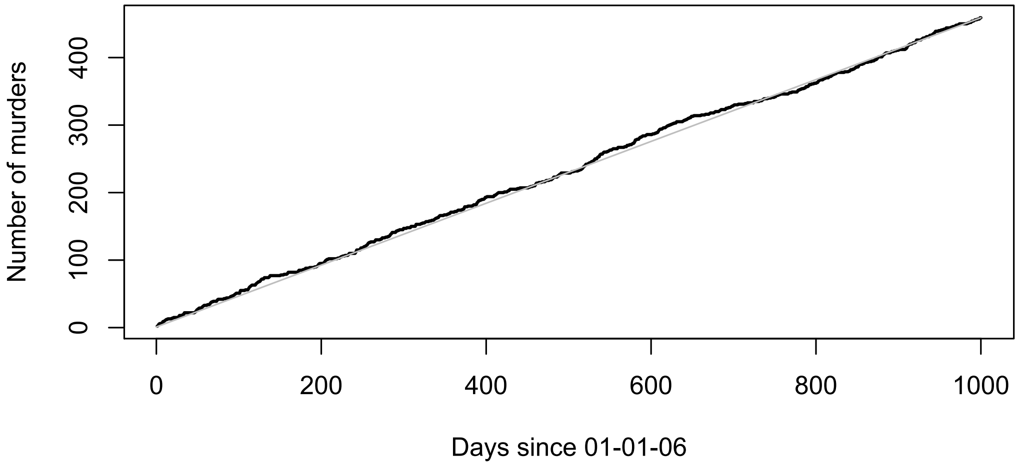
Figure 6: The total number of murders in London since 1st January 2006. The grey line has a slope of 0.458.



In the case of murders in London there is no such discrete unit. We could say a day was the unit, but you can get more than one murder in a day, so the outcome is not binary and hence the binomial does not apply. So what about making an hour the unit. It is less likely that more than one murder occurs in an hour, so this is closer to what we want, but not quite there – because you could still get more than one murder in an hour. So lets make a minute the unit. Closer still, but it would still be possible to get more than one murder in a minute,... OK, fine, what we'll do then is make the unit infinitely small - that is certainly as close as we can get to there being at most one murder per unit.

So how do we make the unit of a binomial distribution infinitely small? This is easy in principle, but it involves a bit more sophisticated mathematics than we can cope with on this module, so we'll say what we do to get the result, and state the result, but not prove it. (You can find proofs on the web and/or on later Maths modules dealing with limits.)

6.1 The Poisson distribution from the binomial

We're starting with a binomial distribution and making the time unit smaller and smaller, so we need some probability to start with. As we worked out the London murder rate in murders per day, we will start with day as our time unit. We calculated from the data that there were 0.458 murders a day on average, so if only one murder can occur a day (as we will assume to start with), the probability of a murder occurring on any particular day is 0.458.

Now if we make hour our unit and at most one murder can occur in any one hour (more realistic than at most one a day, although still not very realistic), then the corresponding probability is 0.458/24 as there are 24 hours in a day. In this case our binomial probability model for the number of murders in our time unit (a day in this case) would be

$$\mathbb{P}(X = n) \quad = \quad \binom{24}{n}\left(\frac{0.458}{24}\right)^n\left(1 - \frac{0.458}{24}\right)^{24-n}$$

If we now consider a minute to be the unit, we have $24 \times 60 = 1,440$ units and the relevant binomial distribution is

$$\mathbb{P}(X = n) \;=\; \binom{1440}{n} \left(\frac{0.458}{1440}\right)^n \left(1 - \frac{0.458}{1440}\right)^{1440-n}$$

And in general, when we divide the day into N units and events occur at a rate of λ per day on average, the relevant binomial distribution is

$$\mathbb{P}(X = n) \;=\; \binom{N}{n} \left(\frac{\lambda}{N}\right)^n \left(1 - \frac{\lambda}{N}\right)^{N-n}$$

By taking the limit as N approaches infinity, we obtain a model that quantifies the probability of n events per day without having to set an arbitrary time unit (i.e. using continuous time):

$$\mathbb{P}(X = n) \;=\; \lim_{N\to\infty} \binom{N}{n} \left(\frac{\lambda}{N}\right)^n \left(1 - \frac{\lambda}{N}\right)^{N-n} \;=\; \frac{\lambda^n e^{-\lambda}}{n!}$$

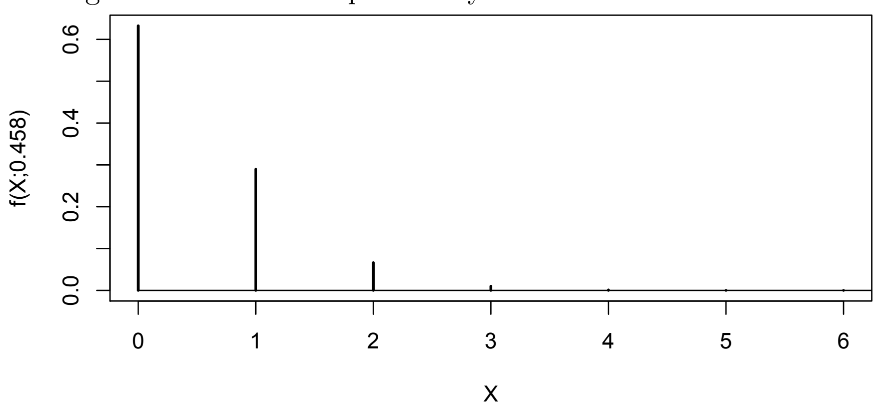
We do not prove this result in this course. A proof can be found on the web if you are interested (it uses the result given in Equation (1)).

The Poisson probability mass function:

$$\begin{aligned} f(n;\lambda) \;=\; \mathbb{P}(X = n) \;&=\; \frac{\lambda^n e^{-\lambda}}{n!} && \text{for } n \in \{0, 1, 2, \ldots\} \\ &=\; 0 && \text{otherwise.} \end{aligned}$$

We say that X "has a Poisson distribution" or "is Poisson distributed". Figure 4 shows the Poisson probability mass function (pmf) for $\lambda = 0.458$.

Figure 7: The Poisson probability mass function for $\lambda = 0.458$.



We can now address the question of how likely it is to have four murders in one day when murders occur at a rate of $\lambda = 0.458$ per day on average. It is $f(4; \lambda = 0.458) = \frac{0.458^4 e^{-0.458}}{4!} \approx 0.00116$, which is very unlikely. So can we conclude Andy Tighe was not exaggerating? See the Exercise on this.

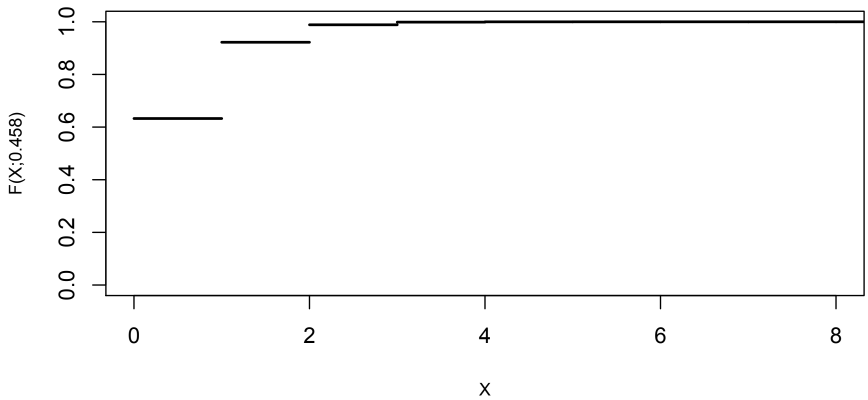
But is this the correct question to ask? The issue that concerns the Metropolitan Police, and the public is surely whether there are unusually many murders, not whether exactly four murders is very unusual. So we should really be asking whether having 4 *or more* murders is very unlikely. To address this question, we need the Poisson cumulative distribution function:

The Poisson cumulative distribution function:

$$F(n;\lambda) \;=\; \mathbb{P}(X \le n) \;=\; \sum_{j=0}^{\lfloor n \rfloor} f(n;\lambda) \;=\; \sum_{j=0}^{\lfloor n \rfloor} \frac{\lambda^n e^{-\lambda}}{n!} \qquad \text{for } n \ge 0$$

The probability of 4 or more murders in a day is $1 - F(n = 3; \lambda = 0.458) \approx 0.127\%$. (Here is an R command to do this calculation: `1-ppois(3,lambda=0.458)`.) So yes, four or more murders in one day is very unusual – we would expect it to happen only about once every $(1/0.00127)/365 \approx 2.16$ years. (But see also the Exercise on this topic.) Figure 8 shows the Poisson probability mass function for $\lambda = 0.458$.

Figure 8: The Poisson cumulative distribution function for $\lambda = 0.458$.



6.2 Waiting for the next murder: the exponential distribution

If you were the head of Homicide at the Metropolitan Police and having to decide on staffing levels you'd be interested in knowing the probability of no murders in the next hour/day/week/etc. The geometric cumulative distribution function quantifies the probability of waiting no more than k trials to the first "success", when events are possible only at discrete occasions (lottery draws, for example). But we need the probability of waiting at least k time units when event can happen in continuous time (murders can happen at any time). We can get this using a similar method to that which we used to get the Poisson distribution, i.e. by taking the limit as we cut time into a larger and larger number of smaller and smaller units.

Consider the geometric probability of waiting no more than k trials until the first success, when the probability of success is p: $\mathbb{P}(X \le k) = 1 - (1-p)^k$. Now we'll do the same thing we did with the binomial and use $p = \lambda/N$, where N is the number of equal-length intervals into which we divide the day. A wait of k days is a wait of Nk of these intervals, so in this case $\mathbb{P}(X \le k) = 1 - (1 - \frac{\lambda}{N})^{Nk}$. If we rewrite this letting

$y = \frac{N}{\lambda}$, and hence $N = \lambda y$, we have $\mathbb{P}(X \le k) = 1 - (1 - \frac{1}{y})^{k\lambda y}$. Noting that as N approaches ∞, y approaches ∞, we can use Equation (1) to see that

$$\lim_{y\to\infty} \left\{ 1 - \left(1 - \frac{1}{y}\right)^{k\lambda y} \right\} \;=\; 1 - e^{-k\lambda}$$

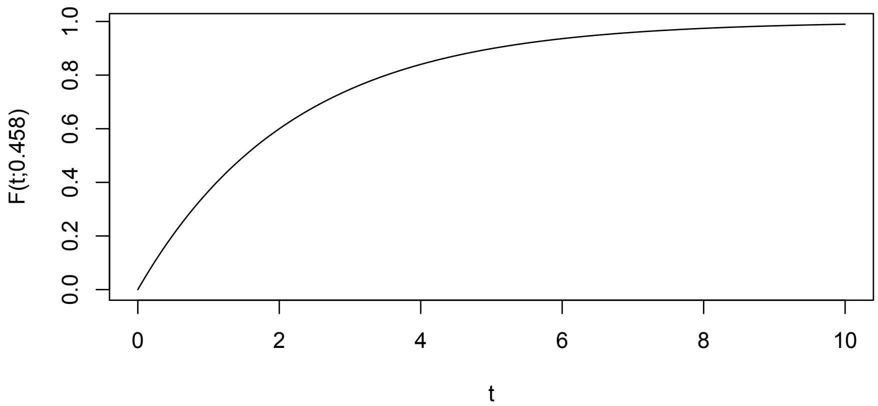
This is the cumulative distribution function for waiting time when time is continuous. (Note that the units of time k (*days* in this case) is the same as the time unit for the rate λ (events per *day* in this case.) We usually use t instead of k, to indicate that we are dealing with continuous time:

The exponential cumulative distribution function:

$$F(t;\lambda) \;=\; \mathbb{P}(X \le t) \;=\; 1 - e^{-\lambda t} \qquad \text{for } t \ge 0$$

Figure 9 shows the exponential cdf when $\lambda = 0.458$.

Figure 9: The exponential cumulative distribution function for $\lambda = 0.458$.



7 Summary

With some problems, like the birthday problem, you just have to work the probabilities out from scratch. But for many problems all you have to do is to *recognise* the problem as one involving a statistical distribution that you know about (not as easy as it sounds) and then use that probability distribution to work out probabilities. So a key part of real-world problem solving when there is randomness involved in the problem, is problem recognition.

We have dealt with four probability distributions in these notes. When does which apply? (In the summaries below "response" is the random variable of interest that you observe: These are some responses we have dealt with in these notes: number of times Paul predicts correctly; how many tries unit his first success; number of murders observed in a day; how long until the next murder.)

Binomial distribution

- Response is a count of events (e.g. number of "successes")
- Counts are associated with discrete units (e.g. football matches)
- Events occur independently
- There are a fixed number (N) of "trials" (occasions on which events occur)
- The probability of a "success" (p) is the same for all events

Geometric distribution

- Response number of discrete "trials" you wait before observing an event.
- "Trials" occur as discrete units (e.g. football matches)
- Events occur independently
- The probability of a "success" (p) is the same for all events

Poisson distribution

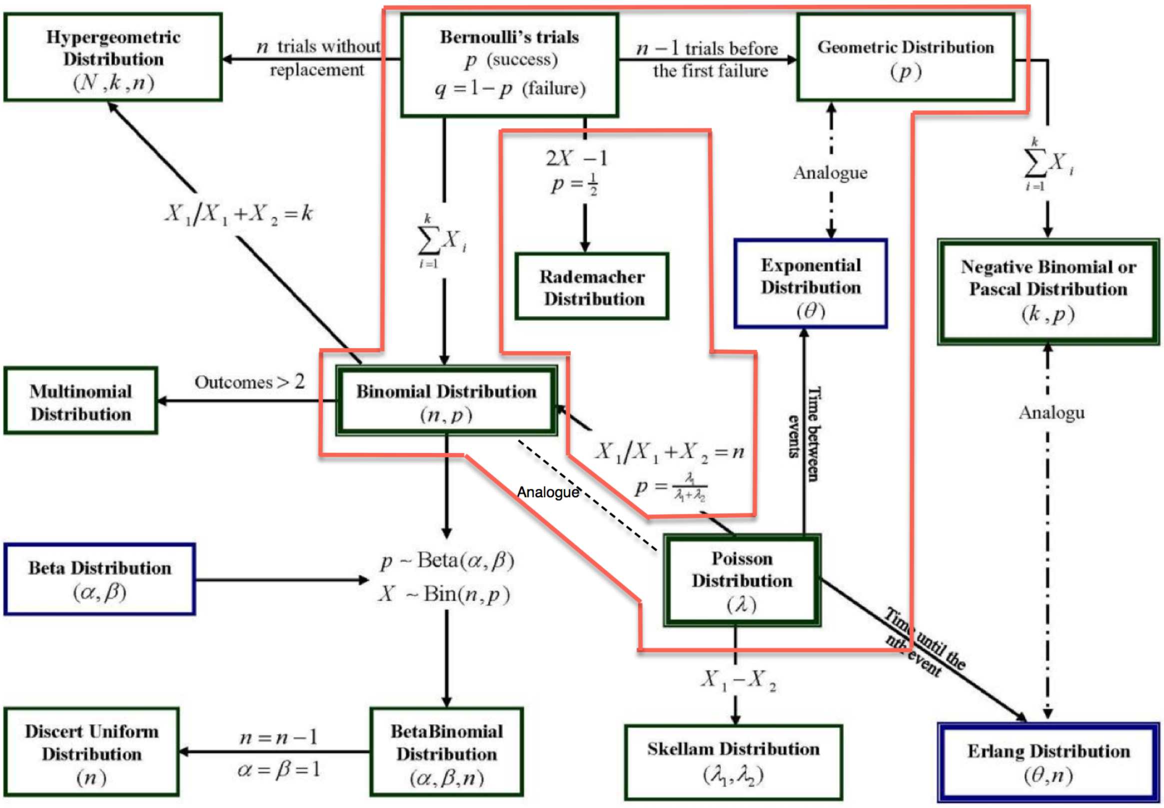
- Response is a count (e.g. number of "successes")
- Counts are associated with continuous intervals (e.g. a period of time)
- Events occur independently
- The average rate that events occur (λ) is the same at all times

Exponential distribution

- Response is a continuous waiting time (or distance, or area or ...) to an event
- Events occur independently
- The average rate that events occur (λ) is the same at all times

Figure 10 shows diagrammatically the relationships between the distributions we have covered in these notes. (The Bernoulli distribution is just the binomial distribution with a single trial: $N = 1$.) It also shows various other distributions that we did not look at (and there are many more besides those shown).

Figure 10: The relationships between the distributions considered in these notes (those inside the red polygon). Note that the notation in the figure is slightly different from that used elsewhere in these notes. For example the figure used θ for the parameter of the geometric distribution where we used p, and the binomial uses n where we used N.



The material below on the exponential pdf is beyond the scope of this module – it is included here for interest/completeness.

The cdf of the geometric distribution is a sum up to the number of events without success (k), but in the case of the exponential distribution we have a continuous variable t instead of a discrete one, so we integrate instead of summing. That is, $F(t;\lambda) = \mathbb{P}(X \leq t) = \int_0^t f(u;\lambda)\, du$, where $f(u;\lambda)$ is the exponential probability density function evaluated at u. Conversely, we can obtain an expression for $f(t;\lambda)$ by differentiating $F(t;\lambda)$ with respect to t:

The exponential probability density function (pdf):

$$f(t;\lambda) \quad = \quad \frac{dF(t;\lambda)}{dt} \quad = \quad \lambda e^{-\lambda t} \qquad \text{for } t \geq 0$$

Note the change in terminology: whereas it is called a "probability **mass** function" (pmf) when you're dealing with a random variable like k that is discrete, it is called a "probability **density** function" (pdf) when you're dealing with a random variable like t that is continuous.

Figure 11 shows the exponential pdf when $\lambda = 0.458$.

Figure 11: The exponential probability density function for $\lambda = 0.458$.