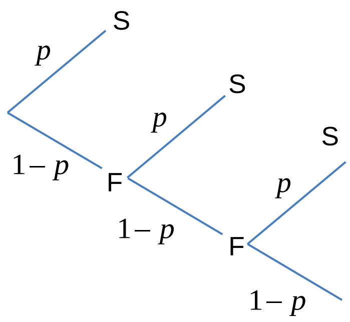
Discrete probability distributions: Geometric, Binomial & Poisson

Discrete probability distributions are essentially formulae for calculating probabilities associated with a random variable *x* which can only take *positive integer values.* e.g. *x* could be the number of rolls of a dice before a six is obtained (**Geometric**), the number of sixes obtained during ten rolls (**Binomial**) or the number of sixes logged per minute, given an 'expected' six scoring rate (**Poisson**) and many rolls of the dice.

The Geometric Distribution

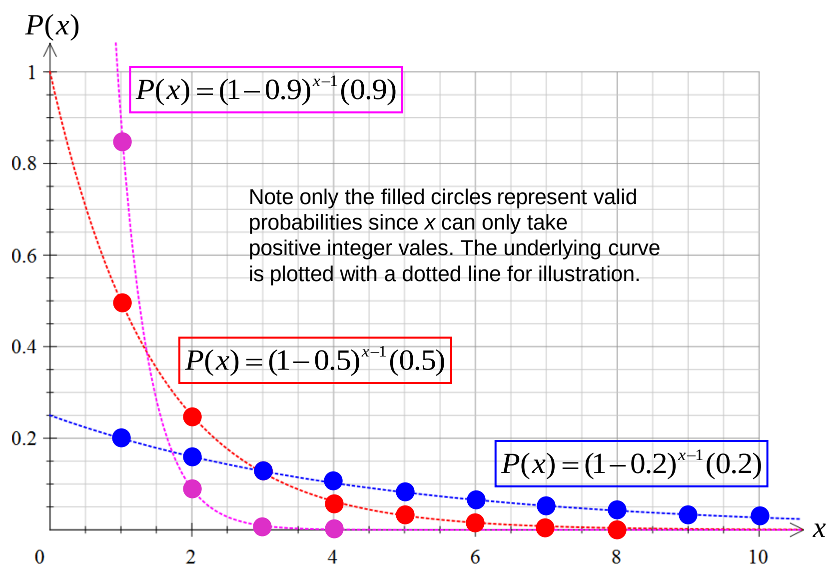
This calculates the probability of success after *x* independent trials. i.e. there are *x -1* failures (F) before success (S) occurs. Each trial is independent with success probability *p*. The random variable *x* is distributed by the Geometric distribution *P(x)* i.e. $\boxed{x \sim \text{Geo}(p)}$



$$P(x=1) = p$$
$$P(x=2) = (1-p)p$$
$$P(x=3) = (1-p)^2 p$$
$$\therefore \boxed{P(x) = (1-p)^{x-1} p}$$

e.g. what is the probability of taking five rolls of a fair six sided dice before a six is rolled?

$$x \sim \text{Geo}(p = \tfrac{1}{6})$$
$$P(x=6) = (1-\tfrac{1}{6})^4 \left(\tfrac{1}{6}\right)$$
$$P(x=6) = \tfrac{625}{7776} \approx 0.0804$$



$$P(x \le N) = P(1) + P(2) + P(3) + ... + P(N)$$
$$P(x \le N) = p + (1-p)p + (1-p)^2 p + ... + (1-p)^{N-1} p$$
$$P(x \le N) = p\frac{1-(1-p)^N}{1-(1-p)}$$
$$\boxed{P(x \le N) = 1-(1-p)^N}$$

Geometric progression

$$S_n = a + ar + ar^2 + ... + ar^{n-1}$$
$$S_n = \sum_{n=1}^{\infty} ar^{n-1} = a\frac{1-r^n}{1-r}$$
$$\text{if } |r| < 1 \quad \therefore S_\infty = \frac{a}{1-r}$$

Hence: $\boxed{\lim_{N \to \infty} P(x \le N) = 1}$

i.e. all the possible geometric probabilities sum to *unity*, which is a fundamental requirement of a probability distribution!

The *Moment Generating Function* of the Geometric Distribution is:

$$M(x,t) = E[e^{tx}]$$
$$M(x,t) = \sum_{n=1}^{\infty} e^{tx} p(1-p)^{x-1}$$
$$M(x,t) = pe^t \sum_{n=1}^{\infty} e^{t(x-1)}(1-p)^{x-1} = pe^t \sum_{n=1}^{\infty} \left(e^t(1-p)\right)^{x-1}$$
$$M(x,t) = \lim_{N \to \infty} \left\{ \frac{pe^t\left(1-\left(e^t(1-p)\right)^N\right)}{1-e^t(1-p)} \right\}$$

Geometric progression again ....

$$\boxed{M(x,t) = \frac{pe^t}{1-e^t(1-p)}}$$

This allows us to calculate the expectation and variance of the geometric distribution

$$\boxed{E[x] = \left.\frac{\partial M}{\partial t}\right|_{t=0} = \frac{1}{p}}$$
$$\boxed{V[x] = \left.\frac{\partial^2 M}{\partial t^2}\right|_{t=0} - \left(E[x]\right)^2 = \frac{1-p}{p^2}}$$

Alternative derivation of the expectation and variance of a geometric distribution using derivatives of the sum of an infinite geometric progression

$$\boxed{\begin{aligned}P(x)&=(1-p)^{x-1}p\\ E[x]&=\sum_{x=1}^{\infty}xP(x)\end{aligned}}$$

$$E[x]=\sum_{x=1}^{\infty}x(1-p)^{x-1}p$$

$$r=1-p\le 1$$

$$\therefore E[x]=p\sum_{x=1}^{\infty}xr^{x-1}$$

$$\sum_{x=1}^{\infty}r^{x-1}=1+r+r^2+...+r^{x-1}=\frac{1-r^x}{1-r}$$

$$\text{if } |r|<1 \;\Rightarrow \sum_{x=1}^{\infty}r^{x-1}=\frac{1}{1-r}$$

$$\therefore \sum_{x=1}^{\infty}r^x=\frac{r}{1-r}$$

$$\frac{d}{dr}\sum_{x=1}^{\infty}r^x=\sum_{x=1}^{\infty}xr^{x-1}$$

$$\frac{d}{dr}\frac{r}{1-r}=\frac{(1-r)(1)-r(-1)}{(1-r)^2}=\frac{1}{(1-r)^2}$$

$$\therefore \boxed{\sum_{x=1}^{\infty}xr^{x-1}=\frac{1}{(1-r)^2}}$$

$$\therefore E[x]=p\sum_{x=1}^{\infty}xr^{x-1}=\frac{p}{(1-(1-p))^2}=\frac{1}{p}$$

$$\boxed{E[x]=\frac{1}{p}}$$

$$\boxed{V[x]=\sum_{x=1}^{\infty}x^2P(x)-\left(E[x]\right)^2}$$

$$\sum_{x=1}^{\infty}x^2P(x)=\sum_{x=1}^{\infty}x^2p(1-p)^{x-1}=p\sum_{x=1}^{\infty}x^2r^{x-1}$$

$$r=1-p$$

$$\sum_{x=1}^{\infty}xr^{x-1}=\frac{1}{(1-r)^2}$$ From the derivation of E[x]

$$\therefore \sum_{x=1}^{\infty}xr^x=\frac{r}{(1-r)^2}$$

$$\frac{d}{dr}\sum_{x=1}^{\infty}xr^x=\sum_{x=1}^{\infty}x^2r^{x-1}$$

$$\frac{d}{dr}\frac{r}{(1-r)^2}=\frac{(1-r)^2(1)-r\left(2(1-r)(-1)\right)}{(1-r)^4}=\frac{(1-r)^2+2r(1-r)}{(1-r)^4}=\frac{1-r+2r}{(1-r)^3}=\frac{1+r}{(1-r)^3}$$

$$\therefore \sum_{x=1}^{\infty}x^2r^{x-1}=\frac{1+r}{(1-r)^3}$$

$$\therefore \sum_{x=1}^{\infty}x^2P(x)=p\sum_{x=1}^{\infty}x^2r^{x-1}=p\frac{1+1-p}{(1-(1-p))^3}=\frac{2-p}{p^2}$$

$$\therefore V[x]=\sum_{x=1}^{\infty}x^2P(x)-\left(E[x]\right)^2=\frac{2-p}{p^2}-\frac{1}{p^2} \longleftarrow E[x]=\frac{1}{p}$$

$$\boxed{V[x]=\frac{1-p}{p^2}}$$

The Binomial Distribution

This calculates the probability of x successes (S) given N independent trials. i.e. there are *N-1* failures (F). Unlike the Geometric Distribution, we do not 'stop' once success occurs. It is assumed the probability of success, p, is the same for each trial.

$$\boxed{x \sim \mathrm{B}(N,p)}$$

$$P(x\,|\,N,p) = P(\underbrace{SSSS...S}_{x}\,\underbrace{FFF...F}_{N-x}) \times \frac{N!}{(N-x)!x!}$$

i.e. there are $\frac{N!}{(N-x)!x!}$ distinct permutations of a sequence of N successes (S) and *N-1* failures (F)

$$\boxed{P(x\,|\,N,p) = \frac{N!}{(N-x)!x!}\,p^x(1-p)^{N-x}}$$

$$\frac{N!}{(N-x)!x!} = \binom{N}{x} = {}^N C_x$$

Binomial coefficients, which are the numbers in the N^{th} row of Pascal's Triangle

1

1 1

1 2 1

1 3 3 1

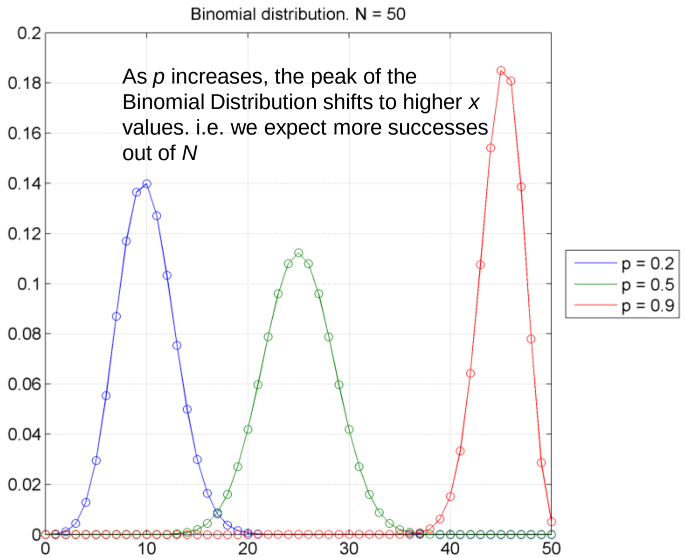
1 4 6 4 1

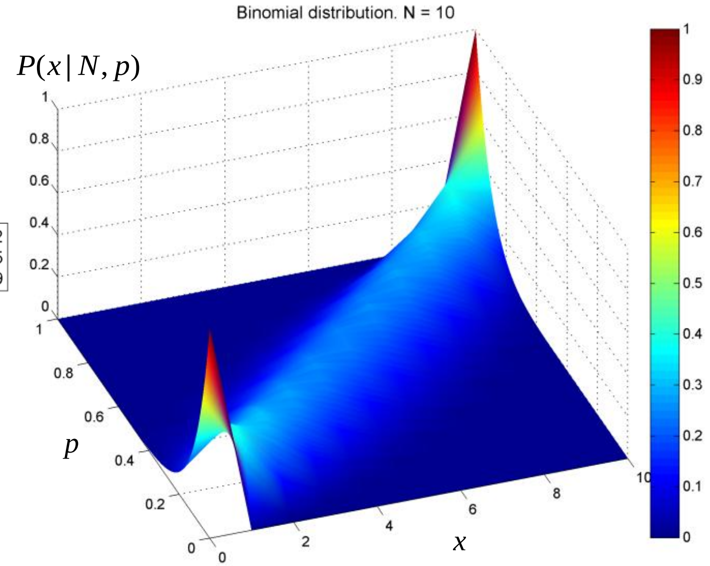
1 5 10 10 5 1

Example: The probability of getting exactly 2 sixes in 10 throws of a fair six sided dice is:

$$x \sim \mathrm{B}(N=10, p=\tfrac{1}{6})$$

$$P(x=2) = \binom{10}{2}(\tfrac{1}{6})^2\left(1-\tfrac{1}{6}\right)^8 = 45 \times \tfrac{390{,}625}{60{,}466{,}176} \approx 0.291$$





The *Moment Generating Function* for the Binomial Distribution is:

$$M(x,t) = E[e^{tx}]$$

$$M(x,t) = \sum_{n=0}^{N} e^{tx}\binom{N}{x}p^x(1-p)^{N-x}$$

$$M(x,t) = \sum_{n=0}^{N}\binom{N}{x}\left(pe^t\right)^x(1-p)^{N-x}$$

$$\boxed{M(x,t) = \left(pe^t + 1 - p\right)^N}$$

Binomial expansion

The MGF allows us to calculate the expectation and variance of the Binomial Distribution rather efficiently

$$\sum_{n=0}^{N} P(n\,|\,N,p) = \binom{N}{0}p^0(1-p)^N + \binom{N}{x}p^x(1-p)^{N-x} + ...\binom{N}{N}p^N(1-p)^0$$

$$\sum_{n=0}^{N} P(n\,|\,N,p) = \left(p+1-p\right)^N = 1$$

Binomial expansion

So the sum of Binomial distribution probabilities is unity as required.

$$\boxed{E[x] = \left.\frac{\partial M}{\partial t}\right|_{t=0} = Np} \qquad \boxed{V[x] = \left.\frac{\partial^2 M}{\partial t^2}\right|_{t=0} - \left(E[x]\right)^2 = Np(1-p)}$$

When N is large in can be shown that the Binomial Distribution is well approximated by a *Normal* (Gaussian) distribution with mean $\mu = Np$ and variance $\sigma^2 = Np$

$$\boxed{x \sim \mathrm{B}(N,p) \approx \mathrm{N}(Np, Np(1-p))}$$

This is a very useful approximation when N is large, as working out the Binomial coefficients may involve ratios of very large numbers and can therefore be hard to compute.

The Poisson Distribution

This calculates the probability of x *events within a defined time interval,* given that there is a known expected event rate λ. For example, x could be the number of goals in a football game. (Based on averages from Premier Leagues in Spain, UK, Germany & Italy, the answer is about $\lambda = 2.7$).

To model this process mathematically, let us break up the time interval into N binary trials, in which the event (e.g. a goal being scored) happens or not. Let us assume the event probability p is very small and the number of trials N is very large. (e.g. probability p of a goal given a random kick, and the numbers of kicks N).

The Poisson Distribution is therefore a *limit of a Binomial Distribution*:

$$\boxed{x \sim \mathrm{Po}(\lambda)}$$

$$P(x\,|\,\lambda) = \lim_{p\to 0, N\to\infty}\left(\frac{N!}{(N-x)!x!} p^x (1-p)^{N-x} \right)$$

Now the expected value of the Binomial Distribution is: $E[x] = Np$

$$\therefore \lambda = Np$$

$$\therefore P(x\,|\,\lambda) = \lim_{p\to 0, N\to\infty}\left(\frac{N!}{(N-x)!x!}\left(\frac{\lambda}{N}\right)^x \left(1-\frac{\lambda}{N}\right)^{N-x} \right)$$

$$\lim_{N\to\infty}\left\{\left(1-\frac{\lambda}{N}\right)^{N-x}\right\} \approx \lim_{N\to\infty}\left\{\left(1-\frac{\lambda}{N}\right)^{N}\right\} = e^{-\lambda}$$

This is a standard result which will not be proven here

$$\lim_{N\to\infty}\left\{\frac{N!}{(N-x)!N^x}\right\} \approx \lim_{N\to\infty}\left\{\frac{N}{N}\times\frac{N-1}{N}\times\ldots\times\frac{N-x+1}{N}\right\} = 1$$

$$\therefore \boxed{P(x\,|\,\lambda) = \frac{\lambda^x}{x!}e^{-\lambda}}$$

The Moment Generating function of the Poisson Distribution is:

$$M(x,t) = E[e^{tx}]$$

$$M(x,t) = \sum_{x=0}^{\infty} e^{tx}\frac{\lambda^x}{x!}e^{-\lambda}$$

$$M(x,t) = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\left(\lambda e^t\right)^x}{x!} = e^{-\lambda}e^{\lambda e^t}$$

$$e^x = \sum_{n=0}^{\infty}\frac{x^n}{n!}$$

Maclaurin Expansion

$$\boxed{M(x,t) = e^{\lambda\left(e^t-1\right)}}$$

Now that the MGF is known:

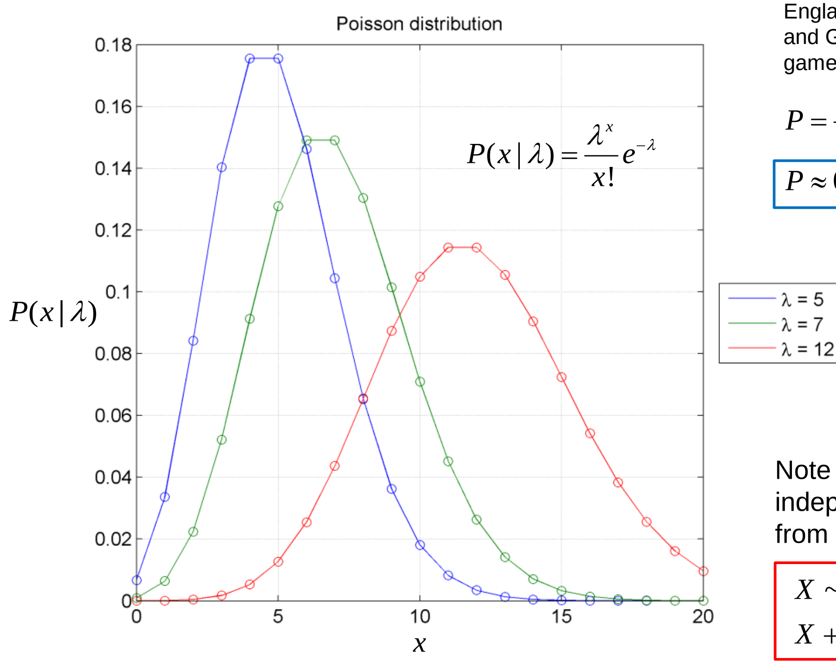
$$\boxed{E[x] = \left.\frac{\partial M}{\partial t}\right|_{t=0} = \left. e^{\lambda\left(e^t-1\right)}\times\lambda e^t\right|_{t=0} = \lambda}$$

$$V[x] = \left.\frac{\partial^2 M}{\partial t^2}\right|_{t=0} - \left(E[x]\right)^2 = \left.\frac{\partial}{\partial t}\lambda e^t e^{\lambda\left(e^t-1\right)}\right|_{t=0} - \lambda^2$$

$$V[x] = \left.\left(\lambda e^t\right)^2 e^{\lambda\left(e^t-1\right)} + \lambda e^{\lambda\left(e^t-1\right)}\right|_{t=0} - \lambda^2$$

$$\boxed{V[x] = \lambda}$$

So for the Poisson Distribution the *mean and variance take the same value*



When λ is large in can be shown that the Poisson Distribution is well approximated by a *Normal* (Gaussian) distribution with mean $\mu = \lambda$ and variance $\sigma^2 = \lambda$

$$\boxed{x \sim \mathrm{Po}(\lambda) \approx \mathrm{N}(\lambda,\lambda)\ ;\ \lambda \gg 1}$$

$$\boxed{P(x \le k) \approx \frac{1}{\sqrt{2\pi\lambda}}\int_{-\infty}^{k} e^{-\frac{1}{2\lambda}(x-\lambda)^2}dx}$$

Example: What is the probability England score 5 goals and Germany score 3, assuming England score 2 goals per game and Germany score 3 goals per game on average?

$$P = \frac{2^5}{5!}e^{-2}\times\frac{3^3}{3!}e^{-3}$$

$$\boxed{P \approx 0.0081}$$

..... whereas a 3:2 win for Germany has probability

$$P = \frac{2^2}{2!}e^{-2}\times\frac{3^3}{3!}e^{-3}$$

$$\boxed{P \approx 0.061}$$

Note if two random variables are independent, and both derive from a Poisson Distribution

$$\boxed{\begin{gathered} X \sim \mathrm{Po}(\lambda_1),\quad Y \sim \mathrm{Po}(\lambda_2) \\ X + Y \sim \mathrm{Po}(\lambda_1 + \lambda_2) \end{gathered}}$$