Discrete Distributions

G. Jay Kerns

September 9, 2011

1 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 resampling and Chapter **??**; it will also be needed in Section **??** 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*[**?**].

What do I want them to know?

- how to choose a reasonable discrete model under a variety of physical circumstances

- item 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

1.1 Discrete Random Variables

1.1.1 Probability Mass Functions

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

$$S_X = \{u_1, u_2, \ldots, u_k\} \text{ or } S_X = \{u_1, u_2, u_3 \ldots\}. \tag{1.1.1}$$

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

$$f_X(x) = \mathbb{P}(X = x), \quad x \in S_X. \tag{1.1.2}$$

Since values of the PMF represent probabilities, we know from Chapter **??** 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 1.1. 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

1.1.2 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):

$$\mu = \mathbb{E}X = \sum_{x \in S} x f_X(x), \tag{1.1.3}$$

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

$$\sigma^2 = \sum_{x \in S} (x - \mu)^2 f_X(x), \tag{1.1.4}$$

which can be computed (see Exercise 1.3) 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 1.2. We will calculate the mean of X in Example 1.1.

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

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 **??**). Another way it is commonly stated is that X is 3.5 "on the average" or "in the long run".

Remark 1.3. 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

$$F_X(t) = \mathbb{P}(X \le t), \quad -\infty < t < \infty. \tag{1.1.5}$$

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 ($t_1 \le t_2$ implies $F_X(t_1) \le F_X(t_2)$).

- 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.

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 1.2. 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 **??**):

```
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
sigma <- sqrt(sigma2)
sigma
```

```
 [1] 0.75
 [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 [?]. 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
```

1.2 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

$$f_X(x) = \frac{1}{m}, \quad x = 1, 2, \ldots, m. \tag{1.2.1}$$

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 $X \sim \mathsf{disunif}(m)$:

$$\mu = \sum_{x=1}^{m} x f_X(x) = \sum_{x=1}^{m} x \cdot \frac{1}{m} = \frac{1}{m}(1 + 2 + \cdots + m) = \frac{m+1}{2}, \tag{1.2.2}$$

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,

$$\sigma^2 = \sum_{x=1}^{m} x^2 f_X(x) - \mu^2 = \frac{(m+1)(2m+1)}{6} - \left(\frac{m+1}{2}\right)^2 = \cdots = \frac{m^2 - 1}{12}. \tag{1.2.3}$$

Example 1.4. Roll a die and let X be the upward face showing. Then $m = 6$, $\mu = 7/2 = 3.5$, and $\sigma^2 = (6^2 - 1)/12 = 35/12$.

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.

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`.

1.3 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

$$X = \begin{cases} 1 & \text{if the outcome is } S, \\ 0 & \text{if the outcome is } F. \end{cases} \tag{1.3.1}$$

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

$$f_X(x) = p^x(1 - p)^{1-x}, \quad x = 0, 1. \tag{1.3.2}$$

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

1.3.1 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

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n. \tag{1.3.3}$$

We say that X has a *binomial distribution* and we write $X \sim \mathtt{binom(size} = n, \mathtt{prob} = p)$. 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). 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 1.5. 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 $4(1/2)(1/2) = 1$.

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

$$0.0625$$
$$0.25$$
$$0.375$$
$$0.25$$
$$0.0625$$

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 1.7. 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 $S = \{$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)
diff(pbinom(c(6,9), size = 12, prob = 1/6))  # same thing
```

«echo = FALSE, fig = true, height = 4.5, width = 6»= plot(0, xlim = c(-1.2, 4.2), ylim = c(-0.04, 1.04), type = "n", xlab = "number of successes", ylab = "cumulative probability") abline(h = c(0,1), lty = 2, col = "grey") lines(stepfun(0:3, pbinom(-1:3, size = 3, prob = 0.5)), verticals = FALSE, do.p = FALSE) points(0:3, pbinom(0:3, size = 3, prob = 0.5), pch = 16, cex = 1.2) points(0:3, pbinom(-1:2, size = 3, prob = 0.5), pch = 1, cex = 1.2)

Figure 1.3.1: Graph of the `binom(size = 3, prob = 1/2)` CDF

```
[1] 0.001291758
[1] 0.001291758
```

Example 1.8. 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 1.3.1.

Example 1.9. Another way to do Example 1.8 is with the `distr` family of packages [**?**]. They use an object oriented approach to random variables, that is, a random variable is stored in an object 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
p(X)(2)    # cdf of X evaluated at x = 2
```

```
[1] 0.375
[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 **??**). See Figure 1.3.2 for an example.

```
plot(X, cex = 0.2)
```

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 $= p$)

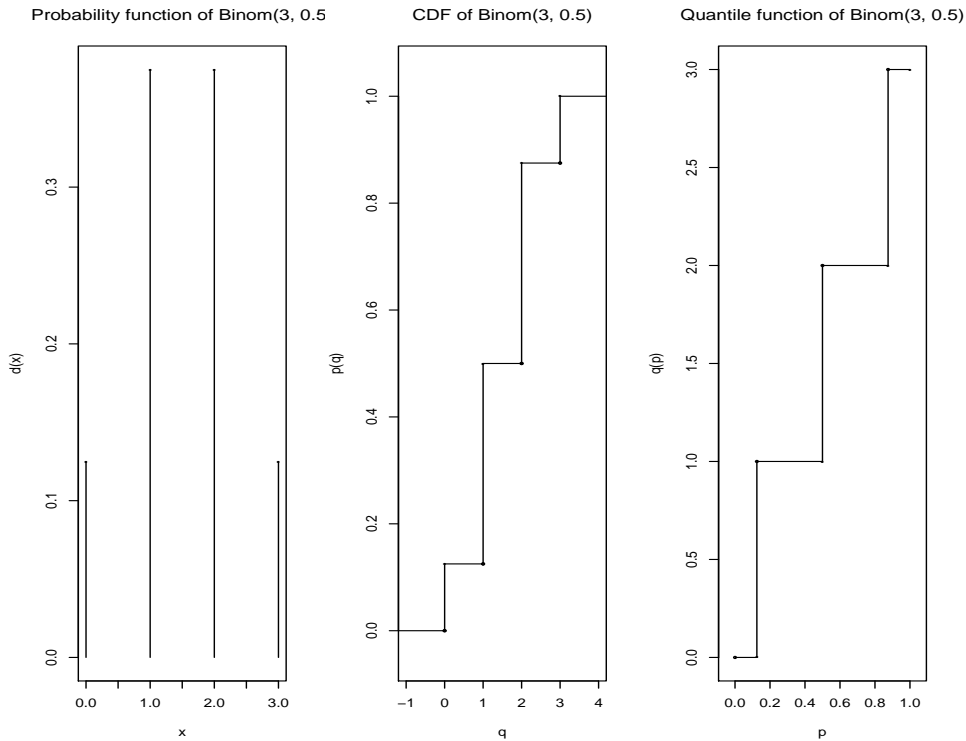Table 1.1: Correspondence between stats and distr

Figure 1.3.2: The binom(size $= 3$, prob $= 0.5$) distribution from the distr package.

1.4 Expectation and Moment Generating Functions

1.4.1 The Expectation Operator

We next generalize some of the concepts from Section 1.1.2. There we saw that every [1] PMF has two important numbers associated with it:

$$\mu = \sum_{x \in S} x f_X(x), \quad \sigma^2 = \sum_{x \in S} (x - \mu)^2 f_X(x). \tag{1.4.1}$$

Intuitively, for repeated observations of X we would expect the sample mean to closely approximate μ as the sample size increases without bound. For this reason we call μ the *expected value* of X and we write $\mu = \mathbb{E}X$, where $\mathbb{E}$ is an *expectation operator*.

Definition 1.10. More generally, given a function g we define the /expected value of/ $g(X)$ by

$$\mathbb{E}\, g(X) = \sum_{x \in S} g(x) f_X(x), \tag{1.4.2}$$

provided the (potentially infinite) series $\sum_x |g(x)| f(x)$ is convergent. We say that $\mathbb{E}g(X)$ /exists/.

In this notation the variance is $\sigma^2 = \mathbb{E}(X - \mu)^2$ and we prove the identity

$$\mathbb{E}(X - \mu)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2 \tag{1.4.3}$$

in Exercise 1.3. Intuitively, for repeated observations of X we would expect the sample mean of the $g(X)$ values to closely approximate $\mathbb{E}\, g(X)$ as the sample size increases without bound.

Let us take the analogy further. If we expect $g(X)$ to be close to $\mathbb{E}g(X)$ on the average, where would we expect $3g(X)$ to be on the average? It could only be $3\mathbb{E}g(X)$. The following theorem makes this idea precise.

Proposition 1.11. *For any functions g and h, any random variable X, and any constant c:*

1. $\mathbb{E}\, c = c,$

2. $\mathbb{E}[c \cdot g(X)] = c\mathbb{E}g(X)$

3. $\mathbb{E}[g(X) + h(X)] = \mathbb{E}g(X) + \mathbb{E}h(X),$

provided $\mathbb{E}g(X)$ and $\mathbb{E}h(X)$ exist.

Proof. Go directly from the definition. For example,

$$\mathbb{E}[c \cdot g(X)] = \sum_{x \in S} c \cdot g(x) f_X(x) = c \cdot \sum_{x \in S} g(x) f_X(x) = c\mathbb{E}g(X).$$

□

---

[1]Not every, only those PMFs for which the (potentially infinite) series converges.

1.4.2 Moment Generating Functions

Definition 1.12. Given a random variable X, its /moment generating function/ (abbreviated MGF) is defined by the formula

$$M_X(t) = \mathbb{E} e^{tX} = \sum_{x \in S} e^{tx} f_X(x), \qquad (1.4.4)$$

provided the (potentially infinite) series is convergent for all t in a neighborhood of zero (that is, for all $-\epsilon < t < \epsilon$, for some $\epsilon > 0$).

Note that for any MGF M_X,

$$M_X(0) = \mathbb{E} e^{0 \cdot X} = \mathbb{E} 1 = 1. \qquad (1.4.5)$$

We will calculate the MGF for the two distributions introduced above.

Example 1.13. Find the MGF for $X \sim \mathsf{disunif}(m)$.
Since $f(x) = 1/m$, the MGF takes the form

$$M(t) = \sum_{x=1}^{m} e^{tx} \frac{1}{m} = \frac{1}{m}(e^t + e^{2t} + \cdots + e^{mt}), \quad \text{for any } t.$$

Example 1.14. Find the MGF for $X \sim \mathsf{binom}(\mathtt{size} = n, \mathtt{prob} = p)$.

$$M_X(t) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x},$$

$$= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x},$$

$$= (pe^t + q)^n, \quad \text{for any } t.$$

Applications   We will discuss three applications of moment generating functions in this book. The first is the fact that an MGF may be used to accurately identify the probability distribution that generated it, which rests on the following:

Theorem 1.15. *The moment generating function, if it exists in a neighborhood of zero, determines a probability distribution /uniquely/.*

Proof. Unfortunately, the proof of such a theorem is beyond the scope of a text like this one. Interested readers could consult Billingsley [**?**].   □

We will see an example of Theorem 1.15 in action.

Example 1.16. Suppose we encounter a random variable which has MGF

$$M_X(t) = (0.3 + 0.7e^t)^{13}.$$

Then $X \sim \mathsf{binom}(\mathtt{size} = 13, \mathtt{prob} = 0.7)$.

An MGF is also known as a "Laplace Transform" and is manipulated in that context in many branches of science and engineering.

Why is it called a Moment Generating Function? This brings us to the second powerful application of MGFs. Many of the models we study have a simple MGF, indeed, which permits us to determine the mean, variance, and even higher moments very quickly. Let us see why. We already know that

$$M(t) = \sum_{x \in S} e^{tx} f(x).$$

Take the derivative with respect to t to get

$$M'(t) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{x \in S} e^{tx} f(x)\right) = \sum_{x \in S} \frac{\mathrm{d}}{\mathrm{d}t}\left(e^{tx} f(x)\right) = \sum_{x \in S} x e^{tx} f(x), \tag{1.4.6}$$

and so if we plug in zero for t we see

$$M'(0) = \sum_{x \in S} x e^0 f(x) = \sum_{x \in S} x f(x) = \mu = \mathbb{E}X. \tag{1.4.7}$$

Similarly, $M''(t) = \sum x^2 e^{tx} f(x)$ so that $M''(0) = \mathbb{E}X^2$. And in general, we can see [2] that

$$M_X^{(r)}(0) = \mathbb{E}X^r = r^{\text{th}} \text{ moment of } X \text{ about the origin.} \tag{1.4.8}$$

These are also known as *raw moments* and are sometimes denoted μ'_r. In addition to these are the so called *central moments* μ_r defined by

$$\mu_r = \mathbb{E}(X - \mu)^r, \quad r = 1, 2, \ldots \tag{1.4.9}$$

Example 1.17. Let $X \sim \mathsf{binom}(\mathtt{size} = n, \mathtt{prob} = p)$ with $M(t) = (q + pe^t)^n$.

We calculated the mean and variance of a binomial random variable in Section 1.3 by means of the binomial series. But look how quickly we find the mean and variance with the moment generating function.

$$\begin{aligned} M'(t) &= n(q + pe^t)^{n-1} pe^t \big|_{t=0}, \\ &= n \cdot 1^{n-1} p, \\ &= np. \end{aligned}$$

---

[2] We are glossing over some significant mathematical details in our derivation. Suffice it to say that when the MGF exists in a neighborhood of $t = 0$, the exchange of differentiation and summation is valid in that neighborhood, and our remarks hold true.

And

$$M''(0) = n(n-1)[q + pe^t]^{n-2}(pe^t)^2 + n[q + pe^t]^{n-1}pe^t \big|_{t=0} ,$$
$$\mathbb{E}X^2 = n(n-1)p^2 + np.$$

Therefore

$$\sigma^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2,$$
$$= n(n-1)p^2 + np - n^2p^2,$$
$$= np - np^2 = npq.$$

See how much easier that was?

Remark 1.18. We learned in this section that $M^{(r)}(0) = \mathbb{E}X^r$. We remember from Calculus II that certain functions f can be represented by a Taylor series expansion about a point a, which takes the form

$$f(x) = \sum_{r=0}^{\infty} \frac{f^{(r)}(a)}{r!}(x - a)^r, \quad \text{for all } |x - a| < R, \tag{1.4.10}$$

where R is called the /radius of convergence/ of the series (see Appendix **??**). We combine the two to say that if an MGF exists for all t in the interval $(-\epsilon, \epsilon)$, then we can write

$$M_X(t) = \sum_{r=0}^{\infty} \frac{\mathbb{E}X^r}{r!}t^r, \quad \text{for all } |t| < \epsilon. \tag{1.4.11}$$

How to do it with R The `distrEx` package provides an expectation operator `E` which can be used on random variables that have been defined in the ordinary `distr` sense:

```
X <- Binom(size = 3, prob = 0.45)
library(distrEx)
E(X)
E(3*X + 4)
```

```
 [1] 1.35
 [1] 8.05
```

For discrete random variables with finite support, the expectation is simply computed with direct summation. In the case that the random variable has infinite support and the function is crazy, then the expectation is not computed directly, rather, it is estimated by first generating a random sample from the underlying model and next computing a sample mean of the function of interest.

There are methods for other population parameters:

```
var(X)
sd(X)
```

```
 [1] 0.7425
 [1] 0.8616844
```

There are even methods for `IQR`, `mad`, `skewness`, and `kurtosis`.

1.5 The Empirical Distribution

Do an experiment n times and observe n values $x_1, x_2, \ldots, x_n$ of a random variable X. For simplicity in most of the discussion that follows it will be convenient to imagine that the observed values are distinct, but the remarks are valid even when the observed values are repeated.

Definition 1.19. The /empirical cumulative distribution function/ F_n (written ECDF) is the probability distribution that places probability mass $1/n$ on each of the values $x_1, x_2, \ldots, x_n$. The empirical PMF takes the form

$$f_X(x) = \frac{1}{n}, \quad x \in \{x_1, x_2, ..., x_n\}. \tag{1.5.1}$$

If the value x_i is repeated k times, the mass at x_i is accumulated to k/n.

The mean of the empirical distribution is

$$\mu = \sum_{x \in S} x f_X(x) = \sum_{i=1}^{n} x_i \cdot \frac{1}{n} \tag{1.5.2}$$

and we recognize this last quantity to be the sample mean, $\overline{x}$. The variance of the empirical distribution is

$$\sigma^2 = \sum_{x \in S} (x - \mu)^2 f_X(x) = \sum_{i=1}^{n} (x_i - \overline{x})^2 \cdot \frac{1}{n} \tag{1.5.3}$$

and this last quantity looks very close to what we already know to be the sample variance.

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2. \tag{1.5.4}$$

The *empirical quantile function* is the inverse of the ECDF. See Section **??**.

How to do it with R   The empirical distribution is not directly available as a distribution in the same way that the other base probability distributions are, but there are plenty of resources available for the determined investigator. Given a data vector of observed values x, we can see the empirical CDF with the `ecdf` function:

```
x <- c(4, 7, 9, 11, 12)
ecdf(x)
```

```
 Empirical CDF
 Call: ecdf(x)
  x[1:5] =       4,       7,       9,      11,      12
```

The above shows that the returned value of `ecdf(x)` is not a *number* but rather a *function*. The ECDF is not usually used by itself in this form, by itself. More commonly it is used as an intermediate step in a more complicated calculation, for instance, in hypothesis testing (see Chapter **??**) or resampling (see Chapter **??**). It is nevertheless instructive to see what the `ecdf` looks like, and there is a special plot method for `ecdf` objects.
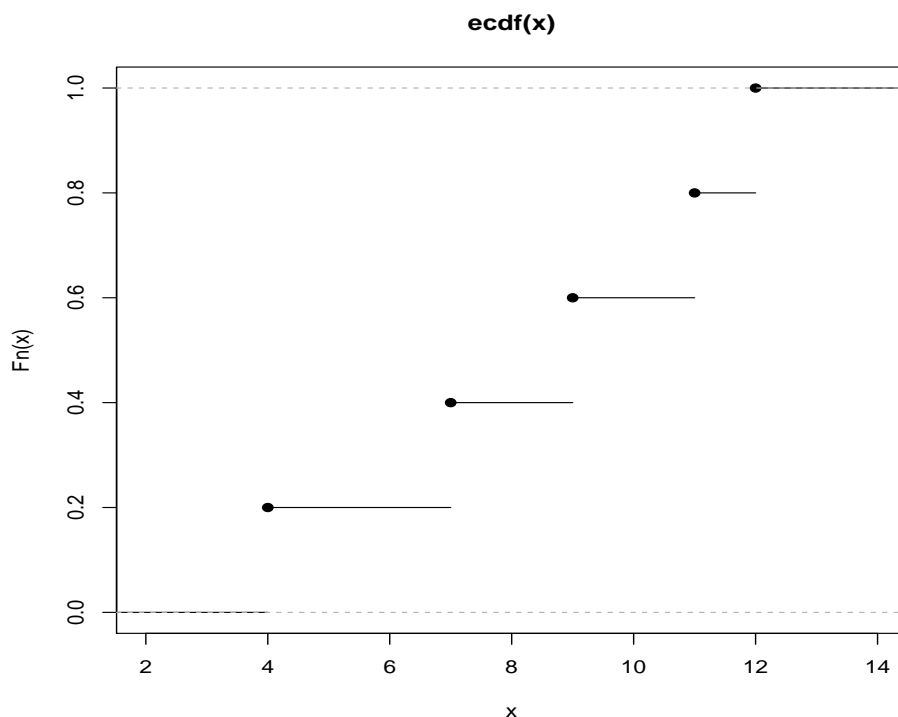
```
plot(ecdf(x))
```

ecdf(x)

Figure 1.5.1: The empirical CDF.

See Figure 1.5.1. The graph is of a right-continuous function with jumps exactly at the locations stored in x. There are no repeated values in x so all of the jumps are equal to $1/5 = 0.2$.

The empirical PDF is not usually of particular interest in itself, but if we really wanted we could define a function to serve as the empirical PDF:

```
epdf <- function(x) function(t){sum(x %in% t)/length(x)}
x <- c(0,0,1)
epdf(x)(0)        # should be 2/3
```

```
 [1] 0.6666667
```

To simulate from the empirical distribution supported on the vector x, we use the sample function.

```
x <- c(0,0,1)
sample(x, size = 7, replace = TRUE)
```

```
[1] 1 1 0 1 0 0 1
```

We can get the empirical quantile function in R with `quantile(x, probs = p, type = 1)`; see Section **??**.

As we hinted above, the empirical distribution is significant more because of how and where it appears in more sophisticated applications. We will explore some of these in later chapters – see, for instance, Chapter **??**.

1.6 Other Discrete Distributions

The binomial and discrete uniform distributions are popular, and rightly so; they are simple and form the foundation for many other more complicated distributions. But the particular uniform and binomial models only apply to a limited range of problems. In this section we introduce situations for which we need more than what the uniform and binomial offer.

1.6.1 Dependent Bernoulli Trials

The Hypergeometric Distribution   Consider an urn with 7 white balls and 5 black balls. Let our random experiment be to randomly select 4 balls, without replacement, from the urn. Then the probability of observing 3 white balls (and thus 1 black ball) would be

$$\mathbb{P}(3W, 1B) = \frac{\binom{7}{3}\binom{5}{1}}{\binom{12}{4}}. \tag{1.6.1}$$

More generally, we sample without replacement K times from an urn with M white balls and N black balls. Let X be the number of white balls in the sample. The PMF of X is

$$f_X(x) = \frac{\binom{M}{x}\binom{N}{K-x}}{\binom{M+N}{K}}. \tag{1.6.2}$$

We say that X has a *hypergeometric distribution* and write $X \sim \mathsf{hyper}(\mathtt{m} = M, \mathtt{n} = N, \mathtt{k} = K)$.

The support set for the hypergeometric distribution is a little bit tricky. It is tempting to say that x should go from 0 (no white balls in the sample) to K (no black balls in the sample), but that does not work if $K > M$, because it is impossible to have more white balls in the sample than there were white balls originally in the urn. We have the same trouble if $K > N$. The good news is that the majority of examples we study have $K \leq M$ and $K \leq N$ and we will happily take the support to be $x = 0, 1, \ldots, K$.

It is shown in Exercise 1.5 that

$$\mu = K\frac{M}{M+N}, \quad \sigma^2 = K\frac{MN}{(M+N)^2}\frac{M+N-K}{M+N-1}. \tag{1.6.3}$$

1 Discrete Distributions

The associated R functions for the PMF and CDF are `dhyper(x, m, n, k)` and `phyper`, respectively. There are two more functions: `qhyper`, which we will discuss in Section **??**, and `rhyper`, discussed below.

Example 1.20. Suppose in a certain shipment of 250 Pentium processors there are 17 defective processors. A quality control consultant randomly collects 5 processors for inspection to determine whether or not they are defective. Let X denote the number of defectives in the sample.

1. Find the probability of exactly 3 defectives in the sample, that is, find $\mathbb{P}(X = 3)$.

   /Solution:/ We know that $X \sim \text{hyper}(m = 17, n = 233, k = 5)$. So the required probability is just

   $$f_X(3) = \frac{\binom{17}{3}\binom{233}{2}}{\binom{250}{5}}.$$

   To calculate it in R we just type

   ```
   dhyper(3, m = 17, n = 233, k = 5)
   ```

   ```
   [1] 0.002351153
   ```

   To find it with the R Commander we go Probability ▷ Discrete Distributions ▷ Hypergeometric distribution ▷ Hypergeometric probabilities... . We fill in the parameters $m = 17$, $n = 233$, and $k = 5$. Click OK, and the following table is shown in the window.

   ```
   A <- data.frame(Pr=dhyper(0:4, m = 17, n = 233, k = 5))
   rownames(A) <- 0:4
   A
   ```

   ```
              Pr
   0 7.011261e-01
   1 2.602433e-01
   2 3.620776e-02
   3 2.351153e-03
   4 7.093997e-05
   ```

   We wanted $\mathbb{P}(X = 3)$, and this is found from the table to be approximately 0.0024. The value is rounded to the fourth decimal place.

We know from our above discussion that the sample space should be $x = 0, 1, 2, 3, 4, 5$, yet, in the table the probabilities are only displayed for $x = 1, 2, 3$, and 4. What is happening? As it turns out, the R Commander will only display probabilities that are 0.00005 or greater. Since $x = 5$ is not shown, it suggests that the outcome has a tiny probability. To find its exact value we use the =dhyper= function:

```
dhyper(5, m = 17, n = 233, k = 5)
```

```
[1] 7.916049e-07
```

In other words, $\mathbb{P}(X = 5) \approx 0.0000007916049$, a small number indeed.

2. Find the probability that there are at most 2 defectives in the sample, that is, compute $\mathbb{P}(X \leq 2)$.

   /Solution:/ Since $\mathbb{P}(X \leq 2) = \mathbb{P}(X = 0, 1, 2)$, one way to do this would be to add the 0, 1, and 2 entries in the above table. this gives $0.7011 + 0.2602 + 0.0362 = 0.9975$. Our answer should be correct up to the accuracy of 4 decimal places. However, a more precise method is provided by the R Commander. Under the Hypergeometric distribution menu we select Hypergeometric tail probabilities.... We fill in the parameters m, n, and k as before, but in the Variable value(s) dialog box we enter the value 2. We notice that the =Lower tail= option is checked, and we leave that alone. Click OK.

   ```
   phyper(2, m = 17, n = 233, k = 5)
   ```

   ```
   [1] 0.9975771
   ```

   And thus $\mathbb{P}(X \leq 2) \approx 0.9975771$. We have confirmed that the above answer was correct up to four decimal places.

3. Find $\mathbb{P}(X > 1)$.

   The table did not give us the explicit probability $\mathbb{P}(X = 5)$, so we can not use the table to give us this probability. We need to use another method. Since $\mathbb{P}(X > 1) = 1 - \mathbb{P}(X \leq 1) = 1 - F_X(1)$, we can find the probability with Hypergeometric tail probabilities.... We enter 1 for Variable Value(s), we enter the parameters as before, and in this case we choose the =Upper tail= option. This results in the following output.

   ```
   phyper(1, m = 17, n = 233, k = 5, lower.tail = FALSE)
   ```

```
[1] 0.03863065
```

In general, the =Upper tail= option of a tail probabilities dialog computes $\mathbb{P}(X > x)$ for all given Variable Value(s) x.

4. Generate $100,000$ observations of the random variable X.

We can randomly simulate as many observations of X as we want in R Commander. Simply choose Simulate hypergeometric variates... in the Hypergeometric distribution dialog.

In the Number of samples dialog, type 1. Enter the parameters as above. Under the Store Values section, make sure New Data set is selected. Click OK.

A new dialog should open, with the default name =Simset1=. We could change this if we like, according to the rules for R object names. In the sample size box, enter 100000. Click OK.

In the Console Window, R Commander should issue an alert that `Simset1` has been initialized, and in a few seconds, it should also state that 100,000 hypergeometric variates were stored in =hyper.sim1=. We can view the sample by clicking the View Data Set button on the R Commander interface.

We know from our formulas that $\mu = K \cdot M/(M + N) = 5 * 17/250 = 0.34$. We can check our formulas using the fact that with repeated observations of X we would expect about 0.34 defectives on the average. To see how our sample reflects the true mean, we can compute the sample mean

```
Rcmdr> mean(Simset2$hyper.sim1, na.rm=TRUE)
[1] 0.340344

Rcmdr> sd(Simset2$hyper.sim1, na.rm=TRUE)
[1] 0.5584982
...
```

We see that when given many independent observations of X, the sample mean is very close to the true mean μ. We can repeat the same idea and use the sample standard deviation to estimate the true standard deviation of X. From the output above our estimate is 0.5584982, and from our formulas we get

$$\sigma^2 = K\frac{MN}{(M+N)^2}\frac{M+N-K}{M+N-1} \approx 0.3117896,$$

with $\sigma = \sqrt{\sigma^2} \approx 0.5583811944$. Our estimate was pretty close.

From the console we can generate random hypergeometric variates with the =rhyper= function, as demonstrated below.

```
rhyper(10, m = 17, n = 233, k = 5)
```

```
[1] 0 0 1 0 1 1 0 0 1 2
```

Sampling With and Without Replacement   Suppose that we have a large urn with, say, M white balls and N black balls. We take a sample of size n from the urn, and let X count the number of white balls in the sample. If we sample

without replacement, then $X \sim$ hyper(m $=M$, n $= N$, k $= n$) and has mean and variance

$$
\begin{aligned}
\mu &= n\frac{M}{M+N}, \\
\sigma^2 &= n\frac{MN}{(M+N)^2}\frac{M+N-n}{M+N-1}, \\
&= n\frac{M}{M+N}\left(1 - \frac{M}{M+N}\right)\frac{M+N-n}{M+N-1}.
\end{aligned}
$$

On the other hand, if we sample

with replacement, then $X \sim$ binom(size $= n$, prob $= M/(M + N)$) with mean and variance

$$
\begin{aligned}
\mu &= n\frac{M}{M+N}, \\
\sigma^2 &= n\frac{M}{M+N}\left(1 - \frac{M}{M+N}\right).
\end{aligned}
$$

We see that both sampling procedures have the same mean, and the method with the larger variance is the "with replacement" scheme. The factor by which the variances differ,

$$
\frac{M+N-n}{M+N-1}, \tag{1.6.4}
$$

is called a *finite population correction*. For a fixed sample size n, as $M, N \rightarrow \infty$ it is clear that the correction goes to 1, that is, for infinite populations the sampling schemes are essentially the same with respect to mean and variance.

1.6.2 Waiting Time Distributions

Another important class of problems is associated with the amount of time it takes for a specified event of interest to occur. For example, we could flip a coin repeatedly until we observe Heads. We could toss a piece of paper repeatedly until we make it in the trash can.

The Geometric Distribution Suppose that we conduct Bernoulli trials repeatedly, noting the successes and failures. Let X be the number of failures before a success. If $\mathbb{P}(S) = p$ then X has PMF

$$f_X(x) = p(1-p)^x, \quad x = 0, 1, 2, \ldots \tag{1.6.5}$$

(Why?) We say that X has a *Geometric distribution* and we write $X \sim \text{geom}(\text{prob} = p)$. The associated R functions are `dgeom(x, prob)`, `pgeom`, `qgeom`, and `rhyper`, which give the PMF, CDF, quantile function, and simulate random variates, respectively.

Again it is clear that $f(x) \geq 0$ and we check that $\sum f(x) = 1$ (see Equation **??** in Appendix **??**):

$$\sum_{x=0}^{\infty} p(1-p)^x = p \sum_{x=0}^{\infty} q^x = p \frac{1}{1-q} = 1.$$

We will find in the next section that the mean and variance are

$$\mu = \frac{1-p}{p} = \frac{q}{p} \text{ and } \sigma^2 = \frac{q}{p^2}. \tag{1.6.6}$$

Example 1.21. The Pittsburgh Steelers place kicker, Jeff Reed, made 81.2% of his attempted field goals in his career up to 2006. Assuming that his successive field goal attempts are approximately Bernoulli trials, find the probability that Jeff misses at least 5 field goals before his first successful goal.

Solution: If $X =$ the number of missed goals until Jeff's first success, then $X \sim \text{geom}(\text{prob} = 0.812)$ and we want $\mathbb{P}(X \geq 5) = \mathbb{P}(X > 4)$. We can find this in R with

```
pgeom(4, prob = 0.812, lower.tail = FALSE)
```

```
[1] 0.0002348493
```

Note 1.22. Some books use a slightly different definition of the geometric distribution. They consider Bernoulli trials and let Y count instead the number of trials until a success, so that Y has PMF

$$f_Y(y) = p(1-p)^{y-1}, \quad y = 1, 2, 3, \ldots \tag{1.6.7}$$

When they say "geometric distribution", this is what they mean. It is not hard to see that the two definitions are related. In fact, if X denotes our geometric and Y theirs, then $Y = X + 1$. Consequently, they have $\mu_Y = \mu_X + 1$ and $\sigma_Y^2 = \sigma_X^2$.

The Negative Binomial Distribution We may generalize the problem and consider the case where we wait for *more* than one success. Suppose that we conduct Bernoulli trials repeatedly, noting the respective successes and failures. Let X count the number of failures before r successes. If $\mathbb{P}(S) = p$ then X has PMF

$$f_X(x) = \binom{r + x - 1}{r - 1} p^r (1 - p)^x, \quad x = 0, 1, 2, \ldots \tag{1.6.8}$$

We say that X has a *Negative Binomial distribution* and write $X \sim$ nbinom(size = r, prob = p). The associated R functions are dnbinom(x, size, prob), pnbinom, qnbinom, and rnbinom, which give the PMF, CDF, quantile function, and simulate random variates, respectively.

As usual it should be clear that $f_X(x) \geq 0$ and the fact that $\sum f_X(x) = 1$ follows from a generalization of the geometric series by means of a Maclaurin's series expansion:

$$\frac{1}{1 - t} = \sum_{k=0}^{\infty} t^k, \quad \text{for } -1 < t < 1, \text{ and} \tag{1.6.9}$$

$$\frac{1}{(1 - t)^r} = \sum_{k=0}^{\infty} \binom{r + k - 1}{r - 1} t^k, \quad \text{for } -1 < t < 1. \tag{1.6.10}$$

Therefore

$$\sum_{x=0}^{\infty} f_X(x) = p^r \sum_{x=0}^{\infty} \binom{r + x - 1}{r - 1} q^x = p^r (1 - q)^{-r} = 1, \tag{1.6.11}$$

since $|q| = |1 - p| < 1$.

Example 1.23. We flip a coin repeatedly and let X count the number of Tails until we get seven Heads. What is $\mathbb{P}(X = 5)$?

Solution: We know that $X \sim$ nbinom(size = 7, prob = 1/2).

$$\mathbb{P}(X = 5) = f_X(5) = \binom{7 + 5 - 1}{7 - 1}(1/2)^7(1/2)^5 = \binom{11}{6}2^{-12}$$

and we can get this in R with

```
dnbinom(5, size = 7, prob = 0.5)
```

```
[1] 0.112793
```

Let us next compute the MGF of $X \sim$ nbinom(size = r, prob = p).

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx} \binom{r+x-1}{r-1} p^r q^x$$

$$= p^r \sum_{x=0}^{\infty} \binom{r+x-1}{r-1} [qe^t]^x$$

$$= p^r(1-qe^t)^{-r}, \quad \text{provided } |qe^t| < 1,$$

and so

$$M_X(t) = \left(\frac{p}{1-qe^t}\right)^r, \quad \text{for } qe^t < 1. \tag{1.6.12}$$

We see that $qe^t < 1$ when $t < -\ln(1-p)$.

Let $X \sim \text{nbinom}(\text{size} = r, \text{prob} = p)$ with $M(t) = p^r(1-qe^t)^{-r}$. We proclaimed above the values of the mean and variance. Now we are equipped with the tools to find these directly.

$$M'(t) = p^r(-r)(1-qe^t)^{-r-1}(-qe^t),$$

$$= rqe^t p^r(1-qe^t)^{-r-1},$$

$$= \frac{rqe^t}{1-qe^t} M(t), \text{ and so}$$

$$M'(0) = \frac{rq}{1-q} \cdot 1 = \frac{rq}{p}.$$

Thus $\mu = rq/p$. We next find $\mathbb{E}X^2$.

$$M''(0) = \left. \frac{rqe^t(1-qe^t) - rqe^t(-qe^t)}{(1-qe^t)^2} M(t) + \frac{rqe^t}{1-qe^t} M'(t) \right|_{t=0},$$

$$= \frac{rqp + rq^2}{p^2} \cdot 1 + \frac{rq}{p}\left(\frac{rq}{p}\right),$$

$$= \frac{rq}{p^2} + \left(\frac{rq}{p}\right)^2.$$

Finally we may say $\sigma^2 = M''(0) - [M'(0)]^2 = rq/p^2$.

Example 1.24. A random variable has MGF

$$M_X(t) = \left(\frac{0.19}{1-0.81e^t}\right)^{31}.$$

Then $X \sim \text{nbinom}(\text{size} = 31, \text{prob} = 0.19)$.

28

Note 1.25. As with the Geometric distribution, some books use a slightly different defini-
tion of the Negative Binomial distribution. They consider Bernoulli trials and let Y be the
number of trials until r successes, so that Y has PMF

$$f_Y(y) = \binom{y-1}{r-1} p^r (1-p)^{y-r}, \quad y = r, r+1, r+2, \ldots \tag{1.6.13}$$

It is again not hard to see that if X denotes our Negative Binomial and Y theirs, then
$Y = X + r$. Consequently, they have $\mu_Y = \mu_X + r$ and $\sigma_Y^2 = \sigma_X^2$.

1.6.3 Arrival Processes

The Poisson Distribution   This is a distribution associated with "rare events", for rea-
sons which will become clear in a moment. The events might be:

- traffic accidents,

- typing errors, or

- customers arriving in a bank.

Let λ be the average number of events in the time interval $[0, 1]$. Let the random variable
X count the number of events occurring in the interval. Then under certain reasonable
conditions it can be shown that

$$f_X(x) = \mathbb{P}(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots \tag{1.6.14}$$

We use the notation $X \sim \mathsf{pois}(\mathtt{lambda} = \lambda)$. The associated R functions are $\mathtt{dpois(x,}$
$\mathtt{lambda)}$, $\mathtt{ppois}$, $\mathtt{qpois}$, and $\mathtt{rpois}$, which give the PMF, CDF, quantile function, and
simulate random variates, respectively.

What are the reasonable conditions?   Divide $[0, 1]$ into subintervals of length $1/n$.
A *Poisson process* satisfies the following conditions:

- the probability of an event occurring in a particular subinterval is $\approx \lambda/n$.

- the probability of two or more events occurring in any subinterval is ≈ 0.

- occurrences in disjoint subintervals are independent.

Remark 1.26. If X counts the number of events in the interval $[0, t]$ and λ is the average
number that occur in unit time, then $X \sim \mathsf{pois}(\mathtt{lambda} = \lambda t)$, that is,

$$\mathbb{P}(X = x) = e^{-\lambda t} \frac{(\lambda t)^x}{x!}, \quad x = 0, 1, 2, 3 \ldots \tag{1.6.15}$$

Example 1.27. On the average, five cars arrive at a particular car wash every hour. Let X count the number of cars that arrive from 10AM to 11AM. Then $X \sim$ pois(lambda = 5). Also, $\mu = \sigma^2 = 5$. What is the probability that no car arrives during this period?

   Solution: The probability that no car arrives is

$$\mathbb{P}(X = 0) = e^{-5}\frac{5^0}{0!} = e^{-5} \approx 0.0067.$$

Example 1.28. Suppose the car wash above is in operation from 8AM to 6PM, and we let Y be the number of customers that appear in this period. Since this period covers a total of 10 hours, from Remark 1.26 we get that $Y \sim$ pois(lambda = 5 * 10 = 50). What is the probability that there are between 48 and 50 customers, inclusive?

   Solution: We want $\mathbb{P}(48 \le Y \le 50) = \mathbb{P}(X \le 50) - \mathbb{P}(X \le 47)$.

```
diff(ppois(c(47, 50), lambda = 50))
```

```
[1] 0.1678485
```

1.7 Functions of Discrete Random Variables

We have built a large catalogue of discrete distributions, but the tools of this section will give us the ability to consider infinitely many more. Given a random variable X and a given function h, we may consider $Y = h(X)$. Since the values of X are determined by chance, so are the values of Y. The question is, what is the PMF of the random variable Y? The answer, of course, depends on h. In the case that h is one-to-one (see Appendix **??**), the solution can be found by simple substitution.

Example 1.29. Let $X \sim$ nbinom(size = r, prob = p). We saw in 1.6 that X represents the number of failures until r successes in a sequence of Bernoulli trials. Suppose now that instead we were interested in counting the number of trials (successes and failures) until the rth success occurs, which we will denote by Y. In a given performance of the experiment, the number of failures (X) and the number of successes (r) together will comprise the total number of trials (Y), or in other words, $X + r = Y$. We may let h be defined by $h(x) = x + r$ so that $Y = h(X)$, and we notice that h is linear and hence one-to-one. Finally, X takes values $0, 1, 2, \ldots$ implying that the support of Y would be $\{r, r + 1, r + 2, \ldots\}$. Solving for X we get $X = Y - r$. Examining the PMF of X

$$f_X(x) = \binom{r + x - 1}{r - 1} p^r (1 - p)^x, \tag{1.7.1}$$

we can substitute $x = y - r$ to get

$$
\begin{aligned}
f_Y(y) &= f_X(y - r), \\
&= \binom{r + (y - r) - 1}{r - 1} p^r (1 - p)^{y-r}, \\
&= \binom{y - 1}{r - 1} p^r (1 - p)^{y-r}, \quad y = r,\, r + 1, \ldots
\end{aligned}
$$

Even when the function h is not one-to-one, we may still find the PMF of Y simply by accumulating, for each y, the probability of all the x's that are mapped to that y.

Proposition 1.30. *Let X be a discrete random variable with PMF f_X supported on the set S_X. Let $Y = h(X)$ for some function h. Then Y has PMF f_Y defined by*

$$
f_Y(y) = \sum_{\{x \in S_X \mid h(x) = y\}} f_X(x) \tag{1.7.2}
$$

Example 1.31. Let $X \sim \mathtt{binom}(\mathtt{size} = 4,\ \mathtt{prob} = 1/2)$, and let $Y = (X - 1)^2$. Consider the following table:

x	0	1	2	3	4
$f_X(x)$	1/16	1/4	6/16	1/4	1/16
$y = (x - 2)^2$	1	0	1	4	9

From this we see that Y has support $S_Y = \{0, 1, 4, 9\}$. We also see that $h(x) = (x - 1)^2$ is not one-to-one on the support of X, because both $x = 0$ and $x = 2$ are mapped by h to $y = 1$. Nevertheless, we see that $Y = 0$ only when $X = 1$, which has probability 1/4; therefore, $f_Y(0)$ should equal 1/4. A similar approach works for $y = 4$ and $y = 9$. And $Y = 1$ exactly when $X = 0$ or $X = 2$, which has total probability 7/16. In summary, the PMF of Y may be written:

y	0	1	4	9
$f_Y(y)$	1/4	7/16	1/4	1/16

There is not a special name for the distribution of Y, it is just an example of what to do when the transformation of a random variable is not one-to-one. The method is the same for more complicated problems.

Proposition 1.32. *If X is a random variable with $\mathbb{E}X = \mu$ and $Var(X) = \sigma^2$, then the mean and variance of $Y = mX + b$ is*

$$
\mu_Y = m\mu + b, \quad \sigma_Y^2 = m^2 \sigma^2, \quad \sigma_Y = |m|\sigma. \tag{1.7.3}
$$

1.8 Chapter Exercises

Exercise 1.1. A recent national study showed that approximately 44.7% of college students have used Wikipedia as a source in at least one of their term papers. Let X equal the number of students in a random sample of size $n = 31$ who have used Wikipedia as a source.

1. How is X distributed?

$$X \sim \mathsf{binom}(\mathtt{size} = 31, \mathtt{prob} = 0.447)$$

2. Sketch the probability mass function (roughly).

```
library(distr)
X <- Binom(size = 31, prob = 0.447)
plot(X, to.draw.arg = "d")
```
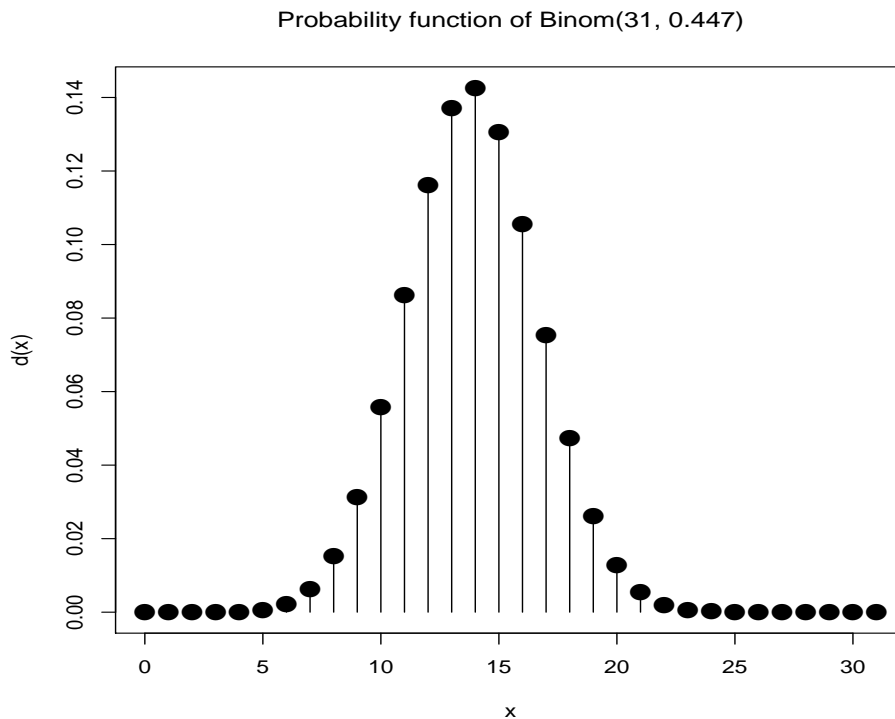
Probability function of Binom(31, 0.447)

Figure 1.8.1: Plot of a binomial PMF.

1. Sketch the cumulative distribution function (roughly).

```
library(distr)
X <- Binom(size = 31, prob = 0.447)
plot(X, to.draw.arg = "p")
```
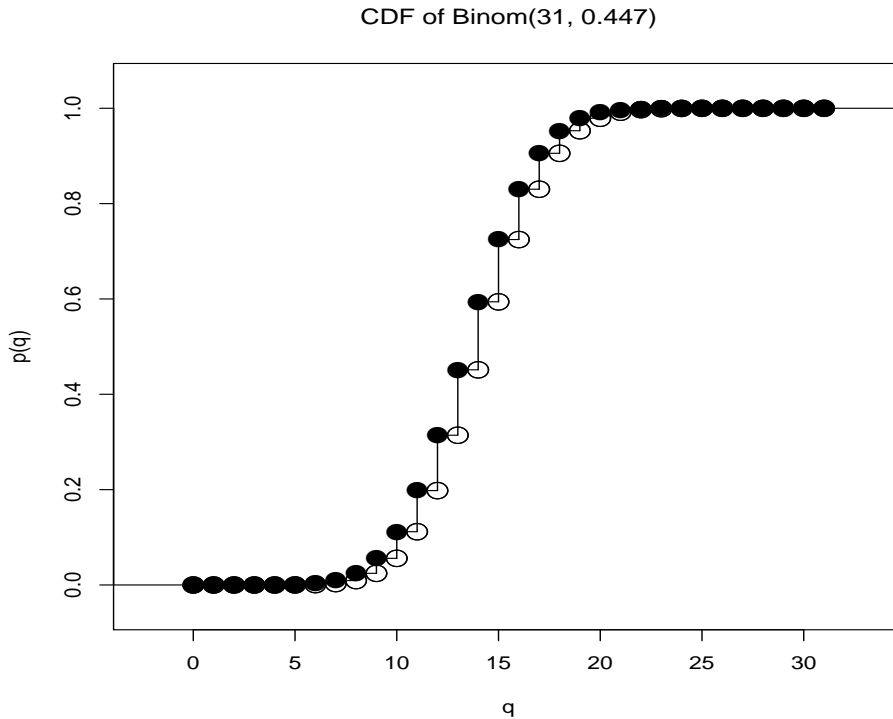
CDF of Binom(31, 0.447)

Figure 1.8.2: Plot of a binomial CDF.

1. Find the probability that *X* is equal to 17.

```
dbinom(17, size = 31, prob = 0.447)
```

```
[1] 0.07532248
```

1. Find the probability that *X* is at most 13.

```
pbinom(13, size = 31, prob = 0.447)
```

```
[1] 0.451357
```

1 Discrete Distributions

1. Find the probability that *X* is bigger than 11.

```
pbinom(11, size = 31, prob = 0.447, lower.tail = FALSE)
```

```
  [1] 0.8020339
```

1. Find the probability that *X* is at least 15.

```
pbinom(14, size = 31, prob = 0.447, lower.tail = FALSE)
```

```
  [1] 0.406024
```

1. Find the probability that *X* is between 16 and 19, inclusive.

```
sum(dbinom(16:19, size = 31, prob = 0.447))
diff(pbinom(c(19,15), size = 31, prob = 0.447, lower.tail = FALSE))
```

```
  [1] 0.2544758
  [1] 0.2544758
```

1. Give the mean of *X*, denoted $\mathbb{E}X$.

```
library(distrEx)
X <- Binom(size = 31, prob = 0.447)
E(X)
```

```
  [1] 13.857
```

1. Give the variance of *X*.

```
var(X)
```

```
  [1] 7.662921
```

1. Give the standard deviation of *X*.

```
sd(X)
```

```
  [1] 2.768198
```

1. Find $\mathbb{E}(4X + 51.324)$

```
E(4*X + 51.324)
```

```
[1] 106.752
```

ORG-LIST-END-MARKER

Exercise 1.2. For the following situations, 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 hold in practice.)

- We shoot basketballs at a basketball hoop, and count the number of shots until we make a goal. Let X denote the number of missed shots. On a normal day we would typically make about 37% of the shots.

- In a local lottery in which a three digit number is selected randomly, let X be the number selected.

- We drop a Styrofoam cup to the floor twenty times, each time recording whether the cup comes to rest perfectly right side up, or not. Let X be the number of times the cup lands perfectly right side up.

- We toss a piece of trash at the garbage can from across the room. If we miss the trash can, we retrieve the trash and try again, continuing to toss until we make the shot. Let X denote the number of missed shots.

- Working for the border patrol, we inspect shipping cargo as when it enters the harbor looking for contraband. A certain ship comes to port with 557 cargo containers. Standard practice is to select 10 containers randomly and inspect each one very carefully, classifying it as either having contraband or not. Let X count the number of containers that illegally contain contraband.

- At the same time every year, some migratory birds land in a bush outside for a short rest. On a certain day, we look outside and let X denote the number of birds in the bush.

- We count the number of rain drops that fall in a circular area on a sidewalk during a ten minute period of a thunder storm.

- We count the number of moth eggs on our window screen.

- We count the number of blades of grass in a one square foot patch of land.

- We count the number of pats on a baby's back until (s)he burps.

Exercise 1.3. Show that $\mathbb{E}(X - \mu)^2 = \mathbb{E}X^2 - \mu^2$. *Hint*: expand the quantity $(X - \mu)^2$ and distribute the expectation over the resulting terms.

Exercise 1.4. If $X \sim \text{binom}(\text{size} = n, \text{prob} = p)$ show that $\mathbb{E}X(X - 1) = n(n - 1)p^2$.

Exercise 1.5. Calculate the mean and variance of the hypergeometric distribution. Show that

$$\mu = K\frac{M}{M+N}, \quad \sigma^2 = K\frac{MN}{(M+N)^2}\frac{M+N-K}{M+N-1}. \tag{1.8.1}$$