Chapter 2

Discrete Random Variables

2.1 Introduction

A *random variable*, denoted by a capital letter such as X, is a function mapping from each outcome in a sample space to the real line:

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

Random variable is *discrete* if its range is either finite or countably infinite.

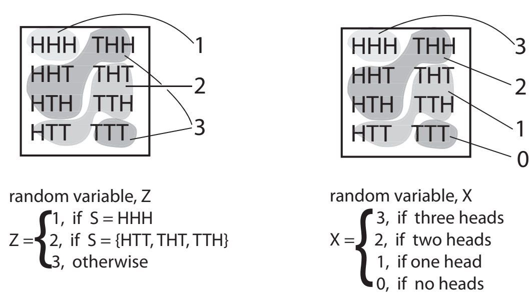

Figure 2.1: Examples of random variables

Exercise 2.1 (Introduction)

1. Number of heads, X in three coin tosses: (i) **discrete** (ii) **continuous**
   domain: (i) $\boldsymbol{S \in \{\text{TTT, TTH, \ldots, HHH}\}}$ (ii) $\boldsymbol{X \in \{0, 1, 2, 3\}}$
   range: (i) $\boldsymbol{S \in \{\text{TTT, TTH, \ldots, HHH}\}}$ (ii) $\boldsymbol{X \in \{0, 1, 2, 3\}}$
   Random variable X (i) **is** (ii) **not** a probability

2. Number of dots, X, in roll of a die: (i) **discrete** (ii) **continuous**
   domain: (i) $\boldsymbol{S \in}$ **dots on die faces** (ii) $\boldsymbol{X \in \{1, 2, 3, 4, 5, 6\}}$

range: (i) $\boldsymbol{S} \in$ **dots on die faces** (ii) $\boldsymbol{X \in \{1, 2, 3, 4, 5, 6\}}$
Value of random variable X denoted (i) $\boldsymbol{X = x}$ (ii) $\boldsymbol{X = X}$

3. Number of seizures, Z, in a year: (i) **discrete** (ii) **continuous**
   domain: (i) $\boldsymbol{S} \in$ **patients in study** (ii) $\boldsymbol{Z \in \{0, 1, 2, \ldots\}}$
   range: (i) $\boldsymbol{S} \in$ **patients in study** (ii) $\boldsymbol{Z \in \{0, 1, 2, \ldots\}}$

   Patients, W, sick (1 or more seizures) or not (0 seizures):
   (i) **discrete** (ii) **continuous**
   domain: (i) $\boldsymbol{S} \in$ **patients in study** (ii) $\boldsymbol{W \in \{0, 1\}}$
   range: (i) $\boldsymbol{S} \in$ **patients in study** (ii) $\boldsymbol{W \in \{0, 1\}}$

4. Waiting time, T, at Burger King up to 1 hour: (i) **discrete** (ii) **continuous**
   domain: (i) $\boldsymbol{S} \in$ **waiting customers** (ii) $\boldsymbol{\{T : 0 \leq T \leq 1\}}$
   range: (i) $\boldsymbol{S} \in$ **waiting customers** (ii) $\boldsymbol{\{T : 0 \leq T \leq 1\}}$

2.2 Probability Mass Functions

If R is the range of discrete random variable X, the *probability mass function (pmf)* of X is function $f : R \to \mathbb{R}$ which satisfies:

- $f(x) > 0$, for all $x \in R$
- $\sum_{x \in R}$ f(x) $= 1$
- If $X \subset R$, then $P(X \in A) = \sum_{x \in A} f(x)$

The pmf can be described by table, graph or function, these descriptions are called the *distribution* of the random variable.

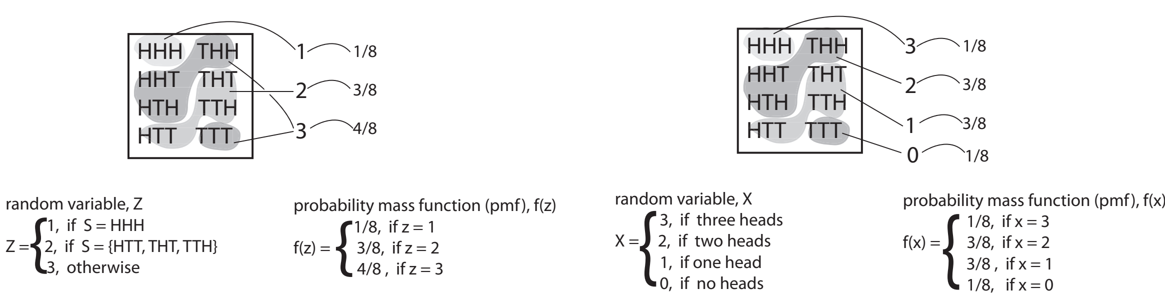

Figure 2.2: Examples of random variable and probability mass function (pmf)

The *cumulative distribution function (cdf)* (*distribution function*) of discrete random variable X is

$$F(b) = P(X \leq b) = \sum_{x \leq b} f(x), \quad \text{for all real numbers } b.$$

If range R of discrete random variable X has k elements, X has a *uniform distribution* with pmf

$$f(x) = \frac{1}{k}, \quad \text{for each } x \in R.$$

The *mode* of discrete random variable X is the value of X where the pmf is a maximum. The *median* is the smallest number m such that

$$P(X \leq m) \geq 0.5 \quad \text{and} \quad P(X \geq m) \geq 0.5.$$

A random variable with two modes is *bimodel*, with two or more modes, *multimodal*.

Exercise 2.2 (Probability Mass Functions)

1. *Probability mass function for seizures.* The number of seizures, X, of a typical epileptic person in any given year is given by the following pmf.

x	0	2	4	6	8	10
$f(x)$	0.17	0.21	0.18	0.11	0.16	0.17

   (a) The chance a person has 8 epileptic seizures is $f(8) =$
   (i) **0.11** (ii) **0.16** (iii) **0.17** (iv) **0.21**.

   (b) The chance a person has *at most* 4 seizures is
   (i) **0.17** (ii) **0.21** (iii) **0.56** (iv) **0.67**.

   (c) $P(X \leq 4) =$ (i) **0.17** (ii) **0.21** (iii) **0.56** (iv) **0.67**.

   (d) $f(2) =$ (i) **0.17** (ii) **0.21** (iii) **0.56** (iv) **0.67**.

   (e) $f(2.1) =$ (i) **0** (ii) **0.21** (iii) **0.56** (iv) **0.67**.

   (f) $P(X > 2.1) =$ (i) **0.21** (ii) **0.38** (iii) **0.56** (iv) **0.62**.

   (g) $\sum_{x=0}^{10} f(x) = P(X = 0) + P(X = 2) + \cdots + P(X = 10) =$
   (i) **0.97** (ii) **0.98** (iii) **0.99** (iv) **1**.

   (h) *Graph of distribution.* Which of the three graphs *best* describes the probability distribution of the number of seizures?
   (i) **(a)** (ii) **(b)** (iii) **(c)**.

   (i) *Probability mass function.* Which one of the following functions describes the pmf of the number of seizures?

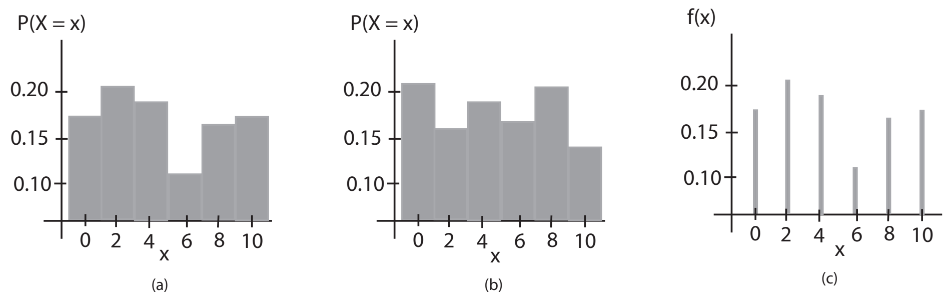

Figure 2.3: Probability mass function: seizures

i. Function (a).

$$P(X = x) = \begin{cases} 0.17, & \text{if } x = 0 \\ 0.21, & \text{if } x = 2 \end{cases}$$

ii. Function (b).

$$P(X = x) = \begin{cases} 0.18, & \text{if } x = 4 \\ 0.11, & \text{if } x = 6 \end{cases}$$

iii. Function (c).

$$P(X = x) = \begin{cases} 0.17, & \text{if } x = 0 \\ 0.21, & \text{if } x = 2 \\ 0.18, & \text{if } x = 4 \\ 0.11, & \text{if } x = 6 \\ 0.16, & \text{if } x = 8 \\ 0.17, & \text{if } x = 10 \end{cases}$$

2. *Rolling a pair of dice: number of fours rolled.*
   Let X be the *number* of 4's rolled. Assume the dice are fair.

   (a) $P(X = 2) = P\{(4,4)\} =$ (i) $\mathbf{\frac{1}{36}}$ (ii) $\mathbf{\frac{20}{36}}$ (iii) $\mathbf{\frac{25}{36}}$ (iv) $\mathbf{\frac{30}{36}}$.

   (b) Since $P(X = 1) =$
   (i) $\boldsymbol{P\{(1,4),(2,4),(3,4),(5,4),(6,4)\}}$
   (ii) $\boldsymbol{P\{(4,1),(4,2),(4,3),(4,5),(4,6)\}}$
   (iii) $\boldsymbol{P\{(1,4),(2,4),(3,4),(5,4),(6,4),(4,1),(4,2),(4,3),(4,5),(4,6)\}}$

   (c) Then $P(X = 1) =$ (i) $\mathbf{\frac{1}{36}}$ (ii) $\mathbf{\frac{10}{36}}$ (iii) $\mathbf{\frac{25}{36}}$ (iv) $\mathbf{\frac{30}{36}}$.

   (d) $P(X = 0) = 1 - P(X = 1) - P(X = 2) =$
   (i) $\mathbf{\frac{11}{36}}$ (ii) $\mathbf{\frac{20}{36}}$ (iii) $\mathbf{\frac{25}{36}}$ (iv) $\mathbf{\frac{30}{36}}$.

   (e) The probability distribution of X is

x	0	1	2
$P(X = x)$	$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$

   (i) **True** (ii) **False**

3. *Flipping until a head comes up.* A (weighted) coin has a probability of $p = 0.7$ of coming up heads (and so a probability of $q = 1 - p = 0.3$ of coming up tails). Let X be the number of flips until a head comes up or until a total of 4 flips are made.

   (a) Head may come up on first flip, $f(1) = P\{H\} =$
   (i) **0.3** (ii) **0.7** (iii) **0.3(0.7)** (iv) $\mathbf{0.3^2(0.7)}$.

   (b) Head may come up on second flip, $f(2) = P\{TH\} =$
   (i) **0.3** (ii) **0.7** (iii) **0.3(0.7)** (iv) $\mathbf{0.3^2(0.7)}$.

   (c) $f(3) = P\{TTH\} =$
   (i) **0.3** (ii) **0.7** (iii) **0.3(0.7)** (iv) $\mathbf{0.3^2(0.7)}$.

   (d) $f(4) = P\{TTTT, TTTH\} = 1 - p(1) - p(2) - p(3) =$
   (i) **0.027** (ii) **0.063** (iii) **0.210** (iv) **0.700**.

   (e) The probability distribution of X is

x	1	2	3	4
$f(x)$	0.700	0.210	0.063	0.027

   (i) **True** (ii) **False**

   (f) *Conditional probability.*

   $$P(X \geq 3 | X \geq 2) = \frac{P(X \geq 3 \cap X \geq 2)}{P(X \geq 2)} = \frac{P(X \geq 3)}{P(X \geq 2)} = \frac{0.063 + 0.027}{1 - 0.7} =$$

   (i) **0.2** (ii) **0.3** (iii) **0.4** (iv) **0.5**.

4. *Geometric probability mass function: bull's eye.* There is a 15% ($p = 0.15$) chance of hitting a bull's eye on a dart board. Throws are independent of one another.

   (a) The chance the first bull's eye occurs on the *first* try is, of course, 15%. The chance the first bull's eye occurs on the *second* try equals the chance a miss occurs on the first try and a bull's eye occurs on the second try,
   $f(2) = (0.85)0.15 =$ (i) **0.1155** (ii) **0.1275** (iii) **0.1385** (iv) **0.2515**.

   (b) The chance the first bull's eye occurs on the *third* try is equal to the chance of two misses and then a bull's eye occurs on the third try,
   $f(3) = (0.85)(0.85)0.15 = (0.85)^2 0.15 \approx$
   (i) **0.078** (ii) **0.099** (iii) **0.108** (iv) **0.158**.

   (c) The chance the first bull's eye occurs on the *fourth* try is
   $f(4) = (0.85)^3 0.15 \approx$
   (i) **0.078** (ii) **0.092** (iii) **0.108** (iv) **0.151**.

(d) The chance the first bull's eye occurs on the *eleventh* try, $x = 11$, is $f(11) = q^{x-1}p = (0.85)^{11-1}0.15 \approx$
(i) **0.01** (ii) **0.02** (iii) **0.03** (iv) **0.04**.

(e) Since the chance the first bull's eye occurs on the xth try is $f(x) = q^{x-1}p$, then the sum of this series

$$\sum_{x \in R} f(x) = \sum_{x=1}^{\infty} f(x) = \sum_{x=1}^{\infty} q^{x-1}p = \frac{p}{1-q} =$$

(i) $\boldsymbol{p}$ (ii) $\boldsymbol{q}$ (iii) $\boldsymbol{\frac{p}{q}}$ (iv) **1**.

5. *Probability mass function and cumulative distribution function: flipping a coin twice.* Let the number of heads flipped in two flips of a coin be a discrete random variable X with pmf $f(x)$ is given in table and Figure 2.4(a), and cdf $F(x) = P(X \leq b)$ given in Figure 2.4(b).

x	0	1	2
$f(x) = P(X = x)$	0.25	0.5	0.25

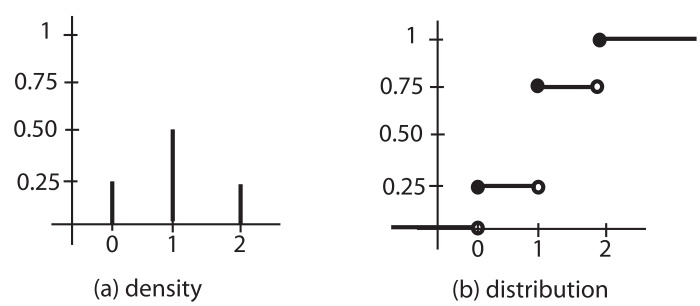

Figure 2.4: Density and distribution function: flipping a coin twice

(a) $f(0) = P(X = 0) =$
(i) **0** (ii) **0.25** (iii) **0.50** (iv) **0.75**.

(b) $f(1) = P(X = 1) =$
(i) **0** (ii) **0.25** (iii) **0.50** (iv) **0.75**.

(c) $f(2) = P(X = 2) =$
(i) **0** (ii) **0.25** (iii) **0.50** (iv) **0.75**.

(d) $F(0) = P(X \leq 0) = P(X = 0) =$
(i) **0** (ii) **0.25** (iii) **0.75** (iv) **1**.

(e) $F(1) = P(X \leq 1) = P(X = 0) + P(X = 1) =$
(i) **0** (ii) **0.25** (iii) **0.75** (iv) **1**.

(f) $F(2) = P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) =$
(i) **0** (ii) **0.25** (iii) **0.75** (iv) **1**.

(g) Also, if $b < 0$, $F(b) = P(X \leq b) =$
(i) **0** (ii) **0.25** (iii) **0.75** (iv) **1**.

(h) And, if $b \geq 2$, $F(b) = P(X \leq b) =$
(i) **0** (ii) **0.25** (iii) **0.75** (iv) **1**.

(i) So, in this case,

$$F(b) = \begin{cases} 0, & b < 0 \\ 0.25, & 0 \leq b < 1 \\ 0.75, & 1 \leq b < 2 \\ 1, & b \geq 2. \end{cases}$$

(i) **True** (ii) **False**

(j) The discontinuous "step function" graph of this $F(b)$ is given in Figure 2.4(b). Notice that $F(b)$ is right continuous, which is indicated by the solid and empty endpoints on the graph of this distribution function. The height of a "step" in distribution $F(b)$ is equal to the height of the corresponding "stick" in density $p(b)$.
(i) **True** (ii) **False**

(k) $P(X < 1) = P(X = 0) = 0.25 =$
(i) $\boldsymbol{F(0)}$ (ii) $\boldsymbol{F(1)}$ (iii) $\boldsymbol{F(2)}$ (iv) $\boldsymbol{F(3)}$.

(l) $P(X < 2) = P(X = 0) + P(X = 1) = 0.75 =$
(i) $\boldsymbol{F(0)}$ (ii) $\boldsymbol{F(1)}$ (iii) $\boldsymbol{F(2)}$ (iv) $\boldsymbol{F(3)}$.

(m) $P(X > 1) = 1 - P(X \leq 1) = 1 - F(1) = 1 - 0.75 =$
(i) **0** (ii) **0.25** (iii) **0.75** (iv) **1**.

6. *Another discrete cumulative distribution function.*
Let random variable X have distribution,

$$F(b) = \begin{cases} 0, & b < 0 \\ \frac{1}{3}, & 0 \leq b < 1 \\ \frac{1}{2}, & 1 \leq b < 2 \\ 1, & b \geq 2. \end{cases}$$

(a) $f(0) = P(X = 0) =$
(i) **0** (ii) $\boldsymbol{\frac{1}{6}}$ (iii) $\boldsymbol{\frac{1}{3}}$ (iv) $\boldsymbol{\frac{1}{2}}$.

(b) $f(1) = P(X = 1) = P(X \leq 1) - P(X \leq 0) = F(1) - F(0) =$
(i) **0** (ii) $\boldsymbol{\frac{1}{6}}$ (iii) $\boldsymbol{\frac{1}{3}}$ (iv) $\boldsymbol{\frac{1}{2}}$.

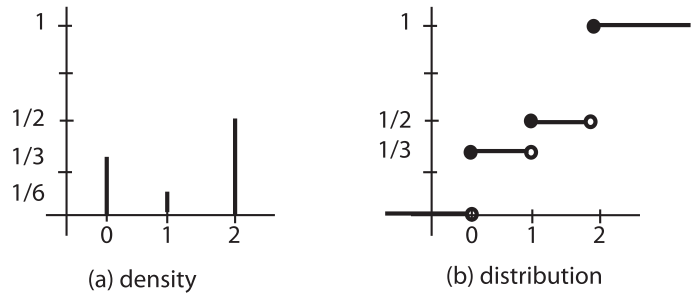

Figure 2.5: pmf and cdf

(c) $f(2) = P(X = 2) = P(X \leq 2) - P(X \leq 1) = F(2) - F(1) =$
(i) **0** (ii) $\frac{1}{6}$ (iii) $\frac{1}{3}$ (iv) $\frac{1}{2}$.

(d) Random variable X is discrete, not continuous, because the associated $F(b)$ is a discontinuous ("step", in this case) function.
(i) **True** (ii) **False**

(e) Notice that
(1) $\lim_{b \to -\infty} F(b) = 0$,
(2) $\lim_{b \to \infty} F(b) = 1$,
(3) if $b_1 < b_2$, then $F(b_1) \leq F(b_2)$; that is, F is nondecreasing.
(i) **True** (ii) **False**

(f) $F(1) = P(X \leq 1) = P(X = 0) + P(X = 1) =$
(i) **0** (ii) $\frac{1}{6}$ (iii) $\frac{1}{3}$ (iv) $\frac{1}{2}$.

(g) $P(X < 2) = P(X = 0) + P(X = 1) =$
(i) **0** (ii) $\frac{1}{6}$ (iii) $\frac{1}{3}$ (iv) $\frac{1}{2}$.

(h) $P(X \leq 1.5) = P(X = 0) + P(X = 1) =$
(i) **0** (ii) $\frac{1}{6}$ (iii) $\frac{1}{3}$ (iv) $\frac{1}{2}$.

7. *Mode and Median, Discrete.*

(a) *Number of Ears of Corn.* The number of ears of corn, Y, on a typical corn plant has the following probability distribution.

x	0	2	4	6	8	10
$f(x)$	0.17	0.21	0.18	0.11	0.16	0.17

The *mode* number of ears of corn with largest probability,
(i) **0** (ii) **2** (iii) **4** (iv) **6**.

The *median* number of ears of corn, m, is equal to the *smallest* number of ears of corn where there is at least a 50% chance of getting less

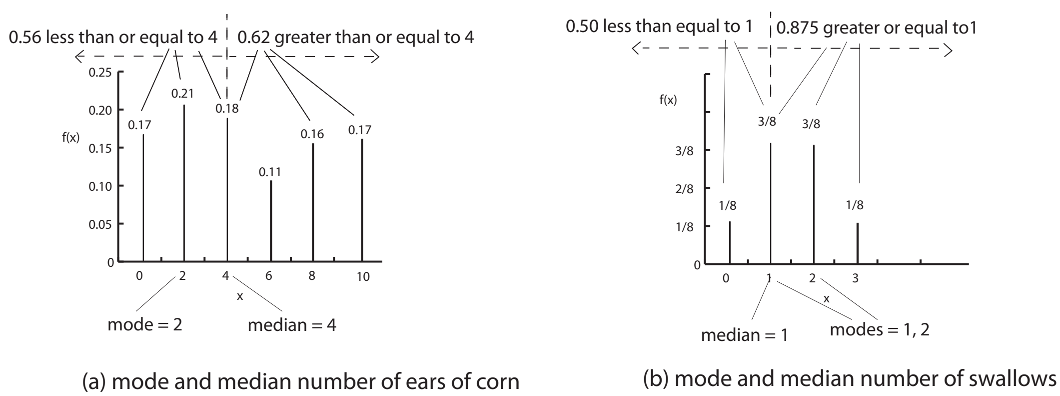

Figure 2.6: Mode and Median for Discrete Probability Distributions

than or equal to this number and also at least a 50% of getting more than this number), in other words the median $m =$

i. 2 ears of corn since there is a 17% + 21% = 38% chance of getting less than or equal to 2 ears and a 21% + 18% + 11% + 16% + 17% = 83% chance of getting 2 or more.

ii. between 2 and 4 ears of corn since there is a 17% + 21% = 38% chance of getting less than or equal to 2 ears and a 18% + 11% + 16% + 17% = 62% chance of getting 4 or more.

iii. 4 ears of corn since there is a 17% + 21% + 18% = 56% chance of getting less than or equal to 4 ears and a 18% + 11% + 16% + 17% = 62% chance of getting 4 or more

(b) *Swallows.* The number of swallows, X, in any group of three birds is given by the following probability distribution.

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

The *mode* number of swallows is the number with largest probability,
(i) **0** (ii) **1** (iii) **2** (iv) **3**.

The *median* number of swallows, m, is *smallest* number

i. $m = 0$ because

$$P(X \leq 0) = \frac{1}{8} < 0.5 \quad \text{and} \quad P(X \geq 0) = \frac{8}{8} \geq 0.5.$$

ii. $m = 1$ because

$$P(X \leq 1) = \frac{4}{8} \geq 0.5 \quad \text{and} \quad P(X \geq 1) = \frac{7}{8} \geq 0.5.$$

iii. $m = 2$ because

$$P(X \leq 2) = \frac{7}{8} \geq 0.5 \quad \text{and} \quad P(X \geq 2) = \frac{4}{8} \geq 0.5.$$

2.3 The Hypergeometric and Binomial Distributions

The *hypergeometric* probability mass function, the probability of selecting n items at random with*out* replacement from a population of size N with N_1 type I objects and N_2 type II objects, $N_1 + N_2 = N$, where x of the n objects are type I objects, is

$$f(x) = P\left(X = x\right) = \frac{\left(\begin{array}{c} N_1 \\ x \end{array}\right)\left(\begin{array}{c} N_2 \\ n-x \end{array}\right)}{\left(\begin{array}{c} N \\ n \end{array}\right)}, \ x = 1, \ldots, n.$$

Random variable X is the number of type I objects from the sample of size n; N_1, N_2, N, and n are all parameters of this pmf.

The *Bernoulli experiment* has the following properties:
(1) An outcome from any Bernoulli experiment trial is either a success or failure,
(2) under repetition, trial outcomes are independent, sampling is *with* replacement,
(3) the probability of success in any trial is constant, is p.

Assuming a sequence of n Bernoulli trials, each with probability of success p, the *binomial* probability mass function is,

$$f(x) = P(X = x) = \left(\begin{array}{c} n \\ x \end{array}\right) p^x (1-p)^{n-x}, \ x = 0, 1, \ldots, n,$$

where random variable X is the number of successes in n trials; n and p are parameters of this pmf and "X is $b(n,p)$" means random variable X is a binomial with parameters n and p.

The *hypergeometric can be approximated by the binomial* as long as only a small sample, at most 5%, of the population, $n \leq 0.05N$, is selected without replacement, in which case, let

$$p = \frac{N_1}{N}.$$

The *unusual-event principle* says if we make an assumption and this assumption leads to only a small chance of an event occurring, the event is unusual, then the assumption is probably incorrect. For example, if we assume X is $b(n,p)$, but find out the *observed* event of x successes in n trials is *unusual* because

$$P(X \leq x) \leq 0.05 \quad \text{or} \quad P(X \geq x) \leq 0.05,$$

then our assumption X is $b(n,p)$ is possibly incorrect.

Exercise 2.3 (The Hypergeometric and Binomial Distributions)

1. *Hypergeometric: sampling red marbles from an urn.* Two marbles are taken, $n = 2$, one at a time, with*out* replacement, from an urn which has $N_1 = 6$ red and $N_2 = 10$ blue marbles. We win \$2 for each red marble chosen, lose \$1 for each blue marble chosen. Let X be number of red marbles and Y be winnings.

   (a) The chance both marbles are red, $X = 2$, is

   $$P\{RR\} = P(X=2) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}} = \frac{\binom{6}{2}\binom{10}{2-2}}{\binom{16}{2}} =$$

   $$\text{(i)}\ \frac{\binom{\mathbf{6}}{\mathbf{2}}\binom{\mathbf{10}}{\mathbf{0}}}{\binom{\mathbf{16}}{\mathbf{2}}} \quad \text{(ii)}\ \frac{\binom{\mathbf{9}}{\mathbf{1}}\binom{\mathbf{8}}{\mathbf{2}}}{\binom{\mathbf{16}}{\mathbf{3}}} \quad \text{(iii)}\ \frac{\binom{\mathbf{8}}{\mathbf{1}}\binom{\mathbf{11}}{\mathbf{2}}}{\binom{\mathbf{16}}{\mathbf{3}}} = 0.125.$$

   ```
   dhyper(2,6,10,2) # hypergeometric pmf
   ```

   ```
   [1] 0.125
   ```

   (b) Alternatively, $P\{RR\} =$
   (i) $\frac{\mathbf{6}}{\mathbf{16}} \times \frac{\mathbf{5}}{\mathbf{15}}$ (ii) $\frac{\mathbf{10}}{\mathbf{16}} \times \frac{\mathbf{9}}{\mathbf{15}}$ (iii) $\frac{\mathbf{12}}{\mathbf{16}} \times \frac{\mathbf{11}}{\mathbf{15}}$ (iv) $\frac{\mathbf{16}}{\mathbf{16}} \times \frac{\mathbf{15}}{\mathbf{15}} = 0.125$.

   (c) Since the winnings are $Y = \$4$ when both marbles are red,
   $P(Y = \$4) = P\{RR\} = P(X = 2) =$
   (i) **0.025** (ii) **0.125** (iii) **0.225** (iv) **0.500**.

   (d) The chance both marbles are blue, $X = 0$, is

   $$P\{BB\} = P(X=0) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}} = \frac{\binom{6}{0}\binom{10}{2-0}}{\binom{16}{2}} =$$

   $$\text{(i)}\ \frac{\binom{\mathbf{6}}{\mathbf{2}}\binom{\mathbf{10}}{\mathbf{0}}}{\binom{\mathbf{16}}{\mathbf{2}}} \quad \text{(ii)}\ \frac{\binom{\mathbf{9}}{\mathbf{1}}\binom{\mathbf{8}}{\mathbf{2}}}{\binom{\mathbf{16}}{\mathbf{3}}} \quad \text{(iii)}\ \frac{\binom{\mathbf{6}}{\mathbf{0}}\binom{\mathbf{10}}{\mathbf{2}}}{\binom{\mathbf{16}}{\mathbf{2}}} = 0.375.$$

```
dhyper(0,6,10,2) # hypergeometric pmf
```

```
[1] 0.375
```

(e) Also $P\{BB\} = P(Y = -\$2) =$
(i) $\frac{\mathbf{6}}{\mathbf{16}} \times \frac{\mathbf{5}}{\mathbf{15}}$ (ii) $\frac{\mathbf{10}}{\mathbf{16}} \times \frac{\mathbf{9}}{\mathbf{15}}$ (iii) $\frac{\mathbf{12}}{\mathbf{16}} \times \frac{\mathbf{11}}{\mathbf{15}}$ (iv) $\frac{\mathbf{16}}{\mathbf{16}} \times \frac{\mathbf{15}}{\mathbf{15}} = 0.375$.

(f) $P\{RB, BR\} = P(Y = \$1) = 1 - P\{RR\} - P\{BB\} =$
(i) **0.025** (ii) **0.125** (iii) **0.225** (iv) **0.500**.

(g) The hypergeometric pmf of number of red marbles, X, is

x	0	1	2
$P(X = x)$	0.375	0.500	0.125

(i) **True** (ii) **False**

(h) The hypergeometric pmf of payoffs, Y, is

y	-\$2	\$1	\$4
$P(Y = y)$	0.375	0.500	0.125

(i) **True** (ii) **False**

2. *Hypergeometric: televisions.* Seven television ($n = 7$) tubes are chosen at random from a shipment of $N = 240$ television tubes of which $N_1 = 15$ are defective.

(a) The probability that $X = 4$ of the chosen televisions are defective is

$$f(4) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}} =$$

(i) $\frac{\binom{\mathbf{15}}{\mathbf{4}} \times \binom{\mathbf{225}}{\mathbf{3}}}{\binom{\mathbf{240}}{\mathbf{4}}}$ (ii) $\frac{\binom{\mathbf{15}}{\mathbf{3}} \times \binom{\mathbf{225}}{\mathbf{3}}}{\binom{\mathbf{240}}{\mathbf{7}}}$ (iii) $\frac{\binom{\mathbf{15}}{\mathbf{4}} \times \binom{\mathbf{225}}{\mathbf{3}}}{\binom{\mathbf{240}}{\mathbf{7}}}$

(b) The probability $X = 4$ of the chosen televisions are defective is
$f(4) =$
(i) **0.0003069** (ii) **0.0005069** (iii) **0.0006069** (iv) **0.0007069**.

```
dhyper(4,15,225,7) # hypergeometric pmf
```

```
[1] 0.0003069143
```

(c) The probability $X = 5$ of the chosen televisions are defective is
$f(5) =$
(i) **0.000007069** (ii) **0.000009084** (iii) **0.00010069** (iv) **0.00013059**.

```
dhyper(5,15,225,7) # hypergeometric pmf
```

```
[1] 9.083563e-06
```

(d) The probability *at most* $X = 5$ of the chosen televisions are defective is $P(X \leq 5) =$
(i) **0.900** (ii) **0.925** (iii) **0.950** (iv) **0.999**.

```
phyper(5,15,225,7,lower.tail=TRUE) # hypergeometric cdf
```

```
[1] 0.9999999
```

(e) The probability *at least* $X = 1$ of the chosen televisions are defective is $P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) =$
(i) **0.367** (ii) **0.435** (iii) **0.545** (iv) **0.633**.

```
phyper(0,15,225,7,lower.tail=FALSE) # 1 - hypergeometric cdf
```

```
[1] 0.3672717
```

3. *Hypergeometric: capture–recapture.* To determine approximate number of perch, N, in Lake Fishalot, $n = 45$ are captured at random from the lake, tagged and let go back into the lake. A short while later, another $n = 32$ perch are captured, of which $x = 2$ are found to be tagged. Approximately how many perch are in Lake Fishalot?

(a) The chance two of the second group of captured fish are tagged is

$$f(2) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}} =$$

(i) $\dfrac{\binom{\mathbf{32}}{\mathbf{2}} \times \binom{\boldsymbol{N}}{\mathbf{43}}}{\binom{\boldsymbol{N}}{\mathbf{3200}}}$ (ii) $\dfrac{\binom{\mathbf{45}}{\mathbf{2}} \times \binom{\boldsymbol{N}-\mathbf{45}}{\mathbf{32}-\mathbf{2}}}{\binom{\boldsymbol{N}}{\mathbf{32}}}$ (iii) $\dfrac{\binom{\mathbf{45}}{\mathbf{43}} \times \binom{\boldsymbol{N}}{\mathbf{2}}}{\binom{\boldsymbol{N}}{\mathbf{3200}}}$,

(b) *Guess* $N = 500$. In this case, chance two of 32 fish chosen are tagged is

$$f(2) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}} = \frac{\binom{45}{2} \times \binom{500-45}{32-2}}{\binom{500}{32}} =$$

(i) **0.24** (ii) **0.26** (iii) **0.29** (iv) **0.32**.

```
dhyper(2,45,455,32) # hypergeometric pmf
```

```
[1] 0.2405091
```

(c) *Guess* $N = 750$. In this case, chance two of 32 fish chosen are tagged is

$$f(2) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}} = \frac{\binom{45}{2} \times \binom{750-45}{32-2}}{\binom{750}{32}} =$$

(i) **0.24** (ii) **0.26** (iii) **0.29** (iv) **0.32**.

```
dhyper(2,45,705,32) # hypergeometric pmf
```

```
[1] 0.2853607
```

(d) *Guess* $N = 1000$. In this case, chance two of 32 fish chosen are tagged is $f(2) =$ (i) **0.24** (ii) **0.26** (iii) **0.29** (iv) **0.32**.

```
dhyper(2,45,955,32) # hypergeometric pmf
```

```
[1] 0.2570794
```

(e) A summary of results are given in the following table.

N	500	750	1000
$f(2)$	0.24	0.29	0.26

Since the *largest* chance that 2 of 32 fish chosen are tagged is 0.29, then it seems from the choices given, the number of fish in Lake Fishalot is $N =$ (i) **500** (ii) **750** (iii) **1000** (iv) **1250**.

(f) The approximation to N would improve if we had more than three $f(2)$ to choose from; however, more effort would be required in calculating the extra $f(2)$. Differentiating

$$\frac{\binom{45}{2} \times \binom{N-45}{32-2}}{\binom{N}{32}}$$

with respect to N and then setting to zero, to locate the maximum N is also possible, but difficult to do.
(i) **True** (ii) **False**

4. *Bernoulli: Flipping a coin.* The number of heads, X, in one flip of a coin, is given by the following probability function,

$$f(x) = (0.25)^x(0.75)^{1-x}, \quad x = 0, 1.$$

(a) The chance of flipping 1 head ($X = 1$) is $f(1) = (0.25)^1(0.75)^{1-1} =$ (i) **0** (ii) **0.25** (iii) **0.50** (iv) **0.75**.

(b) This coin is (i) **fair** (ii) **unfair**.

(c) The chance of flipping no heads ($X = 0$) is
$f(0) = (0.25)^0(0.75)^{1-0} =$ (i) **0** (ii) **0.25** (iii) **0.50** (iv) **0.75**.

(d) A "tabular" version of this probability distribution of flipping a coin is

i. Distribution A.

x	0	1
$f(x)$	0.25	0.75

ii. Distribution B.

x	0	1
$f(x)$	0.75	0.25

iii. Distribution C.

x	0	1
$f(x)$	0.50	0.50

(e) The number of different ways of describing a distribution include (choose one or more) (i) **function** (ii) **table** (iii) **graph**.

5. *Airplane engines.* Each engine of four ($n = 4$) on an airplane fails 11% ($p = 0.11, q = 1 - p = 0.89$) of the time. Assume this problem obeys the conditions of a binomial experiment.

(a) The chance two engines fail is
$f(2) = \left(\begin{array}{c} 4 \\ 2 \end{array} \right) 0.11^2 0.89^2 =$ (i) **0.005** (ii) **0.011** (iii) **0.058** (iv) **0.157**.

```
dbinom(2,4,0.11) # binomial pmf
```

```
[1] 0.05750646
```

(b) The chance three engines fail is
$f(2) = \left(\begin{array}{c} 4 \\ 3 \end{array} \right) 0.11^3 0.89^1 =$ (i) **0.005** (ii) **0.011** (iii) **0.040** (iv) **0.057**.

```
dbinom(3,4,0.11) # binomial pmf
```

```
[1] 0.00473836
```

(c) The chance *at most* two engines fail is
$P(X \leq 2) = \sum_{x=0}^{2} \left(\begin{array}{c} 4 \\ x \end{array} \right) 0.11^x 0.89^{4-x} \approx$
(i) **0.991** (ii) **0.995** (iii) **0.997** (iv) **0.999**.

```
pbinom(2,4,0.11) # binomial cdf
```

```
[1] 0.9951152
```

(d) The chance at most *three* engines fail is
$P(X \leq 3) = \sum_{x=0}^{3} \left(\begin{array}{c} 4 \\ x \end{array} \right) 0.11^x 0.89^{4-x} \approx$
(i) **0.991** (ii) **0.995** (iii) **0.997** (iv) **0.999**.

```
pbinom(3,4,0.11,lower.tail=TRUE) # binomial cdf
```

```
[1] 0.9998536
```

(e) The chance *at least* three engines fail is
$P(X \geq 3) = 1 - P(X \leq 2) \approx$
(i) **0.005** (ii) **0.010** (iii) **0.016** (iv) **0.023**.

```
pbinom(2,4,0.11,lower.tail=FALSE) # 1 - binomial cdf
1 - pbinom(2,4,0.11) # 1 - binomial cdf
```

```
[1] 0.00488477
```

(f) Since

$$P(X \leq 3) \approx 0.999 > 0.05 \quad \text{and} \quad P(X \geq 3) \approx 0.005 < 0.05,$$

it (i) **is** (ii) **is not** unusual 3 engines fail, assuming 11% engine failure.

6. *Binomial distribution: shaded region.*
Let $0 \leq x \leq 1, 0 \leq y \leq 1$, where part of the square, $x < y$, is shaded.

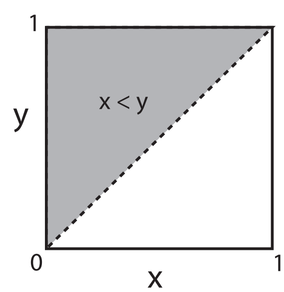

Figure 2.7: Area of $x < y$

(a) If points are taken at random from this square, determine the probability of choosing a point from the shaded region, which is half the square.
$p =$ (i) **0.025** (ii) **0.050** (iii) **0.069** (iv) **0.500**.

(b) If two points are taken at random from a square, what is the chance both are from the shaded region? Clearly,
$p^2 = 0.5^2 =$ (i) **0.25** (ii) **0.50** (iii) **0.069** (iv) **0.119**
or, using binomial pmf, since $p = 0.5$, $n = 2$, and $x = 2$, then
$f(2) = \binom{2}{2} 0.5^2 0.5^0 =$ (i) **0.25** (ii) **0.050** (iii) **0.069** (iv) **0.119**.

```
dbinom(2,2,0.5) # binomial pmf
```

```
[1] 0.25
```

(c) If ten points are taken at random from a square, what is the chance seven are from the shaded region? Using binomial pmf, since $p = 0.5$, $n = 10$, and $x = 7$, then
$f(7) = \binom{10}{7} 0.5^7 0.5^3 =$ (i) **0.117** (ii) **0.252** (iii) **0.369** (iv) **0.419**.

```
dbinom(7,10,0.5) # binomial pmf
[1] 0.1171875
```

(d) If $n = 11$, $p = 0.5$, $x = 3$, then
$f(3) = \left(\begin{array}{c} 11 \\ 3 \end{array} \right) 0.5^3 0.5^8 =$ (i) **0.081** (ii) **0.231** (iii) **0.258** (iv) **0.319**.

```
dbinom(3,11,0.5) # binomial pmf
[1] 0.08056641
```

7. *Multiple choice questions.* On a multiple–choice exam with 4 possible answers for each of the 5 questions, what is the probability that a student should get at most 3 correct answers just by guessing?

   (a) Since there are five questions,
   $n =$ (i) **2** (ii) **3** (iii) **4** (iv) **5**.

   (b) Since a student wants 3 or more correct answers,
   $x =$ (i) **2** (ii) **2, 3** (iii) **2, 3, 4** (iv) **3, 4, 5**.

   (c) Since the student is choosing from 4 answers at random,
   $p =$ (i) $\mathbf{\frac{1}{2}}$ (ii) $\mathbf{\frac{1}{3}}$ (iii) $\mathbf{\frac{1}{4}}$ (iv) $\mathbf{\frac{1}{5}}$.

   (d) The chance a student should get *at most* 3 correct answers is
   $P(X \leq 3) = \sum_{x=0}^{3} \left(\begin{array}{c} 5 \\ x \end{array} \right) \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{5-x} \approx$
   (i) **0.697** (ii) **0.704** (iii) **0.812** (iv) **0.984**.

```
pbinom(3,5,0.25) # binomial cdf
[1] 0.984375
```

   (e) The chance a student should get *at least* 3 correct answers is
   $P(X \geq 3) = 1 - P(X \leq 2) \approx$
   (i) **0.097** (ii) **0.104** (iii) **0.112** (iv) **0.284**.

```
1 - pbinom(2,5,0.25) # 1 - binomial cdf
[1] 0.1035156
```

   (f) Since

$$P(X \leq 3) = 0.984 > 0.05 \quad \text{and} \quad P(X \geq 3) = 0.104 > 0.05,$$

   it (i) **is** (ii) **is not** unusual to score 3 correct answers, guessing at random.

8. *Lawyer.* A lawyer estimates she wins 40% ($p = 0.4$) of her cases. Assume each trial is independent of one another and, in general, this problem obeys the conditions of a binomial experiment. The lawyer presently represents 10 ($n = 10$) defendants.

   (a) The tabular form of this distribution is given by,

r	0	1	2	3	4	5	6	7	8	9	10
$P(R = r)$	0.006	0.040	0.121	0.215	0.251	0.201	0.111	0.043	0.011	0.002	0.000

where, for example, the chance of her winning 6 of 10 cases is 0.111. In a similar way, the chance of her winning 4 of 10 cases is
(i) **0.121** (ii) **0.215** (iii) **0.251** (iv) **0.351**.

(b) The graphical form of the probability distribution is given in Figure 2.8.

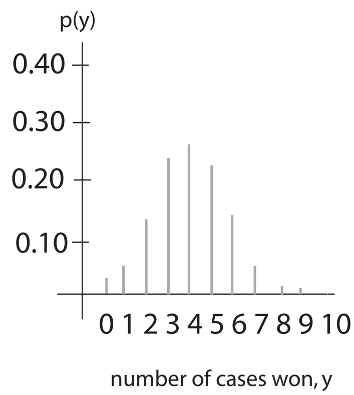

Figure 2.8: Binomial distribution: lawyer wins

The number of cases the lawyer has most chance of winning is
(i) **one** (ii) **two** (iii) **three** (iv) **four**.

(c) The binomial distribution, in this case, is
(i) **skewed left** (ii) **skewed right** (iii) **more or less symmetric** .

(d) In general, the binomial distribution will sometimes, but not always, be symmetric. If her chance of winning was $p = 0.1$ instead of $p = 0.4$, the binomial distribution would be
(i) **skewed left** (ii) **skewed right** (iii) **more or less symmetric**.

(e) *Understanding the binomial formula.*
Chance of nine wins and then a lose would be $p^9q = 0.4^9 0.6^1$,
eight wins, one lose, one win would be: $p^8qp = p^9q = 0.4^9 0.6^1$,
seven wins, one lose, two wins would be: $p^7qp^2 = p^9q = 0.4^9 0.6^1, \ldots$.
Since nine wins occurs ten different ways, chance nine wins would be:
$10 \times p^9q = 10 \times 0.4^9 0.6^1 = \binom{10}{9} 0.4^9 0.6^1$.
In general, chance of x wins in n trials is binomial formula,
$\binom{n}{x} p^x q^{n-x},\ x = 0, 1, \ldots, n$
(i) **True** (ii) **False**

(f) *Conditional versus unconditional binomial.*

Since $P(X > 3) = 1 - P(X \leq 3) \approx$
(i) **0.46** (ii) **0.58** (iii) **0.62** (iv) **0.67**,

```
1 - pbinom(3,10,0.4) # 1 - binomial cdf
```

```
[1] 0.6177194
```

and

$$P(X > 4{+}3|X > 4) = \frac{P(X > 4+3, X > 4)}{P(X > 4)} = \frac{P(X > 4+3)}{P(X > 4)} = \frac{P(X > 7)}{P(X > 4)} \approx$$

(i) **0.034** (ii) **0.048** (iii) **0.075** (iv) **0.089**;

```
(1 - pbinom(7,10,0.4))/(1 - pbinom(4,10,0.4)) # conditional probability
```

```
[1] 0.03350957
```

in other words, $P(X > 3) = 0.62 \neq P(X > 4 + 3|X > 4) = 0.03$. The chance the lawyer wins at least 4 cases is not equal to the chance she wins at least 8 cases, given winning at least 5 cases.

9. *Binomial approximation to the hypergeometric: televisions.* Seven television ($n = 7$) tubes are chosen at random from a shipment of $N = 240$ television tubes of which $N_1 = 15$ are defective.

   (a) We sample with*out* replacement; that is, every time a TV is chosen, we do *not* replace it to be potentially chosen again. In other words, the chance of choosing a defective TV, every time a TV is chosen, *changes* or *depends* on the number of defective TVs that were chosen before it.
   (i) **True** (ii) **False**

   (b) If the sample size, n, is small relative to the number of televisions, N, $n < 0.05N$, say, the hypergeometric can be approximated by a binomial. The chance, $p = \frac{N_1}{N}$, of choosing a defective TV, every time a TV is chosen, does not change "that much" when $\frac{n}{N} < 0.05$. Since $\frac{n}{N} = \frac{7}{240} \approx 0.029 < 0.05$, the binomial will probably approximate the hypergeometric
   (i) **very closely.** (ii) **somewhat closely.** (iii) **not closely at all.**

   (c) *(Exact) hypergeometric.* Probability $x = 4$ of $n = 7$ chosen TVs are defective is

$$f(4) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}} = \frac{\binom{15}{4} \times \binom{225}{3}}{\binom{240}{7}} =$$

   (i) **0.0003069** (ii) **0.0004400** (iii) **0.0006069** (iv) **0.0007069**.

```
dhyper(4,15,225,7) # exact hypergeometric pmf
```

```
[1] 0.0003069143
```

(d) *Approximate binomial.* Since $p = \frac{N_1}{N} = \frac{15}{240} = 0.0625$ and so a binomial approximation to probability $x = 4$ of $n = 7$ chosen TVs are defective is:

$$f(4) = \binom{7}{4} 0.0625^4 (1 - 0.0625)^{7-4} =$$

(i) **0.0003069** (ii) **0.0004400** (iii) **0.0006069** (iv) **0.0007069**.

```
dbinom(4,7,0.0625) # approximate binomial pmf
```

```
[1] 0.0004400499
```

10. *Multinomial: faculty and subjects.* There are 9 different faculty members and 3 subjects: mathematics, statistics and physics. There is a 50%, 35% and 15% chance a faculty member teaches mathematics, statistics and physics, respectively.

(a) Chance 4, 3 and 2 faculty members teach mathematics, statistics and physics, respectively, is

$$p(4, 3, 2) = \frac{9!}{4!3!2!} 0.5^4 0.35^3 0.15^2 \approx$$

(i) **0.055** (ii) **0.067** (iii) **0.076** (iv) **0.111**

```
dmultinom(c(4,3,2), prob = c(0.5,0.35,0.15)) # multinomial pmf
```

```
[1] 0.07596914
```

(b) Chance 4, 4 and 1 faculty members teach mathematics, statistics and physics, respectively, is

$$p(4, 4, 1) = \frac{9!}{4!4!1!} 0.5^4 0.35^4 0.15^1 \approx$$

(i) **0.089** (ii) **0.098** (iii) **0.108** (iv) **0.131**

```
dmultinom(c(4,4,1), prob = c(0.5,0.35,0.15)) # multinomial pmf
```

```
[1] 0.08863066
```

2.4 The Poisson Distribution

Poisson probability mass function, used to describe probabilities of count, X, of occurrences (successes) of an event in a specified time (or space) interval, is

$$f(x) = P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}, \; x = 0, 1, \ldots,$$

where $\lambda > 0$ is *average* number of occurrences of an event; also, random variable X obeys conditions of Poisson process:

- zero probability of two or more successes in sufficiently small subinterval,
- probability of success same if two intervals are of equal length,
- number successes in one interval independent of number of successes in any other nonoverlapping interval.

If $n \geq 100$, $np \leq 100$, Poisson pmf can be used to approximate binomial pmf where $\lambda = np$.

Exercise 2.4 (The Poisson Distribution)

1. *Poisson: number of photon hits.*
   Piece of iron bombarded with electrons releases a number of photons, X, to a surrounding magnetic detection field. An average of $\lambda = 5$ particles hit magnetic detection field *per (one) microsecond*. Assume this is a Poisson process.

   (a) Chance $x = 0$ photons hit field in one microsecond

   $$f(0) = \frac{\lambda^x}{x!}e^{-\lambda} = \frac{5^0}{0!}e^{-5} \approx$$

   (i) **0.007** (ii) **0.008** (iii) **0.009**.

   ```
   dpois(0,5) # Poisson pmf
   ```

   ```
   [1] 0.006737947
   ```

   (b) Chance $x = 2$ photons hit field in one microsecond

   $$f(2) = \frac{\lambda^x}{x!}e^{-\lambda} = \frac{5^2}{2!}e^{-5} \approx$$

   (i) **0.06** (ii) **0.07** (iii) **0.08**.

   ```
   dpois(2,5) # Poisson pmf
   ```

   ```
   [1] 0.08422434
   ```

   (c) Chance *at most* $x = 2$ photons hit field in one microsecond

   $$P(X \leq 2) = \frac{5^0}{0!}e^{-5} + \frac{5^1}{1!}e^{-5} + \frac{5^2}{2!}e^{-5} \approx$$

   (i) **0.08** (ii) **0.12** (iii) **0.18**.

   ```
   ppois(2,5) # Poisson cdf
   ```

   ```
   [1] 0.124652
   ```

(d) Chance *at least* $x = 2$ photons hit field in one microsecond

$$\begin{aligned} P(X \geq 2) &= \frac{5^2}{2!}e^{-5} + \frac{5^3}{3!}e^{-5} + \cdots \text{ forever} \\ &= 1 - P(X < 2) \ = \ 1 - P(X \leq 1) \\ &= 1 - [P(0) + P(1)] \\ &= 1 - \left[\frac{5^0}{0!}e^{-5} + \frac{5^1}{1!}e^{-5}\right] \ \approx \end{aligned}$$

(i) **0.91** (ii) **0.93** (iii) **0.96**.

```
1 - ppois(1,5) # 1 - Poisson cdf
ppois(1,5,lower.tail=FALSE) # 1 - Poisson cdf
```

```
[1] 0.9595723
```

(e) Chance $x = 2$ photons hit field in *two* microseconds

$$P(2) = \frac{(5 \cdot 2)^2}{2!}e^{-5 \cdot 2} \approx$$

(i) **0.002** (ii) **0.005** (iii) **0.006**.

```
dpois(2,10) # Poisson pmf, 2 x 5 = 10
```

```
[1] 0.002269996
```

(f) Chance *at most* $x = 3$ photons hit field in *two* microseconds

$$P(X \leq 3) \approx$$

(i) **0.01** (ii) **0.06** (iii) **0.08**.

```
ppois(3,10) # Poisson cdf, 2 x 5 = 10
```

```
[1] 0.01033605
```

(g) Chance *at least* $x = 21$ photons hit field in *four* microseconds

$$P(X \geq 21) = 1 - P(X < 21) = 1 - P(X \leq 20) \approx$$

(i) **0.44** (ii) **0.52** (iii) **0.68**.

```
1 - ppois(20,20) # 1 Poisson cdf, 4 x 5 = 20
```

```
[1] 0.4409074
```

(h) *Poisson process?* Match columns.

Poisson process	photon example
(a) zero chance more than one success in small subinterval	(A) *assume* "zero" chance two hits at same time
(b) chance of success same if two intervals are of equal length	(B) *assume* chance number hits same per microsecond
(c) independence of successes in different intervals	(C) *assume* hits independent of one another

Poisson process	(a)	(b)	(c)
photon example			

2. *Poisson: Bryozoan count.*
During a biology study, 250 bryozoans are found attached on a submerged 10 centimeter by 10 centimeter (100 centimeters^2) plate, an average of $\lambda = \frac{250}{100} = 2.5$ bryozoans per one centimeter^2. Assume this is a Poisson process.

   (a) Chance $x = 0$ bryozoans attached in one centimeter^2

$$f(0) = \frac{\lambda^x}{x!}e^{-\lambda} = \frac{2.5^0}{0!}e^{-2.5} \approx$$

   (i) **0.07** (ii) **0.08** (iii) **0.09**.

```
dpois(0,2.5) # Poisson pmf
```

```
[1] 0.082085
```

   (b) Chance *at most* $x = 2$ bryozoans attached in one centimeter^2

$$P(X \leq 2) \approx$$

   (i) **0.54** (ii) **0.62** (iii) **0.78**.

```
ppois(2,2.5) # Poisson cdf
```

```
[1] 0.5438131
```

   (c) Chance *between* $x = 2$ and $x = 4$ bryozoans in one centimeter^2, inclusive:

$$P(2 \leq X \leq 4) = P(X \leq 4) - P(X < 2) = P(X \leq 4) - P(X \leq 1) \approx$$

   (i) **0.60** (ii) **0.73** (iii) **0.76**.

```
ppois(4,2.5) - ppois(1,2.5) # between Poisson cdfs
```

```
[1] 0.6038805
```

   (d) Chance *between* $x = 2$ and $x = 4$ bryozoans in *three* centimeters^2, inclusive:

$$P(2 \leq X \leq 4) = P(X \leq 4) - P(X < 2) = P(X \leq 4) - P(X \leq 1) \approx$$

   (i) **0.13** (ii) **0.23** (iii) **0.36**.

```
ppois(4,7.5) - ppois(1,7.5) # between Poisson cdfs, 3 x 2.5 = 7.5
```

```
[1] 0.1273606
```

3. *Poisson approximation of the binomial: photons.* What is the chance $x = 15$ particles hit in one microsecond? The Poisson distribution can be used to approximate the binomial distribution by letting $\lambda = np$. This is a fairly good approximation if $n \geq 100$ and $np \leq 100$.

(a) Assume $n = 2000$ particles are released by iron per microsecond, and there is a chance $p = 0.005$ that a particle hits the surrounding field per microsecond. Since $n = 2000$ and $p = 0.005$, then $n = 2000 > 100$ and $np = 2000(0.005) = 10 \leq 100$ and so the conditions for approximation are **satisfied** (ii) **violated**.

(Exact) binomial

$$f(15) = \frac{2000!}{15!(2000-15)!} \times (0.005)^{15} \times (0.995)^{1985} \approx$$

(i) **0.00234** (ii) **0.0346** (iii) **0.0445** (iv) **0.0645**.

```
dbinom(15,2000,0.005) # exact binomial pmf
```

```
[1] 0.03463059
```

Approximate Poisson.
since $\lambda = np = 2000(0.005) =$ (i) **5** (ii) **10** (iii) **15** (iv) **20**, then

$$f(15) = \frac{10^{15}}{15!} e^{-10} \approx$$

(i) **0.00234** (ii) **0.0347** (iii) **0.0445** (iv) **0.0645**

```
dpois(15,10) # approximate Poisson pmf
```

```
[1] 0.03471807
```

(b) If $n = 1500$ and $p = 0.01$, then $n = 1500 > 100$ and $np = 1500(0.01) = 15 \leq 100$ and so the conditions for approximation are **satisfied** (ii) **violated**.

(Exact) binomial

$$f(15) = \frac{1500!}{15!(1500-15)!} \times (0.01)^{15} \times (0.99)^{985} \approx$$

(i) **0.1024** (ii) **0.1030** (iii) **0.1245** (iv) **0.1345**.

```
dbinom(15,1500,0.01) # exact binomial pmf
```

```
[1] 0.1029519
```

Approximate Poisson
since $\lambda = np = 1500(0.01) =$ (i) **5** (ii) **10** (iii) **15** (iv) **20**, then

$$f(15) = \frac{15^{15}}{15!} e^{-15} \approx$$

(i) **0.1024** (ii) **0.1030** (iii) **0.1345** (iv) **0.1445**

```
dpois(15,15) # approximate Poisson pmf
```

```
[1] 0.1024359
```

4. *Poisson process: photons.* A random variable has a Poisson distribution if the Poisson process conditions are satisfied. For the photons example, this would mean the following assumptions are satisfied.

   - P(no "hit" occurs in time subinterval)$= 1 - p$
   - P(one "hit" occurs in time subinterval)$= p$
   - P(two or more "hits" occurs in time subinterval)$= 0$
   - Occurrence of hit in each time subinterval is independent of occurrence of event in other nonoverlapping subintervals.

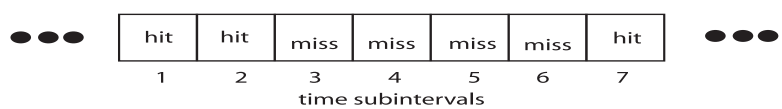

Figure 2.9: Poisson process: sequence of photon hits

   (a) The Poisson process assumes all time subintervals are created so small *one and only one* photon *could* hit the magnetic detection field during each subinterval. During seven time subintervals shown in Figure 2.9, three hits occur in time subintervals 1, 2 and 7 and four misses occur in time subintervals 3, 4, 5 and 6.
   (i) **True** (ii) **False**

   (b) The Poisson process assumes the chance a photon hits the magnetic detection field during any one time subinterval is p and the chance it misses is $q = 1 - p$. It is impossible to have more than one photon hit the magnetic detection field during any time subinterval. Consequently, the time subintervals are *infinitesimally small*.
   (i) **True** (ii) **False**

   (c) The Poisson process assumes a hit in each time subinterval is independent of any other hit in other nonoverlapping subintervals.
   (i) **True** (ii) **False**

   (d) If there was a 20% chance, $p = 0.2$, a photon hit the magnetic detection field during any time subinterval then the chance of $y = 3$ hits in the $n = 7$ time subintervals is given by the *binomial*,

$$f(3) = \left(\begin{array}{c} n \\ x \end{array} \right) p^x q^{n-x} = \left(\begin{array}{c} 7 \\ 3 \end{array} \right) 0.2^3 0.8^{7-3} \approx$$

   (i) **0.089** (ii) **0.115** (iii) **0.124** (iv) **0.134**.

```
dbinom(3,7,0.2) # binomial pmf
```

```
[1] 0.114688
```

(e) If a large number of time subintervals are considered; in other words, as, n gets bigger, in fact, as $n \to \infty$, it becomes increasingly difficult to calculate the probability of a number of hits using the binomial. In this case, the poisson is used instead.
(i) **True** (ii) **False**

(f) In general,

$$\lim_{n\to\infty} \binom{n}{x} p^x q^{n-x} = \frac{\lambda^x}{x!}e^{-\lambda}$$

where $\lambda = np$.
(i) **True** (ii) **False**

5. *Poisson: stars on tiles.*
Twenty-four (24) stars are placed on a tiled floor with 90 tiles and it is found, for example, 70 of tiles had no (0) stars, 17 tiles had 1 star and so on. Does this distribution of stars approximately follow a Poisson distribution?

x	0	1	2	3
frequency	70	17	2	1

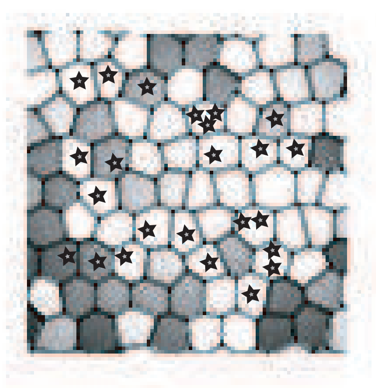

Figure 2.10: Stars on tiles

(a) Fill in the blank for the relative frequency of stars per tile table:

x	0	1	2	3
frequency	70	17	2	1
relative frequency	$\frac{70}{90} \approx 0.78$	0.19	0.02	_

(b) Average number of stars per tile is

$$\lambda = \frac{24}{90}$$

(i) **0.21** (ii) **0.24** (iii) **0.27** (iv) **0.29**.

(c) Assuming Poisson where $\lambda = 0.27$, chance $x = 0$ stars on one tile is

$$f(0) = \frac{\lambda^x}{x!}e^{-\lambda} = \frac{0.27^0}{0!}e^{-0.27} \approx$$

(i) **0.01** (ii) **0.02** (iii) **0.19** (iv) **0.76**.

```
dpois(0,0.27) # Poisson pmf
```

```
[1] 0.7633795
```

(d) Fill in the blank for the Poisson pmf $f(x)$ of stars per tile table:

x	0	1	2	3
frequency	70	17	2	1
relative frequency	$\frac{70}{90} \approx 0.78$	0.19	0.02	0.01
$f(x)$	0.76	____	0.03	0.00

The Poisson pmf (i) **is** (ii) **is not** a close approximation to the relative frequency of stars per tiles.

```
dpois(0:3,0.27) # Poisson pmf for 0,1,2,3
```

```
[1] 0.763379494 0.206112463 0.027825183 0.002504266
```