

# **Introduction to Probability and Statistics Using R**

**Second Edition**

G. Jay Kerns

October 3, 2011



IP<sub>S</sub>UR: Introduction to Probability and Statistics Using R  
Copyright © 2011 G. Jay Kerns ISBN: 978-0-557-24979-4

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled “GNU Free Documentation License”.

Date: October 3, 2011

# Contents

<b>Preface to the Second Edition</b>	<b>v</b>
<b>Preface to the First Edition</b>	<b>vii</b>
<b>List of Figures</b>	<b>xiii</b>
<b>List of Tables</b>	<b>xv</b>
<b>1 Continuous Distributions</b>	<b>xvii</b>
1.1 Continuous Random Variables . . . . .	xvii
1.2 The Continuous Uniform Distribution . . . . .	xxiii
1.3 The Normal Distribution . . . . .	xxiv
1.4 Functions of Continuous Random Variables . . . . .	xxviii
1.5 Other Continuous Distributions . . . . .	xxxiii
1.6 Exercises . . . . .	xxxix



# **Preface to the Second Edition**



# Preface to the First Edition

This book was expanded from lecture materials I use in a one semester upper-division undergraduate course entitled *Probability and Statistics* at Youngstown State University. Those lecture materials, in turn, were based on notes that I transcribed as a graduate student at Bowling Green State University. The course for which the materials were written is 50-50 Probability and Statistics, and the attendees include mathematics, engineering, and computer science majors (among others). The catalog prerequisites for the course are a full year of calculus.

The book can be subdivided into three basic parts. The first part includes the introductions and elementary *descriptive statistics*; I want the students to be knee-deep in data right out of the gate. The second part is the study of *probability*, which begins at the basics of sets and the equally likely model, journeys past discrete/continuous random variables, and continues through to multivariate distributions. The chapter on sampling distributions paves the way to the third part, which is *inferential statistics*. This last part includes point and interval estimation, hypothesis testing, and finishes with introductions to selected topics in applied statistics.

I usually only have time in one semester to cover a small subset of this book. I cover the material in Chapter 2 in a class period that is supplemented by a take-home assignment for the students. I spend a lot of time on Data Description, Probability, Discrete, and Continuous Distributions. I mention selected facts from Multivariate Distributions in passing, and discuss the meaty parts of Sampling Distributions before moving right along to Estimation (which is another chapter I dwell on considerably). Hypothesis Testing goes faster after all of the previous work, and by that time the end of the semester is in sight. I normally choose one or two final chapters (sometimes three) from the remaining to survey, and regret at the end that I did not have the chance to cover more.

In an attempt to be correct I have included material in this book which I would normally not mention during the course of a standard lecture. For instance, I normally do not highlight the intricacies of measure theory or integrability conditions when speaking to the class. Moreover, I often stray from the matrix approach to multiple linear regression because many of my students have not yet been formally trained in linear algebra. That being said, it is important to me for the students to hold something in their hands which acknowledges the world of mathematics and statistics beyond the classroom, and which may be useful to them for many semesters to come. It also mirrors my own experience as a student.

The vision for this document is a more or less self contained, essentially complete,



correct, introductory textbook. There should be plenty of exercises for the student, with full solutions for some, and no solutions for others (so that the instructor may assign them for grading). By Sweave's dynamic nature it is possible to write randomly generated exercises and I had planned to implement this idea already throughout the book. Alas, there are only 24 hours in a day. Look for more in future editions.

Seasoned readers will be able to detect my origins: *Probability and Statistical Inference* by Hogg and Tanis [?], *Statistical Inference* by Casella and Berger [?], and *Theory of Point Estimation* and *Testing Statistical Hypotheses* by Lehmann [?, ?]. I highly recommend each of those books to every reader of this one. Some R books with “introductory” in the title that I recommend are *Introductory Statistics with R* by Dalgaard [?] and *Using R for Introductory Statistics* by Verzani [?]. Surely there are many, many other good introductory books about R, but frankly, I have tried to steer clear of them for the past year or so to avoid any undue influence on my own writing.

I would like to make special mention of two other books: *Introduction to Statistical Thought* by Michael Lavine [?] and *Introduction to Probability* by Grinstead and Snell [?]. Both of these books are *free* and are what ultimately convinced me to release *IP&SUR* under a free license, too.

Please bear in mind that the title of this book is “Introduction to Probability and Statistics Using R”, and not “Introduction to R Using Probability and Statistics”, nor even “Introduction to Probability and Statistics and R Using Words”. The people at the party are Probability and Statistics; the handshake is R. There are several important topics about R which some individuals will feel are underdeveloped, glossed over, or wantonly omitted. Some will feel the same way about the probabilistic and/or statistical content. Still others will just want to learn R and skip all of the mathematics.

Despite any misgivings: here it is, warts and all. I humbly invite said individuals to take this book, with the GNU Free Documentation License (GNU-FDL) in hand, and make it better. In that spirit there are at least a few ways in my view in which this book could be improved.

**Better data.** The data analyzed in this book are almost entirely from the `datasets` package in base R, and here is why:

- I made a conscious effort to minimize dependence on contributed packages,
- The data are instantly available, already in the correct format, so we need not take time to manage them, and
- The data are *real*.

I made no attempt to choose data sets that would be interesting to the students; rather, data were chosen for their potential to convey a statistical point. Many of the data sets are decades old or more (for instance, the data used to introduce simple linear regression are the speeds and stopping distances of cars in the 1920's).

In a perfect world with infinite time I would research and contribute recent, *real* data in a context crafted to engage the students in *every* example. One day I hope to stumble over said time. In the meantime, I will add new data sets incrementally as time permits.

**More proofs.** I would like to include more proofs for the sake of completeness (I understand that some people would not consider more proofs to be improvement). Many proofs have been skipped entirely, and I am not aware of any rhyme or reason to the current omissions. I will add more when I get a chance.

**More and better~graphics.** I have not used the `ggplot2` package [?] because I do not know how to use it yet. It is on my to-do list.

**More and better exercises.** There are only a few exercises in the first edition simply because I have not had time to write more. I have toyed with the `exams` package [?] and I believe that it is a right way to move forward. As I learn more about what the package can do I would like to incorporate it into later editions of this book.

## About This Document

IPSUR contains many interrelated parts: the *Document*, the *Program*, the *Package*, and the *Ancillaries*. In short, the *Document* is what you are reading right now. The *Program* provides an efficient means to modify the Document. The *Package* is an R package that houses the Program and the Document. Finally, the *Ancillaries* are extra materials that reside in the Package and were produced by the Program to supplement use of the Document. We briefly describe each of them in turn.

## The Document

The *Document* is that which you are reading right now – IPSUR’s *raison d’être*. There are transparent copies (nonproprietary text files) and opaque copies (everything else). See the GNU-FDL in Appendix ?? for more precise language and details.

**IPSUR.tex** is a transparent copy of the Document to be typeset with a L<sup>A</sup>T<sub>E</sub>X distribution such as MikTeX or T<sub>E</sub>X Live. Any reader is free to modify the Document and release the modified version in accordance with the provisions of the GNU-FDL. Note that this file cannot be used to generate a randomized copy of the Document. Indeed, in its released form it is only capable of typesetting the exact version of IPSUR which you are currently reading. Furthermore, the `.tex` file is unable to generate any of the ancillary materials.

**IPSUR-xxx.eps, IPSUR-xxx.pdf** are the image files for every graph in the Document. These are needed when typesetting with L<sup>A</sup>T<sub>E</sub>X.

## Contents

**IPSUR.pdf** is an opaque copy of the Document. This is the file that instructors would likely want to distribute to students.

**IPSUR.dvi** is another opaque copy of the Document in a different file format.

## The Program

The *Program* includes `IPSUR.lyx` and its nephew `IPSUR.Rnw`; the purpose of each is to give individuals a way to quickly customize the Document for their particular purpose(s).

**IPSUR.lyx** is the source LyX file for the Program, released under the GNU General Public License (GNU GPL) Version 3. This file is opened, modified, and compiled with LyX, a sophisticated open-source document processor, and may be used (together with Sweave) to generate a randomized, modified copy of the Document with brand new data sets for some of the exercises and the solution manuals (in the Second Edition). Additionally, LyX can easily activate/deactivate entire blocks of the document, *e.g.* the *proofs* of the theorems, the student *solutions* to the exercises, or the instructor *answers* to the problems, so that the new author may choose which sections (s)he would like to include in the final Document (again, Second Edition). The `IPSUR.lyx` file is all that a person needs (in addition to a properly configured system – see Appendix ??) to generate/compile/export to all of the other formats described above and below, which includes the ancillary materials `IPSUR.Rdata` and `IPSUR.R`.

**IPSUR.Rnw** is another form of the source code for the Program, also released under the GNU GPL Version 3. It was produced by exporting `IPSUR.lyx` into R/Sweave format (`.Rnw`). This file may be processed with Sweave to generate a randomized copy of `IPSUR.tex` – a transparent copy of the Document – together with the ancillary materials `IPSUR.Rdata` and `IPSUR.R`. Please note, however, that `IPSUR.Rnw` is just a simple text file which does not support many of the extra features that LyX offers such as WYSIWYM editing, instantly (de)activating branches of the manuscript, and more.

## The Package

There is a contributed package on CRAN, called `IPSUR`. The package affords many advantages, one being that it houses the Document in an easy-to-access medium. Indeed, a student can have the Document at his/her fingertips with only three commands:

Another advantage goes hand in hand with the Program's license; since `IPSUR` is free, the source code must be freely available to anyone that wants it. A package hosted on CRAN allows the author to obey the license by default.

A much more important advantage is that the excellent facilities at R-Forge are building and checking the package daily against patched and development versions of the absolute latest pre-release of R. If any problems surface then I will know about it within 24 hours.

And finally, suppose there is some sort of problem. The package structure makes it *incredibly* easy for me to distribute bug-fixes and corrected typographical errors. As an author I can make my corrections, upload them to the repository at R-Forge, and they will be reflected *worldwide* within hours. We aren't in Kansas anymore, Toto.

## Ancillary Materials

These are extra materials that accompany `IPSUR`. They reside in the `/etc` subdirectory of the package source.

**IPSUR.RData** is a saved image of the R workspace at the completion of the Sweave processing of `IPSUR`. It can be loaded into memory with `File ▶ Load Workspace` or with the command `load("/path/to/IPSUR.Rdata")`. Either method will make every single object in the file immediately available and in memory. In particular, the data `BLANK` from Exercise `BLANK` in Chapter `BLANK` on page `BLANK` will be loaded. Type `BLANK` at the command line (after loading `IPSUR.RData`) to see for yourself.

**IPSUR.R** is the exported R code from `IPSUR.Rnw`. With this script, literally every R command from the entirety of `IPSUR` can be resubmitted at the command line.

## Notation

We use the notation `x` or `stem.leaf` notation to denote objects, functions, *etc.*. The sequence `Statistics ▶ Summaries ▶ Active Dataset` means to click the `Statistics` menu item, next click the `Summaries` submenu item, and finally click `Active Dataset`.

## Acknowledgements

This book would not have been possible without the firm mathematical and statistical foundation provided by the professors at Bowling Green State University, including Drs. Gábor Székely, Craig Zirbel, Arjun K. Gupta, Hanfeng Chen, Truc Nguyen, and James Albert. I would also like to thank Drs. Neal Carothers and Kit Chan.

I would also like to thank my colleagues at Youngstown State University for their support. In particular, I would like to thank Dr. G. Andy Chang for showing me what it means to be a statistician.

I would like to thank Richard Heiberger for his insightful comments and improvements to several points and displays in the manuscript.

## *Contents*

Finally, and most importantly, I would like to thank my wife for her patience and understanding while I worked hours, days, months, and years on a *free book*. Looking back, I can't believe I ever got away with it.

# List of Figures

1.5.1	Chi square distribution for various degrees of freedom . . . . .	xxxvi
1.5.2	Plot of the <code>gamma(shape = 13, rate = 1)</code> MGF . . . . .	xxxix



# List of Tables





# 1 Continuous Distributions

The focus of the last chapter was on random variables whose support can be written down in a list of values (finite or countably infinite), such as the number of successes in a sequence of Bernoulli trials. Now we move to random variables whose support is a whole range of values, say, an interval  $(a, b)$ . It is shown in later classes that it is impossible to write all of the numbers down in a list; there are simply too many of them.

This chapter begins with continuous random variables and the associated PDFs and CDFs. The continuous uniform distribution is highlighted, along with the Gaussian, or normal, distribution. Some mathematical details pave the way for a catalogue of models.

The interested reader who would like to learn more about any of the assorted discrete distributions mentioned below should take a look at *Continuous Univariate Distributions, Volumes 1 and 2* by Johnson *et al* [?, ?].

## What do I want them to know?

- how to choose a reasonable continuous model under a variety of physical circumstances
- basic correspondence between continuous versus discrete random variables
- the general tools of the trade for manipulation of continuous random variables, integration, *etc.*
- some details on a couple of continuous models, and exposure to a bunch of other ones
- how to make new continuous random variables from old ones

## 1.1 Continuous Random Variables

### 1.1.1 Probability Density Functions

Continuous random variables have supports that look like

$$S_X = [a, b] \text{ or } (a, b), \tag{1.1.1}$$

or unions of intervals of the above form. Examples of random variables that are often taken to be continuous are:

## 1 Continuous Distributions

- the height or weight of an individual,
- other physical measurements such as the length or size of an object, and
- durations of time (usually).

Every continuous random variable  $X$  has a *probability density function* (PDF) denoted  $f_X$  associated with it<sup>1</sup> that satisfies three basic properties:

1.  $f_X(x) > 0$  for  $x \in S_X$ ,
2.  $\int_{x \in S_X} f_X(x) dx = 1$ , and
3.  $\mathbb{P}(X \in A) = \int_{x \in A} f_X(x) dx$ , for an event  $A \subset S_X$ .

*Remark 1.1.* We can say the following about continuous random variables:

- Usually, the set  $A$  in ?? takes the form of an interval, for example,  $A = [c, d]$ , in which case

$$\mathbb{P}(X \in A) = \int_c^d f_X(x) dx. \quad (1.1.2)$$

- It follows that the probability that  $X$  falls in a given interval is simply the *area under the curve* of  $f_X$  over the interval.
- Since the area of a line  $x = c$  in the plane is zero,  $\mathbb{P}(X = c) = 0$  for any value  $c$ . In other words, the chance that  $X$  equals a particular value  $c$  is zero, and this is true for any number  $c$ . Moreover, when  $a < b$  all of the following probabilities are the same:

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b). \quad (1.1.3)$$

- The PDF  $f_X$  can sometimes be greater than 1. This is in contrast to the discrete case; every nonzero value of a PMF is a probability which is restricted to lie in the interval  $[0, 1]$ .

We met the cumulative distribution function,  $F_X$ , in Chapter ??. Recall that it is defined by  $F_X(t) = \mathbb{P}(X \leq t)$ , for  $-\infty < t < \infty$ . While in the discrete case the CDF is unwieldy, in the continuous case the CDF has a relatively convenient form:

$$F_X(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^t f_X(x) dx, \quad -\infty < t < \infty. \quad (1.1.4)$$

*Remark 1.2.* For any continuous CDF  $F_X$  the following are true.

---

<sup>1</sup>Not true. There are pathological random variables with no density function. (This is one of the crazy things that can happen in the world of measure theory). But in this book we will not get even close to these anomalous beasts, and regardless it can be proved that the CDF always exists.

- $F_X$  is nondecreasing, that is,  $t_1 \leq t_2$  implies  $F_X(t_1) \leq F_X(t_2)$ .
- $F_X$  is continuous (see Appendix ??). Note the distinction from the discrete case: CDFs of discrete random variables are not continuous, they are only right continuous.
- $\lim_{t \rightarrow -\infty} F_X(t) = 0$  and  $\lim_{t \rightarrow \infty} F_X(t) = 1$ .

There is a handy relationship between the CDF and PDF in the continuous case. Consider the derivative of  $F_X$ :

$$F'_X(t) = \frac{d}{dt}F_X(t) = \frac{d}{dt} \int_{-\infty}^t f_X(x) dx = f_X(t), \quad (1.1.5)$$

the last equality being true by the Fundamental Theorem of Calculus, part (2) (see Appendix ??). In short,  $(F_X)' = f_X$  in the continuous case<sup>2</sup>.

### 1.1.2 Expectation of Continuous Random Variables

For a continuous random variable  $X$  the expected value of  $g(X)$  is

$$\mathbb{E}g(X) = \int_{x \in S} g(x)f_X(x) dx, \quad (1.1.6)$$

provided the (potentially improper) integral  $\int_S |g(x)|f(x)dx$  is convergent. One important example is the mean  $\mu$ , also known as  $\mathbb{E}X$ :

$$\mu = \mathbb{E}X = \int_{x \in S} xf_X(x) dx, \quad (1.1.7)$$

provided  $\int_S |x|f(x)dx$  is finite. Also there is the variance

$$\sigma^2 = \mathbb{E}(X - \mu)^2 = \int_{x \in S} (x - \mu)^2 f_X(x) dx, \quad (1.1.8)$$

which can be computed with the alternate formula  $\sigma^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2$ . In addition, there is the standard deviation  $\sigma = \sqrt{\sigma^2}$ . The moment generating function is given by

$$M_X(t) = \mathbb{E} e^{tX} = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx, \quad (1.1.9)$$

provided the integral exists (is finite) for all  $t$  in a neighborhood of  $t = 0$ .

---

<sup>2</sup>In the discrete case,  $f_X(x) = F_X(x) - \lim_{t \rightarrow x^-} F_X(t)$ .

## 1 Continuous Distributions

**Example 1.3.** Let the continuous random variable  $X$  have PDF

$$f_X(x) = 3x^2, \quad 0 \leq x \leq 1.$$

We will see later that  $f_X$  belongs to the *Beta* family of distributions. It is easy to see that  $\int_{-\infty}^{\infty} f(x)dx = 1$ .

$$\begin{aligned} \int_{-\infty}^{\infty} f_X(x)dx &= \int_0^1 3x^2 dx \\ &= x^3 \Big|_{x=0}^1 \\ &= 1^3 - 0^3 \\ &= 1. \end{aligned}$$

This being said, we may find  $\mathbb{P}(0.14 \leq X < 0.71)$ .

$$\begin{aligned} \mathbb{P}(0.14 \leq X < 0.71) &= \int_{0.14}^{0.71} 3x^2 dx, \\ &= x^3 \Big|_{x=0.14}^{0.71} \\ &= 0.71^3 - 0.14^3 \\ &\approx 0.355167. \end{aligned}$$

We can find the mean and variance in an identical manner.

$$\begin{aligned} \mu &= \int_{-\infty}^{\infty} x f_X(x)dx = \int_0^1 x \cdot 3x^2 dx, \\ &= \frac{3}{4} x^4 \Big|_{x=0}^1, \\ &= \frac{3}{4}. \end{aligned}$$

It would perhaps be best to calculate the variance with the shortcut formula  $\sigma^2 = \mathbb{E}X^2 - \mu^2$ :

$$\begin{aligned} \mathbb{E}X^2 &= \int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_0^1 x^2 \cdot 3x^2 dx \\ &= \frac{3}{5} x^5 \Big|_{x=0}^1 \\ &= 3/5. \end{aligned}$$

which gives  $\sigma^2 = 3/5 - (3/4)^2 = 3/80$ .

**Example 1.4.** We will try one with unbounded support to brush up on improper integration. Let the random variable  $X$  have PDF

$$f_X(x) = \frac{3}{x^4}, \quad x > 1.$$

We can show that  $\int_{-\infty}^{\infty} f(x)dx = 1$ :

$$\begin{aligned} \int_{-\infty}^{\infty} f_X(x)dx &= \int_1^{\infty} \frac{3}{x^4} dx \\ &= \lim_{t \rightarrow \infty} \int_1^t \frac{3}{x^4} dx \\ &= \lim_{t \rightarrow \infty} 3 \frac{1}{-3} x^{-3} \Big|_{x=1}^t \\ &= - \left( \lim_{t \rightarrow \infty} \frac{1}{t^3} - 1 \right) \\ &= 1. \end{aligned}$$

We calculate  $\mathbb{P}(3.4 \leq X < 7.1)$ :

$$\begin{aligned} \mathbb{P}(3.4 \leq X < 7.1) &= \int_{3.4}^{7.1} 3x^{-4} dx \\ &= 3 \frac{1}{-3} x^{-3} \Big|_{x=3.4}^{7.1} \\ &= -1(7.1^{-3} - 3.4^{-3}) \\ &\approx 0.0226487123. \end{aligned}$$

We locate the mean and variance just like before.

$$\begin{aligned} \mu &= \int_{-\infty}^{\infty} x f_X(x) dx = \int_1^{\infty} x \cdot \frac{3}{x^4} dx \\ &= 3 \frac{1}{-2} x^{-2} \Big|_{x=1}^{\infty} \\ &= -\frac{3}{2} \left( \lim_{t \rightarrow \infty} \frac{1}{t^2} - 1 \right) \\ &= \frac{3}{2}. \end{aligned}$$

## 1 Continuous Distributions

Again we use the shortcut  $\sigma^2 = \mathbb{E}X^2 - \mu^2$ :

$$\begin{aligned}\mathbb{E}X^2 &= \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_1^{\infty} x^2 \cdot \frac{3}{x^4} dx \\ &= 3 \left. \frac{1}{-1} x^{-1} \right|_{x=1}^{\infty} \\ &= -3 \left( \lim_{t \rightarrow \infty} \frac{1}{t^2} - 1 \right) \\ &= 3,\end{aligned}$$

which closes the example with  $\sigma^2 = 3 - (3/2)^2 = 3/4$ .

### How to do it with R

There exist utilities to calculate probabilities and expectations for general continuous random variables, but it is better to find a built-in model, if possible. Sometimes it is not possible. We show how to do it the long way, and the `distr` package way.

**Example 1.5.** Let  $X$  have PDF  $f(x) = 3x^2$ ,  $0 < x < 1$  and find  $\mathbb{P}(0.14 \leq X \leq 0.71)$ . (We will ignore that  $X$  is a beta random variable for the sake of argument.)

```
f <- function(x) 3*x^2
integrate(f, lower = 0.14, upper = 0.71)
```

0.355167 with absolute error < 3.9e-15

Compare this to the answer we found in Example ???. We could integrate the function  $xf(x) = 3x^3$  from zero to one to get the mean, and use the shortcut  $\sigma^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2$  for the variance.

**Example 1.6.** Let  $X$  have PDF  $f(x) = 3/x^4$ ,  $x > 1$ . We may integrate the function  $xf(x) = 3/x^3$  from zero to infinity to get the mean of  $X$ .

```
g <- function(x) 3/x^3
integrate(g, lower = 1, upper = Inf)
```

1.5 with absolute error < 1.7e-14

Compare this to the answer we got in Example ??. Use `-Inf` for  $-\infty$ .

**Example 1.7.** Let us redo Example ?? with the `distr` package. The method is similar to that encountered in Section ?? in Chapter ??. We define an absolutely continuous random variable:

```
library(distr)
f <- function(x) 3*x^2
X <- AbscontDistribution(d = f, low1 = 0, up1 = 1)
p(X)(0.71) - p(X)(0.14)
```

```
[1] 0.355167
```

Compare this to the answer we found earlier. Now let us try expectation with the `distrEx` package [?]:

```
library(distrEx)
E(X); var(X); 3/80
```

```
[1] 0.7496337
[1] 0.03768305
[1] 0.0375
```

Compare these answers to the ones we found in Example ???. Why are they different? Because the `distrEx` package resorts to numerical methods when it encounters a model it does not recognize. This means that the answers we get for calculations may not exactly match the theoretical values. Be careful.

## 1.2 The Continuous Uniform Distribution

A random variable  $X$  with the continuous uniform distribution on the interval  $(a, b)$  has PDF

$$f_X(x) = \frac{1}{b-a}, \quad a < x < b. \quad (1.2.1)$$

The associated R function is `dunif(min = a, max = b)`. We write  $X \sim \text{unif}(\min = a, \max = b)$ . Due to the particularly simple form of this PDF we can also write down explicitly a formula for the CDF  $F_X$ :

$$F_X(t) = \begin{cases} 0, & t < a, \\ \frac{t-a}{b-a}, & a \leq t < b, \\ 1, & t \geq b. \end{cases} \quad (1.2.2)$$

The continuous uniform distribution is the continuous analogue of the discrete uniform distribution; it is used to model experiments whose outcome is an interval of numbers that are “equally likely” in the sense that any two intervals of equal length in the support have the same probability associated with them.



**Example 1.8.** Choose a number in  $[0, 1]$  at random, and let  $X$  be the number chosen. Then  $X \sim \text{unif}(\min = 0, \max = 1)$ . The mean of  $X \sim \text{unif}(\min = a, \max = b)$  is relatively simple to calculate:

$$\begin{aligned}\mu = \mathbb{E}X &= \int_{-\infty}^{\infty} x f_X(x) dx, \\ &= \int_a^b x \frac{1}{b-a} dx, \\ &= \frac{1}{b-a} \left. \frac{x^2}{2} \right|_{x=a}^b, \\ &= \frac{1}{b-a} \frac{b^2 - a^2}{2}, \\ &= \frac{b+a}{2},\end{aligned}$$

using the popular formula for the difference of squares. The variance is left to Exercise ??.

### 1.3 The Normal Distribution

We say that  $X$  has a *normal distribution* if it has PDF

$$f_X(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < \infty. \quad (1.3.1)$$

We write  $X \sim \text{norm}(\text{mean} = \mu, \text{sd} = \sigma)$ , and the associated R function is `dnorm(x, mean = 0, sd = 1)`.

The familiar bell-shaped curve, the normal distribution is also known as the *Gaussian distribution* because the German mathematician C. F. Gauss largely contributed to its mathematical development. This distribution is by far the most important distribution, continuous or discrete. The normal model appears in the theory of all sorts of natural phenomena, from to the way particles of smoke dissipate in a closed room, to the journey of a bottle in the ocean to the white noise of cosmic background radiation.

When  $\mu = 0$  and  $\sigma = 1$  we say that the random variable has a *standard normal* distribution and we typically write  $Z \sim \text{norm}(\text{mean} = 0, \text{sd} = 1)$ . The lowercase Greek letter phi ( $\phi$ ) is used to denote the standard normal PDF and the capital Greek letter phi  $\Phi$  is used to denote the standard normal CDF: for  $-\infty < z < \infty$ ,

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \text{ and } \Phi(t) = \int_{-\infty}^t \phi(z) dz. \quad (1.3.2)$$

**Proposition 1.9.** *If  $X \sim \text{norm}(\text{mean} = \mu, \text{sd} = \sigma)$  then*

$$Z = \frac{X - \mu}{\sigma} \sim \text{norm}(\text{mean} = 0, \text{sd} = 1). \quad (1.3.3)$$

The MGF of  $Z \sim \text{norm}(\text{mean} = 0, \text{sd} = 1)$  is relatively easy to derive:

$$\begin{aligned} M_Z(t) &= \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz, \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(z^2 + 2tz + t^2) + \frac{t^2}{2}\right\} dz, \\ &= e^{t^2/2} \left( \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-[z-(-t)]^2/2} dz \right), \end{aligned}$$

and the quantity in the parentheses is the total area under a  $\text{norm}(\text{mean} = -t, \text{sd} = 1)$  density, which is one. Therefore,

$$M_Z(t) = e^{-t^2/2}, \quad -\infty < t < \infty. \quad (1.3.4)$$

**Example 1.10.** The MGF of  $X \sim \text{norm}(\text{mean} = \mu, \text{sd} = \sigma)$  is then not difficult either because

$$Z = \frac{X - \mu}{\sigma}, \text{ or rewriting, } X = \sigma Z + \mu.$$

Therefore

$$M_X(t) = \mathbb{E}e^{tX} = \mathbb{E}e^{t(\sigma Z + \mu)} = \mathbb{E}e^{\sigma t Z} e^{\mu} = e^{t\mu} M_Z(\sigma t),$$

and we know that  $M_Z(t) = e^{-t^2/2}$ , thus substituting we get

$$M_X(t) = e^{t\mu} e^{(\sigma t)^2/2} = \exp\left\{\mu t + \sigma^2 t^2/2\right\},$$

for  $-\infty < t < \infty$ .

**Fact 1.11.** The same argument above shows that if  $X$  has MGF  $M_X(t)$  then the MGF of  $Y = a + bX$  is

$$M_Y(t) = e^{ta} M_X(bt). \quad (1.3.5)$$

**Example 1.12.** The 68-95-99.7 Rule. We saw in Section ?? that when an empirical distribution is approximately bell shaped there are specific proportions of the observations which fall at varying distances from the (sample) mean. We can see where these come from – and obtain more precise proportions – with the following:

```
pnorm(1:3) - pnorm(-(1:3))
```

```
[1] 0.6826895 0.9544997 0.9973002
```

**Example 1.13.** Let the random experiment consist of a person taking an IQ test, and let  $X$  be the score on the test. The scores on such a test are typically standardized to have a mean of 100 and a standard deviation of 15, and IQ tests have (approximately and notoriously) a bell-shaped distribution. What is  $\mathbb{P}(85 \leq X \leq 115)$ ?

*Solution:* this one is easy because the limits 85 and 115 fall exactly one standard deviation (below and above, respectively) from the mean of 100. The answer is therefore approximately 68%.

### 1.3.1 Normal Quantiles and the Quantile Function

Until now we have been given two values and our task has been to find the area under the PDF between those values. In this section, we go in reverse: we are given an area, and we would like to find the value(s) that correspond to that area.

**Example 1.14.** Assuming the IQ model of Example ??, what is the lowest possible IQ score that a person can have and still be in the top 1% of all IQ scores? *Solution:* If a person is in the top 1%, then that means that 99% of the people have lower IQ scores. So, in other words, we are looking for a value  $x$  such that  $F(x) = \mathbb{P}(X \leq x)$  satisfies  $F(x) = 0.99$ , or yet another way to say it is that we would like to solve the equation  $F(x) - 0.99 = 0$ . For the sake of argument, let us see how to do this the long way. We define the function  $g(x) = F(x) - 0.99$ , and then look for the root of  $g$  with the `uniroot` function. It uses numerical procedures to find the root so we need to give it an interval of  $x$  values in which to search for the root. We can get an educated guess from the Empirical Rule ??; the root should be somewhere between two and three standard deviations (15 each) above the mean (which is 100).

```
g <- function(x) pnorm(x, mean = 100, sd = 15) - 0.99
uniroot(g, interval = c(130, 145))
```

```
$root
[1] 134.8952

$f.root
[1] -4.873083e-09

$iter
[1] 6

$estim.prec
[1] 6.103516e-05
```

The answer is shown in `$root` which is approximately 134.8952, that is, a person with this IQ score or higher falls in the top 1% of all IQ scores.

The discussion in Example ?? was centered on the search for a value  $x$  that solved an equation  $F(x) = p$ , for some given probability  $p$ , or in mathematical parlance, the search for  $F^{-1}$ , the inverse of the CDF of  $X$ , evaluated at  $p$ . This is so important that it merits a definition all its own.

**Definition 1.15.** The *quantile function*<sup>3</sup> of a random variable  $X$  is the inverse of its cumulative distribution function:

$$Q_X(p) = \min \{x : F_X(x) \geq p\}, \quad 0 < p < 1. \quad (1.3.6)$$

*Remark 1.16.* Here are some properties of quantile functions:

1. The quantile function is defined and finite for all  $0 < p < 1$ .
2.  $Q_X$  is left-continuous (see Appendix ??). For discrete random variables it is a step function, and for continuous random variables it is a continuous function.
3. In the continuous case the graph of  $Q_X$  may be obtained by reflecting the graph of  $F_X$  about the line  $y = x$ . In the discrete case, before reflecting one should: 1) connect the dots to get rid of the jumps – this will make the graph look like a set of stairs, 2) erase the horizontal lines so that only vertical lines remain, and finally 3) swap the open circles with the solid dots. Please see Figure ?? for a comparison.
4. The two limits

$$\lim_{p \rightarrow 0^+} Q_X(p) \quad \text{and} \quad \lim_{p \rightarrow 1^-} Q_X(p)$$

always exist, but may be infinite (that is, sometimes  $\lim_{p \rightarrow 0} Q(p) = -\infty$  and/or  $\lim_{p \rightarrow 1} Q(p) = \infty$ ).

As the reader might expect, the standard normal distribution is a very special case and has its own special notation.

**Definition 1.17.** For  $0 < \alpha < 1$ , the symbol  $z_\alpha$  denotes the unique solution of the equation  $\mathbb{P}(Z > z_\alpha) = \alpha$ , where  $Z \sim \text{norm}(\text{mean} = 0, \text{sd} = 1)$ . It can be calculated in one of two equivalent ways: `qnorm(1 -  $\alpha$ )=` and `qnorm( $\alpha$ , lower.tail = FALSE)=`.

There are a few other very important special cases which we will encounter in later chapters.

## How to do it with R

Quantile functions are defined for all of the base distributions with the `q` prefix to the distribution name, except for the ECDF whose quantile function is exactly the  $Q_x(p) = \text{quantile}(x, \text{probs} = p, \text{type} = 1)$  function.

**Example 1.18.** Back to Example ??, we are looking for  $Q_X(0.99)$ , where  $X \sim \text{norm}(\text{mean} = 100, \text{sd} = 15)$ . It could not be easier to do with R.

---

<sup>3</sup>The precise definition of the quantile function is  $Q_X(p) = \inf\{x : F_X(x) \geq p\}$ , so at least it is well defined (though perhaps infinite) for the values  $p = 0$  and  $p = 1$ .

## 1 Continuous Distributions

```
qnorm(0.99, mean = 100, sd = 15)
```

```
[1] 134.8952
```

Compare this answer to the one obtained earlier with `uniroot`.

**Example 1.19.** Find the values  $z_{0.025}$ ,  $z_{0.01}$ , and  $z_{0.005}$  (these will play an important role from Chapter ?? onward).

```
qnorm(c(0.025, 0.01, 0.005), lower.tail = FALSE)
```

```
[1] 1.959964 2.326348 2.575829
```

Note the `lower.tail` argument. We would get the same answer with

```
qnorm(c(0.975, 0.99, 0.995))
```

## 1.4 Functions of Continuous Random Variables

The goal of this section is to determine the distribution of  $U = g(X)$  based on the distribution of  $X$ . In the discrete case all we needed to do was back substitute for  $x = g^{-1}(u)$  in the PMF of  $X$  (sometimes accumulating probability mass along the way). In the continuous case, however, we need more sophisticated tools. Now would be a good time to review Appendix ??.

### 1.4.1 The PDF Method

**Proposition 1.20.** Let  $X$  have PDF  $f_X$  and let  $g$  be a function which is one-to-one with a differentiable inverse  $g^{-1}$ . Then the PDF of  $U = g(X)$  is given by

$$f_U(u) = f_X[g^{-1}(u)] \left| \frac{d}{du} g^{-1}(u) \right|. \quad (1.4.1)$$

*Remark 1.21.* The formula in Equation 1.4.1 is nice, but does not really make any sense. It is better to write in the intuitive form

$$f_U(u) = f_X(x) \left| \frac{dx}{du} \right|. \quad (1.4.2)$$

**Example 1.22.** Let  $X \sim \text{norm}(\text{mean} = \mu, \text{sd} = \sigma)$ , and let  $Y = e^X$ . What is the PDF of  $Y$ ? **Solution:** Notice first that  $e^x > 0$  for any  $x$ , so the support of  $Y$  is  $(0, \infty)$ . Since the

transformation is monotone, we can solve  $y = e^x$  for  $x$  to get  $x = \ln y$ , giving  $dx/dy = 1/y$ . Therefore, for any  $y > 0$ ,

$$f_Y(y) = f_X(\ln y) \cdot \left| \frac{1}{y} \right| = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ \frac{(\ln y - \mu)^2}{2\sigma^2} \right\} \cdot \frac{1}{y},$$

where we have dropped the absolute value bars since  $y > 0$ . The random variable  $Y$  is said to have a *lognormal distribution*; see Section 1.5.

**Example 1.23.** Suppose  $X \sim \text{norm}(\text{mean} = 0, \text{sd} = 1)$  and let  $Y = 4 - 3X$ . What is the PDF of  $Y$ ?

The support of  $X$  is  $(-\infty, \infty)$ , and as  $x$  goes from  $-\infty$  to  $\infty$ , the quantity  $y = 4 - 3x$  also traverses  $(-\infty, \infty)$ . Solving for  $x$  in the equation  $y = 4 - 3x$  yields  $x = -(y - 4)/3$  giving  $dx/dy = -1/3$ . And since

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty,$$

we have

$$\begin{aligned} f_Y(y) &= f_X\left(\frac{y-4}{3}\right) \cdot \left| -\frac{1}{3} \right|, \quad -\infty < y < \infty, \\ &= \frac{1}{3\sqrt{2\pi}} e^{-(y-4)^2/2 \cdot 3^2}, \quad -\infty < y < \infty. \end{aligned}$$

We recognize the PDF of  $Y$  to be that of a  $\text{norm}(\text{mean} = 4, \text{sd} = 3)$  distribution. Indeed, we may use an identical argument as the above to prove the following fact:

**Fact 1.24.** If  $X \sim \text{norm}(\text{mean} = \mu, \text{sd} = \sigma)$  and if  $Y = a + bX$  for constants  $a$  and  $b$ , with  $b \neq 0$ , then  $Y \sim \text{norm}(\text{mean} = a + b\mu, \text{sd} = |b|\sigma)$ .

Note that it is sometimes easier to *postpone* solving for the inverse transformation  $x = x(u)$ . Instead, leave the transformation in the form  $u = u(x)$  and calculate the derivative of the *original* transformation

$$du/dx = g'(x). \quad (1.4.3)$$

Once this is known, we can get the PDF of  $U$  with

$$f_U(u) = f_X(x) \left| \frac{1}{du/dx} \right|. \quad (1.4.4)$$

In many cases there are cancellations and the work is shorter. Of course, it is not always true that

$$\frac{dx}{du} = \frac{1}{du/dx}, \quad (1.4.5)$$

but for the well-behaved examples in this book the trick works just fine.

*Remark 1.25.* In the case that  $g$  is not monotone we cannot apply Proposition ?? directly. However, hope is not lost. Rather, we break the support of  $X$  into pieces such that  $g$  is monotone on each one. We apply Proposition ?? on each piece, and finish up by adding the results together.

### 1.4.2 The CDF method

We know from Section 1.1 that  $f_X = F'_X$  in the continuous case. Starting from the equation  $F_Y(y) = \mathbb{P}(Y \leq y)$ , we may substitute  $g(X)$  for  $Y$ , then solve for  $X$  to obtain  $\mathbb{P}[X \leq g^{-1}(y)]$ , which is just another way to write  $F_X[g^{-1}(y)]$ . Differentiating this last quantity with respect to  $y$  will yield the PDF of  $Y$ .

**Example 1.26.** Suppose  $X \sim \text{unif}(\min = 0, \max = 1)$  and suppose that we let  $Y = -\ln X$ . What is the PDF of  $Y$ ?

The support set of  $X$  is  $(0, 1)$ , and  $y$  traverses  $(0, \infty)$  as  $x$  ranges from 0 to 1, so the support set of  $Y$  is  $S_Y = (0, \infty)$ . For any  $y > 0$ , we consider

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(-\ln X \leq y) = \mathbb{P}(X \geq e^{-y}) = 1 - \mathbb{P}(X < e^{-y}),$$

where the next to last equality follows because the exponential function is *monotone* (this point will be revisited later). Now since  $X$  is continuous the two probabilities  $\mathbb{P}(X < e^{-y})$  and  $\mathbb{P}(X \leq e^{-y})$  are equal; thus

$$1 - \mathbb{P}(X < e^{-y}) = 1 - \mathbb{P}(X \leq e^{-y}) = 1 - F_X(e^{-y}).$$

Now recalling that the CDF of a  $\text{unif}(\min = 0, \max = 1)$  random variable satisfies  $F(u) = u$  (see Equation 1.2.2), we can say

$$F_Y(y) = 1 - F_X(e^{-y}) = 1 - e^{-y}, \quad \text{for } y > 0.$$

We have consequently found the formula for the CDF of  $Y$ ; to obtain the PDF  $f_Y$  we need only differentiate  $F_Y$ :

$$f_Y(y) = \frac{d}{dy} (1 - e^{-y}) = 0 - e^{-y}(-1),$$

or  $f_Y(y) = e^{-y}$  for  $y > 0$ . This turns out to be a member of the exponential family of distributions, see Section 1.5.

**Example 1.27. The Probability Integral Transform.** Given a continuous random variable  $X$  with strictly increasing CDF  $F_X$ , let the random variable  $Y$  be defined by  $Y = F_X(X)$ . Then the distribution of  $Y$  is  $\text{unif}(\min = 0, \max = 1)$ .

*Proof.* We employ the CDF method. First note that the support of  $Y$  is  $(0, 1)$ . Then for any  $0 < y < 1$ ,

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(F_X(X) \leq y).$$

Now since  $F_X$  is strictly increasing, it has a well defined inverse function  $F_X^{-1}$ . Therefore,

$$\mathbb{P}(F_X(X) \leq y) = \mathbb{P}(X \leq F_X^{-1}(y)) = F_X[F_X^{-1}(y)] = y.$$

Summarizing, we have seen that  $F_Y(y) = y$ ,  $0 < y < 1$ . But this is exactly the CDF of a  $\text{unif}(\min = 0, \max = 1)$  random variable.  $\square$

**Fact 1.28.** *The Probability Integral Transform is true for all continuous random variables with continuous CDFs, not just for those with strictly increasing CDFs (but the proof is more complicated). The transform is **not** true for discrete random variables, or for continuous random variables having a discrete component (that is, with jumps in their CDF).*

**Example 1.29.** Let  $Z \sim \text{norm}(\text{mean} = 0, \text{sd} = 1)$  and let  $U = Z^2$ . What is the PDF of  $U$ ? Notice first that  $Z^2 \geq 0$ , and thus the support of  $U$  is  $[0, \infty)$ . And for any  $u \geq 0$ ,

$$F_U(u) = \mathbb{P}(U \leq u) = \mathbb{P}(Z^2 \leq u).$$

But  $Z^2 \leq u$  occurs if and only if  $-\sqrt{u} \leq Z \leq \sqrt{u}$ . The last probability above is simply the area under the standard normal PDF from  $-\sqrt{u}$  to  $\sqrt{u}$ , and since  $\phi$  is symmetric about 0, we have

$$\mathbb{P}(Z^2 \leq u) = 2\mathbb{P}(0 \leq Z \leq \sqrt{u}) = 2[F_Z(\sqrt{u}) - F_Z(0)] = 2\Phi(\sqrt{u}) - 1,$$

because  $\Phi(0) = 1/2$ . To find the PDF of  $U$  we differentiate the CDF recalling that  $\Phi' = \phi$ .

$$f_U(u) = (2\Phi(\sqrt{u}) - 1)' = 2\phi(\sqrt{u}) \cdot \frac{1}{2\sqrt{u}} = u^{-1/2}\phi(\sqrt{u}).$$

Substituting,

$$f_U(u) = u^{-1/2} \frac{1}{\sqrt{2\pi}} e^{-(\sqrt{u})^2/2} = (2\pi u)^{-1/2} e^{-u}, \quad u > 0.$$

This is what we will later call a *chi-square distribution with 1 degree of freedom*. See Section 1.5.

## How to do it with R

The `distr` package has functionality to investigate transformations of univariate distributions. There are exact results for ordinary transformations of the standard distributions, and `distr` takes advantage of these in many cases. For instance, the `distr` package can handle the transformation in Example ?? quite nicely:

```
library(distr)
X <- Norm(mean = 0, sd = 1)
Y <- 4 - 3*X
Y
```



## 1 Continuous Distributions

Distribution Object of Class: Norm

mean: 4

sd: 3

So `distr` “knows” that a linear transformation of a normal random variable is again normal, and it even knows what the correct mean and sd should be. But it is impossible for `distr` to know everything, and it is not long before we venture outside of the transformations that `distr` recognizes. Let us try Example ??:

```
Y <- exp(X)
```

```
Y
```

Distribution Object of Class: AbscontDistribution

The result is an object of class `AbscontDistribution`, which is one of the classes that `distr` uses to denote general distributions that it does not recognize (it turns out that  $Z$  has a *lognormal* distribution; see Section 1.5). A simplified description of the process that `distr` undergoes when it encounters a transformation  $Y = g(X)$  that it does not recognize is

1. Randomly generate many, many copies  $X_1, X_2, \dots, X_n$  from the distribution of  $X$ ,
2. Compute  $Y_1 = g(X_1), Y_2 = g(X_2), \dots, Y_n = g(X_n)$  and store them for use.
3. Calculate the PDF, CDF, quantiles, and random variates using the simulated values of  $Y$ .

As long as the transformation is sufficiently nice, such as a linear transformation, the exponential, absolute value, *etc.*, the d-p-q functions are calculated analytically based on the d-p-q functions associated with  $X$ . But if we try a crazy transformation then we are greeted by a warning:

```
W <- sin(exp(X) + 27)
```

```
W
```

Distribution Object of Class: AbscontDistribution

The warning confirms that the d-p-q functions are not calculated analytically, but are instead based on the randomly simulated values of  $Y$ . *We must be careful to remember this.* The nature of random simulation means that we can get different answers to the same question: watch what happens when we compute  $\mathbb{P}(W \leq 0.5)$  using the  $W$  above, then define  $W$  again, and compute the (supposedly) same  $\mathbb{P}(W \leq 0.5)$  a few moments later.

```

p(W) (0.5)
W <- sin(exp(X) + 27)
p(W) (0.5)

[1] 0.5793242
[1] 0.5793242

```

The answers are not the same! Furthermore, if we were to repeat the process we would get yet another answer for  $\mathbb{P}(W \leq 0.5)$ .

The answers were close, though. And the underlying randomly generated  $X$ 's were not the same so it should hardly be a surprise that the calculated  $W$ 's were not the same, either. This serves as a warning (in concert with the one that `distr` provides) that we should be careful to remember that complicated transformations computed by R are only approximate and may fluctuate slightly due to the nature of the way the estimates are calculated.

## 1.5 Other Continuous Distributions

### 1.5.1 Waiting Time Distributions

In some experiments, the random variable being measured is the time until a certain event occurs. For example, a quality control specialist may be testing a manufactured product to see how long it takes until it fails. An efficiency expert may be recording the customer traffic at a retail store to streamline scheduling of staff.

#### The Exponential Distribution

We say that  $X$  has an *exponential distribution* and write  $X \sim \text{exp}(\text{rate} = \lambda)$ .

$$f_X(x) = \lambda e^{-\lambda x}, \quad x > 0 \quad (1.5.1)$$

The associated R functions are `dexp(x, rate = 1)`, `pexp`, `qexp`, and `rexp`, which give the PDF, CDF, quantile function, and simulate random variates, respectively.

The parameter  $\lambda$  measures the rate of arrivals (to be described later) and must be positive. The CDF is given by the formula

$$F_X(t) = 1 - e^{-\lambda t}, \quad t > 0. \quad (1.5.2)$$

The mean is  $\mu = 1/\lambda$  and the variance is  $\sigma^2 = 1/\lambda^2$ .

The exponential distribution is closely related to the Poisson distribution. If customers arrive at a store according to a Poisson process with rate  $\lambda$  and if  $Y$  counts the number of customers that arrive in the time interval  $[0, t]$ , then we saw in Section ?? that  $Y \sim$

## 1 Continuous Distributions

`pois(lambda =  $\lambda t$ )`. Now consider a different question: let us start our clock at time 0 and stop the clock when the first customer arrives. Let  $X$  be the length of this random time interval. Then  $X \sim \text{exp}(\text{rate} = \lambda)$ . Observe the following string of equalities:

$$\begin{aligned}\mathbb{P}(X > t) &= \mathbb{P}(\text{first arrival after time } t), \\ &= \mathbb{P}(\text{no events in } [0, t]), \\ &= \mathbb{P}(Y = 0), \\ &= e^{-\lambda t},\end{aligned}$$

where the last line is the PMF of  $Y$  evaluated at  $y = 0$ . In other words,  $\mathbb{P}(X \leq t) = 1 - e^{-\lambda t}$ , which is exactly the CDF of an  $\text{exp}(\text{rate} = \lambda)$  distribution.

The exponential distribution is said to be *memoryless* because exponential random variables “forget” how old they are at every instant. That is, the probability that we must wait an additional five hours for a customer to arrive, given that we have already waited seven hours, is exactly the probability that we needed to wait five hours for a customer in the first place. In mathematical symbols, for any  $s, t > 0$ ,

$$\mathbb{P}(X > s + t | X > t) = \mathbb{P}(X > s). \quad (1.5.3)$$

See Exercise ??.

### The Gamma Distribution

This is a generalization of the exponential distribution. We say that  $X$  has a gamma distribution and write  $X \sim \text{gamma}(\text{shape} = \alpha, \text{rate} = \lambda)$ . It has PDF

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0. \quad (1.5.4)$$

The associated R functions are `dgamma(x, shape, rate = 1)`, `pgamma`, `qgamma`, and `rgamma`, which give the PDF, CDF, quantile function, and simulate random variates, respectively. If  $\alpha = 1$  then  $X \sim \text{exp}(\text{rate} = \lambda)$ . The mean is  $\mu = \alpha/\lambda$  and the variance is  $\sigma^2 = \alpha/\lambda^2$ .

To motivate the gamma distribution recall that if  $X$  measures the length of time until the first event occurs in a Poisson process with rate  $\lambda$  then  $X \sim \text{exp}(\text{rate} = \lambda)$ . If we let  $Y$  measure the length of time until the  $\alpha^{\text{th}}$  event occurs then  $Y \sim \text{gamma}(\text{shape} = \alpha, \text{rate} = \lambda)$ . When  $\alpha$  is an integer this distribution is also known as the *Erlang* distribution.

**Example 1.30.** At a car wash, two customers arrive per hour on the average. We decide to measure how long it takes until the third customer arrives. If  $Y$  denotes this random time then  $Y \sim \text{gamma}(\text{shape} = 3, \text{rate} = 1/2)$ .

## 1.5.2 The Chi square, Student's $t$ , and Snedecor's $F$ Distributions

### The Chi square Distribution

A random variable  $X$  with PDF

$$f_X(x) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{p/2-1} e^{-x/2}, \quad x > 0, \quad (1.5.5)$$

is said to have a *chi-square distribution* with  $p$  *degrees of freedom*. We write  $X \sim \text{chisq}(\text{df} = p)$ . The associated R functions are `dchisq(x, df)`, `pchisq`, `qchisq`, and `rchisq`, which give the PDF, CDF, quantile function, and simulate random variates, respectively. See Figure 1.5.1. In an obvious notation we may define  $\chi^2_\alpha(p)$  as the number on the  $x$ -axis such that there is exactly  $\alpha$  area under the  $\text{chisq}(\text{df} = p)$  curve to its right.

The code to produce Figure 1.5.1 is

```
curve(dchisq(x, df = 3), from = 0, to = 20, ylab = "y")
ind <- c(4, 5, 10, 15)
for (i in ind) curve(dchisq(x, df = i), 0, 20, add = TRUE)
```

*Remark 1.31.* Here are some useful things to know about the chi-square distribution.

1. If  $Z \sim \text{norm}(\text{mean} = 0, \text{sd} = 1)$ , then  $Z^2 \sim \text{chisq}(\text{df} = 1)$ . We saw this in Example ??, and the fact is important when it comes time to find the distribution of the sample variance,  $S^2$ . See Theorem ?? in Section ??.
2. The chi-square distribution is supported on the positive  $x$ -axis, with a right-skewed distribution.
3. The  $\text{chisq}(\text{df} = p)$  distribution is the same as a  $\text{gamma}(\text{shape} = p/2, \text{rate} = 1/2)$  distribution.
4. The MGF of  $X \sim \text{chisq}(\text{df} = p)$  is

$$M_X(t) = (1 - 2t)^{-p/2}, \quad t < 1/2. \quad (1.5.6)$$

### Student's $t$ distribution

A random variable  $X$  with PDF

$$f_X(x) = \frac{\Gamma[(r+1)/2]}{\sqrt{r\pi} \Gamma(r/2)} \left(1 + \frac{x^2}{r}\right)^{-(r+1)/2}, \quad -\infty < x < \infty \quad (1.5.7)$$

is said to have *Student's  $t$  distribution* with  $r$  *degrees of freedom*, and we write  $X \sim t(\text{df} = r)$ . The associated R functions are `dt`, `pt`, `qt`, and `rt`, which give the PDF, CDF, quantile function, and simulate random variates, respectively. See Section ??.

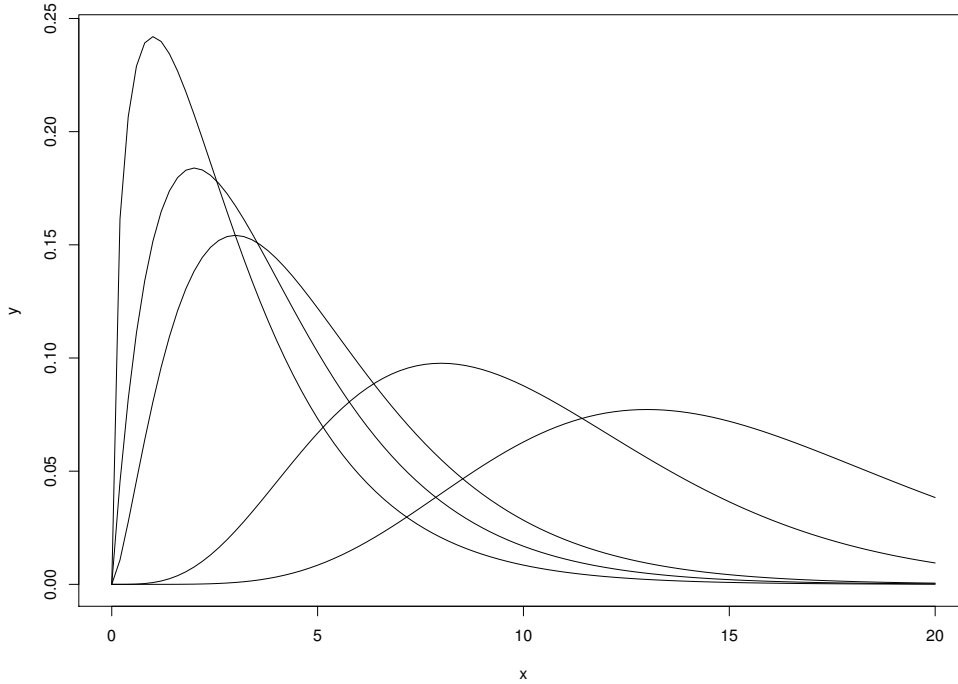


Figure 1.5.1: The chi square distribution for various degrees of freedom.

### Snedecor's $F$ distribution

A random variable  $X$  with PDF

$$f_X(x) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{m/2} x^{m/2-1} \left(1 + \frac{m}{n}x\right)^{-(m+n)/2}, \quad x > 0. \quad (1.5.8)$$

is said to have an  $F$  distribution with  $(m, n)$  degrees of freedom. We write  $X \sim f(\text{df1} = m, \text{df2} = n)$ . The associated R functions are `df(x, df1, df2)`, `pf`, `qf`, and `rf`, which give the PDF, CDF, quantile function, and simulate random variates, respectively. We define  $F_\alpha(m, n)$  as the number on the  $x$ -axis such that there is exactly  $\alpha$  area under the  $f(\text{df1} = m, \text{df2} = n)$  curve to its right.

*Remark 1.32.* Here are some notes about the  $F$  distribution.

1. If  $X \sim f(\text{df1} = m, \text{df2} = n)$  and  $Y = 1/X$ , then  $Y \sim f(\text{df1} = n, \text{df2} = m)$ . Historically, this fact was especially convenient. In the old days, statisticians used printed tables for their statistical calculations. Since the  $F$  tables were symmetric in  $m$  and  $n$ , it meant that publishers could cut the size of their printed tables in half. It plays less of a role today now that personal computers are widespread.

2. If  $X \sim t(df = r)$ , then  $X^2 \sim f(df1 = 1, df2 = r)$ . We will see this again in Section ??.

### 1.5.3 Other Popular Distributions

#### The Cauchy Distribution

This is a special case of the Student's  $t$  distribution. It has PDF

$$f_X(x) = \frac{1}{\beta\pi} \left[ 1 + \left( \frac{x-m}{\beta} \right)^2 \right]^{-1}, \quad -\infty < x < \infty. \quad (1.5.9)$$

We write  $X \sim \text{cauchy}(\text{location} = m, \text{scale} = \beta)$ . The associated R function is `dcauchy(x, location = 0, scale = 1)`.

It is easy to see that a `cauchy(location = 0, scale = 1)` distribution is the same as a  $t(df = 1)$  distribution. The cauchy distribution looks like a norm distribution but with very heavy tails. The mean (and variance) do not exist, that is, they are infinite. The median is represented by the `location` parameter, and the `scale` parameter influences the spread of the distribution about its median.

#### The Beta Distribution

This is a generalization of the continuous uniform distribution.

$$f_X(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 < x < 1. \quad (1.5.10)$$

We write  $X \sim \text{beta}(\text{shape1} = \alpha, \text{shape2} = \beta)$ . The associated R function is `dbeta(x, shape1, shape2)`. The mean and variance are

$$\mu = \frac{\alpha}{\alpha + \beta} \text{ and } \sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}. \quad (1.5.11)$$

See Example ??. This distribution comes up a lot in Bayesian statistics because it is a good model for one's prior beliefs about a population proportion  $p$ ,  $0 \leq p \leq 1$ .

#### The Logistic Distribution

$$f_X(x) = \frac{1}{\sigma} \exp\left(-\frac{x-\mu}{\sigma}\right) \left[ 1 + \exp\left(-\frac{x-\mu}{\sigma}\right) \right]^{-2}, \quad -\infty < x < \infty. \quad (1.5.12)$$

We write  $X \sim \text{logis}(\text{location} = \mu, \text{scale} = \sigma)$ . The associated R function is `dlogis(x, location = 0, scale = 1)`. The logistic distribution comes up in differential equations as a model for population growth under certain assumptions. The mean is  $\mu$  and the variance is  $\pi^2\sigma^2/3$ .

## The Lognormal Distribution

This is a distribution derived from the normal distribution (hence the name). If  $U \sim \text{norm}(\text{mean} = \mu, \text{sd} = \sigma)$ , then  $X = e^U$  has PDF

$$f_X(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right], \quad 0 < x < \infty. \quad (1.5.13)$$

We write  $X \sim \text{lnorm}(\text{meanlog} = \mu, \text{sdlog} = \sigma)$ . The associated R function is `dlnorm(x, meanlog = 0, sdlog = 1)`. Notice that the support is concentrated on the positive  $x$  axis; the distribution is right-skewed with a heavy tail. See Example ??.

## The Weibull Distribution

This has PDF

$$f_X(x) = \frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{x}{\beta}\right)^\alpha\right), \quad x > 0. \quad (1.5.14)$$

We write  $X \sim \text{weibull}(\text{shape} = \alpha, \text{scale} = \beta)$ . The associated R function is `dweibull(x, shape, scale = 1)`.

## How to do it with R

There is some support of moments and moment generating functions for some continuous probability distributions included in the `actuar` package [?]. The convention is `m` in front of the distribution name for raw moments, and `mgf` in front of the distribution name for the moment generating function. At the time of this writing, the following distributions are supported: gamma, inverse Gaussian, (non-central) chi-squared, exponential, and uniform.

**Example 1.33.** Calculate the first four raw moments for  $X \sim \text{gamma}(\text{shape} = 13, \text{rate} = 1)$  and plot the moment generating function.

We load the `actuar` package and use the functions `mgamma` and `mgfgamma`:

```
library(actuar)
mgamma(1:4, shape = 13, rate = 1)
```

```
[1]    13   182  2730 43680
```

For the plot we can use the function in the following form:

```
plot(function(x){mgfgamma(x, shape = 13, rate = 1)},
      from=-0.1, to=0.1, ylab = "gamma_mgf")
```

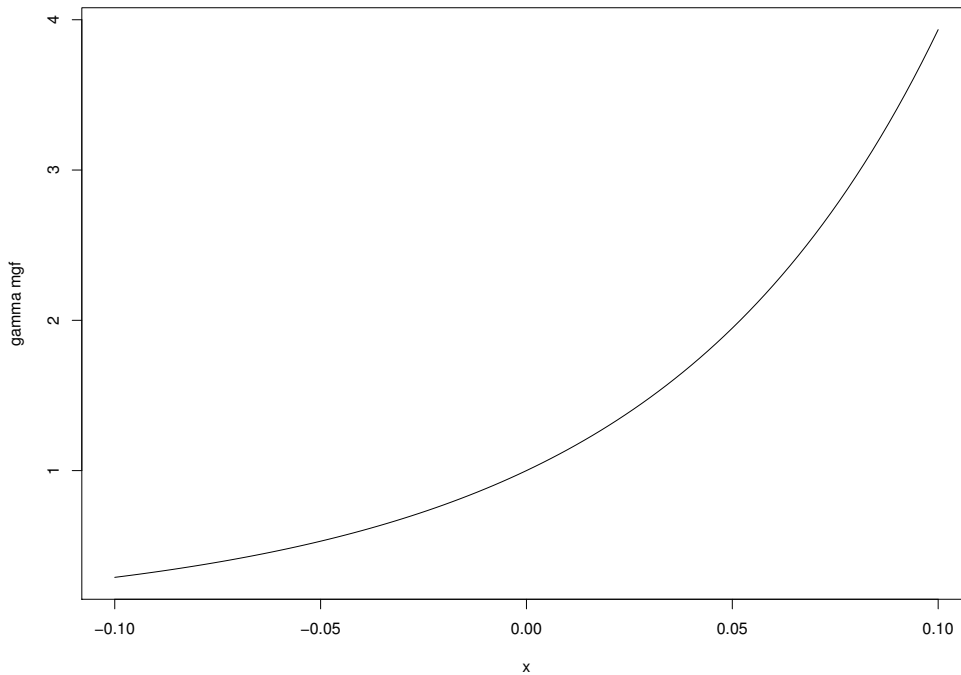


Figure 1.5.2: A plot of the `gamma(shape = 13, rate = 1)` MGF.

## 1.6 Exercises

**Exercise 1.1.** Find the constant  $c$  so that the given function is a valid PDF of a random variable  $X$ .

1.  $f(x) = Cx^n$ ,  $0 < x < 1$ .
2.  $f(x) = Cxe^{-x}$ ,  $0 < x < \infty$ .
3.  $f(x) = e^{-(x-C)}$ ,  $7 < x < \infty$ .
4.  $f(x) = Cx^3(1-x)^2$ ,  $0 < x < 1$ .
5.  $f(x) = C(1 + x^2/4)^{-1}$ ,  $-\infty < x < \infty$ .

**Exercise 1.2.** For the following random experiments, decide what the distribution of  $X$  should be. In nearly every case, there are additional assumptions that should be made for the distribution to apply; identify those assumptions (which may or may not strictly hold in practice).



## 1 Continuous Distributions

1. We throw a dart at a dart board. Let  $X$  denote the squared linear distance from the bulls-eye to the where the dart landed.
2. We randomly choose a textbook from the shelf at the bookstore and let  $P$  denote the proportion of the total pages of the book devoted to exercises.
3. We measure the time it takes for the water to completely drain out of the kitchen sink.
4. We randomly sample strangers at the grocery store and ask them how long it will take them to drive home.

**Exercise 1.3.** If  $Z$  is `norm(mean = 0, sd = 1)`, find

1.  $\mathbb{P}(Z > 2.64)$

`pnorm(2.64, lower.tail = FALSE)`

1.  $\mathbb{P}(0 \leq Z < 0.87)$

`pnorm(0.87) - 1/2`

1.  $\mathbb{P}(|Z| > 1.39)$  (Hint: draw a picture!)

`2 * pnorm(-1.39)`

**Exercise 1.4.** Calculate the variance of  $X \sim \text{unif}(\min = a, \max = b)$ . *Hint:* First calculate  $\mathbb{E}X^2$ .

**Exercise 1.5.** Prove the memoryless property for exponential random variables. That is, for  $X \sim \text{exp}(\text{rate} = \lambda)$  show that for any  $s, t > 0$ ,

$$\mathbb{P}(X > s + t | X > t) = \mathbb{P}(X > s).$$