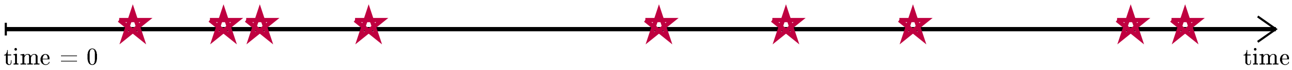
1 Poisson Distribution

Suppose that there is a pizza restaurant that receives, on average, 90 calls per hour. Let X be a random variable describing the number of calls the restaurant will receive in the next hour. What is the probability distribution of X? As time goes on, the calls can come in at any time. They're described on a continuous timeline, but their number is discrete.



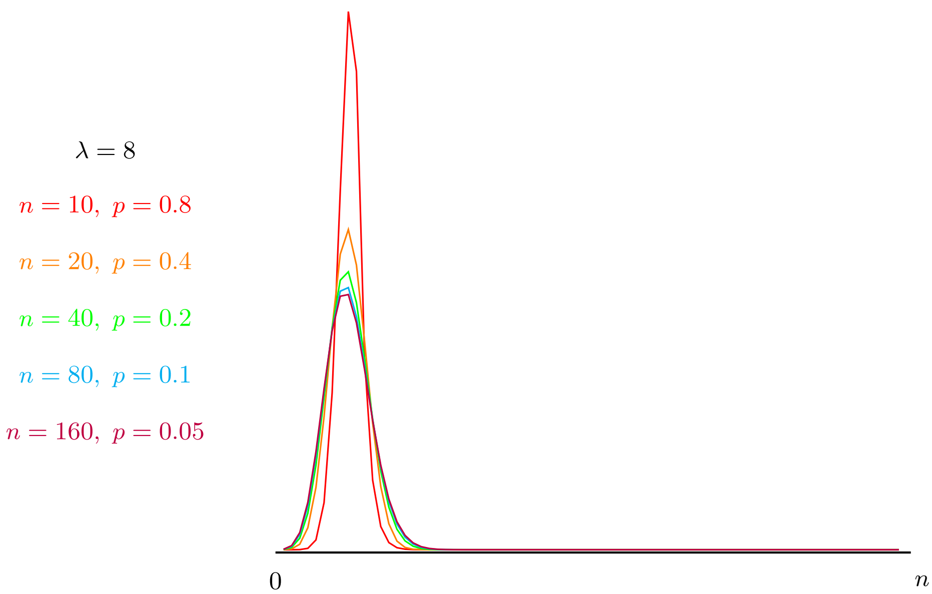
Suppose we divide the hour up into seconds and say that in every second a call might come in or not. That's a Bernoulli trial for every second. We want to keep the average number of calls over the hour fixed. If we set it up as a binomial distribution with $n = 3600$, the number of seconds in a minute, we want $np = 90$. So in every second, a call comes in with probability $p = 90/3600$. If we divide the hour up into even smaller intervals for better accuracy, we still need $np = 90$, so as n grows p will get smaller. The random variable X for the number of calls in the next hour is distributed in a way described by the limit of this as $n \to \infty$.

Definition 1 The *Poisson Distribution* $P(X = k)$ *with parameter* λ is the continuous limit of the binomial distribution $b(n, p, k)$ with $pn = \lambda$ as $n \to \infty$.

$$P(X = k) = \lim_{n\to\infty} b(n, \frac{\lambda}{n}, k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

The next section goes through the derivation of $\lim_{n\to\infty} b(n, \frac{\lambda}{n}, k) = e^{-\lambda}\frac{\lambda^k}{k!}$.

The picture illustrates a few binomial distributions with $\lambda = np$ constant and ascending n.



2 Deriving the Poisson distribution

We have $b(n,p,k) = \binom{n}{k} p^k (1-p)^{n-k}$. Then:

$$P(X=k) = \lim_{n\to\infty} \binom{n}{k} (\frac{\lambda}{n})^k (1-\frac{\lambda}{k})^{n-k}$$

Let us first look at $P(X=0)$. This expression is:

$$P(X=0) = \lim_{n\to\infty} (1-\frac{\lambda}{k})^n = 1 - \lambda + \frac{\lambda^2}{2!} - \frac{\lambda^3}{3!} + \dots,$$

which is the Taylor series for $e^{-\lambda}$. So $P(X=0) = e^{-\lambda}$. The next step is to figure out how to make $P(X=k)$ out of $P(X=k-1.)$

$$\frac{P(X=k)}{P(X=k-1)} = \frac{\lim\limits_{n\to\infty} b(n,p,k)}{\lim\limits_{n\to\infty} b(n,p,k-1)} = \lim_{n\to\infty} \frac{\frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}}{\frac{n!}{(k-1)!(n-k+1)!} (1-p)^{n-k+1}}$$

Most of these terms cancel out. We eventually get:

$$P(X=k) = \lim_{n\to\infty} \frac{(n-k+1)p}{k(1-p)} = \lim_{n\to\infty} \frac{np-(k-1)p}{k(1-p)} = \lim_{n\to\infty} \frac{\lambda - (k-1)\frac{\lambda}{n}}{k(1-\frac{\lambda}{n})} = \frac{\lambda - 0}{k \times 1} = \frac{\lambda}{k}.$$

So, the two ingredients we got that are enough to construct $P(X=k)$ for every k are:

$$P(X=0) = e^{-\lambda}$$

$$\frac{P(X=k)}{P(X=k-1)} = \frac{\lambda}{k}$$

From these we can derive:

$P(X=1) = P(X=0)\frac{\lambda}{1} = e^{-\lambda}\frac{\lambda}{1}$
$P(X=2) = P(X=1)\frac{\lambda}{2} = e^{-\lambda}\frac{\lambda^2}{2}$
$P(X=3) = P(X=2)\frac{\lambda}{3} = e^{-\lambda}\frac{\lambda^3}{6}$
$\vdots$
$P(X=k) = P(X=k-1)\frac{\lambda}{k} = e^{-\lambda}\frac{\lambda^k}{k!}$

So the number of calls to the pizza restaurant in the next hour obeys the distribution:

$$P(X=k) = e^{-90}\frac{90^k}{k!}.$$

The number of call in the next minute obeys the distribution:

$$P(Y=k) = e^{-1.5}\frac{1.5^k}{k!}.$$

3 Poisson is much simpler to compute, so it's often used to approximate binomial distribution.

With Poisson, **as soon as you know what the average λ is supposed to be you can figure out the whole distribution.**

One of the best uses for Poisson distribution is approximating binomial probabilities. For large n, it's much easier to compute than complicated $\binom{n}{k}$ factors!

Example 1 In a fixed amount A of blood an average human has 40 white blood cells. Each blood cell in the body represents a trial and an average of 40 of them will be in a sample of A blood. This can be approximated with a Poisson distribution for $\lambda = 40$:

$$P(X=k) = e^{-40}\frac{40^k}{k!}.$$

This could also be evaluated exactly with a binomial distribution for the right n =the number of white blood cells in the body, but in this case Poisson approximation is much simpler.

4 Negative Binomial Distribution

You have a biased coin that comes up “tails" with probability p. You toss it until it comes up tails a total of k times, then stop. Let the random variable x describe the total number of times you tossed the coin. What is the distribution $u(x = n, k, p)$ of x? Suppose that the total number of tosses is n. The sequences that result in that look as follows:

$$\underbrace{\text{all the tosses that come before the } k\text{th time you toss tails}}_{n-1 \text{ tosses, } k-1 \text{ of which are tails}}\ \mathrm{T}$$

The probability of each of those occurrences is $p^k(1-p)^{n-k}$, and there are $\binom{n-1}{k-1}$ sequences of this type. So:

$$u(x = n, k, p) = \binom{n-1}{k-1} p^k(1-p)^{n-k}.$$

5 Hypergeometric Distribution

You have a set of N balls, where k are red and $N-k$ are blue. Pick n of these balls at random without replacement. Let the random variable x describe the number of red balls in the picked set. What is the probability distribution $h(N, k, n, x)$ of x?

First of all, unless $0 \leq x \leq n, k$ the probability is 0. Otherwise, what is the probability of picking exactly x red and $n-x$ blue balls? There are $\binom{N}{n}$ ways in total to pick the set of n balls. There are $\binom{k}{x}$ ways to pick x red alls out of the set of k. There are $\binom{N-k}{n-x}$ ways to choose $n-x$ blue balls out of a set of $N-k$. The overall probaility is:

$$h(N, k, n, x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}.$$

This can be generalized to any number of colours. For example, if we have k red, l blue and $N-k-l$ yellow balls, the probability of picking exactly x red and y blue balls in a set of n is:

$$h(N, k, n, x, y) = \frac{\binom{k}{x}\binom{l}{y}\binom{N-k-l}{n-x-y}}{\binom{N}{n}}.$$