UNIVERSITY OF LONDON

Probability and Statistics: To p, or not to p?

Module Leader: Dr James Abdey

2.6 The distribution zoo

Suppose we carry out n **Bernoulli trials** such that:

- at each trial, the probability of success is π
- different trials are statistically independent events.

Let X denote the total number of successes in these n trials, then X follows a **binomial distribution** with parameters n and π, where $n \geq 1$ is a known integer and $0 \leq \pi \leq 1$. This is often written as:

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

If $X \sim \text{Bin}(n,\, \pi)$, then:

$$\text{E}(X) = n\,\pi.$$

Example

A multiple choice test has 4 questions, each with 4 possible answers. James is taking the test, but has no idea at all about the correct answers. So he guesses every answer and, therefore, has the probability of 1/4 of getting any individual question correct.

Let X denote the number of correct answers in James' test. X follows the binomial distribution with $n = 4$ and $\pi = 0.25$, i.e. we have:

$$X \sim \text{Bin}(4,\, 0.25).$$

For example, what is the probability that James gets 3 of the 4 questions correct?

Here it is assumed that the guesses are independent, and each has the probability $\pi = 0.25$ of being correct.

The probability of any particular sequence of 3 correct and 1 incorrect answers, for example 1110, is $\pi^3 (1-\pi)^1$, where '1' denotes a correct answer and '0' denotes an incorrect answer.

However, we do not care about the order of the 0s and 1s, only about the number of 1s. So 1101 and 1011, for example, also count as 3 correct answers. Each of these also has the probability $\pi^3 (1-\pi)^1$.

The total number of sequences with three 1s (and, therefore, one 0) is the number of locations for the three 1s which can be selected in the sequence of 4 answers. This is $\binom{4}{3} = 4$ (see below). Therefore, the probability of obtaining three 1s is:

$$\binom{4}{3} \pi^3 (1-\pi)^1 = 4 \times (0.25)^3 \times (0.75)^1 \approx 0.0469.$$

In general, the **probability function** of $X \sim \text{Bin}(n, \pi)$ is:

$$\boldsymbol{P(X = x) = \begin{cases} \binom{n}{x} \pi^x (1-\pi)^{n-x} & \text{for } x = 0, 1, \ldots, n \\ 0 & \text{otherwise} \end{cases}}$$

where $\binom{n}{x}$ is the **binomial coefficient** – in short, the number of ways of choosing x objects out of n when sampling without replacement when the order of the objects does not matter.

$\binom{n}{x}$ can be calculated as:

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

where $k! = k \times (k-1) \times \cdots \times 3 \times 2 \times 1$, for an integer $k > 0$. Also note that $0! = 1$. For example:

$$\binom{4}{3} = \frac{4!}{3!\,(4-3)!} = \frac{4!}{3!\,1!} = \frac{4 \times 3 \times 2 \times 1}{(3 \times 2 \times 1) \times 1} = \frac{24}{6 \times 1} = 4.$$

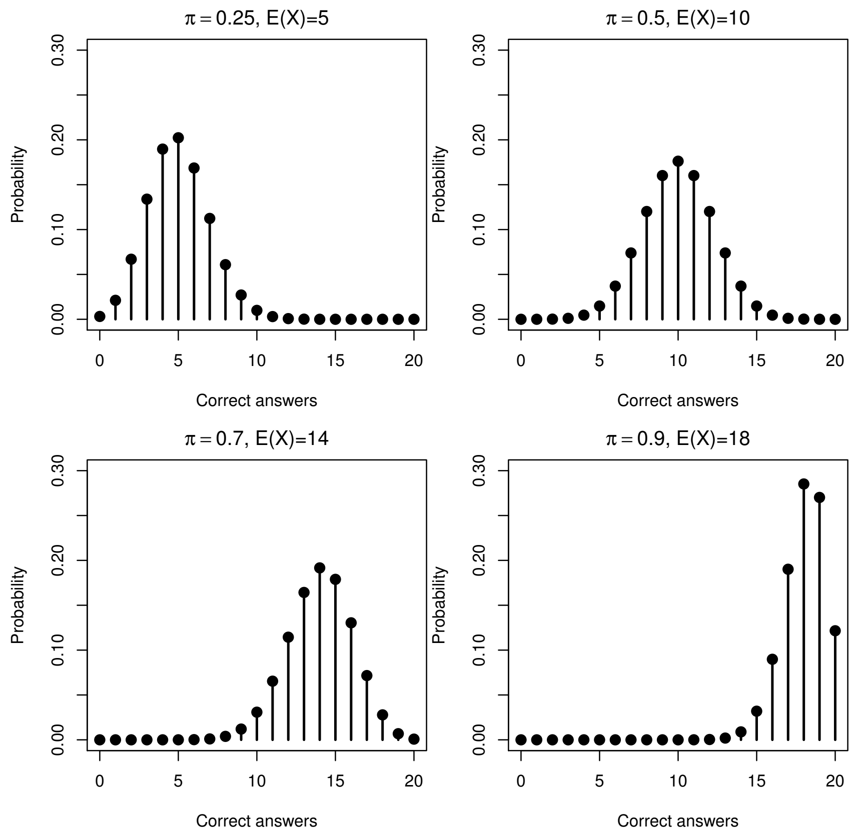
Example

Now assume there are $n = 20$ questions, each with 4 possible answers.

More generally, consider a student who has the same probability π of the correct answer for every question, so that $X \sim \text{Bin}(20, \pi)$.

The figure below shows plots of the probabilities for $\pi = 0.25$, 0.5, 0.7 and 0.9 (reflecting students of differing abilities, i.e. the better the student the more likely s/he is to get the answer correct and hence a higher π).

Note that as π increases, the probability of obtaining a large number of correct answers increases (and hence the probability of obtaining a small number of correct answers decreases) as we would expect because better-prepared students tend to score higher marks. Of course, there is an opportunity cost: it takes more time and effort to prepare, but this on average is rewarded with high marks!



Poisson distribution

The possible values of the **Poisson distribution** are the non-negative integers $0, 1, 2, \ldots$.

The **probability function** of the Poisson distribution is:

$$P(X = x) = \begin{cases} e^{-\lambda}\,\lambda^x/x! & \text{for } x = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

where $\lambda > 0$ is a parameter, e is the exponential constant ($e \approx 2.71828$) and $x!$ is **'x factorial'**, defined earlier as:

$$x! = x \times x - 1 \times x - 2 \times \cdots \times 3 \times 2 \times 1.$$

If a random variable X has a Poisson distribution with parameter λ, this is often denoted by:

$$X \sim \text{Poisson}(\lambda) \quad \text{or} \quad X \sim \text{Pois}(\lambda).$$

If $X \sim \text{Poisson}(\lambda)$, then:

$$E(X) = \lambda$$

Poisson distributions are used for *counts* of occurrences of various kinds. To give a formal motivation, suppose that we consider the number of occurrences of some phenomenon in time, and that the process which generates the occurrences satisfies the following conditions:

1. The numbers of occurrences in any two *disjoint* intervals of time are independent of each other.

2. The probability of two or more occurrences at the *same* time is negligibly small.

3. The probability of one occurrence in any short time interval of length t is λt for some constant $\lambda > 0$.

In essence, these state that individual occurrences should be independent, sufficiently rare, and happen at a constant rate λ per unit of time. A process like this is a **Poisson process**.

If occurrences are generated by a Poisson process, then the number of occurrences in a randomly selected time interval of length $t = 1$, X, follows a Poisson distribution with mean λ, i.e. $X \sim \text{Poisson}(\lambda)$.

The single parameter λ of the Poisson distribution is, therefore, the *rate* of occurrences per unit of time.

Example

Examples of variables for which we might use a Poisson distribution include the following.

- The number of telephone calls received at a call centre *per* minute.
- The number of accidents on a stretch of motorway *per* week.
- The number of customers arriving at a checkout *per* minute.
- The number of misprints *per* page of newsprint.

Because λ is the rate *per* unit of time, its value also depends on the unit of time (that is, the length of interval) we consider.

Example

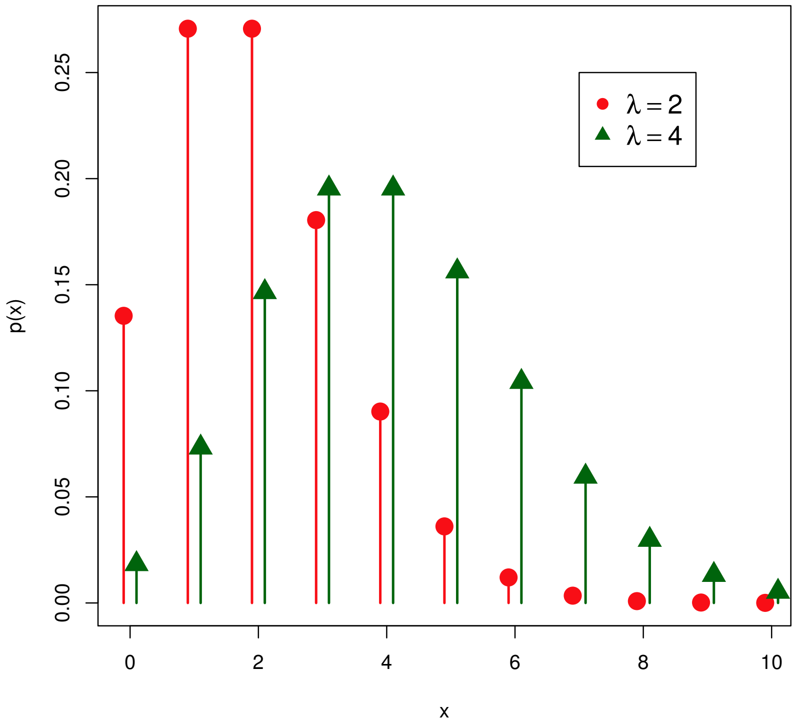
If X is the number of arrivals per hour and $X \sim \text{Poisson}(1.5)$, then if Y is the number of arrivals per *two* hours, $Y \sim \text{Poisson}(2 \times 1.5) = \text{Poisson}(3)$.

λ is also the mean of the distribution, i.e. $\text{E}(X) = \lambda$.

Both motivations suggest that distributions with higher values of λ have higher probabilities of large values of X.

Example

The figure below shows the probabilities $P(X = x)$ for $x = 0, 1, 2, \ldots, 10$ for $X \sim \text{Poisson}(2)$ and $X \sim \text{Poisson}(4)$.



Example

Customers arrive at a bank on weekday afternoons randomly at an average rate of 1.6 customers per minute. Let X denote the number of arrivals per minute and Y denote the number of arrivals per 5 minutes.

We assume a Poisson distribution for both, such that:

$$X \sim \text{Poisson}(1.6) \quad \text{and} \quad Y \sim \text{Poisson}(5 \times 1.6) = \text{Poisson}(8).$$

1. What is the probability that no customer arrives in a one-minute interval?

   For $X \sim \text{Poisson}(1.6)$, the probability $P(X = 0)$ is:

   $$P(X = 0) = \frac{\mathrm{e}^{-\lambda}\,\lambda^0}{0!} = \frac{\mathrm{e}^{-1.6}\,(1.6)^0}{0!} = \mathrm{e}^{-1.6} = 0.2019.$$

2. What is the probability that more than two customers arrive in a one-minute interval?

   $P(X > 2) = 1 - P(X \leq 2) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$ which is:

$$1 - \frac{e^{-1.6}\,(1.6)^0}{0!} - \frac{e^{-1.6}\,(1.6)^1}{1!} - \frac{e^{-1.6}\,(1.6)^2}{2!} = 1 - e^{-1.6} - 1.6e^{-1.6} - 1.28e^{-1.6}$$
$$= 1 - 3.88e^{-1.6}$$
$$= 0.2167.$$

3. What is the probability that no more than 1 customer arrives in a five-minute interval?

   For $Y \sim \text{Poisson}(8)$, the probability $P(Y \leq 1)$ is:

$$P(Y = 0) + P(Y = 1) = \frac{e^{-8}\,(8)^0}{0!} + \frac{e^{-8}\,(8)^1}{1!} = e^{-8} + 8e^{-8} = 9e^{-8} = 0.0030.$$

Connections between probability distributions

There are close connections between some probability distributions, even across different families of them. Some connections are **exact**, i.e. one distribution is exactly equal to another, for particular values of the parameters. For example, Bernoulli(π) is the same distribution as Bin(1, π).

Some connections are **approximate** (or **asymptotic**), i.e. one distribution is closely approximated by another under some limiting conditions. We next discuss one of these, the Poisson approximation of the binomial distribution.

Poisson approximation of the binomial distribution

Suppose that:

- $X \sim \text{Bin}(n,\, \pi)$.
- n is large and π is small.

Under such circumstances, the distribution of X is well-approximated by a Poisson(λ) distribution with $\lambda = n\,\pi$.

The connection is exact at the limit, i.e. $\text{Bin}(n,\, \pi) \to \text{Poisson}(\lambda)$ if $n \to \infty$ and $\pi \to 0$ in such a way that $n\,\pi = \lambda$ remains constant.

This **'law of small numbers'** provides another motivation for the Poisson distribution.

Example – Bortkiewicz's horses

A classic example (from Bortkiewicz (1898) *Das Gesetz der kleinen Zahlen*) helps to remember the key elements of the 'law of small numbers'.

Below shows the numbers of soldiers killed by horsekick in each of 14 Army Corps of the Prussian army in each of the years spanning 1875–94.

	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94
G	—	2	2	1	—	—	1	1	—	3	—	2	1	—	—	1	—	1	—	1
I	—	—	—	2	—	3	—	2	—	—	—	1	1	1	—	2	—	3	1	—
II	—	—	—	2	—	2	—	—	1	1	—	—	2	1	1	—	—	2	—	—
III	—	—	—	1	1	1	2	—	2	—	—	—	1	—	1	2	1	—	—	—
IV	—	1	—	1	1	1	1	—	—	—	—	1	—	—	—	—	1	1	—	—
V	—	—	—	—	2	1	—	—	1	—	—	1	—	1	1	1	1	1	1	—
VI	—	—	1	—	2	—	—	1	2	—	1	1	3	1	1	1	—	3	—	—
VII	1	—	1	—	—	—	1	—	1	1	—	—	2	—	—	2	1	—	2	—
VIII	1	—	—	—	1	—	—	1	—	—	—	—	1	—	—	—	1	1	—	1
IX	—	—	—	—	—	2	1	1	1	—	2	1	1	—	1	2	—	1	—	—
X	—	—	1	1	—	1	—	2	—	2	—	—	—	—	2	1	3	—	1	1
XI	—	—	—	—	2	4	—	1	3	—	1	1	1	1	2	1	3	1	3	1
XIV	1	1	2	1	1	3	—	4	—	1	—	3	2	1	—	2	1	1	—	—
XV	—	1	—	—	—	—	—	1	—	1	1	—	—	—	2	2	—	—	—	—

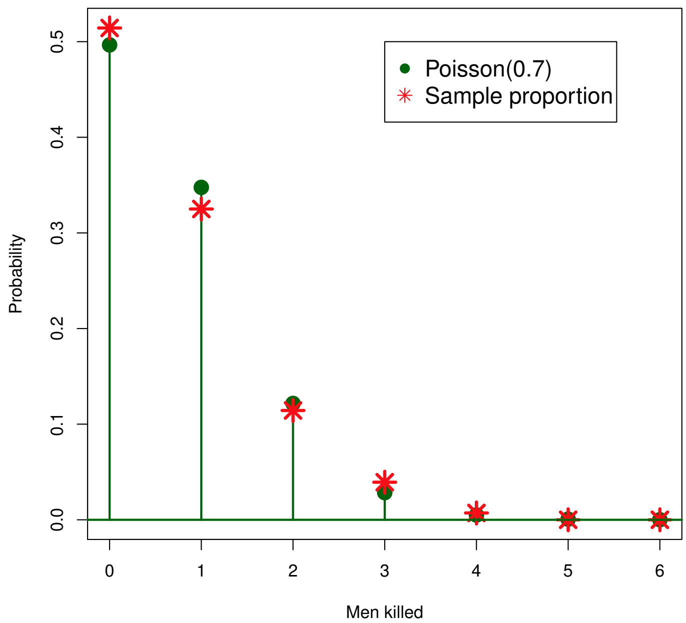
Suppose that the number of men killed by horsekicks in one corps in one year is $X \sim \text{Bin}(n, \pi)$, where:

- n is large – the number of men in a corps (perhaps 50,000)
- π is small – the probability that a man is killed by a horsekick.

X should be well-approximated by a Poisson distribution with some mean λ. The sample frequencies and proportions of different counts are as follows:

Number killed	0	1	2	3	4	More
Count	144	91	32	11	2	0
%	51.4	32.5	11.4	3.9	0.7	0

The sample mean (formally introduced in week 3) of the counts is $\bar{x} = 0.7$, which we use as λ for the Poisson distribution. $X \sim \text{Poisson}(0.7)$ is indeed a good fit to the data, as shown below.



This is an excellent example where we can use a known probability distribution (here, the Poisson distribution) to *model* some real-world phenomenon (here, deaths by horsekick). Recall that although we care about reality, models are *not equal to* reality:

$$\textbf{model} \neq \textbf{reality.}$$

However, a *good* model is one which provides a *very close approximation* to reality:

$$\textbf{good model} \approx \textbf{reality.}$$