Discrete probability distributions

Dr Gavin M Abernethy

Contents

Today we will cover. . .

- Discrete probability distributions.
- The Binomial distribution.
- The Poisson distribution.
- Poisson processes and applications to failure rates.

Discrete versus continuous probability distributions

The normal/Gaussian and log-normal distributions are examples of **continuous probability distributions** as they describe the probable values of a continuous random variable (e.g. height, length, mass, temperature).

A continuous random variable, such as the height h of a student, can take any value in some interval. Thus we have been interested in the probability that it falls within a particular range $P(a < h < b)$ using a continuous probability distribution.

Discrete versus continuous probability distributions

There are also **discrete probability distributions** that show the probability of a discrete random variable (e.g. number of customers, number of vehicles passing over a bridge, number of mechanical failures of machinery) taking a particular value. These can be useful if we need the probability of whether an event occurs or not, or if we wanted to know *how many times* a discrete event is likely to occur.

A discrete random variable, such as the number of students X in a class, can only take values from some discrete set. A discrete probability distribution would therefore tell us the likelihood of X taking specific discrete values:

$$P(X = 0), \quad P(X = 1), \quad P(X = 2), \quad \ldots$$

Discrete distributions

We will consider two examples of discrete probability distributions that have applications in engineering:

- Binomial distributions.
- Poisson distributions.

Binomial distribution

Consider an experiment with n independent trials, where there are exactly two outcomes of each trial: success with probability p, or failure with probability $1 - p$.

For example, tossing a (potentially unfair) coin n times, with probability p of getting a head each time.

- If you conducted $n = 100$ coin tosses, what is the probability of getting exactly 17 heads?
- What is the mean number of heads that you can expect?
- What is the variance of the distribution of the number of heads?

Binomial distribution

Because there can only be an integer number of heads, the variable H "number of heads" has a discrete distribution.

In particular, for an experiment with n independent trials with two outcomes, the probability that the number X of "successes" is equal to k is given by the ...

Binomial distribution

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0, 1, \ldots, n$$

where

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

is the binomial coefficient and ! is the factorial symbol.

Binomial distribution

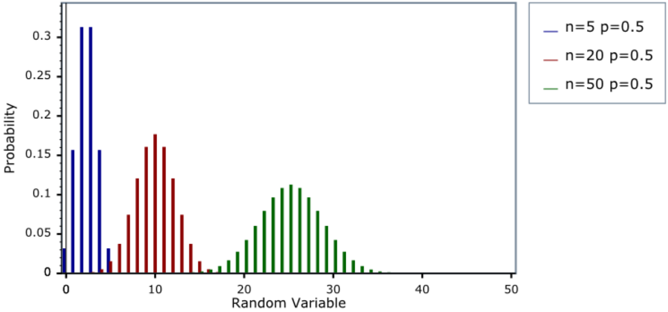
A binomial distribution with n trials and p probability of success has mean (i.e. the "expected number of successes"):

$$E(X) = np$$

and variance

$$Var(X) = np(1-p)$$

The shape of the distribution depends on n and p:



Heads and tails

Returning to the coin toss problem then:

The probability of 17 heads is:

$$\begin{aligned} P(H = 17) &= \binom{100}{17} p^{17}(1-p)^{83} \\ &= \frac{100!}{17!\,83!} p^{17}(1-p)^{83} \end{aligned}$$

The fraction is the number of different combinations of choosing 17 heads from 100 coins: about 1 with 268 zeros written after it!

If the coin is *fair* ($p = 0.5$), this becomes:

$$P(H = 17) = \frac{100!}{17!\,83!}(0.5)^{17}(0.5)^{83} = 5.246 \times 10^{-12}$$

Heads and tails

The expected (mean) number of heads from 100 tosses is:

$$E(H) = np = 100p$$

For a fair coin:

$$E(H) = 100 \times 0.5 = 50$$

as we would probably have guessed!

The variance in the number of heads is:

$$Var(H) = np(1 - p) = 100p(1 - p)$$

For a fair coin:

$$Var(H) = 100 \times 0.5(1 - 0.5) = 25$$

Example 1

A component supplier claims that 95% of its catalogue items are in stock at any time.

An order is placed for 20 different random components from this supplier. How likely is it that at least three of the items order will be out of stock, if the supplier's claim is accurate?

Example 1 - Solution

Each item is either in stock or not, and the probability of each item being out of stock is 5%, so the binomial distribution applies:

$$P(k \text{ of 20 items out of stock}) = \binom{20}{k} 0.05^k 0.95^{20-k}$$

So the probability that *at least three* items are out of stock is:

$$\begin{aligned} P(3+ \text{ items out of stock}) &= P(3) + P(4) + \cdots + P(20) \\ &= 1 - P(0) - P(1) - P(2) \\ &= 1 - 0.3585 - 0.3774 - 0.1887 \\ &= 0.0755 \end{aligned}$$

So this is unlikely to occur.

Poisson processes

A Poisson process describes a situation where discrete events occur from time to time.

They may be characterised in terms of either:

- The probability of an event happening at each moment in time
- The distribution of the amount of time between two subsequent events.

The specific criteria are:

- Events are independent of each other. The occurrence of one event does not affect the probability another event will occur.
- The **average** rate of events (the number of events per a given time period, on average) is constant.
- Two events cannot occur at the exact same time.

Poisson processes

Typical examples include:

- The arrival of customers to a queue.
- The arrival of jobs to a printer.
- The arrival of telephone calls to an exchange.
- Breakdowns of a machine.
- The arrival of claims to an insurance company.
- The arrival of vehicles at a traffic intersection.

Poisson processes and the Poisson distribution

For such a process, consider a particular interval of time. If λ is the rate at which the event can be expected to occur over that interval, then the probability that the *actual* number of events which occur X in that time period is equal to k is given by:

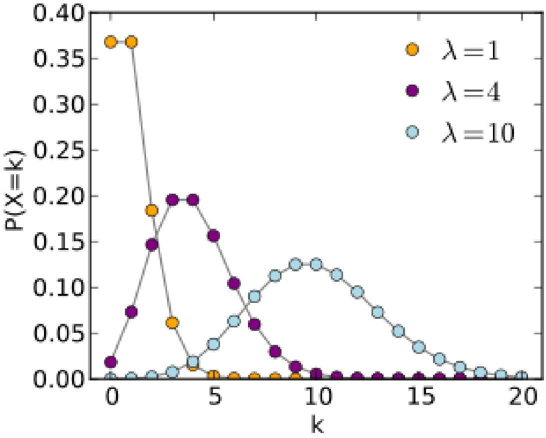
Poisson distribution

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

Both the mean and the variance of a Poisson distribution are equal to the rate λ.

The Poisson distribution

The distribution of the different values (the number of events that occur in that time interval) is given by the Possion distribution.



Example 2

If you knew that (on average) my internet connection disconnects three times in a 30-minute match of *Rainbow Six: Siege* and that this is a Poisson process, then the probability that it will drop *exactly twice* during the next match is:

$$P(X = 2) \;=\; \frac{3^2\,\mathrm{e}^{-3}}{2!} \;=\; \frac{9}{2}\,\mathrm{e}^{-3} \;\approx\; 0.224$$

where the variable X is defined as the number of internet disconnects during a 30-minute match.

Example 3

In a remote location, the frequency with which vehicles arrive at a particular intersection is modelled as a Poisson process. On average, 12 cars pass through every hour.

What is the probability that:

1. Exactly 8 cars arrive in a given hour?
2. Exactly 6 cars arrive in a given 20 minute period?
3. No more than 2 cars arrive in a given hour?

Example 3 - Solution (I/III)

Let X be the number of cars that arrive in one hour.

In this case, the rate is $\lambda = 12$, and we want to find $P(X = 8)$:

$$P(X = 8) = \frac{12^8 \, e^{-12}}{8!} \approx 0.0655$$

Example 3 - Solution (II/III)

Let Y be the number of cars that arrive in a 20 minute interval.

This time we are interested in a smaller interval. If the rate is constant, and 12 cars are expected in one hour, then in 20 minutes we would expect the rate to be:

$$\lambda = \frac{12}{3} = 4$$

Then we want to determine $P(Y = 6)$:

$$P(Y = 6) = \frac{4^6 \, e^{-4}}{6!} \approx 0.1042$$

Example 3 - Solution (III/III)

This time we return to $\lambda = 12$, but we want to find $P(X \leq 2)$:

$$P(X = 0) = \frac{12^0 \, e^{-12}}{0!} \approx 0.00000614$$

$$P(X = 1) = \frac{12^1 \, e^{-12}}{1!} \approx 0.00007373$$

$$P(X = 2) = \frac{12^2 \, e^{-12}}{2!} \approx 0.00044238$$

Then,

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= 0.00000614 + 0.00007373 + 0.00044238 \\ &= 0.000522 \end{aligned}$$

Poisson processes

For engineers, poisson processes have important applications in statistical quality control and planning for extreme events.

- Determining the likelihood that a product (e.g. the components of a jet engine) will fail in a given time, and then how likely it is that multiple components (e.g. all four jet engines) will suffer independent simultaneous failure.
- Designing computer and internet networks to handle search requests that may occur according to a Poisson process.
- Designing road networks for unusually high volumes of traffic.
- Analysing clusters of rare events (e.g. aircraft near misses) to determine if it is likely to be a coincidence, or indicative of an underlying cause.