Chapter 5

Discrete Distributions

In this chapter we introduce discrete random variables, those who take values in a finite or countably infinite support set. We discuss probability mass functions and some special expectations, namely, the mean, variance and standard deviation. Some of the more important discrete distributions are explored in detail, and the more general concept of expectation is defined, which paves the way for moment generating functions.

We give special attention to the empirical distribution since it plays such a fundamental role with respect to re sampling and Chapter 13; it will also be needed in Section 10.5.1 where we discuss the Kolmogorov-Smirnov test. Following this is a section in which we introduce a catalogue of discrete random variables that can be used to model experiments.

There are some comments on simulation, and we mention transformations of random variables in the discrete case. The interested reader who would like to learn more about any of the assorted discrete distributions mentioned here should take a look at *Univariate Discrete Distributions* by Johnson *et al* [50].

What do I want them to know?

- how to choose a reasonable discrete model under a variety of physical circumstances
- the notion of mathematical expectation, how to calculate it, and basic properties
- moment generating functions (yes, I want them to hear about those)
- the general tools of the trade for manipulation of continuous random variables, integration, *etc*.
- some details on a couple of discrete models, and exposure to a bunch of other ones
- how to make new discrete random variables from old ones

5.1 Discrete Random Variables

5.1.1 Probability Mass Functions

Discrete random variables are characterized by their supports which take the form

$$S_X = \{u_1, u_2, \ldots, u_k\} \text{ or } S_X = \{u_1, u_2, u_3 \ldots\}. \tag{5.1.1}$$

Every discrete random variable X has associated with it a probability mass function (PMF) $f_X : S_X \to [0, 1]$ defined by

$$f_X(x) = \mathbb{P}(X = x), \quad x \in S_X. \tag{5.1.2}$$

Since values of the PMF represent probabilities, we know from Chapter 4 that PMFs enjoy certain properties. In particular, all PMFs satisfy

1. $f_X(x) > 0$ for $x \in S$,
2. $\sum_{x \in S} f_X(x) = 1$, and
3. $\mathbb{P}(X \in A) = \sum_{x \in A} f_X(x)$, for any event $A \subset S$.

Example 5.1. Toss a coin 3 times. The sample space would be

$$S = \{HHH,\ HTH,\ THH,\ TTH,\ HHT,\ HTT,\ THT,\ TTT\}.$$

Now let X be the number of Heads observed. Then X has support $S_X = \{0, 1, 2, 3\}$. Assuming that the coin is fair and was tossed in exactly the same way each time, it is not unreasonable to suppose that the outcomes in the sample space are all equally likely. What is the PMF of X? Notice that X is zero exactly when the outcome TTT occurs, and this event has probability $1/8$. Therefore, $f_X(0) = 1/8$, and the same reasoning shows that $f_X(3) = 1/8$. Exactly three outcomes result in $X = 1$, thus, $f_X(1) = 3/8$ and $f_X(3)$ holds the remaining $3/8$ probability (the total is 1). We can represent the PMF with a table:

$x \in S_X$	0	1	2	3	Total
$f_X(x) = \mathbb{P}(X = x)$	1/8	3/8	3/8	1/8	1

5.1.2 Mean, Variance, and Standard Deviation

There are numbers associated with PMFs. One important example is the mean μ, also known as $\mathbb{E}\, X$ (which we will discuss later):

$$\mu = \mathbb{E}\, X = \sum_{x \in S} x f_X(x), \tag{5.1.3}$$

provided the (potentially infinite) series $\sum |x| f_X(x)$ is convergent. Another important number is the variance:

$$\sigma^2 = \sum_{x \in S} (x - \mu)^2 f_X(x), \tag{5.1.4}$$

which can be computed (see Exercise 5.3) with the alternate formula $\sigma^2 = \sum x^2 f_X(x) - \mu^2$. Directly defined from the variance is the standard deviation $\sigma = \sqrt{\sigma^2}$.

Example 5.2. We will calculate the mean of X in Example 5.1.

$$\mu = \sum_{x=0}^{3} x f_X(x) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 3.5.$$

We interpret $\mu = 3.5$ by reasoning that if we were to repeat the random experiment many times, independently each time, observe many corresponding outcomes of the random variable X, and take the sample mean of the observations, then the calculated value would fall close to 3.5. The approximation would get better as we observe more and more values of X (another form of the Law of Large Numbers; see Section 4.3). Another way it is commonly stated is that X is 3.5 "on the average" or "in the long run".

Remark 5.3. Note that although we say X is 3.5 on the average, we must keep in mind that our X never actually equals 3.5 (in fact, it is impossible for X to equal 3.5).

Related to the probability mass function $f_X(x) = \mathbb{P}(X = x)$ is another important function called the cumulative distribution function (CDF), F_X. It is defined by the formula

$$F_X(t) = \mathbb{P}(X \leq t), \quad -\infty < t < \infty. \tag{5.1.5}$$

We know that all PMFs satisfy certain properties, and a similar statement may be made for CDFs. In particular, any CDF F_X satisfies

- F_X is nondecreasing ($t_1 \leq t_2$ implies $F_X(t_1) \leq F_X(t_2)$).
- F_X is right-continuous ($\lim_{t \to a^+} F_X(t) = F_X(a)$ for all $a \in \mathbb{R}$).
- $\lim_{t \to -\infty} F_X(t) = 0$ and $\lim_{t \to \infty} F_X(t) = 1$.

We say that X has the distribution F_X and we write $X \sim F_X$. In an abuse of notation we will also write $X \sim f_X$ and for the named distributions the PMF or CDF will be identified by the family name instead of the defining formula.

5.1.3 How to do it with R

The mean and variance of a discrete random variable is easy to compute at the console. Let's return to Example 5.2. We will start by defining a vector `x` containing the support of X, and a vector `f` to contain the values of f_X at the respective outcomes in `x`:

```
> x <- c(0,1,2,3)
> f <- c(1/8, 3/8, 3/8, 1/8)
```

To calculate the mean μ, we need to multiply the corresponding values of `x` and `f` and add them. This is easily accomplished in R since operations on vectors are performed *element-wise* (see Section 2.3.4):

```
> mu <- sum(x * f)
> mu
[1] 1.5
```

To compute the variance σ^2, we subtract the value of `mu` from each entry in `x`, square the answers, multiply by `f`, and `sum`. The standard deviation σ is simply the square root of σ^2.

```
> sigma2 <- sum((x-mu)^2 * f)
> sigma2
[1] 0.75
> sigma <- sqrt(sigma2)
> sigma
[1] 0.8660254
```

Finally, we may find the values of the CDF F_X on the support by accumulating the probabilities in f_X with the `cumsum` function.

```
> F = cumsum(f)
> F
[1] 0.125 0.500 0.875 1.000
```

As easy as this is, it is even easier to do with the `distrEx` package [74]. We define a random variable `X` as an object, then compute things from the object such as mean, variance, and standard deviation with the functions `E`, `var`, and `sd`:

```
> library(distrEx)
> X <- DiscreteDistribution(supp = 0:3, prob = c(1,3,3,1)/8)
> E(X); var(X); sd(X)
[1] 1.5
[1] 0.75
[1] 0.8660254
```

5.2 The Discrete Uniform Distribution

We have seen the basic building blocks of discrete distributions and we now study particular models that statisticians often encounter in the field. Perhaps the most fundamental of all is the *discrete uniform* distribution.

A random variable X with the discrete uniform distribution on the integers $1, 2, \ldots, m$ has PMF

$$f_X(x) = \frac{1}{m}, \quad x = 1, 2, \ldots, m. \tag{5.2.1}$$

We write $X \sim \mathsf{disunif}(m)$. A random experiment where this distribution occurs is the choice of an integer at random between 1 and 100, inclusive. Let X be the number chosen. Then $X \sim \mathsf{disunif}(m = 100)$ and

$$\mathbb{P}(X = x) = \frac{1}{100}, \quad x = 1, \ldots, 100.$$

We find a direct formula for the mean of $X \sim \mathsf{disunif}(m)$:

$$\mu = \sum_{x=1}^{m} x f_X(x) = \sum_{x=1}^{m} x \cdot \frac{1}{m} = \frac{1}{m}(1 + 2 + \cdots + m) = \frac{m+1}{2}, \tag{5.2.2}$$

where we have used the famous identity $1 + 2 + \cdots + m = m(m+1)/2$. That is, if we repeatedly choose integers at random from 1 to m then, on the average, we expect to get $(m+1)/2$. To get the variance we first calculate

$$\sum_{x=1}^{m} x^2 f_X(x) = \frac{1}{m} \sum_{x=1}^{m} x^2 = \frac{1}{m} \frac{m(m+1)(2m+1)}{6} = \frac{(m+1)(2m+1)}{6},$$

and finally,

$$\sigma^2 = \sum_{x=1}^{m} x^2 f_X(x) - \mu^2 = \frac{(m+1)(2m+1)}{6} - \left(\frac{m+1}{2}\right)^2 = \cdots = \frac{m^2-1}{12}. \tag{5.2.3}$$

Example 5.4. Roll a die and let X be the upward face showing. Then $m = 6$, $\mu = 7/2 = 3.5$, and $\sigma^2 = (6^2 - 1)/12 = 35/12$.

5.2.1 How to do it with R

From the console: One can choose an integer at random with the `sample` function. The general syntax to simulate a discrete uniform random variable is `sample(x, size, replace = TRUE)`.

The argument `x` identifies the numbers from which to randomly sample. If `x` is a number, then sampling is done from 1 to `x`. The argument `size` tells how big the sample size should be, and `replace` tells whether or not numbers should be replaced in the urn after having been sampled. The default option is `replace = FALSE` but for discrete uniforms the sampled values should be replaced. Some examples follow.

5.2.2 Examples

- To roll a fair die 3000 times, do `sample(6, size = 3000, replace = TRUE)`.
- To choose 27 random numbers from 30 to 70, do `sample(30:70, size = 27, replace = TRUE)`.
- To flip a fair coin 1000 times, do `sample(c("H","T"), size = 1000, replace = TRUE)`.

With the R Commander: Follow the sequence Probability ▹ Discrete Distributions ▹ Discrete Uniform distribution ▹ Simulate Discrete uniform variates. . . .

Suppose we would like to roll a fair die 3000 times. In the `Number of samples` field we enter 1. Next, we describe what interval of integers to be sampled. Since there are six faces numbered 1 through 6, we set `from = 1`, we set `to = 6`, and set `by = 1` (to indicate that we travel from 1 to 6 in increments of 1 unit). We will generate a list of 3000 numbers selected from among 1, 2, . . . , 6, and we store the results of the simulation. For the time being, we select `New Data set`. Click OK.

Since we are defining a new data set, the R Commander requests a name for the data set. The default name is `Simset1`, although in principle you could name it whatever you like (according to R's rules for object names). We wish to have a list that is 3000 long, so we set `Sample Size = 3000` and click OK.

In the R Console window, the R Commander should tell you that `Simset1` has been initialized, and it should also alert you that `There was 1 discrete uniform variate sample stored in Simset 1.`. To take a look at the rolls of the die, we click View data set and a window opens.

The default name for the variable is `disunif.sim1`.

5.3 The Binomial Distribution

The binomial distribution is based on a *Bernoulli trial*, which is a random experiment in which there are only two possible outcomes: success (S) and failure (F). We conduct the Bernoulli trial and let

$$X = \begin{cases} 1 & \text{if the outcome is } S, \\ 0 & \text{if the outcome is } F. \end{cases} \tag{5.3.1}$$

If the probability of success is p then the probability of failure must be $1 - p = q$ and the PMF of X is

$$f_X(x) = p^x (1-p)^{1-x}, \quad x = 0, 1. \tag{5.3.2}$$

It is easy to calculate $\mu = \mathbb{E}\, X = p$ and $\mathbb{E}\, X^2 = p$ so that $\sigma^2 = p - p^2 = p(1-p)$.

5.3.1 The Binomial Model

The Binomial model has three defining properties:

- Bernoulli trials are conducted n times,
- the trials are independent,
- the probability of success p does not change between trials.

If X counts the number of successes in the n independent trials, then the PMF of X is

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n. \tag{5.3.3}$$

We say that X has a *binomial distribution* and we write $X \sim \mathsf{binom}(\mathtt{size} = n, \mathtt{prob} = p)$. It is clear that $f_X(x) \geq 0$ for all x in the support because the value is the product of nonnegative numbers. We next check that $\sum f(x) = 1$:

$$\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = [p + (1-p)]^n = 1^n = 1.$$

We next find the mean:

$$
\begin{aligned}
\mu =& \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}, \\
=& \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x q^{n-x}, \\
=& n \cdot p \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x}, \\
=& np \sum_{x-1=0}^{n-1} \binom{n-1}{x-1} p^{(x-1)} (1-p)^{(n-1)-(x-1)}, \\
=& np.
\end{aligned}
$$

A similar argument shows that $\mathbb{E}\, X(X-1) = n(n-1)p^2$ (see Exercise 5.4). Therefore

$$
\begin{aligned}
\sigma^2 =& \mathbb{E}\, X(X-1) + \mathbb{E}\, X - [\mathbb{E}\, X]^2, \\
=& n(n-1)p^2 + np - (np)^2, \\
=& n^2 p^2 - np^2 + np - n^2 p^2, \\
=& np - np^2 = np(1-p).
\end{aligned}
$$

Example 5.5. A four-child family. Each child may be either a boy (B) or a girl (G). For simplicity we suppose that $\mathbb{P}(B) = \mathbb{P}(G) = 1/2$ and that the genders of the children are determined independently. If we let X count the number of B's, then $X \sim$ binom(`size` = 4, `prob` = 1/2). Further, $\mathbb{P}(X = 2)$ is

$$
f_X(2) = \binom{4}{2} (1/2)^2 (1/2)^2 = \frac{6}{2^4}.
$$

The mean number of boys is $4(1/2) = 2$ and the variance of X is $4(1/2)(1/2) = 1$.

5.3.2 How to do it with R

The corresponding R function for the PMF and CDF are `dbinom` and `pbinom`, respectively. We demonstrate their use in the following examples.

Example 5.6. We can calculate it in R Commander under the Binomial Distribution menu with the Binomial probabilities menu item.

```
      Pr
0 0.0625
1 0.2500
2 0.3750
3 0.2500
4 0.0625
```

We know that the binom(`size` = 4, `prob` = 1/2) distribution is supported on the integers 0, 1, 2, 3, and 4; thus the table is complete. We can read off the answer to be $\mathbb{P}(X = 2) = 0.3750$.

Example 5.7. Roll 12 dice simultaneously, and let X denote the number of 6's that appear. We wish to find the probability of getting seven, eight, or nine 6's. If we let $S = \{\text{get a 6 on one roll}\}$, then $\mathbb{P}(S) = 1/6$ and the rolls constitute Bernoulli trials; thus $X \sim$ binom(`size` =12, `prob` =1/6) and our task is to find $\mathbb{P}(7 \leq X \leq 9)$. This is just

$$\mathbb{P}(7 \leq X \leq 9) = \sum_{x=7}^{9} \binom{12}{x} (1/6)^x (5/6)^{12-x}.$$

Again, one method to solve this problem would be to generate a probability mass table and add up the relevant rows. However, an alternative method is to notice that $\mathbb{P}(7 \leq X \leq 9) = \mathbb{P}(X \leq 9) - \mathbb{P}(X \leq 6) = F_X(9) - F_X(6)$, so we could get the same answer by using the Binomial tail probabilities... menu in the R Commander or the following from the command line:

```
> pbinom(9, size=12, prob=1/6) - pbinom(6, size=12, prob=1/6)
[1] 0.001291758
> diff(pbinom(c(6,9), size = 12, prob = 1/6))  # same thing
[1] 0.001291758
```

Example 5.8. Toss a coin three times and let X be the number of Heads observed. We know from before that $X \sim$ binom(`size` = 3, `prob` = 1/2) which implies the following PMF:

x = #of Heads	0	1	2	3
$f(x) = \mathbb{P}(X = x)$	1/8	3/8	3/8	1/8

Our next goal is to write down the CDF of X explicitly. The first case is easy: it is impossible for X to be negative, so if $x < 0$ then we should have $\mathbb{P}(X \leq x) = 0$. Now choose a value x satisfying $0 \leq x < 1$, say, $x = 0.3$. The only way that $X \leq x$ could happen would be if $X = 0$, therefore, $\mathbb{P}(X \leq x)$ should equal $\mathbb{P}(X = 0)$, and the same is true for any $0 \leq x < 1$. Similarly, for any $1 \leq x < 2$, say, $x = 1.73$, the event $\{X \leq x\}$ is exactly the event $\{X = 0 \text{ or } X = 1\}$. Consequently, $\mathbb{P}(X \leq x)$ should equal $\mathbb{P}(X = 0 \text{ or } X = 1) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1)$. Continuing in this fashion, we may figure out the values of $F_X(x)$ for all possible inputs $-\infty < x < \infty$, and we may summarize our observations with the following piecewise defined function:

$$F_X(x) = \mathbb{P}(X \leq x) = \begin{cases} 0, & x < 0, \\ \frac{1}{8}, & 0 \leq x < 1, \\ \frac{1}{8} + \frac{3}{8} = \frac{4}{8}, & 1 \leq x < 2, \\ \frac{4}{8} + \frac{3}{8} = \frac{7}{8}, & 2 \leq x < 3, \\ 1, & x \geq 3. \end{cases}$$

In particular, the CDF of X is defined for the entire real line, $\mathbb{R}$. The CDF is right continuous and nondecreasing. A graph of the binom(`size` = 3, `prob` = 1/2) CDF is shown in Figure 5.3.1.

Example 5.9. Another way to do Example 5.8 is with the `distr` family of packages [74]. They use an object oriented approach to random variables, that is, a random variable is stored in an object

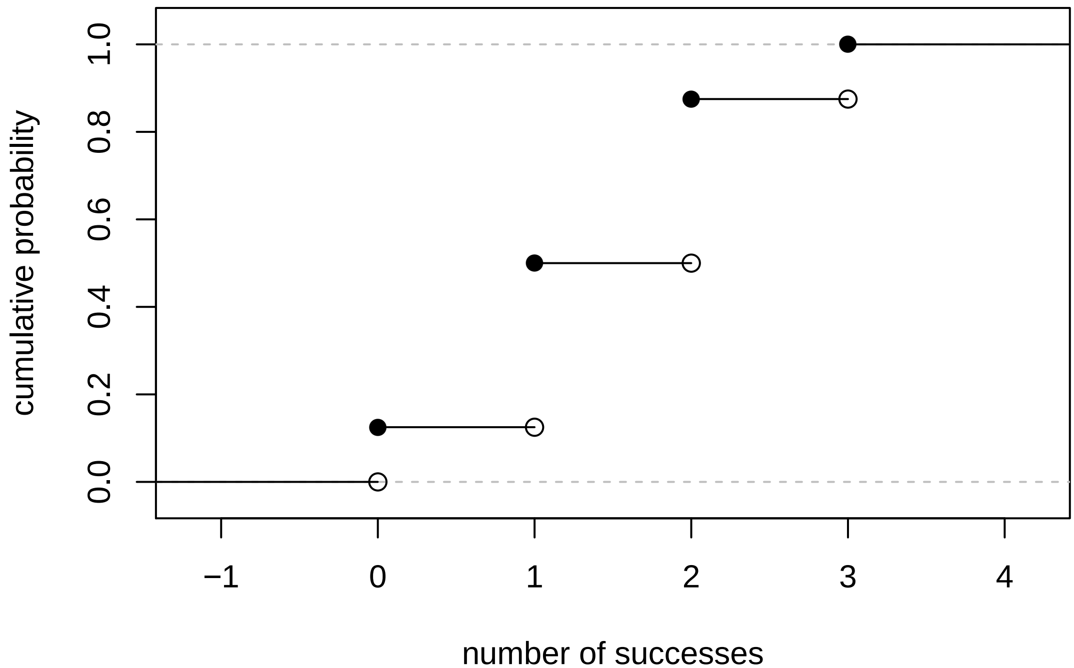

Figure 5.3.1: Graph of the binom(`size` = 3, `prob` = 1/2) CDF

X, and then questions about the random variable translate to functions on and involving X. Random variables with distributions from the base package are specified by capitalizing the name of the distribution.

```
> library(distr)
> X <- Binom(size = 3, prob = 1/2)
> X
```

```
Distribution Object of Class: Binom
 size: 3
 prob: 0.5
```

The analogue of the dbinom function for X is the d(X) function, and the analogue of the pbinom function is the p(X) function. Compare the following:

```
> d(X)(1)   # pmf of X evaluated at x = 1
```

```
[1] 0.375
```

```
> p(X)(2)   # cdf of X evaluated at x = 2
```

```
[1] 0.875
```

Random variables defined via the distr package may be *plotted*, which will return graphs of the PMF, CDF, and quantile function (introduced in Section 6.3.1). See Figure 5.3.2 for an example.

Given $X \sim$ dbinom(`size` $= n$, `prob` $= p$).

How to do:	with `stats` (default)	with `distr`
PMF: $\mathbb{P}(X = x)$	`dbinom(x, size = n, prob = p)`	`d(X)(x)`
CDF: $\mathbb{P}(X \leq x)$	`pbinom(x, size = n, prob = p)`	`p(X)(x)`
Simulate k variates	`rbinom(k, size = n, prob = p)`	`r(X)(k)`

For `distr` need `X <- Binom(size =`n`, prob =`p`)`

Table 5.1: Correspondence between `stats` and `distr`

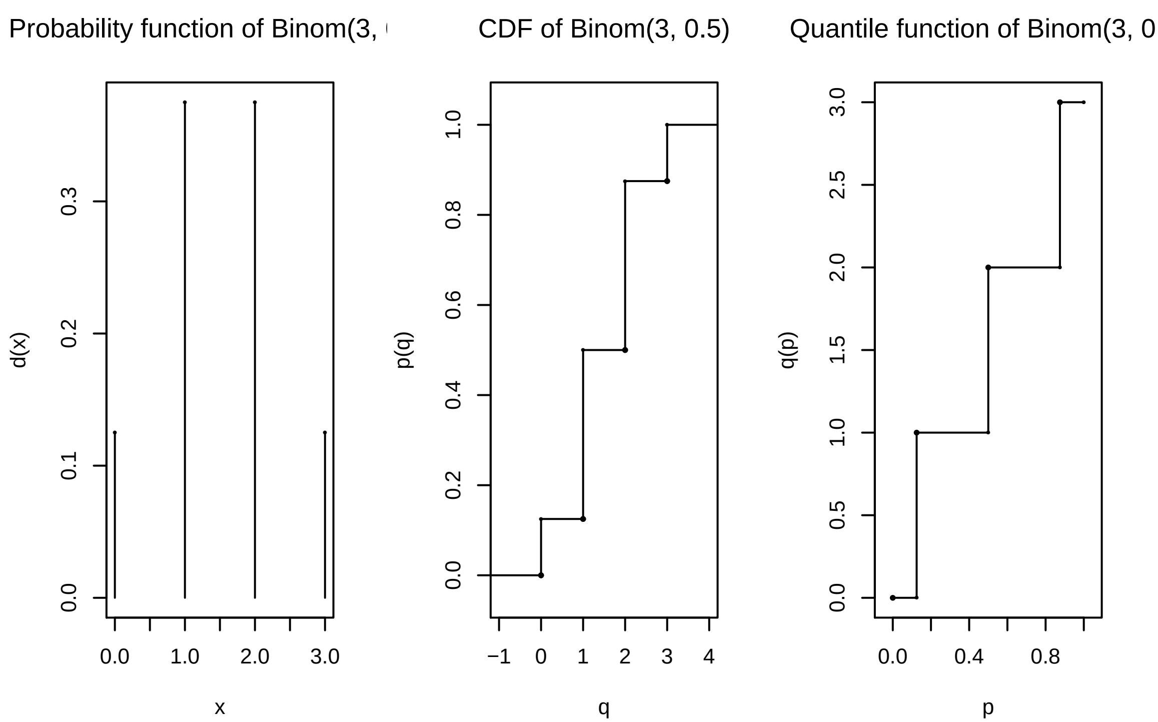

Figure 5.3.2: The binom(`size` = 3, `prob` = 0.5) distribution from the `distr` package

Chapter 6

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* [47, 48].

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

6.1 Continuous Random Variables

6.1.1 Probability Density Functions

Continuous random variables have supports that look like

$$S_X = [a, b] \text{ or } (a, b), \tag{6.1.1}$$

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

- 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[1] that satisfies three basic properties:

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

Remark 6.1. We can say the following about continuous random variables:

- Usually, the set A in 3 takes the form of an interval, for example, $A = [c, d]$, in which case
$$\mathbb{P}(X \in A) = \int_c^d f_X(x)\,\mathrm{d}x. \tag{6.1.2}$$
- 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:
$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b). \tag{6.1.3}$$
- 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 5. 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:

$$F_X(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^t f_X(x)\,\mathrm{d}x, \quad -\infty < t < \infty. \tag{6.1.4}$$

Remark 6.2. For any continuous CDF F_X the following are true.

- F_X is nondecreasing , that is, $t_1 \leq t_2$ implies $F_X(t_1) \leq F_X(t_2)$.

[1] 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 continuous (see Appendix E.2). Note the distinction from the discrete case: CDFs of discrete random variables are not continuous, they are only right continuous.
- $\lim_{t\to-\infty} F_X(t) = 0$ and $\lim_{t\to\infty} F_X(t) = 1$.

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

$$F'_X(t) = \frac{\mathrm{d}}{\mathrm{d}t}F_X(t) = \frac{\mathrm{d}}{\mathrm{d}t}\int_{-\infty}^{t} f_X(x)\,\mathrm{d}x = f_X(t), \tag{6.1.5}$$

the last equality being true by the Fundamental Theorem of Calculus, part (2) (see Appendix E.2). In short, $(F_X)' = f_X$ in the continuous case[2].

6.1.2 Expectation of Continuous Random Variables

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

$$\mathbb{E}\, g(X) = \int_{x\in S} g(x) f_X(x)\,\mathrm{d}x, \tag{6.1.6}$$

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

$$\mu = \mathbb{E}\, X = \int_{x\in S} x f_X(x)\,\mathrm{d}x, \tag{6.1.7}$$

provided $\int_S |x| f(x)\mathrm{d}x$ is finite. Also there is the variance

$$\sigma^2 = \mathbb{E}(X-\mu)^2 = \int_{x\in S} (x-\mu)^2 f_X(x)\,\mathrm{d}x, \tag{6.1.8}$$

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

$$M_X(t) = \mathbb{E}\ \mathrm{e}^{tX} = \int_{-\infty}^{\infty} \mathrm{e}^{tx} f_X(x)\,\mathrm{d}x, \tag{6.1.9}$$

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

Example 6.3. Let the continuous random variable X have PDF

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

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

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

[2] In the discrete case, $f_X(x) = F_X(x) - \lim_{t\to x^-} F_X(t)$.

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

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

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

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

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

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

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

Example 6.4. 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 $\int_{-\infty}^{\infty} f(x) \mathrm{d}x = 1$:

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

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

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

We locate the mean and variance just like before.

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

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

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

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

6.1.3 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 6.5. 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 6.3. We could integrate the function $xf(x) =$ `3*x^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 6.6. 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 6.4. Use `-Inf` for $-\infty$.

Example 6.7. Let us redo Example 6.3 with the `distr` package. The method is similar to that encountered in Section 5.1.3 in Chapter 5. 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 answers we found earlier. Now let us try expectation with the `distrEx` package [74]:

```
> library(distrEx)
> E(X)
```

```
[1] 0.7496337
```

```
> var(X)
```

```
[1] 0.03768305
```

```
> 3/80
```

```
[1] 0.0375
```

Compare these answers to the ones we found in Example 6.3. 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.

6.2 The Continuous Uniform Distribution

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

$$f_X(x) = \frac{1}{b-a}, \quad a < x < b. \tag{6.2.1}$$

The associated R function is dunif(`min` $= a$, `max` $= b$). We write $X \sim$ 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:

$$F_X(t) = \begin{cases} 0, & t < 0, \\ \frac{t-a}{b-a}, & a \leq t < b, \\ 1, & t \geq b. \end{cases} \tag{6.2.2}$$

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 6.8. Choose a number in [0,1] at random, and let X be the number chosen. Then $X \sim$ `unif(min = 0, max = 1)`.

The mean of $X \sim$ `unif(min = a, max = b)` is relatively simple to calculate:

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

using the popular formula for the difference of squares. The variance is left to Exercise 6.4.

6.3 The Normal Distribution

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

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\},\quad -\infty < x < \infty. \tag{6.3.1}$$

We write $X \sim$ `norm(mean = μ, sd = σ)`, 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 $Z \sim$ `norm(mean = 0, sd = 1)`. The lowercase Greek letter phi (ϕ) is used to denote the standard normal PDF and the capital Greek letter phi Φ is used to denote the standard normal CDF: for $-\infty < z < \infty$,

$$\phi(z) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-z^2/2} \text{ and } \Phi(t) = \int_{-\infty}^{t}\phi(z)\,\mathrm{d}z. \tag{6.3.2}$$

Proposition 6.9. *If* $X \sim \mathsf{norm}(\mathtt{mean} = \mu,\ \mathtt{sd} = \sigma)$ *then*

$$Z = \frac{X - \mu}{\sigma} \sim \mathsf{norm}(\mathtt{mean} = 0,\ \mathtt{sd} = 1). \tag{6.3.3}$$

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

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

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

$$M_Z(t) = \mathrm{e}^{-t^2/2}, \quad -\infty < t < \infty. \tag{6.3.4}$$

Example 6.10. The MGF of $X \sim \mathsf{norm}(\mathtt{mean} = \mu,\ \mathtt{sd} = \sigma)$ is then not difficult either because

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

Therefore:

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

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

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

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

Fact 6.11. *The same argument above shows that if X has MGF $M_X(t)$ then the MGF of $Y = a + bX$ is*

$$M_Y(t) = \mathrm{e}^{ta} M_X(bt). \tag{6.3.5}$$

Example 6.12. The 68-95-99.7 Rule. We saw in Section 3.3.6 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 6.13. 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. 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%.

6.3.1 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 6.14. Assuming the IQ model of Example 6.13, 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 3.13; 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 6.14 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 6.15. The *quantile function*[3] of a random variable X is the inverse of its cumulative distribution function:

$$Q_X(p) = \min\{x: \; F_X(x) \geq p\}, \quad 0 < p < 1. \tag{6.3.6}$$

Remark 6.16. Here are some properties of quantile functions:

1. The quantile function is defined and finite for all $0 < p < 1$.

[3]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 $p = 1$.

2. Q_X is left-continuous (see Appendix E.2). 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 5.3.2 for a comparison.

4. The two limits
$$\lim_{p\to 0^+} Q_X(p) \quad \text{and} \quad \lim_{p\to 1^-} Q_X(p)$$
always exist, but may be infinite (that is, sometimes $\lim_{p\to 0} Q(p) = -\infty$ and/or $\lim_{p\to 1} Q(p) = \infty$).

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

Definition 6.17. For $0 < \alpha < 1$, the symbol z_α denotes the unique solution of the equation $\mathbb{P}(Z > z_\alpha) = \alpha$, where $Z \sim$ norm(mean = 0, sd = 1). It can be calculated in one of two equivalent ways: `qnorm(1 −` α`)` and `qnorm(`α`, lower.tail = FALSE)`.

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

6.3.2 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 $Q_x(p)$ =`quantile(x, probs = ` p `, type = 1)` function.

Example 6.18. Back to Example 6.14, we are looking for $Q_X(0.99)$, where $X \sim$ norm(mean = 100, sd = 15). It could not be easier to do with R.

```
> qnorm(0.99, mean = 100, sd = 15)
[1] 134.8952
```

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

Example 6.19. Find the values $z_{0.025}$, $z_{0.01}$, and $z_{0.005}$ (these will play an important role from Chapter 9 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))
```