5 Discrete Probability Models

Previously we considered finite sample spaces where each outcome was equally likely. In this section we generalize this idea, so that different outcomes have different probabilities, and so that the sample space can potentially be infinite.

5.1 Discrete random variables and mass functions

5.1.1 Random variables

Suppose we are working with a discrete sample space. In this section we introduce the concept of a *random variable*. Formally, a *random variable* is a real valued *function* which assigns a real number to each element of the sample space. Discrete random variables take a countable number of values. Random variables are denoted by upper case letters, e.g. $X(s)$, where s is in the sample space. We usually drop the argument s and just write X. Lower case letters are used for observed values and we write $X = x$ to mean the observed value of X. For example: X=number of heads in 2 tosses; $x = 1$ head observed.

5.1.2 Probability mass function

For any discrete random variable X, we define the *probability mass function* (PMF) to be the function p which gives the probability of each possible value x i.e. $p(x) = \Pr(X = x)$. You can think of this as being like a "table of probabilities": for each possible value of X you just write down the probability with which that value occurs.

The PMF of a random variable is denoted $p(x)$, or sometimes $p_X(x)$. It satisfies:

1. $0 \leq p(x) \leq 1$.
2. $\sum p(x) = 1$ where the sum is over all possible values of x.

Examples:

1. Let X be the number shown on a single die roll. The PMF is:

x	1	2	3	4	5	6
$\Pr(X = x)$	1/6	1/6	1/6	1/6	1/6	1/6

Note that $\sum \Pr(X = x) = 1$. To obtain a plot of the PMF in R, use the following commands:

```
px = c(1/6, 1/6, 1/6, 1/6, 1/6, 1/6)
plot(1:6,px,type="h",xlab="Outcome",ylab="Probability",ylim=c(0,0.2))
# The argument ylim specifies the lower and upper limits of the y-axis
# and the argument type specifies the type of plot. In this case,
# type="h" which joins the (x,y) coordinate of each point to the
# x-axis with a vertical line.
```

which gives Figure 1a.

2. Let Y be the total score when two dice are rolled. The PMF is

y	2	3	4	5	6	7
$\Pr(Y = y)$	1/36	2/36	3/36	4/36	5/36	6/36
y	8	9	10	11	12	
$\Pr(Y = y)$	5/36	4/36	3/36	2/36	1/36	

To plot this in R, use the following commands:

```
py = c(1/36, 2/36, 3/36, 4/36, 5/36, 6/36,
    5/36, 4/36, 3/36, 2/36, 1/36)
plot(2:12,py,type="h",xlab="Outcome",ylab="Probability",ylim=c(0,0.2))
```

This produces the plots in Figure 1b.

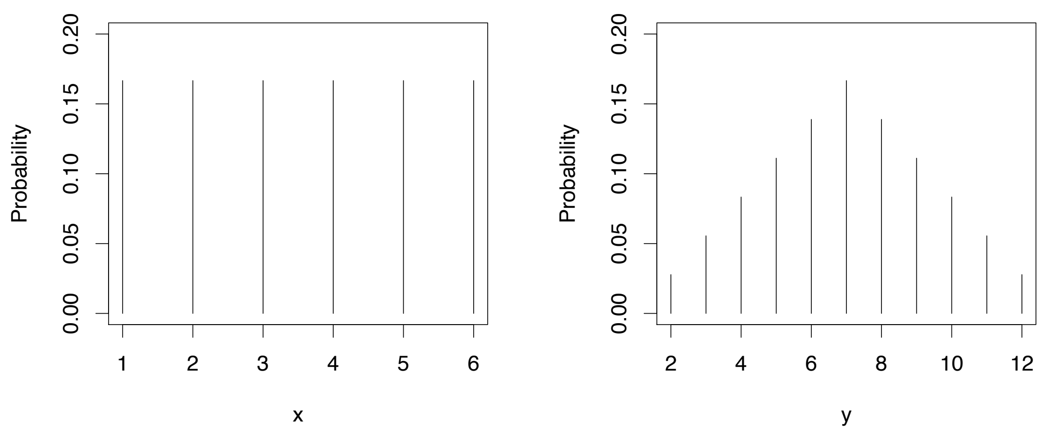

Figure 1: *Probability mass functions for a single die and the total of 2 dice.*

3. Let Z be the number of heads tossed from 3 coins. Write down the PMF of Z:

z (number of heads)	0	1	2	3
PMF				

4. Suppose a function f is defined by:

$$f(j) = \frac{k}{j} \quad \text{for } j = 1, 2, 4.$$

What value of k makes f a PMF? Well,

$$f(1) + f(2) + f(4) = k + \frac{k}{2} + \frac{k}{4} = \frac{7k}{4}$$

so, if we set $k = 4/7$, then $f(j)$ will be a probability mass function.

5.1.3 Cumulative distribution function (CDF)

The cumulative distribution function $F_X(x)$ or just $F(x)$ is defined by

$$F_X(x) = \Pr(X \leq x) = \sum_{z \leq x} \Pr(X = z) \ .$$

In general:

1. $0 \leq F(x) \leq 1$;
2. $F(-\infty) = 0$, $F(\infty) = 1$;
3. If $x_1 \leq x_2$ then $F(x_1) \leq F(x_2)$, i.e. F is monotonic increasing.

Examples:

1. Write down the CDF of X (where X is defined as previously).

x	1	2	3	4	5	6
$F(x)$	1/6	2/6	3/6	4/6	5/6	6/6

To plot this in R, use

```
Fx = c(1/6, 2/6, 3/6, 4/6, 5/6, 6/6)
plot(1:6,Fx,type="s",xlab="x",ylab="Cumulative Probability",ylim=c(0,1))
# The argument type="s" specifies that we want a step-function.
```

This gives Figure 2a.

2. Write down the CDF of Y.

y	2	3	4	5	6	7
$F(y)$	1/36	3/36	6/36	10/36	15/36	21/36
y	8	9	10	11	12	
$F(y)$	26/36	30/36	33/36	35/36	36/36	

To plot this in R, use:

```
Fy = c(1/36, 3/36, 6/36, 10/36, 15/36, 21/36, 26/36,
     30/36, 33/36, 35/36, 36/36)
plot(2:12,Fy,type="s",xlab="y",ylab="Cumulative Probability",ylim=c(0,1))
```

This gives Figure 2b. Note that the CDF is defined *for all real numbers* — not just the possible values of Y. In the example

$$\begin{aligned} F(-3) &= \Pr(Y \leq -3) = 0 \\ F(4.5) &= \Pr(Y \leq 4.5) = \Pr(Y \leq 4) = 6/36 \\ F(25) &= \Pr(Y \leq 25) = 1 \end{aligned}$$

3. Write down the CDF of Z: ✎

# of Heads (z)	0	1	2	3
CDF				

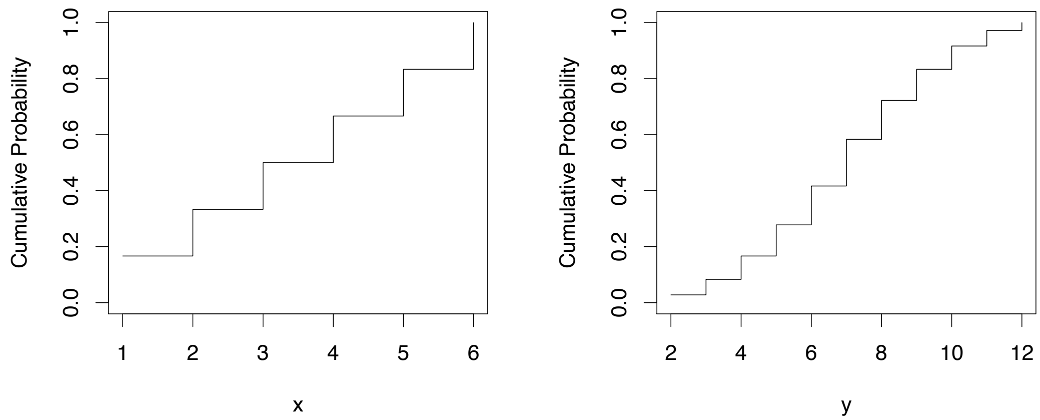

Figure 2: *Cumulative distribution functions for a single die and the total of 2 dice.*

5.2 Expectation

The expectation (or *expected value* or *mean*) of a *discrete* random variable X is defined by

$$\mathrm{E}[X] = \sum_x x\mathrm{Pr}(X = x)\,.$$

It represents an *average value* of the RV. Recall our distinction between the *population* and *sample*. Random variables form our probabilistic models of random quantities, and so the expectation is our model of the population mean. The population mean is often denoted μ_X or μ.

Examples:

1. Calculate $\mathrm{E}[X]$ (X as previously):

$$\mathrm{E}[X] = \left(1 \times \frac{1}{6}\right) + \left(2 \times \frac{1}{6}\right) + \left(3 \times \frac{1}{6}\right) + \left(4 \times \frac{1}{6}\right) + \left(5 \times \frac{1}{6}\right) + \left(6 \times \frac{1}{6}\right) = \frac{21}{6} = 3.5\,.$$

To do this in R, use:

```
x = seq(1,6,1)
px = c(1/6, 1/6, 1/6, 1/6, 1/6, 1/6)
sum(x*px)
```

2. Calculate $\mathrm{E}[Y]$:

$$\mathrm{E}[Y] = \left(2 \times \frac{1}{36}\right) + \left(3 \times \frac{2}{36}\right) + \left(4 \times \frac{3}{36}\right) + \cdots + \left(12 \times \frac{1}{36}\right) = \frac{252}{36} = 7\,.$$

3. Calculate $\mathrm{E}[Z]$. You might find it helpful to refer back to the PMF previously given.

We can also define expectation for any function of X, say $g(X)$:

$$\mathrm{E}[g(X)] = \sum_x g(x)\mathrm{Pr}(X = x)$$

Note that in general $\mathrm{E}[g(X)] \neq g(\mathrm{E}[X])$

Examples:

1. Find $\mathrm{E}[X^2]$.

$$\mathrm{E}\left[X^2\right] = \left(1^2 \times \frac{1}{6}\right)+\left(2^2 \times \frac{1}{6}\right)+\left(3^2 \times \frac{1}{6}\right)+\left(4^2 \times \frac{1}{6}\right)+\left(5^2 \times \frac{1}{6}\right)+\left(6^2 \times \frac{1}{6}\right) \simeq 15.167$$

To do this in R, do:

```
x = seq(1,6,1)
px = c(1/6, 1/6, 1/6, 1/6, 1/6, 1/6)
sum(x^2*px)
```

2. Find $\mathrm{E}[Y^2]$.

$$\mathrm{E}\left[Y^2\right] = \left(2^2 \times \frac{1}{36}\right) + \left(3^2 \times \frac{2}{36}\right) + \left(4^2 \times \frac{3}{36}\right) + \cdots + \left(12^2 \times \frac{1}{36}\right) \simeq 54.8\ .$$

3. Calculate $\mathrm{E}[Z^2]$.

5.3 Variance

Whereas expectation is a measure of location for a random variable, the *variance* is a measure of spread. It is defined by:

$$\text{Var}(X) = \text{E}\left[(X - \text{E}[X])^2\right].$$

We will show later that this is equivalent to:

$$\text{Var}(X) = \text{E}\left[X^2\right] - (E[X])^2.$$

The variance is often denoted σ_X^2, or σ^2. The *standard deviation* of a random variable is

$$\text{SD}(X) = \sqrt{\text{Var}(X)}$$

and this is usually denoted by σ_X or just σ. Like the expectation, the variance and SD defined above are *population* quantities, not to be confused with the sample equivalents.

Examples:

1. Find $\text{Var}(X)$ and $\text{SD}(X)$ for a single roll of a die;

$$\text{Var}(X) = \text{E}\left[X^2\right] - \text{E}[X]^2 = 15.167 - 3.5^2 = 2.917$$

$$\text{SD}(X) = \sqrt{2.917} = 1.708$$

2. Find $\text{Var}(Y)$ and $\text{SD}(Y)$ for the sum of two dice;

$$\text{Var}(Y) = \text{E}\left[Y^2\right] - \text{E}[Y]^2 = 54.8 - 7^2 = 5.8.$$

$$\text{SD}(Y) = \sqrt{5.8} = 2.42$$

3. Find $\text{Var}(Z)$ and $\text{SD}(Z)$ for the number of heads from three coin throws. ✎

Exercise:

Find the mean and variance of random variable X with the following PMF.

x	-2	0	2
$p(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

✎

5.4 Two or more random variables

Often we will deal with two or more random variables. Just as for *events* it is useful to develop the concept of *independence* for random variables. Two random variables X and Y are *independent* if

$$\Pr((X=x)\cap(Y=y)) = \Pr(X=x)\Pr(Y=y) \quad \text{for all possible values of } x \text{ and } y$$

For 3 or more random variables, just as for 3 or more events, there are analogous definitions of *pairwise* and *mutual* independence. In fact, for a set of random variables $X_1,\dots,X_n$ mutual indepence follows if

$$\Pr((X_1=x_1)\cap(X_2=x_2)\cap\cdots\cap(X_n=x_n)) = \Pr(X_1=x_1)\times\Pr(X_2=x_2)\times\cdots\times\Pr(X_n=x_n)$$

for all possible values of $x_1, x_2, \dots x_n$. We usually drop the word "mutual" and simply refer to such random variables as "independent".

Suppose we have two random variables X and Y and $g(X,Y)$ is some function of them, e.g. $g(X,Y)=X+Y$ or $g(X,Y)=XY$. The expectation of $g(X,Y)$ is

$$\mathrm{E}[g(X,Y)] = \sum_{\text{all } x}\sum_{\text{all } y} g(x,y)\Pr((X=x)\cap(Y=y))$$

If X and Y are *independent* this is

$$\mathrm{E}[g(X,Y)] = \sum_{\text{all } x}\sum_{\text{all } y} g(x,y)\Pr(X=x)\Pr(Y=y)\,.$$

5.5 Properties of expectations and variances

The following properties are stated without proof.

- Expectation of a linear transformation: if we have a random variable X, and a linear transformation, $Y=aX+b$, where a and b are known real constants, then we have that
$$\mathrm{E}[aX+b] = a\mathrm{E}[X]+b$$
The equivalent result for variances is
$$\mathrm{Var}\,(aX+b) = a^2\mathrm{Var}\,(X)$$
- For two random variables X and Y, the expectation of their sum is given by
$$\mathrm{E}[X+Y] = \mathrm{E}[X]+\mathrm{E}[Y]$$
whether or not X and Y are independent
- If X and Y are *independent* random variables, then
$$\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y]$$

- If X and Y are *independent* random variables, then

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$$

Note that if X and Y are independent:

$$\text{SD}(X+Y) = \sqrt{\text{SD}(X)^2 + \text{SD}(Y)^2} \neq \text{SD}(X) + \text{SD}(Y) \text{ in general}$$

Exercise:
Use the formulae for properties of expectation and the original definition of $\text{Var}(X)$ to prove that $\text{Var}(X) = \text{E}[X^2] - (\text{E}[X])^2$.

Proof.

□

5.6 Properties of a sample mean

Suppose $X_1, X_2, \ldots, X_n$ are *independent and identically distributed* (IID) with common expectation μ and common variance σ^2. Let

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Then

$$\text{E}\left[\bar{X}\right] = \mu \text{ and } \text{Var}\left(\bar{X}\right) = \frac{\sigma^2}{n}$$

Proof.

$$\begin{aligned}
\text{E}\left[\bar{X}\right] = \text{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] &= \frac{1}{n}\text{E}[X_1 + X_2 + \ldots + X_n] \\
&= \frac{1}{n}\left(\text{E}[X_1] + \text{E}[X_2] + \cdots + \text{E}[X_n]\right) \\
&= \frac{1}{n}\left(\mu + \mu + \cdots + \mu\right) \\
&= \frac{1}{n} \times n\mu \\
&= \mu
\end{aligned}$$

The proof for the variance is similar.

□

Why is this important?

- An observation of $\bar{X}$, often denoted $\bar{x}$, is obtained by calculating the sample mean of n measurements.
- An observed sample mean $\bar{x}$ is a realisation (i.e. observation) of the random variable $\bar{X}$ whose distribution tells us how likely different values are.
- The mean of the random variable $\bar{X}$ is μ and its variance gets smaller as the sample size n increases; in other words, larger samples can be expected to give more precise estimates of the population mean μ.

Example:
Joe Soap takes $X = 3$ minutes to walk to his bus stop and then waits a random time Y which has expected value 4 minutes and standard deviation 1.5 minutes. What is the expected value and standard deviation of the *average* of the total times $T = X + Y$ over a 5 day week?

Solution. ✐

The weak law of large numbers

In addition to the above properties of the expectation and variance of the sample mean, the Weak Law of Large Numbers describes the probability of measurements being far from the sample mean. Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed with common expectation μ and common variance σ^2. Define $\bar{X}$ as previously. Then, for any $\epsilon > 0$,

$$\Pr\left(|\bar{X} - \mu| \geq \epsilon\right) \to 0 \text{ as } n \to \infty$$

In other words, $\bar{X}$ converges to μ in probability.

5.7 The Binomial Distribution

The Binomial distribution is one of the most commonly used distributions in statistics. It specifies the distribution of the number of successes in independent *Bernoulli trials*.

5.7.1 Bernoulli random variables

A Bernoulli trial is an experiment with just two possible outcomes, say success and failure. Assume a success occurs with probability p. A Bernoulli random variable I is an *indicator* of whether success occurs:

$$I(\text{success}) = 1 \quad I(\text{failure}) = 0$$

So $\Pr(I = 1) = p$, $\Pr(I = 0) = 1 - p$ and we write $I \sim \text{Bern}(p)$. Also

$$\mathrm{E}[I] = 0 \times \Pr(I = 0) + 1 \times \Pr(I = 1) = 0 \times (1 - p) + 1 \times p = p$$

and

$$\text{Var}\,(I) = \mathrm{E}\left[I^2\right] - \mathrm{E}[I]^2 = \left\{0^2 \times \Pr(I = 0) + 1^2 \times \Pr(I = 1)\right\} - p^2 = p - p^2 = p(1 - p)\ .$$

5.7.2 Binomial random variables

Suppose we have n independent and identically distributed Bernoulli random variables $I_1, \ldots, I_n$ with probability p of success. If X is defined by

$$X = \sum_{k=1}^{n} I_k$$

then X is a *Binomial* random variable and we write

$$X \sim \text{Bin}(n, p)$$

Examples:

1. A lizard lays 8 eggs, each of which will hatch independently with probability 0.7. Let Y be the number hatching. Then $Y \sim \text{Bin}(8, 0.7)$

2. Let X be the number of heads from 100 throws of a coin. Then $X \sim \text{Bin}(100, 0.5)$

3. 3% of the electronic devices produced by a manufacturer are defective. Let Z be the number of defective items in a shipment of 100 items. ✐

4. X is the number of rainy days in a week. Why is this *not* a Binomial random variable? ✐

Comment: Consider the cell viability example introduced in Chapter 1. In an experiment 100 separate cell cultures were monitored and the number still viable after 14 days was recorded. The experiment was repeated 40 times. Suppose that cell cultures are viable with probability p and denote by X the number of viable cell cultures at the end of an experiment. Then each of the replications of the experiment gives a realisation (i.e. observation) of $X \sim \text{Bin}(100, p)$. We shall see in Chapter **??** how we can use samples like this to *estimate* parameters like p.

5.7.3 Probability mass function

The probability mass function for $X \sim \text{Bin}(n, p)$ is

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, 2, \dots, n$$

The Binomial distribution has $\text{E}[X] = np$ and $\text{Var}\,(X) = np(1-p)$.

Proof. To work out the PMF $\Pr(X = k)$ we need to work out the number of ways of getting k successes and $n-k$ failures. One way is if the first k are successes and the rest are failures.

$$SSS \dots SFFF \dots F$$

another is

$$SS \dots SFFF \dots FS.$$

Because the trials are independent, the probability that the first will occur is

$$p \times p \times p \times \cdots \times p \times (1-p) \times (1-p) \times \cdots \times (1-p) = p^k (1-p)^{n-k}$$

and the probability of the second is

$$p \times p \times \cdots \times p \times (1-p) \times (1-p) \times \cdots \times (1-p) \times p = p^k (1-p)^{n-k}$$

which is the same as the first sequence, so the probability of obtaining any particular sequence of k successes will be the same as the above two sequences.

So, how many ways can n Bernoulli trials give rise to exactly k successes and $n-k$ failures? This is just the number of combinations:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \,,$$

Since each of these sequences are mutually exclusive events we add the probabilities together to get the total probability of k successes in n Bernoulli trials:

$$\binom{n}{k} p^k (1-p)^{n-k} \,.$$

$\square$

The results for the expectation and variance follow from the linearity properties in Section 5.5. ✐

Proof.

□

Examples:

1. For the lizard eggs, $Y \sim \text{Bin}(8, 0.7)$ we have

$$\Pr(Y = k) = \binom{8}{k} 0.7^k 0.3^{8-k}, \qquad k = 0, 1, 2, \ldots, 8.$$

The PMF and CDF are:

k	0	1	2	3	4	5	6	7	8
$\Pr(Y = k)$	0.00	0.00	0.01	0.05	0.14	0.25	0.30	0.20	0.06
$F_Y(k)$	0.00	0.00	0.01	0.06	0.19	0.45	0.74	0.94	1.00

To use R to compute the probabilities associated with $k = 3$, for example, we have:

```
n=8; p=0.7; k=3;
dbinom(k,n,p) # PMF, which gives 0.04667544
pbinom(k,n,p) # CDF, which gives 0.05796765
```

To plot the PMF and CDF in R, use:

```
x = seq(0,8) # Produces a vector with entries 0, 1, ..., 8.
plot(x,dbinom(x,8,0.7),xlab="k",ylab="Prob(Y=k)",type="h")
plot(x,pbinom(x,8,0.7),type="s",xlab="y",ylab="F(y)")
```

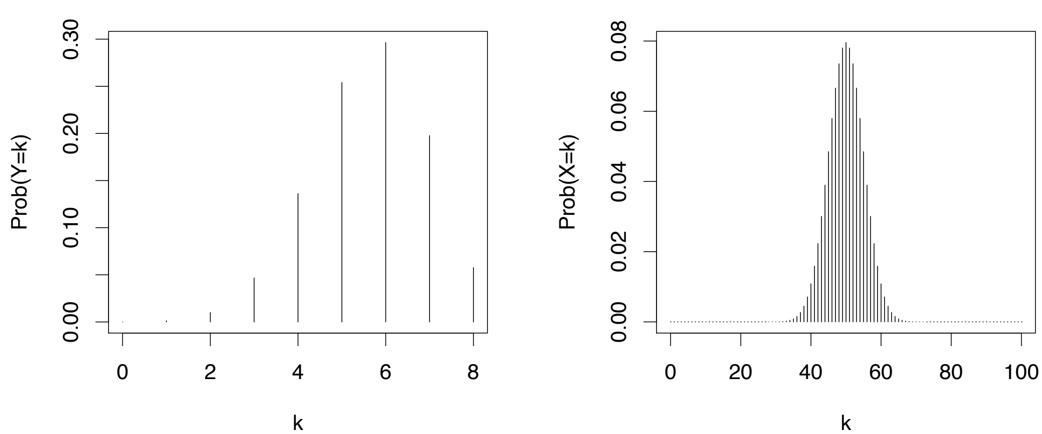

Figure 3: *PMFs for $Y \sim Bin(8, 0.7)$ and $X \sim Bin(100, 0.5)$.*

We can also calculate the mean and variance of Y:

$$\begin{aligned} \mathrm{E}[Y] &= np = 8 \times 0.7 = 5.6 \\ \mathrm{Var}\,(Y) &= np(1-p) = 8 \times 0.7 \times 0.3 = 1.68 \end{aligned}$$

and so

$$\mathrm{SD}(Y) = \sqrt{1.68} \simeq 1.296.$$

2. For the number of heads from 100 throws. Then $X \sim \mathrm{Bin}(100, 0.5)$. See Figure 3. So

$$\begin{aligned}\Pr(X \le 25) &= 2 \times 10^{-7}\\ \Pr(X \le 50) &= 0.54\\ \Pr(X \le 70) &= 0.999984\end{aligned}$$

The above probabilities were obtained using `pbinom(r,100,0.5)` command with `r=25, 50, 70`. The mean and variance of the number of heads obtained from 100 throws are:

$$\begin{aligned}\mathrm{E}[X] &= np = 100 \times 0.5 = 50\\ \mathrm{Var}\,(X) &= np(1-p) = 100 \times 0.5^2 = 25\\ \text{and so}\quad \mathrm{SD}(X) &= 5\end{aligned}$$

3. For the defective devices example we have $Z \sim \mathrm{Bin}(100, 0.03)$. So $\mathrm{E}[Z] = 100 \times 0.03 = 3$. The probability that no items are defective is:

$$\Pr(Z = 0) = \binom{100}{0} 0.03^0 0.97^{100} \simeq 0.048$$

4. A gambler bets £1 on one of the numbers 1 to 6. Three dice are then rolled and if the number appears i times the gambler wins £i, so for example if the gambler picks the number 3 and the dice come up as $(3, 6, 3)$ the reward is £2. Is the game fair? ✎

5.8 The Geometric distribution

The Geometric distribution is used for the number X of independent Bernoulli trials needed until the first success is encountered. We write

$$X \sim \mathrm{Geom}(p)$$

Binomial random variables have a maximum possible value; Geometric random variables are unbounded, and so there is a very small probability of any given arbitrarily large value.

Examples:

- Weeks until first win on lottery;
- Attempts to get through to call centre;
- Number of children until first girl.

But *not* (why not?)

- Attempts until pass driving test;
- Drawing beads without replacement from a jar containing red and blue beads until you select the first red.

5.8.1 PMF and CDF

The PMF of $X \sim \text{Geom}(p)$ is

$$\Pr(X = k) = (1-p)^{k-1}p \qquad k = 1, 2, 3, \ldots.$$

and X has expectation and variance of:

$$\text{E}[X] = \frac{1}{p} \quad \text{and} \quad \text{Var}\,(X) = \frac{1-p}{p^2}\ .$$

Why does the PMF take this form? Suppose the chance of a success was p. Then probability of there being $k-1$ attempts before a success is:

$$(1-p) \times (1-p) \times \ldots \times (1-p)p = (1-p)^{k-1}p.$$

The CDF is

$$F_X(k) = \Pr(X \leq k) = \Pr(\text{at least one success out of } k) = 1 - \Pr(k \text{ failures}) = 1 - (1-p)^k$$

In R, the commands `dgeom` and `pgeom` are used to obtain values of the PMF and CDF respectively. However, R starts the Geometric distribution with $k = 0$ (instead of 1) so you need to subtract 1 from k. For example, to evaluate $\Pr(X = 4)$ type `dgeom(4-1,p)`.

Examples:

1. Play a game where the probability of winning is 0.2 on any turn. If X is the number of turns until the first win, then $X \sim \mathit{Geom}(0.2)$. Find

   (a) $\Pr(X \leq 6)$

(b) $\Pr(3 < X \leq 8)$

(c) $\mathrm{E}[X]$ and $\mathrm{Var}\,(X)$

(d) To plot the PMF of X in R, as shown in Figure 4, we use

```
k = seq(1,30,1); p = 0.2
plot(k,dgeom(k-1, p),type="h",ylab="Prob(X=k)")
```

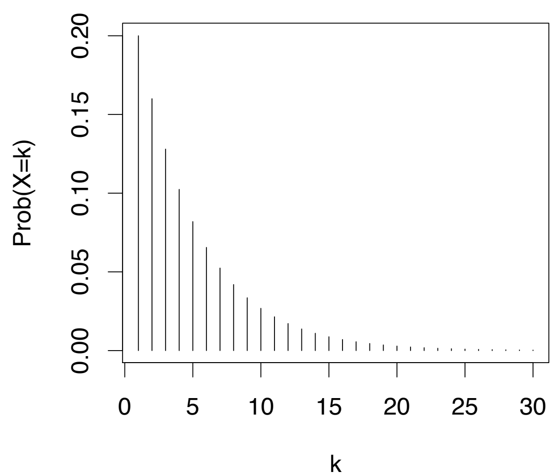

Figure 4: PMF for Geom(0.2).

2. Suppose that on average it takes 2.0 attempts to log on to a computer from a remote terminal at any given time. What is the probability of requiring more than 4 attempts to gain access to the computer? ✎

3. Suppose we align two DNA sequences and a proportion $p = 0.8$ of the letters "match". What is the probability of obtaining a continuous run of 13 or more matching letters at any given starting point in the alignment?

5.9 The Poisson Distribution

The Poisson distribution is a discrete probability distribution with no upper bound which is used to represent the number of events occuring in a fixed period of time or within a fixed region of space. The Poisson distribution has PMF given by

$$\Pr(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}, \quad k = 0, 1, 2, 3, \dots.$$

and we write $X \sim \text{Po}(\lambda)$. The distribution has mean and variance given by:

$$\text{E}[X] = \lambda \quad \text{and} \quad \text{Var}\,(X) = \lambda.$$

The PMF for $\lambda = 5$ is shown in Figure 5. The PMF and CDF can be obtained in R using the `dpois` and `ppois` commands. Figure 5 was created by typing

```
x = seq(0,20,1)
plot(x,dpois(x,lambda=5),type="h",xlab="k",ylab="Prob(X=k)")
```

The Poisson distribution arises in many applications in statistics:

- the number of typing errors per page in a book;
- the number of trees per acre of forest;
- the number of visits per hour to a website;
- the number of cosmic ray showers seen at an observatory per month.

An important property of Poisson random variables is the distribution of their sum. If $X \sim \text{Po}(\lambda)$ independently of $Y \sim \text{Po}(\mu)$ then $Z = X + Y \sim \text{Po}(\lambda + \mu)$.

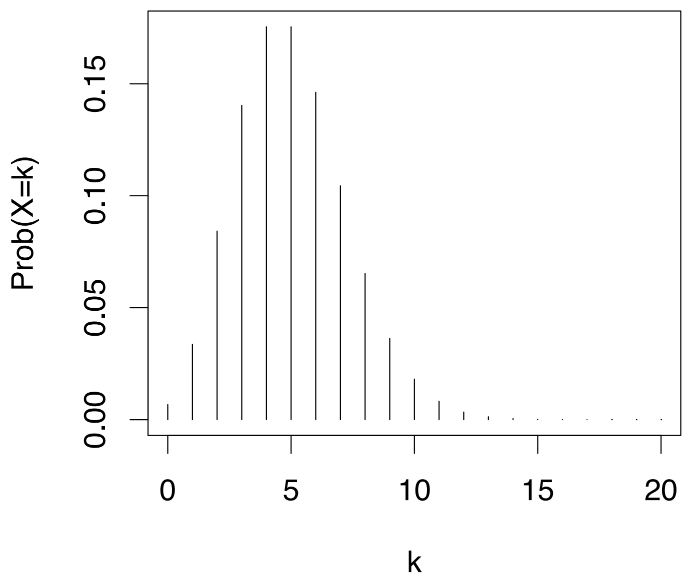

Figure 5: PMF for Po(5).

Example:
A given individual visits social media sites 4 times per day on average. What is the probability of them making 0 visits in a day? What is the distribution of the number of visits in a week?

Solution.

5.9.1 Poisson distribution as the limit of a Binomial

Although the Poisson distribution has no maximum value (it is not a count "out of n"), often Poisson random variables are used in place of Binomial random variables in situations where n is large, p is small, and the expectation np is stable. Mathematically, this is expressed as follows. Suppose $X \sim \text{Bin}(n, p)$, and define λ by $\lambda = E[X] = np$. As n gets very large and p gets very small (keeping λ fixed), it can be shown that

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}.$$

In other words, the Poisson distribution is the large n limit of a Binomial distribution with the same mean. As a rule of thumb, the approximation is accurate if $n > 50$ and $np < 5$.

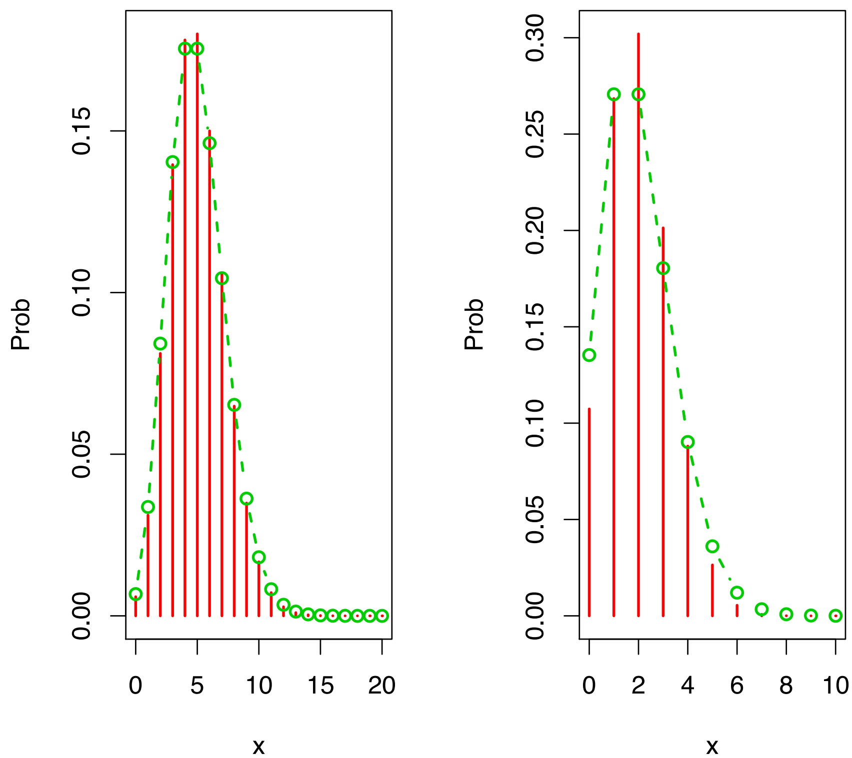

Figure 6: Comparison of Poisson (circles) and Binomial (spikes) PMFs (a). Bin(100,0.05) and Po(5) (b). Bin(10, 0.2) and Po(2)

Figure 6 shows how the Poisson and Binomial distributions are very similar for large n and small p. Notice that when n is large (i.e. $n = 100$) and p is small (i.e. $p = 0.05$) then the Poisson and Binomial are almost identical. To generate Figure 6 we use the following R commands:

```
par(mfrow=c(1,2)) # Creates a multi-panelled plotting
                  # window with 1 row and 2 columns.
x = seq(0,20)

# Left hand plot:
y = dbinom(x,100,0.05) # Binomial PMF.
# Plot Binomial PMF:
plot(x,y,col=2,type="h",lwd=2,ylab="Prob")
# The col=2 argument here means plot in red and the
# lwd=2 argument means make the line width 2 (default
# is 1).
y  =  dpois(x,5)
# Add Poisson PMF:
lines(x,y,col=3,type="b",lwd=2,lty=2)
# The type="b" argument here means plot both a line
# and points and the lty=2 argument means make the
# line type dashed.

# Right hand plot:
x = seq(0,10)
y = dbinom(x,10,0.2)
plot(x,y,col=2,type="h",lwd=2,ylab="Prob")
y  =  dpois(x,2)
lines(x,y,col=3,type="b",lwd=2,lty=2)
par(mfrow=c(1,1))
```

Example:
It was reported in 1993 that 1 in 200 people carry the defective gene that causes inherited colon cancer. In a sample of 500 individuals, what is the (approximate) probability of there being no more than 3 carriers.

Solution. ✐

5.9.2 The Poisson Process

The Poisson distribution is used to model the number of "events" or "successes" in a fixed period of time or in a fixed volume of space e.g. defects per centimetre of wire, visits per hour. In many problems, however, we might be interested in arbitrary intervals of time or parts of space e.g. defects in a stretch of wire of length 0.5 metres. The *Poisson process* models events that occur randomly in time or in a single spatial dimension. A sequence of such events is said to follow a Poisson process with rate λ if the number of events X in any interval of length t is Poisson distributed with:

$$X \sim \mathrm{Po}(\lambda t)$$

and counts in disjoint intervals are independent. Note that this means there are, on average, λ events per unit interval since $\mathrm{E}[X] = \lambda$ if $t = 1$.

Example:
Insurance claims occur as a Poisson process with rate 2 per week on average. Find the probability of at least 3 claims during the next 2 weeks.

Solution. ✐