Chapter 4 Random Variables and Expectation

Probability is expectation founded upon partial knowledge. ~ George Boole

4.1 Random Variables

Often we are not interested in the actual outcome of an experiment but in some function of the outcome. Once the outcome ω is known we can compute the value of the function.

Example 1

Toss a coin 100 times. The events of interest could be the

(a) number of heads;

(b) number of heads − number of tails;

(c) length of longest run of heads.

Example 2

Draw 6 numbers from 1, 2, . . ., 59 without replacement. We could be interested in the

(a) number of matches with given ticket, say with {1, 2, 3, 4, 5, 6};

(b) smallest gap between numbers drawn.

A **random variable** X is a mapping from the sample space Ω into the real line $\mathbb{R}$, i.e. $X:\Omega \to \mathbb{R}$. Hence $X(\omega)\in\mathbb{R}$ for all $\omega\in\Omega$. In fact, several random variables can be defined on the same sample space.

Example 3

The sum of the numbers shown on the top faces of two dice in a roll.

$$\begin{array}{l}
\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),\\
(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),\\
(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),\\
(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\
(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),\\
(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}
\end{array} \quad 7$$

Example 4

Toss a coin 3 times, X is number of heads. List ω and $X(\omega)$.

Sol: $\omega \in \{TTT, TTH, THT, THH, HTT, HTH, HHT, HHH\}$

$$X(\omega)=\begin{cases} 0 & \omega = TTT, \\ 1 & \omega = TTH, THT, HTT, \\ 2 & \omega = THH, HTH, HHT, \\ 3 & \omega = HHH. \end{cases}$$

Usually, we are interested in the probability of $X(\omega) = x$.

The **probability mass function** (pmf) of a random variable X, denoted p_X, is defined by

$$p_X(x) = P(X = x) = P(\{\omega : X(\omega) = x\}).$$

Let $x_1, \dots, x_i, \dots$ be all possible distinct values of $X(\omega)$ and denote $E_i = \{\omega : X(\omega) = x_i\}$. Then $\{E_i\}$ is a partition of Ω, and so

$$\sum_i p_X(x_i) = \sum_i P(E_i) = P\left(\bigcup_i E_i\right) = P(\Omega) = 1,$$

where the summation is over all possible values of x_i.

The **mode**(s) of a discrete random variable is (are) the outcome(s) with the highest probability.

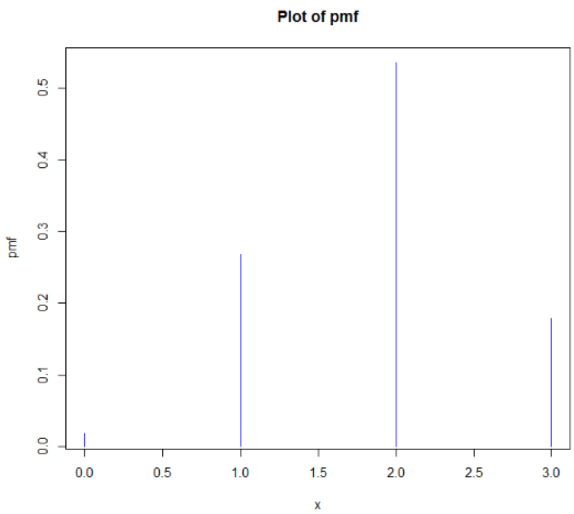
Example 5

A company employs 3 men and 5 women. During a recession, 3 of the workers are made redundant. Let X be the number of women made redundant then, assuming workers picked at random.

(a) Write down the probability mass function of X.

(b) Find the mode of X.

Sol:




```
x<-c(0,1,2,3)
pmf<-c(1/56,15/56,30/56,10/56)
plot(x,pmf,type="h",col="blue",main="Plot of pmf")
```

Example 6

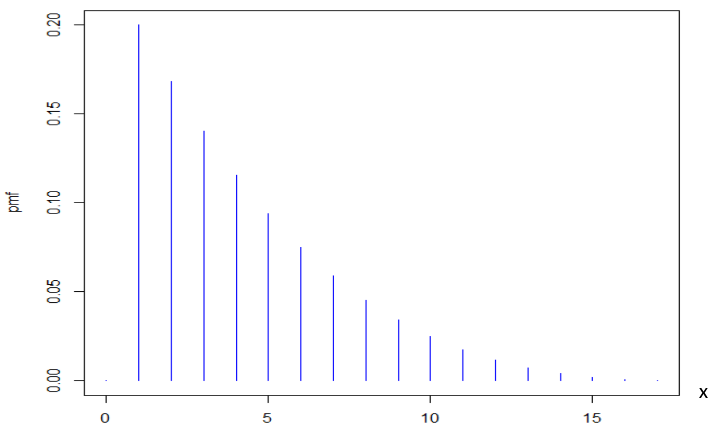
Four balls are to be randomly selected, without replacement, from an urn that contains 20 balls numbered 1 through 20. Let X be the smallest numbered ball selected.

(a) Write down the pmf of X.

(b) What is the mode of X ?

Sol: The smallest number is x when x is selected and there is no balls numbered less than x is selected.

(a)



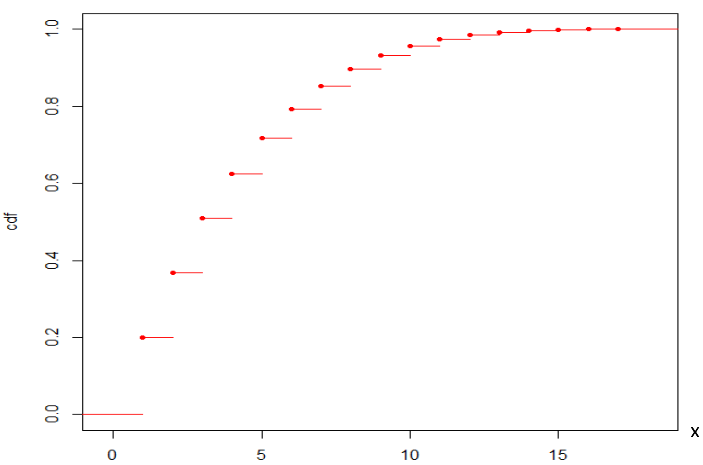
(b) ${}^{20-x}C_3 = \dfrac{(20-x)(19-x)(18-x)}{3!}$ is a decreasing function of x. Hence, it attains the higher value when x is smallest, i.e. 1. The mode is 1.

Can you get the answer for this part without any calculation?

The (**cumulative**) **distribution function** (cdf) defined by

$$F_X(x) = F(x) = P(X \le x) = \sum_{y \le x} p_X(y).$$

Below shows the cdf of the random variable in Example 6.



```
x<-c(1:17)
pmf<-choose(20-x,3)/choose(20,4)
plot(c(0, x), c(0, pmf), type="h", col="blue", main="Plot of pmf", xlab="x",
     ylab="pmf")

cdf<-cumsum(pmf)
cdf<-stepfun(x,c(0,cdf),f=0)
plot(cdf,verticals=FALSE,ylab="cdf",col=2,xlab="x",cex=0,pch=20,
     main="Plot of cdf")
```

The following are some of the important characteristics of a cdf:

1. F_X is non-decreasing, i.e. if $x < y$ then $F_X(x) \le F_X(y)$.
2. $p_X(x_i) = F_X(x_i) - F_X(x_{i-1})$.
3. If $a < b$, $P(a < X \le b) = P(X \le b) - P(X \le a)$
   $= F_X(b) - F_X(a)$.
4. $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to\infty} F_X(x) = 1$.
5. F_X is right-continuous, i.e. $\lim_{x\to a^+} F_X(x) = F(a)$.
6. $F_X(x) = F_Y(x)$ does NOT imply that $X = Y$. For example,

$$P(X = 1) = P(X = -1) = \frac{1}{2} \text{ and } Y = -X \text{. In fact, } P(X = Y) = 0.$$

In fact, if $F_X(x) = F_Y(x)$ for all x, then $P(X \in A) = P(Y \in A)$ for all event A.

Example 7

Independent trials consisting of the flipping of a coin having probability p of coming up heads are continually performed until either a head occurs or a total of n flips is made. Let X denote the number of times the coin is flipped.

(a) Write down the pmf of X.

(b) Find $F_X(x)$.

Sol:

(a)

(b) $$F_X(x) = \begin{cases} 0 & x < 1, \\ \sum_{i=1}^{\lfloor x \rfloor} (1-p)^{i-1} p & x < n, \\ 1 & x \ge n. \end{cases} = \begin{cases} 0 & x < 1, \\ 1-(1-p)^{\lfloor x \rfloor - 1} & x < n, \\ 1 & x \ge n. \end{cases}$$

Note: $\lfloor x \rfloor$ denote the greatest integer less than or equal to x. It is also known as the floor function.

4.2 Expectation

The **expectation** of a (discrete) random variable X is defined by

$$E[X]=\sum_{x} xp_X(x).$$

$E[X]$ is also called the **mean** of X.

Lemma: If X is a discrete random variable that takes one of the values $x_i, i\geq 1$ and g is a function of X, then $E[g(X)]=\sum_{i} g(x_i)p_X(x_i)$.

Proof: Let $y_j, j\geq 1$ represent the different values of $g(x_i), i\geq 1$.

$$\begin{aligned}\sum_{i} g(x_i)p_X(x_i) &= \sum_{j}\sum_{i:g(x_i)=y_j} g(x_i)p_X(x_i)\\ &= \sum_{j} y_j \sum_{i:g(x_i)=y_j} p_X(x_i)\\ &= \sum_{j} y_j P\{g(X)=y_j\}\\ &= E[g(X)]\end{aligned}$$

Example 8

Toss a fair coin three times, X is the number of heads. Find $E[X]$ and $E[X^2]$.

Sol:

Properties of expectation:

1. If $g(X)=b$, a constant, then $E[g(X)]=b$.
2. For all $a,b\in\mathbb{R}$, $E[ag(X)+bh(X)]=aE[g(X)]+bE[h(X)]$.
   In particular, $E[ag(X)+b]=aE[g(X)]+b$.

Proof: 1.

$$\begin{aligned}E[g(X)] &= \sum_{i} g(x_i)p_X(x_i)\\ &= \sum_{i} bp_X(x_i)\\ &= b\sum_{i} p_X(x_i)\\ &= b\end{aligned}$$

2. $E[ag(X)+bh(X)] = \sum_x [ag(x)+bh(x)]p_X(x)$

$$= \sum_x [ag(x)p_X(x)+bh(x)p_X(x)]$$
$$= \sum_x ag(x)p_X(x) + \sum_x bh(x)p_X(x)$$
$$= a\sum_x g(x)p_X(x) + b\sum_x h(x)p_X(x)$$
$$= aE[g(X)]+bE[h(X)]$$

The **variance** of X, $\operatorname{var}(X)$ is defined by

$$\operatorname{var}(X) = E[(X-E[X])^2]$$
$$= \sum_x (x-E[X])^2 p_X(x).$$

The variance is the average of the squared distance from mean. It gives a measure of how spread out the mass function of X is. The **standard deviation** of X is $\sqrt{\operatorname{var}(X)}$.

Properties of variance:

1. $\operatorname{var}(X) = E[X^2]-(E[X])^2$.
2. For all $a,b\in\mathbb{R}$, $\operatorname{var}(aX+b) = a^2\operatorname{var}(X)$.
3. $\operatorname{var}(X)\geq 0$ and $\operatorname{var}(X)=0 \Leftrightarrow P(X=E[X])=1$.

Proof:

1. $\operatorname{var}(X) = E[(X-E[X])^2]$

$$= \sum_x (x-E[X])^2 p_X(x)$$
$$= \sum_x \left[x^2-2xE[X]+(E[X])^2\right]p_X(x)$$
$$= \sum_x x^2 p_X(x) - 2E[X]\sum_x xp_X(x) + (E[X])^2\sum_x p_X(x)$$
$$= E[X^2]-2E[X]\cdot E[X]+E[X]^2$$
$$= E[X^2]-E[X]^2$$

2. $\operatorname{var}(aX+b) = E\left[(aX+b-E[aX+b])^2\right]$

$$= E\left[(aX+b-aE[X]-b)^2\right]$$
$$= E\left[a^2(X-E[X])^2\right]$$
$$= a^2E\left[(X-E[X])^2\right]$$
$$= a^2\operatorname{var}(X)$$

3. Since $(X - E[X])^2 \geq 0$ and $p_X(x) \geq 0$,

$$\mathrm{var}(X) = \sum_x (x - E[X])^2 p_X(x) \geq 0.$$

On the other hand,

$$\begin{aligned}\mathrm{var}(X) = 0 &\Leftrightarrow \sum_x (x - E[X])^2 p_X(x) = 0 \\ &\Leftrightarrow (x - E[X])^2 p_X(x) = 0 \text{ for all } x \\ &\Leftrightarrow x = E[X] \text{ or } p_X(x) = 0 \text{ for all } x\end{aligned}$$

Since $E[X]$ is a constant, there is at most one x, say x_0, such that $x_0 = E[X]$ and for all others $x \neq x_0$, $p_X(x) = 0$. However, $\sum_x p_X(x) = 1 \Rightarrow p_X(x_0) = 1$.

$\therefore \mathrm{var}(X) = 0 \Leftrightarrow P(X = E[X]) = 1$

Example 9

Find the variance and standard deviation of X in Example 8.

Sol:

For $r = 1, 2, 3, \ldots$ the r^{th} **moment** of X is defined as $E[X^r]$.

Example 10

Toss a coin 3 times, X is the number of heads, Y is the number of tails, and $Z = X - Y$ is number of heads – number of tails.

Sol:

ω	X	Y	Z	ω	X	Y	Z
HHH	3	0	3	HTT	1	2	–1
HHT	2	1	1	THT	1	2	–1
HTH	2	1	1	TTH	1	2	–1
THH	2	1	1	TTT	0	3	–3

$$P(X=0) = \frac{1}{8}, P(X=1) = \frac{3}{8}, P(X=2) = \frac{3}{8}, P(X=3) = \frac{1}{8}$$

$$P(Y=0) = \frac{1}{8}, P(Y=1) = \frac{3}{8}, P(Y=2) = \frac{3}{8}, P(Y=3) = \frac{1}{8}$$

$$P(Z=-3) = \frac{1}{8}, P(Z=-1) = \frac{3}{8}, P(Z=1) = \frac{3}{8}, P(Z=3) = \frac{1}{8}$$

Further properties of expectation:

1. $E[X_1 + X_2 + \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n]$
2. If X and Y are independent, then
$$\operatorname{var}(X+Y) = \operatorname{var}(X) + \operatorname{var}(Y)$$
3. If $X_1, X_2, \cdots, X_n$ are mutually independent, then
$$\operatorname{var}(X_1 + X_2 + \cdots + X_n) = \operatorname{var}(X_1) + \operatorname{var}(X_2) + \cdots + \operatorname{var}(X_n)$$

Remarks:

- Independence is not required in 1.
- Independence of random variables will be defined later.

Example 11

Verify that $E[Z] = E[X] - E[Y]$ but $\operatorname{var}(Z) \neq \operatorname{var}(X) + \operatorname{var}(Y)$ in Example 10.

Sol: We observe that, although X and Y are different random variables, they have the same distribution.

Hence, $E[X] = E[Y] = \frac{3}{2}$ and $\operatorname{var}(X) = \operatorname{var}(Y) = \frac{3}{4}$ from the previous examples.

$$E[Z] = \left(-3 \times \frac{1}{8}\right) + \left(-1 \times \frac{3}{8}\right) + \left(1 \times \frac{3}{8}\right) + \left(3 \times \frac{1}{8}\right) = 0$$

$$E[Z^2] = \left(9 \times \frac{1}{8}\right) + \left(1 \times \frac{3}{8}\right) + \left(1 \times \frac{3}{8}\right) + \left(9 \times \frac{1}{8}\right) = 3$$

$$\operatorname{var}(Z) = E[Z^2] - E[Z]^2 = 3$$

$$\operatorname{var}(X) + \operatorname{var}(Y) = \frac{3}{4} + \frac{3}{4} = \frac{3}{2} \neq \operatorname{var}(Z)$$

Two random variables X and Y are said to be **independent and identically distributed** (i.i.d.) if for all $x, y \in \mathbb{R}$, $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$ (independence), and for all $x \in \mathbb{R}$, $P(X = x) = P(Y = x)$ (identically distributed).

♦ ♦ ♦ ♦

Upon completion of this chapter, check that you are able to

- state the definitions and calculate the expectation, median and mode of a random variable;
- calculate probabilities and moments using probability mass functions or the cumulative distribution functions;
- state, prove and apply basic properties of expectation and variance.