10. Let $\alpha > 0, \delta > 0, \lambda > 0$. Let X and Y be two independent Gamma (α, λ) and Gamma (δ, λ) random variables respectively.

(a) Let $W = \frac{Y}{X}$. Find the probability density function of W.

(b) Let $Z = \frac{X}{X+Y}$. Find the probability density function of Z.

(c) Are X and Z independent ?
Hint: Compute the joint density and see if it is a product of the marginals.

Ex. 5.5.11. Suppose X, Y are independent random variables each normally distributed with mean 0 and variance 1.

(a) Find the probability density function of $R = \sqrt{X^2 + Y^2}$

(b) Find the probability density function of $Z = \frac{X}{Y}$

(c) Find the probability density function of $\theta = \arctan\left(\frac{X}{Y}\right)$

(d) Are R, θ independent random variables ?
Hint: Compute the joint density using the change of variable indicated in Exercise 5.1.10. Decide if it is a product of the marginals