Example: *Alternative calculation of mean and variance of binomial distribution*

A r.v. X *has the Bernoulli distribution if it takes the values* 1 *('success') or* 0 *('failure') with probabilities* p *and* $(1-p)$*. This implies that if* $X \sim$ *Bernoulli*(p) *its probability function is given by*

$$f_X(x) = P(X = x) = p^x(1-p)^{1-x}, \qquad x = 0, 1. \tag{14}$$

It can be shown that if $X_1, X_2, \ldots, X_n$ *are independent r.v.'s with* $X_i \sim$ *Bernoulli*(p)*, for* $i = 1, 2, \ldots, n$*, then the r.v.*

$$Y = X_1 + X_2 + ... + X_n \sim \textit{Bin}(n, p).$$

Based on this property, derive the mean and variance of the binomial distribution.

Solution ...

4 The distribution function of a discrete r.v.

We can also determine the distribution of a r.v. X by using its (cumulative) distribution function (CDF), defined as

$$F_X(x) = P(X \leq x) = \sum_{t \leq x} f_X(t), \text{ for all } x \in S_x. \tag{15}$$

The graph of the CDF is a step function:

Properties of the CDF

(i) $\lim_{x \to -\infty} F_X(x) = 0$

(ii) $\lim_{x \to \infty} F_X(x) = 1$

(iii) if $a < b$, then $F_X(a) \leq F_x(b)$, for any real a, b.

Also (theorem):

If $S_x = \{x_1, x_2, \ldots, x_n\}$ with $x_1 < x_2 < \ldots < x_n$, then

$f_X(x_1) = F_X(x_1)$,

$f_X(x_i) = F_X(x_i) - F_X(x_{i-1}), \qquad i = 2, 3, \ldots, n$.

Example

Determine the CDF of the r.v. X if

(a) $X \sim$ Geometric(p)

(b) $X \sim$ Bin(n, p)

Solution ...

4.1 Use of Statistical Tables

The values of the CDF of various distributions are tabulated in Statistical Tables (e.g. *Lindley & Scott, New Cambridge Statistical Tables*). With discrete r.v.'s care should be taken with inequalities, e.g. notice that $P(X \leq 5) = P(X < 6)$.

In general we can use Tables to calculate the probability of X lying between specified bounds:

$$\begin{aligned} P(a \leq X \leq b) &= P(a-1 < X \leq b) \\ &= \sum_{x=a}^{b} f_X(x) = \sum_{x=-\infty}^{b} f_X(x) - \sum_{x=-\infty}^{a-1} f_X(x) \\ &= F_X(b) - F_X(a-1). \end{aligned}$$

For example, the values of $F_X(x) = P(X \leq x)$ of a r.v. $X \sim \text{Bin}(n, p)$ can be obtained for various x, n and p.

Notice however, that most Tables provide the CDF of the binomial distribution only for $p \leq 0.5$!

To compute the CDF of $X \sim$ Bin(n, p) with $p > 0.5$ we can work as follows:

Let $Y = n - X$ [This is the no. of ‘failures’]

Then $Y \sim$ Bin$(n, 1 - p)$, and

$$\begin{aligned} P(X \leq x) &= 1 - P(X > x) \\ &= 1 - P(Y < n - x) \\ &= 1 - P(Y \leq n - x - 1) \\ &= 1 - F_Y(n - x - 1) \end{aligned}$$

Notice that $1 - p < 0.5$, so $F_Y(n - x - 1)$ can be obtained by Tables.

Example: *(Vehicle pollution in Leith)*

The 'Edinburgh Herald and Post' reported in October 1996 that officials from the Department of Transport's Vehicle Inspectorate had tested the amount of carbon monoxide in exhaust emissions from 16 *vehicles in a spot check in Leith.*

What is the probability that 13 *or more vehicles passed the check, if it is believed that the average failure rate for these checks was* 7% *that year.*

Solution ...

An important property

If X, Y are r.v.'s such that $F_X(x) = F_Y(x)$ for all x, then X and Y have the same distribution.

Example

Let the r.v.'s X, Y be independently and identically distributed (i.i.d.) as Geometric(p). Find the distribution of

$$Z = \min\{X, Y\}.$$

Solution

The CDF of the r.v. Z is given for $z = 1, 2, 3, \ldots$, by

$$\begin{aligned} F_Z(z) = P(Z \le z) &= \\ &= \\ &= \end{aligned}$$

Now (from a previous example)

$$P(X > z) = P(Y > z) = (1 - p)^z.$$

Thus,

$$
\begin{aligned}
F_Z(z) &= \\
&= \\
&=
\end{aligned}
$$

This is the CDF of a Geometric(p') r.v. where

$$p' = 1 - q' = 1 - (1-p)^2.$$

Example: *A r.v. with infinite expectation*

Consider the following experiment:

An urn initially contains 1 *black ball (B) and* 1 *white ball (W). The following procedure takes place:*

(i) *a ball is randomly drawn from the urn;*

(ii) *if ball is (B) then stop; if (W), the ball is repalced in the urn along with an additional (W). Then go to step (i).*

Let X *be the number of draws before stopping (i.e. until the first (B) is drawn). We want the probability function and the expectation of the r.v.* X.

Solution:

The probability function is given by

$$f_X(x) = P(X = x) = P(\text{1st B on } x^{th}\text{draw}) = P(\underbrace{\text{W, W, ..., W}}, \text{B})$$

For

$$x = 1 \qquad f_X(1) \quad =$$

$$x = 2 \qquad f_X(2) \quad =$$

$$=$$

$$x = 3 \qquad f_X(3) \quad =$$

$$\vdots$$

In general

$$x = n \qquad f_X(n) \quad =$$

$$=$$

Therefore we have

$$f_x(x) = \frac{1}{x(x+1)}, \qquad x = 1, 2, 3, \ldots$$

We can verify that the probabilities sum up to 1:

$$\begin{aligned}
\sum_{x=1}^{n} f_x(x) &= \sum_{x=1}^{n} \frac{1}{x(x+1)} \\
&= \\
&= \\
&=
\end{aligned}$$

In the limit as $n \to \infty$, $\sum_{x=1}^{\infty} f_x(x) = 1$.

Now

$$
\begin{aligned}
E(X) &= \sum_{x=1}^{\infty} x f_x(x) \\
&= \\
&=
\end{aligned}
$$

The experiment set-up of this example is known as Polya's urn (also Polya's distribution; see tutorial sheet 1).

5 A note on mathematical induction

Induction is a technique that can be useful for carrying out calculations in probability theory. It works as follows. Suppose that we have a mathematical fact or identity that we wish to prove for all $n = 1, 2, 3, \ldots$, e.g.

$$\sum_{i=1}^{n} i = 1 + 2 + 3 + \ldots + n = \frac{n(n+1)}{2}$$

First we identify that the identity holds for $n = 1$.

Next we assume that the identity holds for all $n \leq k$ for some k (this assumption is called the induction hypothesis - I.H.). Then we show that this implies it must also hold for $n = k + 1$.

Suppose

$\sum_{i=1}^{n} i = 1 + 2 + 3 + \ldots + n = \frac{n(n+1)}{2}$ for all $n \leq k$.

Then

$$\sum_{i=1}^{k+1} i \;=\; \sum_{i=1}^{k} i + (k+1) = \frac{k(k+1)}{2} + (k+1) \qquad \text{by I.H.}$$

$$= \frac{k(k+1)+2(k+1)}{2} = \frac{(k+1)(k+2)}{2}.$$

It follows that identity holds for $n = k + 1$. We already checked that it holds for $n = 1$, so that it holds for $n = 2, \Rightarrow n = 3$, and so on ...

Example: *Probability theory*

Let X denote the number of successes in a series of n independent trials (each with probability p of success). Show that

$$f_X(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x},$$

for $x = 0, 1, 2, 3, \ldots, n$.

Solution

First check that it's true for $n = 1$:

$$P(X = 0) = 1 - p \qquad P(X = 1) = p.$$

Now suppose that it's true for $n \leq k$. Consider $n = k + 1$ and let Y be the number of successes scored on the first k trials. In this case

$$P(X = 0) = P(Y = 0) \times P(F \text{ on } (k+1)^{st}\text{trial}) = (1 - p)^k (1 - p) = (1 - p)^{k+1}$$

$$P(X = k + 1) = P(Y = k) \times P(S \text{ on } (k+1)^{st}\text{trial}) = p^k p = p^{k+1}.$$

For $x = 1, 2, 3, \ldots, k$ we have

$$\begin{aligned}
P(X = x) &= P(Y = x-1)P(S \text{ on } (k+1)^{th} \text{ trial}) \\
&\quad + P(Y = x)P(F \text{ on } (k+1)^{th} \text{ trial}) \\
&= \binom{k}{x-1} p^{x-1}(1-p)^{k-x+1} p + \binom{k}{x} p^{x}(1-p)^{k-x}(1-p) \\
&= \left[\binom{k}{x-1} + \binom{k}{x}\right] p^{x}(1-p)^{k+1-x} \\
&= \binom{k+1}{x} p^{x}(1-p)^{k+1-x}.
\end{aligned}$$

Therefore the result holds for $n = k + 1$ and hence for all $n = 1, 2, 3, \ldots$.

Part (B): Inference and continuous random variables

6 Statistical Inference

6.1 Introduction

Entire populations of 'subjects' of interest cannot always be examined. In case where only a sample, i.e. a subset of the population is examined, the aim is still to investigate the properties of the entire population.

Statistical inference is the operation through which information provided by the sample data is used to draw conclusions about the characteristics of the population which the sample comes from.

Mechanisms that generate the data are subject to uncertainty (random variation) and therefore we need to use probability models (distributions) to describe them.

The idea of inference is to use the available observations of some random phenomenon (data), to refine our understanding of the phenomenon. A suitable probability model will help us deal with the involved randomness in a mathematically appropriate manner.

We illustrate this with a simple example.

Example:

You decide to open a coconut-shy at your local fun fair. Contestants are given 10 balls to throw. If they hit the coconut 3 or more times they win a coconut (value $\pounds 1$*). If they hit it 10 times they win a Vauxhall Astra (value* $\pounds 10000$*). You charge each contestant* $\pounds 5$ *to play.*

The first day draws 100 contestants, who achieve scores as follows:

Score (x) :	0	1	2	3	4	5	6	7	8	9	10
Frequency (f_x) :	1	4	7	21	29	23	14	1	0	0	0

Your net income for the day is

Despite your success you wonder whether $\pounds 5$ *is a sensible price to charge?*

Solution

To answer this statistically, we need a statistical model for the score achieved. We treat a contestant's score X as a random variable and propose a suitable distribution for it.

Since our game (experiment) involves counting hits (successes) in 10 throws (trials), we assume that

$$X \sim \text{Bin}(10, p)$$

for some p. [Note that we assume throws are independent.]

Now let $Y = g(X)$ be your net profit for a single contestant. Then

$$g(x) =$$

To decide if your pricing strategy is sensible you consider the expected net profit, $E(Y)$, since this should approximate your average profit over a large number of contestants. We have

$$
\begin{aligned}
E(Y) &= \sum_{x=0}^{10} g(x) f_X(x) = \sum_{x=0}^{10} g(x) \binom{10}{x} p^x (1-p)^{10-x} \\
&=
\end{aligned}
$$

So, our expression for the expected net profit depends on p, the unknown probability of a 'hit' in a single throw, and thus we cannot calculate $E(Y)$ explicitly!

However, we do have the data (the scores of the 100 contestants from day 1) that we can use to estimate p.

[Notice that in this example our 'population' is all possible contestants in the game and the characteristic of interest (our random variable) is the score of hits out of 10 throws. The 'uncertainty' of this score is expressed through the binomial distribution and it depends on an unknown quantity (parameter), the probability p. We want to use a sample of 100 contestants to estimate the unknown population characteristic p (and then X, Y).]

6.2 Parameter estimation using the method of moments

Consider again the problem introduced in the last example. We assume that the random scores of the 100 contestants $X_1, X_2, \ldots, X_{100}$ satisfy:

- $X_i \sim \text{Bin}(10, p), \qquad i = 1, 2, \ldots, 100$
- $X_i \& X_j$ are independent for $i \neq j$.

Then the observed scores $x_1, x_2, \ldots, x_{100}$ are realisations of the r.v. X, i.e. the realised (sampled) values in a random sample of size 100 from a Bin$(10, p)$ distribution. We want to estimate p.

The motivation behind the method of moments estimation is that we expect the average of a large number of independent draws from a Bin(n, p) distribution to be $E(X) = np$. Therefore a sensible 'guess' for p, $\hat{p}$, would satisfy

$$n\hat{p} = \bar{x} \qquad \text{[Notice that } n \text{ refers to \# trials, NOT sample size]}$$

where $\bar{x}$ is the sample mean, given as

$$\bar{x} = \frac{1}{100}\sum_{i=1}^{100} x_i.$$

Then,

$$\hat{p} = \frac{\bar{x}}{n} = \frac{\bar{x}}{10} \text{ [Since } n = 10 \text{ here]}$$

[In effect we select $\hat{p}$ so that it specifies a distribution whose mean matches the average in our sample.

Note that other alternatives are possible, e.g. choose $\hat{p}$ to match the distribution variance to the sample variance.]

Method of moments

In general if $x_1, x_2, \ldots, x_N$ are the observed values in a random sample from a distribution specified by an unknown parameter α, with mean (depending on α)

$$E(X) = \mu(\alpha)$$

then the method of moments estimate of α is obtained by solving

$$\mu(\alpha) = \frac{1}{N} \sum x_i$$

for α.

Note on terminology

If X is a r.v. then $E(X^r)$ is the r^{th} (simple) moment of X.

More generally, we can obtain a method-of-moments estimate if we express a parameter of interest as a function of a population moments, and then set this expression equal to the same function of the sample moments. For example, notice that in the binomial case

$$\text{var}(X) = np(1-p)$$

and if we can find an equivalent expression for the sample variance, $[s^2 = \frac{1}{N-1}\sum(x_i - \bar{x})^2]$, then we can solve for p to obtain a method of moments estimate.

Let's now return to our fun fair example and apply the method to the data. Recall:

Score (x) :	0	1	2	3	4	5	6	7	8	9	10
Frequency (f_x) :	1	4	7	21	29	23	14	1	0	0	0

The sample mean is given by

Therefore we choose $\hat{p}$ such that

What is the expected net profit for $p = 0.4$?

Recall:

$$E(Y) = 5 \times F_X(2) + 4 \times \{F_X(9) - F_X(2)\} - 9996 \times p^{10}$$

where F_X is now the CDF of the Bin$(10, 0.4)$ distribution. From Tables (e.g. *Lindley & Scott*, p.9) we find:

$$\begin{aligned} E(Y) &= \\ &= \end{aligned}$$

Your pricing strategy should yield a healthy profit on average.

Some caveats

1. The assumption of the binomial model is central to our results. Remember we have assumed that $X_1, X_2, \ldots, X_{100} \sim \text{Bin}(n, p)$.

   How realistic is this assumption?

   - p might vary between contestants
   - p might change between first and last attempts.

   We will return to this in a while.

2. Assuming that the binomial model is correct, how well does our estimate $\hat{p} = 0.4$ reflect the true value of p?

   A different sample of 100 would produce a different estimate! The true value of p might be smaller or larger than 0.4.

   An estimate is not very useful unless we have some indication of how precise it is. We can identify a range of plausible values for p (later in this module).

Note on terminology

A single estimate of a parameter (e.g. $\hat{p} = 0.4$ in the example) is referred to as a point estimate of the parameter.

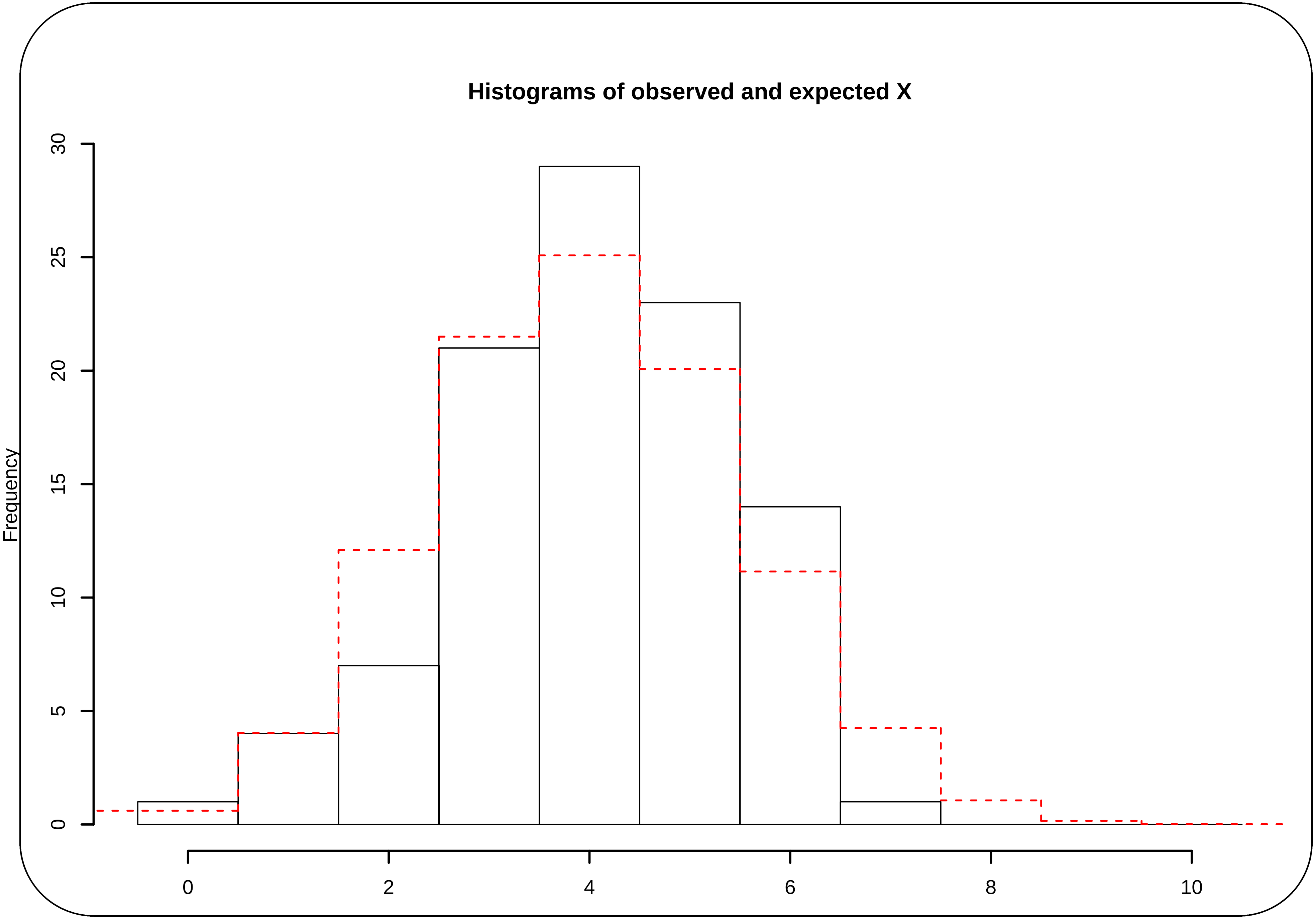
If instead we specify a range of values in which the parameter is thought to lie, this is called a confidence interval for the parameter (interval estimation).

Let's now return to the first issue above. How could we check our assumption that $X \sim \text{Bin}(10, 0.4)$?

We can compare the expected frequencies for each score under our assumption, with the observed frequencies.

Score (x) :	0	1	2	3	4	5
$P(X = x)$:						
Observed (f_x) :	1	4	7	21	29	23
Expected (E_x) :						

Score (x) :	6	7	8	9	10
$P(X = x)$:					
Observed (f_x) :	14	1	0	0	0
Expected (E_x) :					



Histograms of observed and expected X
Frequency
0
5
10
15
20
25
30
0
2
4
6
8
10
x

Things to look out for:

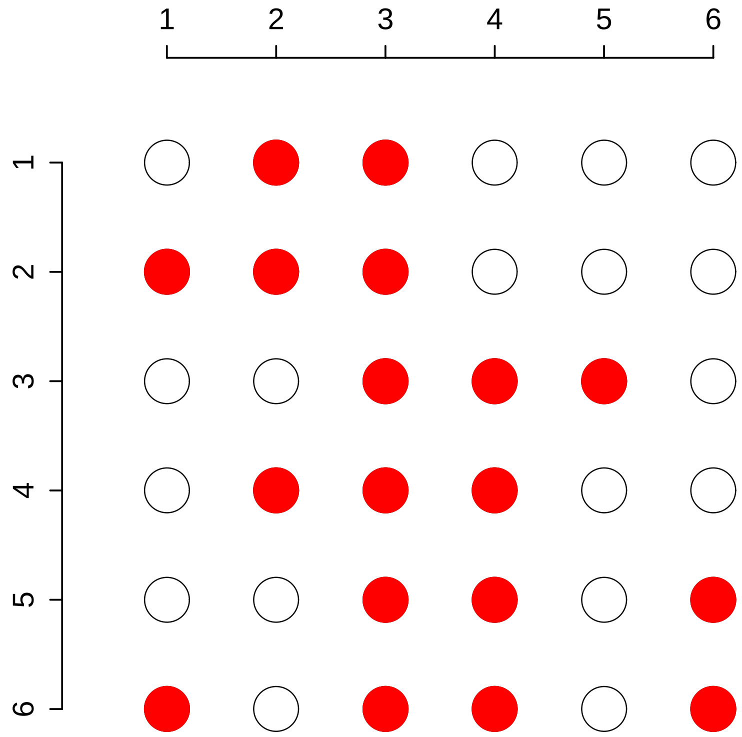
- Over-dispersion: too many extreme observations compared with expected numbers.
- Under-dispersion: not enough extreme observations.

By inspection we have reasonable agreement between observed and expected frequencies. Therefore the Bin$(10, 0.4)$ assumption seems plausible.

[In later courses on data analysis you will learn how to carry out proper statistical tests of assumptions in such situations.]

6.3 Application: Analysis of plant disease data using probability distributions

In an agricultural experiment a population of plants in a field plot are observed during the growing season and the positions of diseased plants noted.



Question: Is the pattern of disease aggregated? i.e. Do diseased individuals tend to be located near to each other? (NB: Red circle = 'diseased plant')

We can investigate this question using probability distributions as follows.

We model the disease status of the individual at position (i, j) as a Bernoulli(p) random variable.

$$I_{ij} = \begin{cases} 1, & \text{if } (i,j) \text{ is diseased;} \\ 0, & \text{if } (i,j) \text{ is healthy;} \end{cases}$$

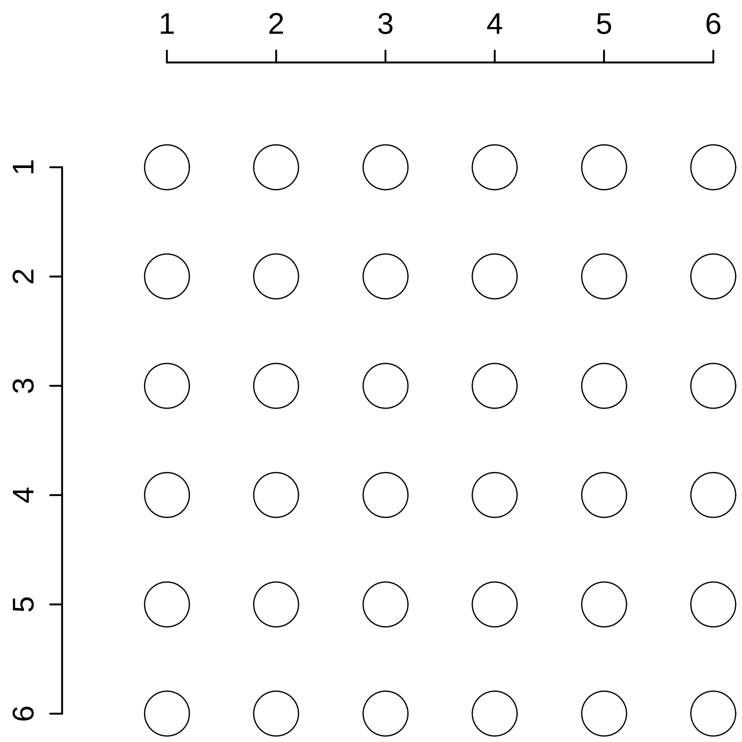
Now if there is <u>no tendency</u> for closely located plants to have similar disease status, then we could propose that the I_{ij} are all independent of each other. Therefore we could regard $\{I_{ij} | 1 \leq i \leq 6, 1 \leq j \leq 6\}$ as a random sample of size 36 from a Bernoulli(p) distribution. It follows that

$$X = \sum_{1 \leq i,j \leq 6} I_{ij} \sim \text{Bin}(36, p).$$

For our particular case $X = 18$ and we can calculate the method of moments estimate for p to be

$$\hat{p} = \frac{18}{36} = \frac{1}{2}.$$

Let us now test the assumption that the I_{ij} are identically independently distributed with distribution Bernoulli$(1/2)$. To do this we follow a continuous path through the 6×6 matrix of sites, consider the random variables I_{ij} in the order encountered and relabel $I_1, I_2, \ldots, I_{36}$.



Now consider those values j for which $I_j = 1$ (i.e. the plant is diseased):

$$j_1 \leq j_2 \leq \ldots \leq j_{18}.$$

If the $I's$ really are i.i.d. Bernoulli$(1/2)$ then

$$j_1, j_2 - j_1, j_3 - j_2, \ldots, j_{k+1} - j_k, \ldots$$

should be i.i.d. random sample of values from a Geometric$(1/2)$ distribution. Therefore we now look at frequency table of these 18 values for our field data to see whether it looks consistent with a Geometric$(1/2)$ distribution.

x :	1	2	3	4	5	6	7
Observed :	10	4	1	2	0	0	1
Expected :	9	4.5	2.25	1.125	0.56	0.28	0.14

[Expected number in category x is $18(1-p)^{x-1}p$, where $p = \frac{1}{2}$.]

There does not seem to be strong evidence that the disease pattern is aggregated since the observed frequencies are not too dissimilar to the expected values. This method of detecting aggregation is called <u>ordinary runs analysis</u>.

7 Continuous random variables

7.1 Introduction

Many characteristics or quantities of interest cannot be modelled with the use of discrete r.v.'s, as they are naturally measured on a continuous scale:

- lifetimes of electronic components
- duration of insurance policies
- birth weights of humans
- annual rainfall at a weather station.

These quantities can be represented using continuous random variables.

The CDF of a continuous r.v.

Recall that CDF of a discrete r.v X, $F_X(x) = P(X \leq x)$, was a step function with properties

(i) $\lim_{x \to -\infty} F_X(x) = 0$

(ii) $\lim_{x \to \infty} F_X(x) = 1$

(iii) F is non-decreasing.

Definition

A continuous r.v. is one whose CDF $F_X(x)$ is a continuous function of x satisfying the conditions (i)-(iii) above.

The graph of $F_X(x) = P(X \leq x)$ for a continuous r.v. X has the form

Just as we could for discrete r.v.'s, we can use the CDF of a continuous r.v. to calculate probabilities of the form

$$P(a < X \leq b) = F_X(b) - F_X(a)$$

7.2 Examples of continuous r.v.'s

7.2.1 The Uniform$(0, 1)$ distribution

This describes continuous r.v.'s that take values uniformly in the interval $(0, 1)$.

It can be viewed as the limit of a sequence of discrete r.v.'s. Let X_n denote the discrete r.v. whose distribution is uniform on $\{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}\}$, i.e.

$$f_{X_n}\left(\frac{i}{n}\right) = \frac{1}{n}, \qquad i = 1, 2, \ldots, n.$$

Now look at the graphs of the CDF's $F_{X_n}, n = 1, 2, 3, \ldots$.

In the limit as n gets very large ($n \to \infty$), we obtain

$$F_X(x) = \begin{cases} 0, & x \leq 0; \\ x, & 0 < x < 1; \\ 1, & x \geq 1 \end{cases}$$

This is the CDF of $X \sim$ Uniform$(0, 1)$.

For any a, b satisfying $0 < a < b < 1$, we have

$$P(a < X \leq b) = F_X(b) - F_X(a) = b - a$$

[i.e. the probability is equal to the length of the interval.]

Note that for a continuous r.v. X

$$P(a < X < b) = P(a \leq X \leq b)$$

[i.e. it doesn't matter if the inequalities are strict on not.]

This is because for a continuous r.v. X and any $a \in \mathbb{R}$:

$$P(X = a) = 0$$

[The intuition behind this is that as there are infinite possible single values that a continuous r.v. can take in any interval in its range, the probability of this happening is zero. We will see later how probabilities are defined for continuous r.v.'s.]

7.2.2 The Exponential(λ) distribution

The exponential distribution is used to describe random variables that take positive values and express waiting times between events e.g.

- emissions of radioactive substance
- appearance of meteors in the sky
- failures of mechanical/electrical equipment.

The CDF of a r.v. $T \sim$ Exp(λ) is

$$F_T(t) = 1 - e^{-\lambda t}, \qquad t > 0, \qquad \lambda > 0.$$

Let's now look at a special case, the Exp(1) distribution, and see how it can be thought as the limit of a sequence of discrete r.v.'s.

Consider an infinite sequence of independent trials occurring at times

$$\frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \ldots$$

with $P(\text{‘success’}) = \frac{1}{n}$.

Let T_n be the r.v. denoting the time of the first success. The range of T_n is

$$\{t_1, t_2, t_3, \ldots\}$$

where $t_i = \frac{i}{n}$.

Let's look at the CDF of T_n:

$$\begin{aligned} F_{T_n} &= P(T_n \leq t_i) = 1 - P(T_n > t_i) \\ &= 1 - P(\text{1st } i \text{ trials are failures}) \\ &= 1 - \left(1 - \frac{1}{n}\right)^i = 1 - \left(1 - \frac{1}{n}\right)^{nt_i} \qquad t_i = i/n \end{aligned}$$

Now, as $n \to \infty$, $\left(1 - \frac{1}{n}\right)^{nt_i} \to e^{-t_i}$.

In the limit we obtain a continuous r.v. T whose CDF is the continuous function

$$F_T(t) = \begin{cases} 1 - e^{-t}, & t > 0; \\ 0, & t \leq 0; \end{cases}$$

The graph of F_T is

7.2.3 Some properties of the exponential distribution

1. Relationship between Exp(λ) and Poisson

Suppose we observe events (e.g. phone calls arriving at a telephone exchange centre), such that the waiting times between successive events are i.i.d. Exp(λ). Let X be the r.v. denoting the number of events occurring in the space of one time unit (same unit as in waiting times).

Then $X \sim$ Poisson(λ).

Proof: not given in this course.

Recall that:

$$f_X(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, 3, \ldots \qquad \lambda > 0$$

$$E(X) = \lambda, \qquad \text{var}(X) = \lambda.$$

2. The re-scaling property

If $X \sim \text{Exp}(\lambda)$, then for $a > 0$

$$Y = aX \sim \text{Exp}\left(\frac{\lambda}{a}\right)$$

Proof ...

3. The lack-of-memory property

As with the geometric distribution, the memoryless property also holds for the exponential distribution.

Suppose $X \sim \text{Exp}(\lambda)$, then

$$P(X > s + t / X > s) = P(X > t)$$

Proof ...

What is the practical implication of this property?

Suppose that no event has occurred in the time interval $[0, s]$. Then the probability that we will have to wait at least for a further time unit (e.g. hour) does not depend on s!

8 The probability density function

For a discrete r.v. we defined the probability (mass) function (p.m.f.) as

$$f_X(x) = P(X = x)$$

for all x in the range of X.

For a continuous r.v. X we have already seen that $P(X = x) = 0$, so we cannot use this to assign probabilities to different values of the r.v.

However, we can consider the probability of X taking values in a very small interval:

$$P(x < X < x + \Delta) = F_X(x + \Delta) - F_X(x)$$

for Δ very small. Now notice that as $\Delta \to 0$,

$$\lim_{\Delta \to 0} \frac{F_X(x + \Delta) - F_X(x)}{\Delta} \approx F'_X(x)$$

or

$$F_X(x + \Delta) - F_X(x) \approx F'_X(x)\Delta.$$

Therefore (for values of x where $F_X(x)$ is differentiable) the derivative $F'_X(x)$ quantifies how likely it is that X takes a value near x.

Definition

The probability density function (p.d.f.) of a continuous r.v. X is defined to be

$$f_X(x) = \lim_{\Delta \to 0} \frac{P(x < X < x + \Delta)}{\Delta} \tag{16}$$

or, if $F_X(x)$ is differentiable

$$f_X(x) = \frac{d}{dx} F_X(x).$$

The p.d.f. of a r.v. X must satisfy the following conditions:

(i) $f_X(x) \geq 0$

(ii) $P(a < X < b) = \int_a^b f_X(x)\, dx$

(iii) $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$

Also, the CDF is given as: $F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$

Intuitive meaning of a p.d.f.

Let X be a continuous r.v. with p.d.f. $f_X(x)$. Suppose we take a very large number, n, of independent realisations (samples) of X and form a histogram with narrow bin-width (Δ).

Suppose n_x samples fall in the bin centred on x. Then

$$\frac{n_x}{n} \approx P\left(x - \frac{\Delta}{2} < X < x + \frac{\Delta}{2}\right) \approx f_X(x)\Delta$$

$$\Rightarrow \frac{n_x}{n\Delta} \approx f_X(x)$$

We can therefore think of the p.d.f. as an 'idealised histogram' which would be formed if we took a sufficiently large number of realisations of X.

8.1 The p.d.f.'s of some distributions

The Uniform(a, b) distribution

Let $X \sim$ U(a, b). Then

CDF :

$$F_X(x) = \begin{cases} 0, & x \le a; \\ \frac{x-a}{b-a}, & a < x < b; \\ 1, & x \ge b \end{cases}$$

PDF :

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b; \\ 0, & \text{elsewhere.} \end{cases}$$

[Show]

The Exp(λ) distribution

Let $X \sim$ Exp(λ). Then

CDF :

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0; \\ 0, & x \leq 0 \end{cases}$$

PDF :

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0; \\ 0, & \text{elsewhere.} \end{cases}$$

[Show]

8.2 Transformations of continuous r.v.'s

Let X be a continuous r.v. with p.d.f. $f_X(x)$. Suppose we define a new r.v. Y such that

$$Y = g(X)$$

where $g()$ is a continuous and differentiable function.

Then, if g is also strictly monotonic, its pdf is given by

$$\begin{aligned} f_Y(y) &= f_X\{g^{-1}(y)\}\left|\frac{d}{dy}g^{-1}(y)\right| \\ &= f_X(x)\left|\frac{dx}{dy}\right| \end{aligned} \tag{17}$$

Proof ...

Example

If $X \sim$ *U*$(0, 1)$*, find the distribution of* $Y = g(X) = -\log(X)$*.*

Solution ...

Example: *(The re-scaling property of the exponential distribution revisited)*

If $X \sim$ *Exp*(λ)*, then find the distribution of* $Y = aX$*, for* $a > 0$*.*

Solution ...

9 Expectations of functions of continuous r.v.'s

We have already seen that for a discrete r.v.

$$E(X) = \sum_x x f_X(x)$$

where $f_X(x) = P(X = x)$ is the p.m.f. of X.

Now for a continuous r.v. X, we have

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx \tag{18}$$

where $f_X(x)$ is the p.d.f. of X.

[For the expectation above to exist, the integral must converge, i.e. $\int_{-\infty}^{\infty} |x| f_X(x)\, dx < \infty$.]

Also, for a function $g(\cdot)$ of X, we have

$$E\{g(X)\} = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$$

Similarly, for the variance of X we have

$$\begin{aligned} \text{var}(X) &= E\{X - E(X)\}^2 \\ &= \int_{-\infty}^{\infty} \{X - E(X)\}^2 f_X(x)\, dx \end{aligned}$$

It is often more convenient to calculate the variance as

$$\begin{aligned} \text{var}(X) &= E(X^2) - E\{X\}^2 \\ &= \int_{-\infty}^{\infty} x^2 f_X(x)\, dx - \left\{ \int_{-\infty}^{\infty} x f_X(x)\, dx \right\}^2. \end{aligned} \qquad (19)$$

9.1 Moments of the U(a, b) distribution

Recall that if $X \sim$ U(a, b), then $f_X(x) = \frac{1}{b-a}$, for $a < X < b$ (or 0 otherwise).

We have

$$E(X) = \frac{b+a}{2}, \qquad \text{var}(X) = \frac{(b-a)^2}{12}$$

Proof ...

9.2 Moments of the Exp(λ) distribution

The p.d.f. of $X \sim$ Exp(λ) is given by $f_X(x) = \lambda e^{-\lambda x}$, for $x > 0$.

Then

$$E(X) = \frac{1}{\lambda}, \qquad \text{var}(X) = \frac{1}{\lambda^2}$$

Proof ...