Chapter 6. Discrete Random Variables

6.1 Random Variables

We are not always interested in an experiment itself, but rather in some consequence of its random outcome. Such consequences, when real valued, may be thought of as functions which map S to $\mathbb{R}$, and these functions are called random variables.

Definition 6.1.1. *A* **random variable** *is a (measurable) mapping*

$$X : S \to \mathbb{R}$$

with the property that $\{s \in S : X(s) \leq x\} \in \mathcal{F}$ *for each* $x \in \mathbb{R}$.

If we denote the unknown outcome of the random experiment as s^*, then the corresponding unknown outcome of the random variable $X(s^*)$ will be generically referred to as X.

The probability measure P already defined on S induces a **probability distribution** on the random variable X in $\mathbb{R}$: For each $x \in \mathbb{R}$, let $S_x \subseteq S$ be the set containing just those elements of S which are mapped by X to numbers no greater than x. Then we see

$$\mathrm{P}_X(\mathrm{X} \leq x) \equiv \mathrm{P}(S_x).$$

Definition 6.1.2. *The image of S under X is called the* **range** *of the random variable:*

$$\mathbb{X} \equiv X(S) = \{x \in \mathbb{R} | \exists s \in S \ s.t. \ X(s) = x\}$$

So as S contains all the possible outcomes of the experiment, $\mathbb{X}$ contains all the possible outcomes for the random variable X.

Example Let our random experiment be tossing a fair coin, with sample space $S = \{\mathrm{H}, \mathrm{T}\}$ and probability measure $\mathrm{P}(\{\mathrm{H}\}) = \mathrm{P}(\{\mathrm{T}\}) = \dfrac{1}{2}$.

We can define a random variable $X : \{\mathrm{H}, \mathrm{T}\} \to \mathbb{R}$ taking values, say,

$$X(\mathrm{T}) = 0 \quad \text{and} \quad X(\mathrm{H}) = 1$$

In this case, we have

$$S_x = \begin{cases} \emptyset & \text{if } x < 0; \\ \{\mathrm{T}\} & \text{if } 0 \leq x < 1; \\ \{\mathrm{H}, \mathrm{T}\} & \text{if } x \geq 1. \end{cases}$$

This defines a range of probabilities P_X on the continuum $\mathbb{R}$

$$\mathrm{P}_X(\mathrm{X} \leq x) = \mathrm{P}(S_x) = \begin{cases} \mathrm{P}(\emptyset) = 0 & \text{if } x < 0; \\ \mathrm{P}(\{\mathrm{T}\}) = \frac{1}{2} & \text{if } 0 \leq x < 1; \\ \mathrm{P}(\{\mathrm{H}, \mathrm{T}\}) = 1 & \text{if } x \geq 1. \end{cases}$$

■

Random variables are important because they provide a compact way of referring to events via their numerical attributes.

Example Consider counting the number of heads in a sequence of 3 coin tosses. The underlying sample space is

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

which contains the 8 possible sequences of tosses. However, since we are only interested in the number of heads in each sequence, we define the random variable X by

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

This mapping is illustrated in Figure 6.1 below

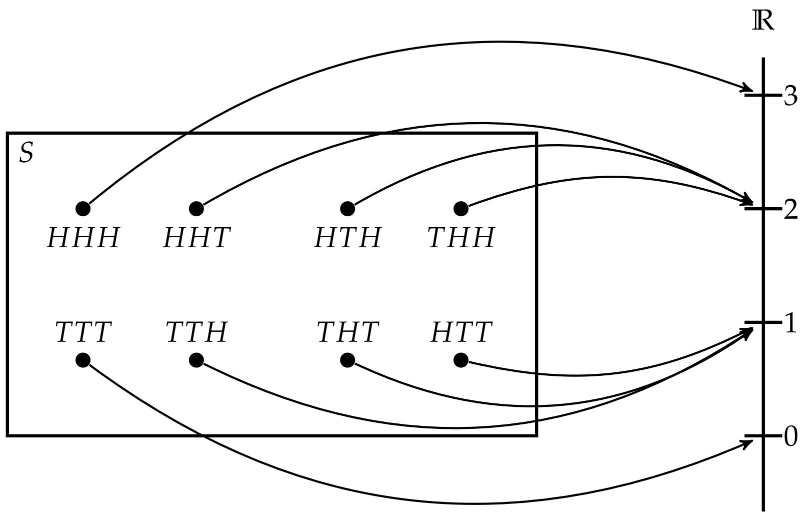

Figure 6.1: Illustration of a random variable X that counts the number of heads in a sequence of 3 coin tosses.

■

Continuing this example, let us assume that the sequences are equally likely. Now lets find the probability that the number of heads X is less than 2. In other words, we want to find $P_X(X < 2)$... but what does this precisely mean? $P_X(X < 2)$ is the shorthand for

$$P_X(\{s \in S : X(s) < 2\}).$$

The first step in calculating the probability is therefore to identify the event $\{s \in S : X(s) < 2\}$. In Figure 6.1, the only lines pointing to the numbers less than 2 are 0 and 1. Tracing these lines backwards from $\mathbb{R}$ into S, we see that

$$\{s \in S : X(s) < 2\} = \{TTT, TTH, THT, HTT\}.$$

Since we have assumed that the sequences are equally likely

$$\mathrm{P}_X(\{TTT, TTH, THT, HTT\}) = \frac{|\{TTT, TTH, THT, HTT\}|}{|S|} = \frac{4}{8} = \frac{1}{2}$$

On the same sample space, we can define another random variable able to describe the event that the number of heads in 3 tosses is even. Define this random variable, Y, as

$$Y(s) = \begin{cases} 0, & s \in \{TTT, THH, HTH, HHT\} \\ 1, & s \in \{TTH, THT, HTT, HHH\} \end{cases}.$$

The probability that the number of heads is less than two and odd is $\mathrm{P}(X < 2, Y = 1)$, by which we mean the probability of the event

$$\{s \in S : X(s) < 2 \text{ and } Y(s) = 1\}.$$

This event is equal to

$$\{s \in S : X(s) < 2\} \cap \{s \in S : Y(s) = 1\}$$

which is

$$\{TTT, TTH, THT, HTT\} \cap \{TTH, THT, HTT, HHH\} = \{TTH, THT, HTT\}.$$

The probability of this event, assuming all sequences are equally likely, is $3/8$.

The shorthand introduced above is standard in probability theory. In general, if $B \subset \mathbb{R}$,

$$\{X \in B\} := \{s \in S : X(s) \in B\}$$

and

$$\mathrm{P}_X(X \in B) := \mathrm{P}_X(\{X \in B\}) = \mathrm{P}_X(\{s \in S : X(s) \in B\}).$$

If B is an interval such as $B = (a, b]$,

$$\{X \in (a, b]\} := \{a < X \le b\} := \{s \in S : a < X(s) \le b\}$$

and

$$\mathrm{P}_X(a < X \le b) = \mathrm{P}_X(\{s \in S : a < X(s) \le b\}).$$

Analogous notation applies to intervals such as $[a, b]$, $[a, b)$, (a, b), $(-\infty, b)$, $(-\infty, b]$, (a, ∞) and $[a, \infty)$.

6.1.1 Cumulative Distribution Function

Given a random variable X, we define the cumulative distribution function (CDF or just distribution function) as follows:

Definition 6.1.3. *The* **cumulative distribution function** *(***CDF***) of a random variable X is the function* $F_X : \mathbb{R} \to [0, 1]$*, defined by*

$$F_X(x) = \mathrm{P}_X(X \leq x)$$

For any random variable X, F_X is right-continuous, meaning if a decreasing sequence of real numbers $x_1, x_2, \ldots \to x$, then $F_X(x_1), F_X(x_2), \ldots \to F_X(x)$.

For a given function $F_X(x)$, to check this is a valid CDF, we need to make sure the following conditions hold.

i) $0 \leq F_X(x) \leq 1, \forall x \in \mathbb{R}$;

ii) Monotonicity: $\forall x_1, x_2 \in \mathbb{R}, x_1 < x_2 \Rightarrow F_X(x_1) \leq F_X(x_2)$;

iii) $F_X(-\infty) = 0$, $F_X(\infty) = 1$.

Some results regarding CDFs.

- For finite intervals $(a, b] \subseteq \mathbb{R}$, it is easy to check that

$$\mathrm{P}_X(a < X \leq b) = F_X(b) - F_X(a).$$

- Unless there is any ambiguity, we generally suppress the subscript of $\mathrm{P}_X(\cdot)$ in our notation and just write $\mathrm{P}(\cdot)$ for the probability measure for the random variable.
  - That is, we forget about the underlying sample space and use random variable directly and its probabilities.
  - Often, it will be most convenient to work this way and consider the random variable directly from the very start, with the range of X being our sample space.

6.2 Discrete Random Variables

Definition 6.2.1. *A random variable X is* **discrete** *if the range of X, denoted by $\mathbb{X}$, is countable, that is*

$$\mathbb{X} = \{x_1, x_2, \ldots, x_n\} \quad \textit{(FINITE)} \quad \textit{or} \quad \mathbb{X} = \{x_1, x_2, \ldots\} \quad \textit{(INFINITE)}.$$

The even numbers, the odd numbers and the rational numbers are countable; the set of real numbers between 0 and 1 is not countable.

Definition 6.2.2. *For a discrete random variable X, we define the* **probability mass function** *(or* **probability function***) as*

$$p_X(x) = \mathrm{P}(X = x), \quad x \in \mathbb{X}.$$

Note For completeness, we define

$$p_X(x) = 0, \quad x \notin \mathbb{X}.$$

for that p_X is defined for all $x \in \mathbb{R}$. Furthermore, we will refer to $\mathbb{X}$ as the support of random variable X, that is, the set of $x \in \mathbb{R}$ such that $p_X > 0$.

6.2.1 Properties of Mass Function p_X

A function p_X is a probability mass function for a discrete random variable X with range $\mathbb{X}$ of the form $\{x_1, x_2, \ldots\}$ if and only if

i) $p_X(x_i) \geq 0$;

ii) $\sum_{x \in \mathbb{X}} p_X(x) = 1$.

6.2.2 Discrete Cumulative Distribution Function

The cumulative distribution function, or CDF, F_X of a discrete random variable X is defined by

$$F_X(x) = \mathrm{P}(X \leq x), \quad x \in \mathbb{R}.$$

6.2.3 Connection between F_X and p_X

Let X be a discrete random with range $\mathbb{X} = \{x_1, x_2, \ldots\}$, where $x_1 < x_2 < \ldots$, and probability mass function p_x and CDF F_X. Then, for any real value x, if $x < x_1$, then $F_X(x) = 0$ and for $x \geq x_1$.

$$F_X(x) = \sum_{x_i \leq x} p_X(x_i) \iff p_X(x_i) = F_X(x_i) - F_X(x_{i-1}), \quad i = 2, 3, \ldots.$$

with, for completeness, $p_X(x_1) = F_X(x_1)$.

6.2.4 Properties of Discrete CDF F_X

i) In the limiting cases,

$$\lim_{x \to -\infty} F_X(x) = 0, \quad \lim_{x \to \infty} F_X(x) = 1.$$

ii) F_X is continuous from the right on $\mathbb{R}$, that is, for $x \in \mathbb{R}$,

$$\lim_{h \to 0^+} F_X(x+h) = F_X(x)$$

iii) F_X is non-decreasing, that is,

$$a < b \implies F_X(a) \leq F_X(b).$$

iv) For $a < b$

$$\mathrm{P}(a < X \leq b) = F_X(b) - F_X(a).$$

Note The key idea is that the functions p_X and/or F_X can be used to describe the probability distribution of the random variable X. A graph of the function p_X is non-zero only at the elements of $\mathbb{X}$. A graph of the function F_X is a step-function which takes the value zero at minus infinity, the value one at infinity, and is non-decreasing with points of discontinuity at the elements of $\mathbb{X}$.

Example Consider a coin tossing experiment where a fair coin is tossed repeatedly under identical experimental conditions, with the sequence of tosses independent, until a Head is obtained. For this experiment, the sample space, S, consists of the set of sequences $(\{H\}, \{TH\}, \{TTH\}, \dots)$ with associated probabilities 1/2, 1/4, 1/8, ...

Define the discrete random variable $X : S \to \mathbb{R}$, by $X(s) = x \iff$ first Head on toss x. Then

$$p_X(x) = \mathrm{P}(X = x) = \left(\frac{1}{2}\right)^x, \quad x = 1, 2, 3, \dots$$

and zero otherwise. For $x \geq 1$, let $k(x)$ be the largest integer not greater than x, then

$$F_X(x) = \sum_{x_i < x} p_X(x_i) = \sum_{i=1}^{k(x)} p_X(i) = 1 - \left(\frac{1}{2}\right)^{k(x)}$$

and $F_X(x) = 0$ for $x < 1$.

Figure 6.2 displays the probability mass function (left) and cumulative distribution function (right). Note that the mass function is only non-zero at points that are elements of $\mathbb{X}$ and the CDF is defined for all real values of x, but is only continuous from the right. F_X is therefore a step-function.

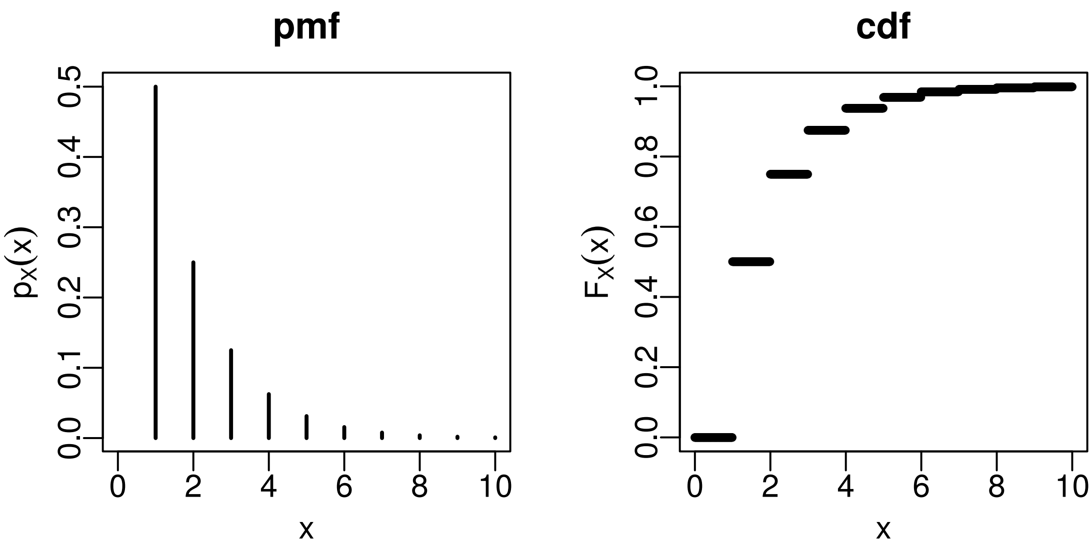

Figure 6.2: pmf $p_X(x) = \left(\frac{1}{2}\right)^x$, $x = 1, 2, \ldots$, and CDF $F_X(x) = 1 - \left(\frac{1}{2}\right)^k(x)$.

■

We are now starting to see our the connections between the numerical summaries and graphical displays we saw in earlier lectures and probability theory.

We can often think of a set of data $(x_1, x_2, \ldots, x_n)$ as n realisations of a random variable X defined on an underlying population for the data.

- Recall the frequency counts we considered for a set of data and their histogram plot. This can be seen as an *empirical estimate* for the pmf of their underlying population.
- Also recall the empirical cumulative distribution function. This too is an empirical estimate, but for the CDF of the underlying population.

6.3 Mean and Variance

6.3.1 Expectation

The mean, or expectation, of a discrete random variable is the average value of X.

Definition 6.3.1. *The* **expected value**, *or* **mean** *of a discrete random variable X is defined to be*

$$\mathrm{E}_X(X) = \sum_x x p_X(x)$$

The expectation is a one-number summary of the distribution and is often just written $\mathrm{E}(X)$ or even μ_X.

$\mathrm{E}(X)$ gives a weighted average of the possible values of the random variable X, with the weights given by the probability of that particular outcome.

1. If X is a random variable taking the integer value scored with a single roll of a fair die, then

$$\begin{aligned}\mathrm{E}(X) &= \sum_{x=1}^{6} x p_X(x) \\ &= 1.\frac{1}{6} + 2.\frac{1}{6} + 3.\frac{1}{6} + 4.\frac{1}{6} + 5.\frac{1}{6} + 6.\frac{1}{6} = \frac{21}{6} = 3.5.\end{aligned}$$

2. If X is a score from a student answering a single multiple choice question with four options, with 3 marks awarded for a correct answer, -1 for a wrong answer and 0 for no answer, what is the expected value if they answer at random?

$$\mathrm{E}(X) = 3.\mathrm{P}_X(\text{Correct}) + (-1).\mathrm{P}_X(\text{Incorrect}) = 3.\frac{1}{4} - 1.\frac{3}{4} = 0.$$

Extension: Let $g : \mathbb{R} \to \mathbb{R}$ be a real-valued (measurable) function of interest of the random variable X; then we have the following result:

Theorem 6.4.

$$\mathrm{E}(g(X)) = \sum_x g(x) p_X(x)$$

Properties of Expectations

Let X be a random variable with pmf p_X. Let g and h be real-valued functions, $g, h : \mathbb{R} \to \mathbb{R}$, and let a and b be constants. Then

$$\mathrm{E}(ag(X) + bh(X)) = a\mathrm{E}(g(X)) + b\mathrm{E}(h(X))$$

Special Cases:

(i) For a linear function, $g(X) = aX + b$ for constants, we have (from Theorem 6.4) that

$$\begin{aligned} \mathrm{E}(g(X)) &= \sum_x (ax+b) p_X(x) \\ &= a \sum_x x p_X(x) + b \sum_x p_X(x) \end{aligned}$$

and since $\sum_x x p_X(x) = \mathrm{E}(X)$ and $\sum_x p_X(x) = 1$ we have

$$\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$$

(ii) Consider $g(x) = (x - \mathrm{E}(X))^2$. The expectation of this function wrt P_X gives a measure of spread or variability of the random variable X around its mean, called the **variance**.

> **Definition 6.4.1.** *Let X be a random variable. The* **variance** *of X, denoted by σ^2 or σ_X^2 or $Var_X(X)$ is defined by*
>
> $$Var_X(X) = E_X[\{X - E_X(X)\}^2].$$

We can expand the expression $\{X - \mathrm{E}(X)\}^2$ and exploit the linearity of expectation to get an alternative formula for the variance.

$$\begin{aligned} \{X - \mathrm{E}(X)\}^2 &= X^2 - 2\mathrm{E}(X)X + \{\mathrm{E}(X)\}^2 \\ \implies \mathrm{Var}(X) &= \mathrm{E}[X^2 - \{2\mathrm{E}(X)\}X + \{\mathrm{E}(X)\}^2] \\ &= \mathrm{E}(X^2) - 2\mathrm{E}(X)\mathrm{E}(X) + \{\mathrm{E}(X)\}^2 \end{aligned}$$

and hence

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \{\mathrm{E}(X)\}^2.$$

It is easy to show that the corresponding result is

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X), \qquad \forall a, b \in \mathbb{R}$$

Related to the variance is the standard deviation, which is defined as follows:

> **Definition 6.4.2.** *The* **standard deviation** *of a random variable X, written $sd_X(X)$ (or sometimes σ_X), is the square root of the variance.*
>
> $$sd_X(X) = \sqrt{Var_X(X)}.$$

Lastly, we can define the skewness of a discrete random variable as follows:

Definition 6.4.3. *The* **skewness** *(γ_1) of a discrete random variable X is given by*

$$\gamma_1 = \frac{E_X[\{X - E_X(X)\}^3]}{sd_X(X)^3}.$$

Example If X is a random variable taking the integer value scored with a single roll of a fair die, then

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}(X^2) - (E(X))^2 \\ &= \sum_{x=1}^{6} x^2 p_X(x) - 3.5^2 \\ &= 1^2.\frac{1}{6} + 2^2.\frac{1}{6} + \ldots + 6^2.\frac{1}{6} - 3.5^2 = 1.25.\end{aligned}$$

■

Example If X is a score from a student answering a single multiple choice question with four options, with 3 marks awarded for a correct answer, -1 for a wrong answer and 0 for no answer, what is the standard deviation if they answer at random?

$$\begin{aligned}\mathrm{E}(X^2) &= 3^2.\mathrm{P}_X(\text{Correct}) + (-1)^2.\mathrm{P}_X(\text{Incorrect}) = 9.\frac{1}{4} + 1.\frac{3}{4} = 3 \\ \Rightarrow \mathrm{sd}(X) &= \sqrt{3 - 0^2} = \sqrt{3}.\end{aligned}$$

■

Note We have met three important quantities for a random variable, defined through expectation – the mean μ, the variance σ^2 and the standard deviation σ.

Again we can see a duality with the corresponding numerical summaries for data which we met – the sample mean $\bar{x}$, the sample variance s^2 and the sample standard deviation s.

The duality is this: If we were to consider the data sample as the *population* and draw a random member from that sample as a *random variable*, this random variable would have CDF $F_n(x)$, the empirical CDF. The mean of the random variable $\mu = \bar{x}$, variance $\sigma^2 = s^2$ and standard deviation $\sigma = s$.

6.4.1 Sums of Random Variables

Let $X_1, X_2, \ldots, X_n$ be n random variables, perhaps with different distributions and not necessarily independent.

Let $S_n = \sum_{i=1}^{n} X_i$ be the sum of those variables, and $\frac{S_n}{n}$ be their average.

Then the mean of S_n is given by

$$\mathrm{E}(S_n) = \sum_{i=1}^{n} \mathrm{E}(X_i), \quad \mathrm{E}\left(\frac{S_n}{n}\right) = \frac{\sum_{i=1}^{n} \mathrm{E}(X_i)}{n}.$$

However, for the variance of S_n, only if $X_1, X_2, \ldots, X_n$ are **independent**, we have

$$\mathrm{Var}(S_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i), \qquad \mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{\sum_{i=1}^{n} \mathrm{Var}(X_i)}{n^2}.$$

So if $X_1, X_2, \ldots, X_n$ are independent and identically distributed with $\mathrm{E}(X_i) = \mu_X$ and $\mathrm{Var}(X_i) = \sigma_X^2$ we get

$$\mathrm{E}\left(\frac{S_n}{n}\right) = \mu_X, \qquad \mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{\sigma_X^2}{n}.$$

6.5 Some Important Discrete Random Variables

6.5.1 Bernoulli Distribution

Consider an experiment with only two possible outcomes, encoded as a random variable X taking value 1, with probability p, or 0, with probability $(1-p)$, accordingly.

Example Tossing a coin, $X = 1$ for a head, $X = 0$ for tails, $p = \frac{1}{2}$. ■

Then we say $X \sim \mathrm{Bernoulli}(p)$ and note the pmf to be

$$p_X(x) = p^x(1-p)^{1-x}, \quad x \in \mathbb{X} = \{0, 1\}, \quad 0 \le p \le 1$$

Note Using the formulae for mean and variance, it follows that

$$\mu \equiv \mathrm{E}(X) = p, \qquad \sigma^2 \equiv \mathrm{Var}(X) = p(1-p).$$

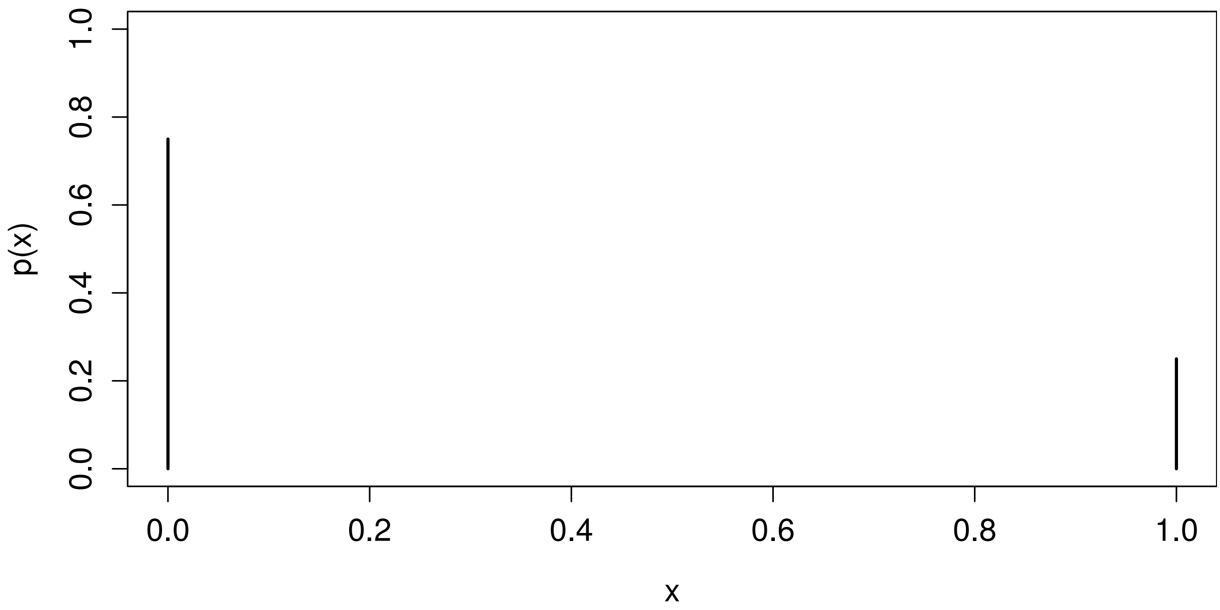

Figure 6.3: Example: pmf of Bernoulli(1/4).

6.5.2 Binomial Distribution

Now consider n identical, independent Bernoulli(p) trials $X_1, \ldots, X_n$. Let $X = \sum_{i=1}^{n} X_i$ be the total number of 1s observed in the n trials.

Example Tossing a coin n times, X is the number of heads obtained, $p = \frac{1}{2}$. ■

Then X is a random variable taking values in $\{0, 1, 2, \ldots, n\}$, and we say $X \sim \text{Binomial}(n, p)$.

From the Binomial Theorem we find the pmf to be

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x \in \mathbb{X} = \{0, 1, 2, \ldots, n\}, \quad n \geq 1, \quad 0 \leq p \leq 1.$$

Notes

- To calculate the Binomial pmf we recall that $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ and $x! = \prod_{i=1}^{x} i$. (Note $0! = 1$.)
- It can be shown, either directly from the pmf or from the results for sums of random variables, that the mean and variance are

$$\mu \equiv \mathrm{E}(X) = np, \qquad \sigma^2 \equiv \mathrm{Var}(X) = np(1-p).$$

- The skewness is given by

$$\gamma_1 = \frac{1-2p}{\sqrt{np(1-p)}}.$$

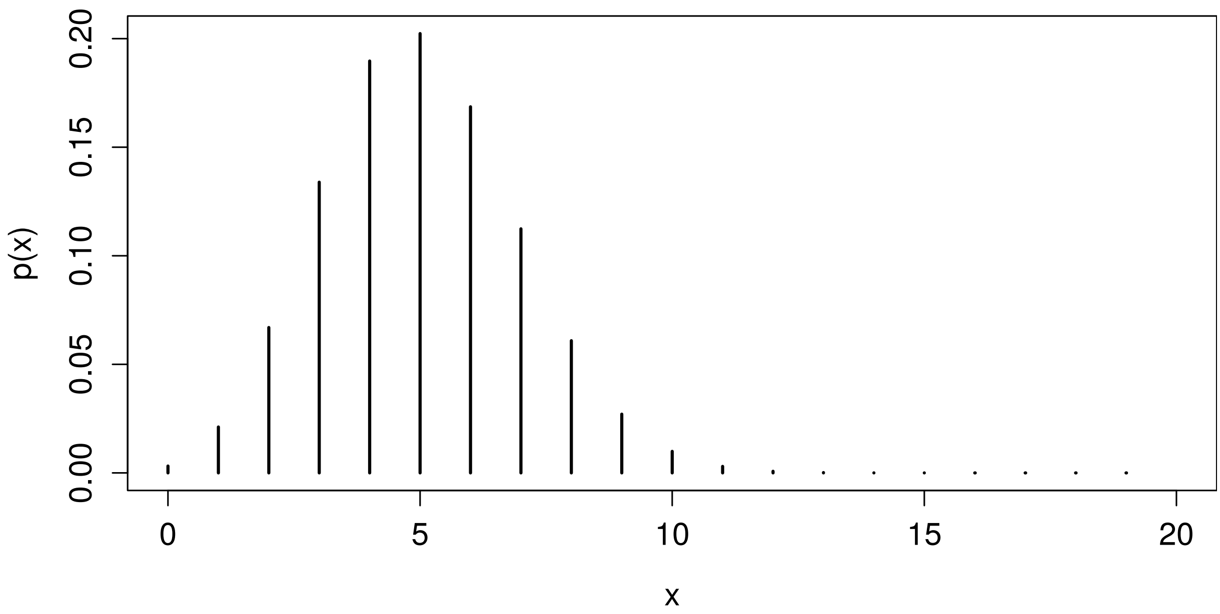

Figure 6.4: Example: pmf of Binomial$(20, 1/4)$.

Example Suppose that 10 users are authorised to use a particular computer system, and that the system collapses if 7 or more users attempt to log on simultaneously. Suppose that each user has the same probability $p = 0.2$ of wishing to log on in each hour.
Question: What is the probability that the system will crash in a given hour?

Solution

The probability that exactly x users will want to log on in any hour is given by Binomial(n, p) = Binomial$(10, 0.2)$.

Hence the probability of 7 or more users wishing to log on in any hour is

$$\begin{aligned} & p_X(7) + p_X(8) + p_X(9) + p_X(10) \\ &= \binom{10}{7} 0.2^7 0.8^3 + \ldots + \binom{10}{10} 0.2^{10} 0.8^0 \\ &= 0.00086. \end{aligned}$$

- A manufacturing plant produces chips with a defect rate of 10%. The quality control procedure consists of checking samples of size 50. Then the distribution of the number of defectives is expected to be Binomial$(50, 0.1)$.

- When transmitting binary digits through a communication channel, the number of digits received correctly out of n transmitted digits, can be modelled by a Binomial(n, p), where p is the probability that a digit is transmitted incorrectly.

Note The independence condition necessary for these models to be reasonable. ■

6.5.3 Geometric Distribution

Consider a potentially infinite sequence of independent Bernoulli(p) random variables $X_1, X_2, \ldots$. Suppose we define a quantity X by

$$X = \min\{i | X_i = 1\}$$

to be the index of the first Bernoulli trial to result in a 1.

Example Tossing a coin, X is the number of tosses until the first head is obtained, $p = \frac{1}{2}$. ■

Then X is a random variable taking values in $\mathbb{Z}^+ = \{1, 2, \ldots\}$, and we say $X \sim \text{Geometric}(p)$.

Clearly the pmf is given by

$$p_X(x) = p(1-p)^{x-1}, \qquad x \in \mathbb{X} = \{1, 2, \ldots\}, \quad 0 \leq p \leq 1.$$

Notes

- The mean and variance are

$$\mu \equiv \mathrm{E}(X) = \frac{1}{p}, \qquad \sigma^2 \equiv \mathrm{Var}(X) = \frac{1-p}{p^2}.$$

- The skewness is given by

$$\gamma_1 = \frac{2-p}{\sqrt{1-p}},$$

  and so is always positive.

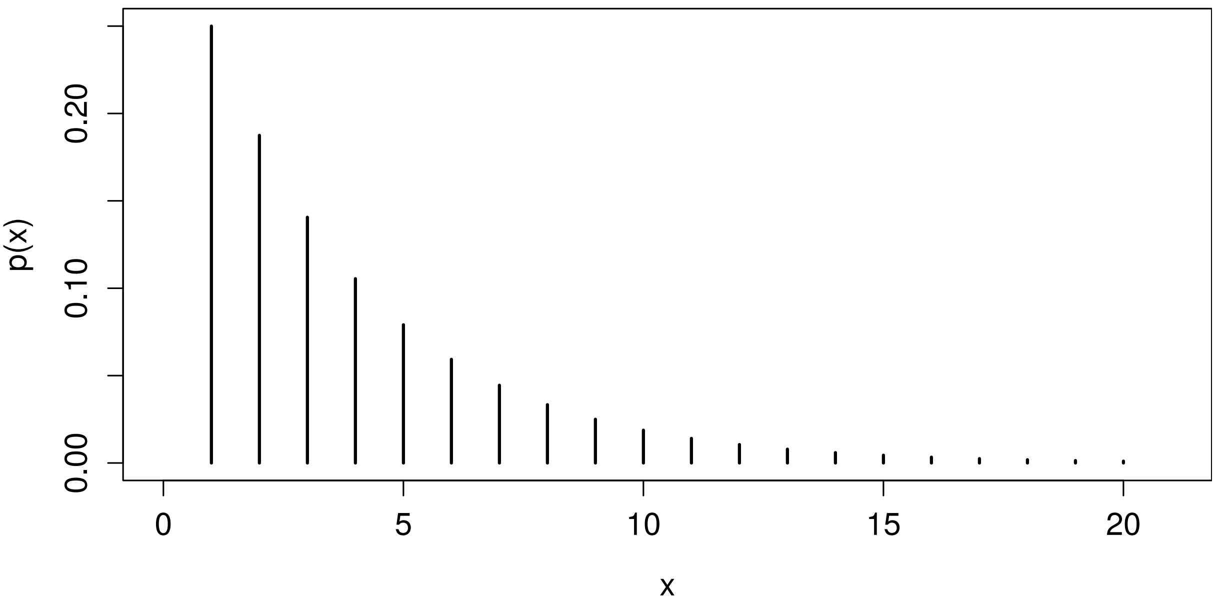

Figure 6.5: Example: pmf of Geometric(1/4).

Alternative Formulation

If $X \sim \text{Geometric}(p)$, let us consider $Y = X - 1$.

Then Y is a random variable taking values in $\mathbb{N} = \{0, 1, 2, \ldots\}$, and corresponds to the number of independent Bernoulli(p) trials *before* we obtain our first 1. (Some texts refer to *this* as the Geometric distribution.)

Note we have pmf

$$p_Y(y) = p(1-p)^y, \qquad y = 0, 1, 2, \ldots,$$

and the mean becomes

$$\mu_Y \equiv \mathrm{E}_Y(Y) = \frac{1-p}{p}.$$

while the variance and skewness are unaffected by the shift.

Example Suppose people have problems logging onto a particular website once every 5 attempts, on average.

1. Assuming the attempts are independent, what is the probability that an individual will not succeed until the 4^{th}?

$$p = \frac{4}{5} = 0.8. \quad p_X(4) = (1-p)^3 p = 0.2^3 0.8 = 0.0064.$$

2. On average, how many trials must one make until succeeding?

$$\text{Mean} = \frac{1}{p} = \frac{5}{4} = 1.25.$$

3. What's the probability that the first successful attempt is the 7^{th} or later?

$$p_X(7) + p_X(8) + p_X(9) + \ldots = \frac{p(1-p)^6}{1-(1-p)} = (1-p)^6 = 0.2^6.$$

Again suppose that 10 users are authorised to use a particular computer system, and that the system collapses if 7 or more users attempt to log on simultaneously. Suppose that each user has the same probability $p = 0.2$ of wishing to log on in each hour.

Using the Binomial distribution we found the probability that the system will crash in any given hour to be 0.00086.

Using the Geometric distribution formulae, we are able to answer questions such as: On average, after how many hours will the system crash?

Mean $= \frac{1}{p} = \frac{1}{0.00086} = 1163$ hours. ■

Example A dictator, keen to maximise the ratio of males to females in his country (so he could build up his all male army) ordered that each couple should keep having children until a boy was born and then stop.

Calculate the number expected number of boys that a couple will have, and the expected number of girls, given that P(boy)=½.

Assume for simplicity that each couple can have arbitrarily many children (although this is not necessary to get the following results). Then since each couple stops when 1 boy is born, the expected number of boys per couple is 1.

On the other hand, if Y is the number of girls given birth to by a couple, Y clearly follows the alternative formulation for the Geometric(½) distribution.

So the expected number of girls for a couple is $\frac{1-\frac{1}{2}}{\frac{1}{2}} = 1$. ■

6.5.4 Poisson Distribution

Let X be a random variable on $\mathbb{N} = \{0, 1, 2, \ldots\}$ with pmf

$$p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x \in \mathbb{X} = \{0, 1, 2, \ldots\}, \quad \lambda > 0.$$

Then X is said to follow a Poisson distribution with *rate* parameter λ and we write $X \sim$ Poisson(λ).

Notes

- Poisson random variables are concerned with the number of random events occurring per unit of time or space, when there is a constant underlying probability 'rate' of events occurring across this unit.
  - the number of minor car crashes per day in the U.K.;
  - the number of potholes in each mile of road;
  - the number of jobs which arrive at a database server per hour;
  - the number of particles emitted by a radioactive substance in a given time.
- An interesting property of the Poisson distribution is that it has equal mean and variance, namely
$$\mu \equiv \mathrm{E}(X) = \lambda, \qquad \sigma^2 \equiv \mathrm{Var}(X) = \lambda.$$
- The skewness is given by
$$\gamma_1 = \frac{1}{\sqrt{\lambda}},$$
so is always positive but decreasing as λ increases.

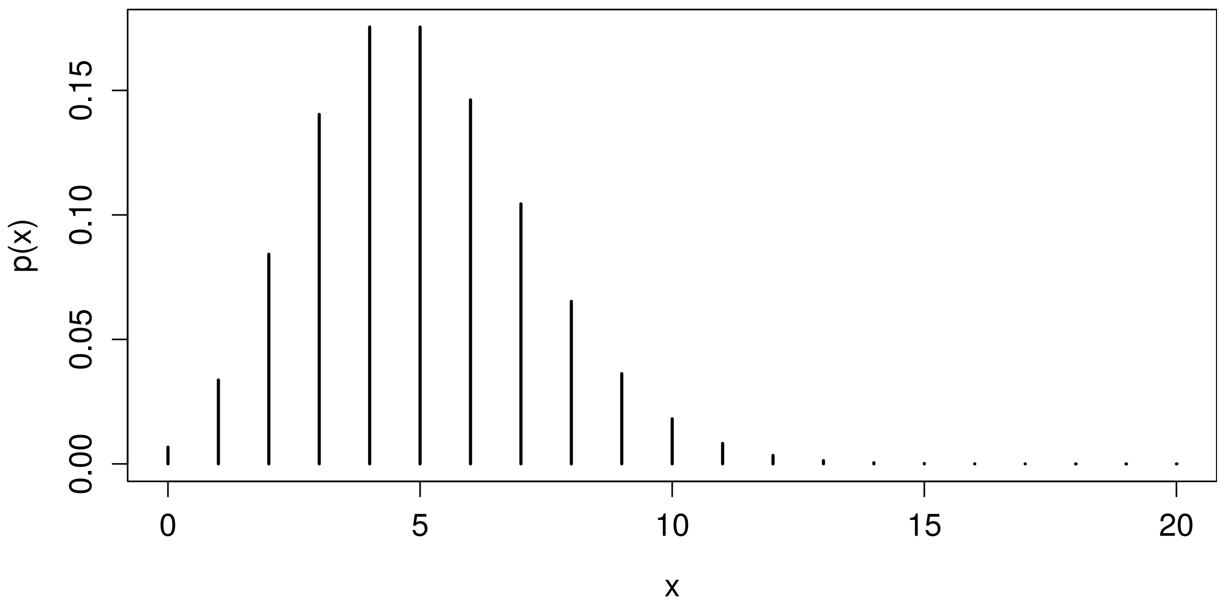

Figure 6.6: Example: pmf of Poisson(5).

Notice the similarity between the pmf plots for Binomial$(20, 1/4)$ and Poisson(5) (Figures 6.4 and 6.6).

It can be shown that for Binomial(n, p), when p is small and n is large, this distribution can be well approximated by the Poisson distribution with rate parameter np, Poison(np).

The value of p in the above is not small, we would typically prefer $p < 0.1$ for the approximation to be useful.

The usefulness of this approximation is in using probability tables; tabulating a single Poisson(λ) distribution encompasses an infinite number of possible corresponding Binomial distributions, Binomial$(n, \frac{\lambda}{n})$.

Example A manufacturer produces VLSI chips, of which 1% are defective. Find the probability that in a box of 100 chips none are defective.

We want $p_X(0)$ from Binomial(100,0.01). Since n is large and p is small, we can approximate this distribution by Poisson$(100 \times 0.01) \equiv$ Poisson(1).

Then $p_X(0) \approx \dfrac{e^{-1}\lambda^0}{0!} = 0.3679$. ■

Example The number of particles emitted by a radioactive substance which reached a Geiger counter was measured for 2608 time intervals, each of length 7.5 seconds.

The (real) data are given in the table below:

x	0	1	2	3	4	5	6	7	8	9	≥10
n_x	57	203	383	525	532	408	273	139	45	27	16

Do these data correspond to 2608 independent observations of an identical Poisson random variable?

The total number of particles, $\sum_x x n_x$, is 10,094, and the total number of intervals observed, $n = \sum_x n_x$, is 2608, so that the average number reaching the counter in an interval is $\frac{10094}{2608} = 3.870$.

Since the mean of Poisson(λ) is λ, we can try setting $\lambda = 3.87$ and see how well this fits the data.

For example, considering the case $x = 0$, for a single experiment interval the probability of observing 0 particles would be $p_X(0) = \frac{e^{-3.87} 3.87^0}{0!} = 0.02086$. So over $n = 2608$ repetitions, our (Binomial) expectation of the number of 0 counts would be $n \times p_X(0) = 54.4$.

Similarly for $x = 1, 2, \ldots$, we obtain the following table of expected values from the Poisson(3.87) model:

x	0	1	2	3	4	5	6	7	8	9	$\geq$10
$O(n_x)$	57	203	383	525	532	408	273	139	45	27	16
$E(n_x)$	54.4	210.5	407.4	525.5	508.4	393.5	253.8	140.3	67.9	29.2	17.1

(O=Observed, E=Expected).

The two sets of numbers appear sufficiently close to suggest the Poisson approximation is a good one. Later, when we come to look at *hypothesis testing*, we will see how to make such judgements quantitatively. ■

6.5.5 Discrete Uniform Distribution

Let X be a random variable on $\{1, 2, \ldots, n\}$ with pmf

$$p_X(x) = \frac{1}{n}, \qquad x \in \mathbb{X} = \{1, 2, \ldots, n\}.$$

Then X is said to follow a discrete uniform distribution and we write $X \sim \mathrm{U}(\{1, 2, \ldots, n\})$.

Note The mean and variance are

$$\mu \equiv \mathrm{E}(X) = \frac{n+1}{2}, \qquad \sigma^2 \equiv \mathrm{Var}(X) = \frac{n^2-1}{12}.$$

and the skewness is clearly zero.