# THE POISSON DISTRIBUTION

It seems that "the first application of the distribution to real data was the now famous example of deaths in the Prussian Army caused by horse kicking." It was bad enough that bullets, swords and cannons were killing the soldiers, now they were being killed by horses…. Someone needed to study this phenomenon.
Prussian soldiers.

The Poisson distribution is used to model events, which occur at random such as deaths per year studied by epidemiologists, suicides per month, accidents per week, cars passing a point in one minute, errors on a page and sea shells in a square metre on the beach.
These are events which occur in a period of time or an area of space. They occur at random since occurrences at one stage do not affect occurrences at the next stage.
The number of cars passing a point would not be Poisson if the drivers were communicating by mobile phone or if the point where the data is taken is just beyond a set of traffic lights.

Data collected would look like this: 4, 2, 5, 0, 3, 3, 7, ……. with no upper limit.
The events are regarded as rare.
The mean number of events per interval, i.e. the rate of occurrence is constant. 2 accidents per week means we can model a fortnight with a Poisson distribution using a mean of 4, or a day using a mean of $\frac{2}{7}$. This does not apply to a die for instance, where the total for two dice is not a rectangular distribution any more.
Another important attribute of the Poisson distribution is that the mean is equal to the variance. Both of these will be denoted by λ and the probabilities calculated from the formula:

$$\Pr(X = x) = \frac{e^{-\lambda}\lambda^{r}}{r!}$$ , λ is the mean and pronounced lambda.

## Example

In a study at a factory over a period of one year, the number of injuries were recorded for each week:
The results were 1, 4, 2, 2, 0, 1, 1, 4, 3, 0, 0, 1, ….and when tabulated:

| No of injuries | 0 | 1 | 2 | 3 | 4 | 5+ |
|---|---|---|---|---|---|---|
| No of weeks. | 29 | 16 | 4 | 2 | 1 | 0 |

Since the events are rare, occur independently of each other and one at a time, a Poisson model can be used. The mean can be found (0x29 + 1x16 + 2x4 +3x2 + 4x1)/52 = 0.65. The variance is 0.84.

Using λ = 0.65 the probabilities can be calculated and hence the expected frequencies.

| Probabilities (p) | 0.522 | 0.339 | 0.110 | 0.024 | 0.004 | 0.001 |
|---|---|---|---|---|---|---|
| Expected frequencies (px52) | 27 | 18 | 6 | 1 | 0 | 0 |

The probabilities can also be obtained from a table of Poisson probabilities remembering that the **cumulative probabilities** are tabulated. $\Pr(x \le r)$

| | | | | | | |
|---|---|---|---|---|---|---|
| Cumulative probabilities | 0.522 | 0.861 | 0.971 | 0.995 | 0.999 | 1.000 |

$\Pr(x = 2) = \Pr(x \le 2) - \Pr(x \le 1)$ 0.971 – 0.861 = 0.110.

## The Poisson approximation to the Binomial.

When n is large and p (or q) is small the Binomial distribution may be approximated by the Poisson distribution putting the mean of the Binomial np equal to $\lambda$, the mean of the Poisson.

**The probability of a bad orange in a crate of 65 is 0.01. Find the probability of more than one bad oranges in a crate.**

Here n is finite so using the Binomial distribution with n = 65 and p = 0.01 gives

$P(0) = {}^{65}c_0\, 0.01^0 0.99^{65}$ and $P(1) = {}^{65}c_1\, 0.01^1 0.99^{64}$ $P(0+1) = 0.862$

If we take $\lambda = np = 0.65$ $P(0+1) = 0.861$ which is a pretty good match. In fact good approximations are obtained when $n > 50$ and $p < 0.1$. Pr(more than one bad) = 1 - Pr(0 + 1) = 0.138

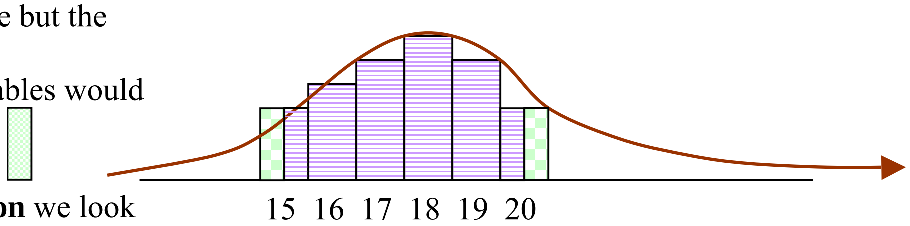
## The Poisson approximation to the Normal.

If a random variable follows the Poisson distribution with mean $\lambda$ i.e. $X \sim Po(\lambda)$**.** Then as long as $\lambda > 10$, we can use the Normal distribution with mean and variance = $\lambda$ **.**
The small drawback is that that Poisson is a discrete variable and the Normal is continuous so we have to allow for that by using a **continuity correction**.

**The number of errors on a page follows a Poisson distribution with mean 1.7. Find the probability that there are between 15 and 20, inclusive, errors on 10 consecutive pages of the book.**

**If we use the additive property of the Poisson distribution, the number of errors X** $\sim Po(17)$.
Using the Normal approximation: $Y \sim N(17, 17)$
A diagram of the probabilities with the Normal distribution overlaid would look like this.

Notice that the bars are discrete but the curve is continuous. Finding $Pr(15 \le x \le 20)$ using Normal tables would lose half the first and last bars.
Using the **continuity correction** we look up $Pr(14.5 \le x \le 20.5)$ for greater accuracy.




$$Pr\left(\frac{14.5-17}{\sqrt{17}} \le z \le \frac{20.5-17}{\sqrt{17}}\right) = Pr(-0.606 \le z \le 0.849) = 0.3023 + 0.2291 = \underline{0.5314}$$

The Binomial distribution may also be approximated to the Normal distribution using a mean of ***np*** and variance ***npq*** provided ***p*** is close to ½ and np > 5.