Stirling's Approximation and Binomial, Poisson and Gaussian distributions
AF 30/7/2014.

These notes describe much of the underpinning mathematics associated with the Binomial, Poisson and Gaussian probability distributions. Key aspects are:

- The use of the Gamma function $\Gamma(n) = \int_0^\infty x^{n-1}e^{-x}dx$ to derive the standard integral $n! = \int_0^\infty x^n e^{-x}dx$
- A derivation of the result $\int_{-\infty}^{\infty} e^{-y^2}dy = \sqrt{\pi}$
- Use of the the above results and Maclaurin expansion of $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \ldots$, for $|x| < 1$, to derive *Stirling's approximation* $n! \approx n^n e^{-n}\sqrt{2\pi n}$, which is valid for large n
- A derivation of the Poisson distribution $P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$ as a limit of the Binomial distribution $P(x) = \frac{n!}{(n-x)!x!}p^x(1-p)^{n-x}$, when n is large and p is small, but the expectation $\lambda = np$ is constant.
- A derivation of the expectation $E[x]$ and variance $V[x]$ of the Binomial and Poisson distributions. This is done directly (for the Poisson case) and also via Moment Generating Functions (MGFs). The MGF of a distribution of random variable x is $M(x,t) = E[e^{xt}]$
- An argument which shows the Poission (and Binomial) distributions tend to a Gaussian in overall shape as, respectively, parameters λ and n become large.
- A derivation of the *Central Limit Theorem.* i.e. *"The distribution of the mean values of a set of independent random values tends towards a Gaussian distribution if the number of samples is large enough."*

  Reference: The essence of most of the arguments presented here have been adapted from from Riley, K.F., Hobson, M.P., Bence, S.J., *Mathematical Methods for Physics and Engineering.* Third Edition, 2006. Cambridge University Press. pp636-640, 1174-1188.

1 The Gamma Function and Stirling's approximation

1.1 Definition of the gamma function

The gamma function $\Gamma(n)$ is defined by

$$\boxed{\Gamma(n) = \int_0^\infty x^{n-1}e^{-x}dx} \tag{1}$$

It is undefined for an n of zero or negative integers. (See figure below).

1.1.1 Special case #1: $\Gamma(1) = 1$

1.1.2

$$\Gamma(1) = \int_0^\infty e^{-x}dx = \left[-e^{-x}\right]_0^\infty = (0) - (-1) \tag{2}$$

$$\therefore \quad \Gamma(1) = 1 \tag{3}$$

Special case #2: $\Gamma(n+1) = n\Gamma(n)$

$$\Gamma(n+1) = \int_0^\infty x^n e^{-x}dx \tag{4}$$

$$= \left[-e^{-x}x^n\right]_0^\infty - \int_0^\infty n\left(-e^{-x}\right)x^{n-1}dx \tag{5}$$

$$= 0 + n\int_0^\infty x^{n-1}e^{-x}dx \tag{6}$$

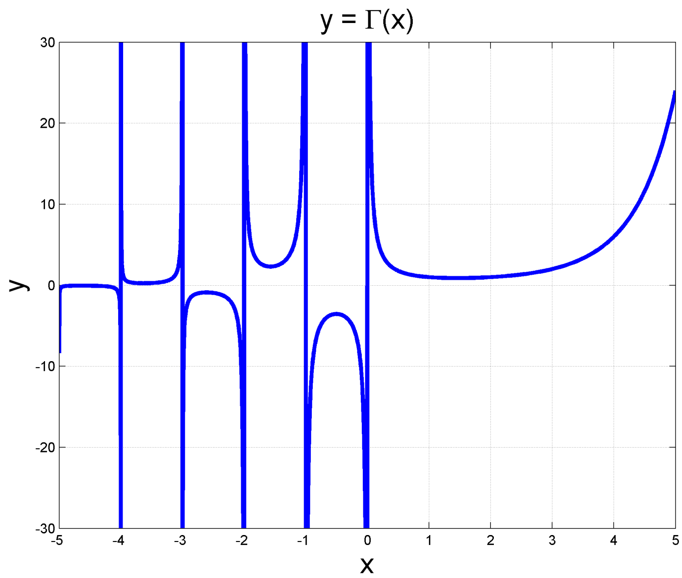

Figure 1: Plot of the gamma function $\Gamma(x) = \int_0^\infty w^{x-1}e^{-w}dw$. Note it is undefined for $x = 0$ and negative integers.

Therefore

$$\Gamma(n+1) = n\Gamma(n) \tag{7}$$

1.1.3 $n! = \int_0^\infty x^n e^{-x}dx$

Using $\Gamma(1) = 1$, and taking n to be a positive integer

$$\begin{aligned}
\Gamma(2) &= \Gamma(1) = 1 && (8)\\
\Gamma(3) &= 2\Gamma(2) = 2 && (9)\\
\Gamma(4) &= 3\Gamma(3) = 3 \times 2 = 6 && (10)\\
\Gamma(5) &= 4\Gamma(4) = 4 \times 3 \times 2 = 24 && (11)
\end{aligned}$$

Hence

$$\boxed{\Gamma(n+1) = n!} \tag{12}$$

We can therefore write

$$\boxed{n! = \int_0^\infty x^n e^{-x}dx} \tag{13}$$

1.2 Stirling's approximation

In the previous section we derived the result

$$n! = \int_0^\infty x^n e^{-x}dx \tag{14}$$

Now $x^n = \left(e^{\ln x}\right)^n = e^{n\ln x}$, hence

$$n! = \int_0^\infty e^{n\ln x - x}dx \tag{15}$$

We can expand this by using the substitution $x = n + y$

$$n! = \int_{-n}^{\infty} \exp\{n\ln(n+y) - n - y\}\,dy \tag{16}$$

Now

$$\ln(n+y) = \ln\left(n\left(1+\frac{y}{n}\right)\right) = \ln n + \ln\left(1+\frac{y}{n}\right) \tag{17}$$

The Maclaurin series for $\ln(1+x)$ is

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + ... \tag{18}$$

which converges if $|x| < 1$. Hence if we can accept the argument that "n is sufficient large such that $\frac{y}{n} < 1$" then we can write

$$\ln\left(1+\frac{y}{n}\right) \approx \frac{y}{n} - \frac{y^2}{2n^2} + \frac{y^3}{3n^3} - \frac{y^4}{4n^4} + ... \tag{19}$$

Hence

$$\begin{aligned} n! &= \int_{-n}^{\infty} \exp\left\{n\left(\ln n + \frac{y}{n} - \frac{y^2}{2n^2} + \frac{y^3}{3n^3} - \frac{y^4}{4n^4} + ...\right) - n - y\right\} dy & (20)\\ &= \int_{-n}^{\infty} e^{n\ln n - n} \exp\left\{y - \frac{y^2}{2n} + \frac{y^3}{3n^2} - \frac{y^4}{4n^3} + ... - y\right\} dy & (21)\\ &= n^n e^{-n} \int_{-n}^{\infty} \exp\left\{-\frac{y^2}{2n} + \frac{y^3}{3n^2} - \frac{y^4}{4n^3} + ...\right\} dy & (22) \end{aligned}$$

where the last step uses the result $e^{n\ln n - n} = \left(e^{\ln n}\right)^n e^{-n} = n^n e^{-n}$

Now if $n \to \infty$

$$n! \approx n^n e^{-n} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2n}}\,dy \tag{23}$$

To evaluate this further consider the integral

$$I = \int_{-\infty}^{\infty} e^{-y^2}\,dy \tag{24}$$

I^2 can be written as

$$I^2 = \int_{-\infty}^{\infty} e^{-y^2}\,dy \int_{-\infty}^{\infty} e^{-x^2}\,dx \tag{25}$$

which can be interpreted as a double-integral, or area-finding integral over the x, y plane

$$I^2 = \int_{x=-\infty}^{\infty}\int_{y=-\infty}^{\infty} e^{-(x^2+y^2)}\,dxdy \tag{26}$$

Changing to plane-polar coordinates r, θ

$$\begin{aligned} x^2 + y^2 &= r^2 & (27)\\ dxdy &= r\,dr\,d\theta & (28) \end{aligned}$$

Hence

$$\begin{aligned} I^2 &= \int_{\theta=0}^{2\pi} d\theta \int_{r=0}^{\infty} re^{-r^2}\,dr & (29)\\ &= 2\pi \int_{r=0}^{\infty} re^{-r^2}\,dr & (30) \end{aligned}$$

Now

$$\frac{d}{dr}\left(e^{-r^2}\right) = -2re^{-r^2} \tag{31}$$

$$\therefore \quad -\frac{1}{2}\frac{d}{dr}\left(e^{-r^2}\right) = re^{-r^2} \tag{32}$$

$$\therefore \quad \int re^{-r^2}dr = -\frac{1}{2}e^{-r^2} + c \tag{33}$$

Hence

$$\int_{r=0}^{\infty} re^{-r^2}dr = -\frac{1}{2}\left[e^{-r^2}\right]_0^{\infty} = \frac{1}{2} \tag{34}$$

Therefore

$$I^2 = 2\pi\int_{r=0}^{\infty} re^{-r^2}dr = \pi \tag{35}$$

Which means

$$\boxed{\int_{-\infty}^{\infty} e^{-y^2}dy = \sqrt{\pi}} \tag{36}$$

Hence, using substitution $k = \frac{y}{\sqrt{2n}}$

$$\int_{-\infty}^{\infty} e^{-\frac{y^2}{2n}}dy = \sqrt{2n}\int_{-\infty}^{\infty} e^{-k^2}dk = \sqrt{2\pi n} \tag{37}$$

Stirling's approximation is therefore, for large n

$$\boxed{n! \approx n^n e^{-n}\sqrt{2\pi n}} \tag{38}$$

2 The Poisson Distribution

2.1 Deriving the Poisson distribution as a limit of the Binomial distribution

Let us firstly consider the *Binomial Distribution*, that is the probability of x successes out of n independent binary outcomes, (i.e. success or failure) where the probability of success in each 'trial' is p

$$\boxed{P(x) = \frac{n!}{(n-x)!x!}p^x(1-p)^{n-x}} \tag{39}$$

The probability of a sequence of x successes followed by $n-x$ failures is $p^x(1-p)^{n-x}$. The number of permutations of x sucesses and $n-x$ failures is $\frac{n!}{(n-x)!x!}$, hence the result above.

For a random variable to be *Poisson distributed*, let us assume success occurs at a average 'rate' λ. If there were n binary trials occuring one after each other within a given time interval then we might expect $\lambda = np$ of them to be successful, since $E[x] = np$ is the *expectation* of the Binomial distribution.[1] In this case x is the number of successes with the time interval, i.e. corresponds to a (random) *success rate*. For example, the probability of a goal resulting from any given kick in a soccer game is fairly low. There are probably thousands of kicks per game. However, the expected number of goals scored is likely to be something like 2 or 3 per game. The goals scored per game are therefore likely to obey Poisson statistics.

Let us assert as a condition for Poisson distribution that $n \to \infty$ and $p \to 0$ such that $np = \lambda$ is constant.

$$P(x,\lambda) = \lim_{p\to 0, n\to\infty} \frac{n!}{(n-x)!x!}p^x(1-p)^{n-x} \tag{40}$$

[1] For the Binomial Distribution $P(x) = \frac{n!}{(n-x)!x!}p^x(1-p)^{n-x}$ the expectation (mean) is $E[x] = np$ and the variance is $V[x] = np(1-p)$. This result is derived in later sections.

Now

$$\lim_{n\to\infty}\frac{n!}{(n-x)!} = \lim_{n\to\infty}\left\{\frac{n(n-1)(n-2)...(n-x+1)(n-x)!}{(n-x)!}\right\} \tag{41}$$

$$= \lim_{n\to\infty}\{n(n-1)(n-2)....(n-x+1)\} \tag{42}$$

$$\to n^x \tag{43}$$

Now

$$p^x(1-p)^{n-x} = \frac{p^x}{(1-p)^x}(1-p)^n \tag{44}$$

And

$$(1-p)^n = \left(1-\frac{\lambda}{n}\right)^n \tag{45}$$

$$= 1+n\left(-\frac{\lambda}{n}\right)+\frac{1}{2!}n(n-1)\left(-\frac{\lambda}{n}\right)^2+\frac{1}{3!}n(n-1)(n-2)\left(-\frac{\lambda}{n}\right)^3+... \tag{46}$$

$$= 1-\lambda+\frac{1}{2}\lambda^2\left(1-\frac{1}{n}\right)-\frac{1}{3!}\lambda^2\left(1-\frac{3}{n}+\frac{2}{n^2}\right)+... \tag{47}$$

Therefore

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = 1-\lambda+\frac{1}{2}\lambda^2-\frac{1}{3!}\lambda^2+... \tag{48}$$

which results in the useful result

$$\boxed{\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}} \tag{49}$$

Hence

$$\lim_{p\to 0,n\to\infty}p^x(1-p)^{n-x} = \lim_{p\to 0,n\to\infty}\frac{p^x}{(1-p)^x}(1-p)^n = \lim_{p\to 0,n\to\infty}\frac{p^x}{(1-p)^x}(1-\frac{\lambda}{n})^n \longrightarrow p^x e^{-\lambda} \tag{50}$$

Putting this all together

$$P(x,\lambda) = \lim_{n\to\infty}\frac{n!}{(n-x)!}\times\frac{1}{x!}\times\lim_{p\to 0,n\to\infty}p^x(1-p)^{n-x} \tag{51}$$

$$= n^x\times\frac{1}{x!}\times p^x e^{-\lambda} \tag{52}$$

Since $\lambda = np$ the Poisson distribution is given by

$$\boxed{P(x,\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}} \tag{53}$$

2.2 Deriving the mean of the Poisson distribution

Note this can be done much more elegantly using a Moment Generating Function (MGF).

λ is correctly associated with the mean success rate if it is the expectation $E[x]$ of the Poisson distribution $P(x,\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$. Note $P(x,\lambda)$ is a *discrete* distribution, i.e. random variable x is restricted to being a positive integer, including zero.

$$E[x] = \sum_{x=0}^{\infty}xP(x,\lambda) = e^{-\lambda}\sum_{x=0}^{\infty}\frac{x\lambda^x}{x!} \tag{54}$$

$$\sum_{x=0}^{\infty}\frac{x\lambda^x}{x!} = \sum_{x=1}^{\infty}\frac{\lambda^x}{(x-1)!} \tag{55}$$

To make further progress, let $y = x - 1$

$$\sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} = \sum_{y=0}^{\infty} \frac{\lambda^{y+1}}{y!} = \lambda \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} \tag{56}$$

Now

$$e^{\lambda} = 1 + \lambda + \frac{1}{2!}\lambda^2 + \frac{1}{3!}\lambda^3 + ... = \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} \tag{57}$$

Hence

$$E\left[x\right] = e^{-\lambda} \sum_{x=1}^{\infty} \frac{x\lambda^x}{x!} = e^{-\lambda}\lambda e^{\lambda} = \lambda \tag{58}$$

Therefore

$$P(x, \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{59}$$

$$E[x] = \lambda \tag{60}$$

2.3 Deriving the variance of the Poisson distribution

Note this can be done much more elegantly using a Moment Generating Function (MGF).

By definition, the variance $V[x]$ of the Poisson distribution is given by

$$V[x] = \sum_{x=0}^{\infty} x^2 P(x, \lambda) - \left(E\left[x\right]\right)^2 \tag{61}$$

Now

$$\sum_{x=0}^{\infty} x^2 P(x, \lambda) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{x^2\lambda^x}{x!} \tag{62}$$

and

$$\sum_{x=0}^{\infty} \frac{x^2\lambda^x}{x!} = \sum_{x=1}^{\infty} \frac{x\lambda^x}{(x-1)!} \tag{63}$$

Let $y = x - 1$

$$\sum_{x=1}^{\infty} \frac{x\lambda^x}{(x-1)!} = \sum_{y=0}^{\infty} \frac{(y+1)\lambda^{y+1}}{y!} = \lambda \sum_{y=0}^{\infty} \frac{y\lambda^y}{y!} + \lambda \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} \tag{64}$$

From above

$$\sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = e^{\lambda} \tag{65}$$

Now

$$\sum_{y=0}^{\infty} \frac{y\lambda^y}{y!} = \sum_{y=1}^{\infty} \frac{\lambda^y}{(y-1)!} \tag{66}$$

If $z = y - 1$

$$\sum_{y=1}^{\infty} \frac{\lambda^y}{(y-1)!} = \sum_{z=0}^{\infty} \frac{\lambda^{z+1}}{z!} = \lambda \sum_{z=0}^{\infty} \frac{\lambda^z}{z!} = \lambda e^{\lambda} \tag{67}$$

Hence

$$\sum_{x=1}^{\infty} \frac{x\lambda^x}{(x-1)!} = \lambda^2 e^{\lambda} + \lambda^2 \tag{68}$$

Hence using $E[x] = \lambda$

$$\begin{aligned} V[x] &= e^{-\lambda} \sum_{x=1}^{\infty} \frac{x\lambda^x}{(x-1)!} - \lambda^2 && (69) \\ &= e^{-\lambda} \left(\lambda^2 e^{\lambda} + \lambda e^{\lambda} \right) - \lambda^2 && (70) \end{aligned}$$

which means

$$V[x] = \lambda \tag{71}$$

In summary:

$$\begin{aligned} P(x,\lambda) &= \frac{\lambda^x e^{-\lambda}}{x!} && (72) \\ E[x] &= \lambda && (73) \\ V[x] &= \lambda && (74) \end{aligned}$$

3 Definition of the Moment Generating Function (MGF)

An elegant way of deriving the expectation and variance of the Binomial distribution (and in-fact any distribution) is to consider the associated *Moment Generating Function* (MGF) $M(x,t)$. This is defined as

$$M(x,t) = E[e^{tx}] \tag{75}$$

where x is a random variable. Using the Maclaurin expansion of e^{tx}

$$e^{tx} = 1 + tx + \frac{1}{2!}(tx)^2 + \frac{1}{3!}(tx)^3 + ... \tag{76}$$

Hence

$$M(x,t) = 1 + E[x]t + E[x^2]\frac{t^2}{2!} + ... \tag{77}$$

We can therefore work out expectations of (integer) powers of x by finding derivatives of the form

$$E[x^n] = \left. \frac{\partial^n M(x,t)}{\partial t^n} \right|_{t=0} \tag{78}$$

The variance is therefore

$$\begin{aligned} V[x] &= E[x^2] - (E[x])^2 && (79) \\ V[x] &= \left. \frac{\partial^2 M(x,t)}{\partial t^2} \right|_{t=0} - \left(\left. \frac{\partial M(x,t)}{\partial t} \right|_{t=0} \right)^2 && (80) \end{aligned}$$

Moment generating functions for various common distributions are as follows:

Distribution	$P(x)$	$M(x,t)$	$E[x]$	$V[x]$
Gaussian	$\exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$	$\exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$	μ	σ^2
Binomial	$\frac{n!}{(n-x)!x!} p^x (1-p)^{n-x}$	$\left(pe^t + 1 - p\right)^n$	np	$np(1-p)$
Poisson	$\frac{\lambda^x e^{-\lambda}}{x!}$	$e^{\lambda(e^t - 1)}$	λ	λ

(81)

4 Gaussian approximations and the central limit theorem

4.1 Gaussian approximation of the Binomial Distribution

Stirling's approximation $n! \approx n^n e^{-n}\sqrt{2\pi n}$ can be used to approximate the Binomial Distribution when both the number of trials n and number of successes x become large.

$$P(x) = \frac{n!}{(n-x)!x!}p^x(1-p)^{n-x} \tag{82}$$

Hence

$$\begin{aligned} n! &\approx n^n e^{-n}\sqrt{2\pi n} && (83)\\ (n-x)! &\approx (n-x)^{n-x}e^{-(n-x)}\sqrt{2\pi(n-x)} && (84)\\ x! &= x^x e^{-x}\sqrt{2\pi x} && (85) \end{aligned}$$

$$\begin{aligned} P(x) &\approx \frac{n^n e^{-n}\sqrt{2\pi n}}{(n-x)^{n-x}e^{-(n-x)}\sqrt{2\pi(n-x)}x^x e^{-x}\sqrt{2\pi x}}p^x(1-p)^{n-x} && (86)\\ &= \frac{e^{-n}\sqrt{2\pi}n^n\sqrt{n}}{\sqrt{2\pi}\sqrt{2\pi}e^{-n}e^x e^{-x}(n-x)^{n-x}\sqrt{(n-x)}x^x\sqrt{x}}p^x(1-p)^{n-x} && (87)\\ &= \frac{n^n\sqrt{n}}{\sqrt{2\pi}(n-x)^{n-x}\sqrt{(n-x)}x^x\sqrt{x}}p^x(1-p)^{n-x} && (88)\\ &= \frac{n^{n+\frac{1}{2}}}{\sqrt{2\pi}(n-x)^{n-x+\frac{1}{2}}x^{x+\frac{1}{2}}}p^x(1-p)^{n-x} && (89)\\ &= \frac{1}{\sqrt{2\pi}(1-\frac{x}{n})^{n-x+\frac{1}{2}}x^{x+\frac{1}{2}}}p^x(1-p)^{n-x} && (90)\\ &= \frac{1}{\sqrt{2\pi}}(1-\frac{x}{n})^{-n+x-\frac{1}{2}} \times x^{-x-\frac{1}{2}} \times p^x(1-p)^{n-x} && (91) \end{aligned}$$

Now using the identities $x^a = e^{a\ln x}$ and $e^A e^B e^C = e^{A+B+C}$

$$\begin{aligned} &(1-\frac{x}{n})^{-n+x-\frac{1}{2}} \times x^{-x-\frac{1}{2}} \times p^x(1-p)^{n-x} && (92)\\ = \;&\exp\left\{-\left(n-x+\frac{1}{2}\right)\ln(1-\frac{x}{n}) - \left(x+\frac{1}{2}\right)\ln x + x\ln p + (n-x)\ln(1-p)\right\} && (93) \end{aligned}$$

Now consider the change of variable $y \gg 1$

$$y = x - np \tag{94}$$

After lots of tedious algebra[2] it can be shown that:

$$\boxed{P(x) \approx \frac{1}{\sqrt{2\pi np(1-p)}}\exp\left[-\frac{1}{2}\frac{(x-np)^2}{np(1-p)}\right]} \tag{95}$$

i.e. Gaussian in form with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$

[2] Riley, Hobson & Bence don't even attempt it!

4.2 Gaussian approximation of the Poisson Distribution

$$P(x, \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{96}$$

Consider the limit when mean success rate $\lambda \to \infty$. If this is true then the discrete nature of x has limited overall bearing on the shape of the curve of $P(x)$, so we can assume it to be *continuous.* We may also assume that $x \to \infty$.

To find the limiting continuous distribution of $P(x)$ let us first find $\ln P$

$$\ln P = x \ln \lambda - \lambda - \ln x! \tag{97}$$

Using Stirling's approximation, valid for $x \to \infty$

$$x! \approx x^x e^{-x} \sqrt{2\pi x} \tag{98}$$

Hence

$$\ln x! \approx x \ln x - x + \ln \sqrt{2\pi x} \tag{99}$$

Therefore

$$\ln P = x \ln \lambda - \lambda - x \ln x + x - \ln \sqrt{2\pi x} \tag{100}$$

Now we assume the Gaussian $P(x)$ ought to fit most accurately around the mean of $P(x)$, which is at $x = \lambda$. Hence define $x = \lambda + \epsilon$, where $\frac{\epsilon}{\lambda} \ll 1$.

Hence

$$\begin{aligned}
\ln P &= (\lambda + \epsilon) \ln \lambda - \lambda - (\lambda + \epsilon) \ln (\lambda + \epsilon) + \lambda + \epsilon - \ln \sqrt{2\pi (\lambda + \epsilon)} && (101) \\
&= \epsilon + (\lambda + \epsilon) \ln \frac{\lambda}{\lambda + \epsilon} - \ln \sqrt{2\pi (\lambda + \epsilon)} && (102) \\
&= \epsilon + (\lambda + \epsilon) \ln \frac{1}{1 + \frac{\epsilon}{\lambda}} - \ln \sqrt{2\pi\lambda \left(1 + \frac{\epsilon}{\lambda}\right)} && (103) \\
&= \epsilon - (\lambda + \epsilon) \ln \left(1 + \frac{\epsilon}{\lambda}\right) - \ln \sqrt{2\pi\lambda} - \frac{1}{2} \ln \left(1 + \frac{\epsilon}{\lambda}\right) && (104) \\
&= \epsilon - \ln \sqrt{2\pi\lambda} - \left(\lambda + \epsilon + \frac{1}{2}\right) \ln \left(1 + \frac{\epsilon}{\lambda}\right) && (105)
\end{aligned}$$

Now since $\frac{\epsilon}{\lambda} \ll 1$, hence we can use the Maclaurin expansion for $\ln\left(1 + \frac{\epsilon}{\lambda}\right) = \frac{\epsilon}{\lambda} - \frac{\epsilon^2}{2\lambda^2} + ...$

Hence

$$\begin{aligned}
\ln P &= \epsilon - \ln \sqrt{2\pi\lambda} - \left(\lambda + \epsilon + \frac{1}{2}\right)\left(\frac{\epsilon}{\lambda} - \frac{\epsilon^2}{2\lambda^2} + ...\right) && (106) \\
\therefore \quad & \ln P \approx \epsilon - \ln \sqrt{2\pi\lambda} - \epsilon - \frac{\epsilon^2}{\lambda} - \frac{\epsilon}{2\lambda} + \frac{\epsilon^2}{2\lambda} + \frac{\epsilon^2}{4\lambda^2} + ... && (107) \\
\ln P &\approx -\frac{\epsilon^2}{2\lambda} - \ln \sqrt{2\pi\lambda} - \frac{\epsilon}{2\lambda} + \frac{\epsilon^2}{4\lambda^2} + .. && (108)
\end{aligned}$$

Now since $\frac{\epsilon}{\lambda} \ll 1$ then we can ignore $\frac{\epsilon^2}{4\lambda^2}$ and $\frac{\epsilon}{2\lambda}$ terms relative to $\frac{\epsilon^2}{2\lambda}$

Hence

$$\ln P \approx -\frac{\epsilon^2}{2\lambda} - \ln \sqrt{2\pi\lambda} \tag{109}$$

which implies

$$P(x) = \frac{1}{\sqrt{2\pi\lambda}} e^{-\frac{\epsilon^2}{2\lambda}} \tag{110}$$

Substituting for $x = \lambda + \epsilon$ we arrive at a Gaussian form for the Poisson distribution, with mean and variance λ

$$\boxed{P(x) = \frac{1}{\sqrt{2\pi\lambda}} e^{-\frac{(x-\lambda)^2}{2\lambda}}} \tag{111}$$

4.3 The Central Limit Theorem

Define x to be a random variable, e.g. a random quantity that can be measured n times. Let us assume each measurement is independent. Let us also assert that the i^{th} measurement x_i derives from a probability distribution with expectation $\mu_i = E[x_i]$ and variance $\sigma_i^2 = V[x_i]$.

Define another random variable

$$Z = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{112}$$

i.e. the mean average of the measurements.

The expectated value of Z is therefore

$$\mu_Z = E[Z] = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \frac{1}{n}\sum_{i=1}^{n}\mu_i \tag{113}$$

The variance is $\sigma_Z^2 = V[x] = E[Z^2] - \mu_Z^2$

$$\sigma_Z^2 = V[x] = E[Z^2] - \mu_Z^2 \tag{114}$$

$$E[Z^2] = E\left[\frac{1}{n^2}\left(\sum_{i=1}^{n} x_i\right)^2\right] \tag{115}$$

Since measurements are deemed independent

$$\sum_{j=1}^{n}\sum_{i=1}^{n} x_i x_j = \sum_{i=1}^{n} x_i^2 \tag{116}$$

Therefore

$$\left(\sum_{i=1}^{n} x_i\right)^2 = \sum_{i=1}^{n} x_i^2 \tag{117}$$

Hence

$$E[Z^2] = E\left[\frac{1}{n^2}\sum_{i=1}^{n} x_i^2\right] = \frac{1}{n^2}\sum_{i=1}^{n} E[x_i^2] \tag{118}$$

Now

$$E[x_i^2] = \sigma_i^2 + \mu_i^2 \tag{119}$$

Therefore

$$\begin{aligned}
\sigma_Z^2 &= V[x] = E[Z^2] - \mu_Z^2 && (120)\\
&= \frac{1}{n^2}\sum_{i=1}^{n} E[x_i^2] - \mu_Z^2 && (121)\\
&= \frac{1}{n^2}\sum_{i=1}^{n} \left(\sigma_i^2 + \mu_i^2\right) - \mu_Z^2 && (122)\\
&= \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2 + \frac{1}{n^2}\sum_{i=1}^{n}\mu_i^2 - \mu_Z^2 && (123)
\end{aligned}$$

Again invoking variable independence, hence

$$\frac{1}{n^2}\sum_{i=1}^{n}\mu_i^2 = \left(\frac{1}{n}\sum_{i=1}^{n}\mu_i\right)^2 = \mu_Z^2 \tag{124}$$

Therefore

$$\sigma_Z^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2 \tag{125}$$

The Moment Generating Function $M(Z,t)$ is defined to be

$$M(Z,t) = E[e^{Zt}] = E\left[\exp\left(\frac{t}{n}\sum_{i=1}^{n}x_i\right)\right] = E\left[\prod_{i=1}^{n}e^{\frac{tx_i}{n}}\right] \tag{126}$$

Now since each measurement x_i is independent

$$E\left[\prod_{i=1}^{n}e^{\frac{tx_i}{n}}\right] = \prod_{i=1}^{n}E[e^{\frac{tx_i}{n}}] \tag{127}$$

Now

$$e^{\frac{tx_i}{n}} = 1 + \frac{tx_i}{n} + \frac{1}{2!}\left(\frac{tx_i}{n}\right)^2 + \frac{1}{3!}\left(\frac{tx_i}{n}\right)^3 + ... \tag{128}$$

Therefore

$$\begin{aligned} E[e^{\frac{tx_i}{n}}] &= 1 + \frac{t}{n}E[x_i] + \frac{1}{2!}\left(\frac{t}{n}\right)^2 E[x_i^2] + \frac{1}{3!}\left(\frac{t}{n}\right)^3 E[x_i^3] + ... & (129)\\ &= 1 + \frac{t}{n}\mu_i + \frac{1}{2}\frac{t^2}{n^2}\left(\sigma_i^2 + \mu_i^2\right) + ... & (130) \end{aligned}$$

Hence

$$M(Z,t) = \prod_{i=1}^{n}E[e^{\frac{tx_i}{n}}] = \prod_{i=1}^{n}\left(1 + \frac{t}{n}\mu_i + \frac{1}{2}\frac{t^2}{n^2}\left(\sigma_i^2 + \mu_i^2\right) + ...\right) \tag{131}$$

Now consider the expansion of

$$\begin{aligned} \exp\left(\frac{t}{n}\mu_i + \frac{1}{2}\frac{t^2}{n^2}\sigma_i^2\right) &= 1 + \frac{t}{n}\mu_i + \frac{1}{2}\frac{t^2}{n^2}\sigma_i^2 + \frac{1}{2!}\left(\frac{t}{n}\mu_i + \frac{1}{2}\frac{t^2}{n^2}\sigma_i^2\right)^2 + ... & (132)\\ &= 1 + \frac{t}{n}\mu_i + \frac{1}{2}\frac{t^2}{n^2}\left(\sigma_i^2 + \mu_i^2\right) + ... & (133) \end{aligned}$$

Hence if n is large

$$\exp\left(\frac{t}{n}\mu_i + \frac{1}{2}\frac{t^2}{n^2}\sigma_i^2\right) \approx 1 + \frac{t}{n}\mu_i + \frac{1}{2}\frac{t^2}{n^2}\left(\sigma_i^2 + \mu_i^2\right) \tag{134}$$

Therefore in this limit

$$\begin{aligned} M(Z,t) &= \prod_{i=1}^{n}\left(1 + \frac{t}{n}\mu_i + \frac{1}{2}\frac{t^2}{n^2}\left(\sigma_i^2 + \mu_i^2\right) + ...\right) \approx \prod_{i=1}^{n}\exp\left(\frac{t}{n}\mu_i + \frac{1}{2}\frac{t^2}{n^2}\sigma_i^2\right) & (135)\\ &= \exp\left(\frac{t}{n}\sum_{i=1}^{n}\mu_i + \frac{1}{2}\frac{t^2}{n^2}\sum_{i=1}^{n}\sigma_i^2\right) & (136) \end{aligned}$$

Using the above results for the expectation and variance of Z

$$\begin{aligned} \mu_Z &= \frac{1}{n}\sum_{i=1}^{n}\mu_i & (137)\\ \sigma_Z^2 &= \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2 & (138) \end{aligned}$$

This means

$$\boxed{M(Z,t) = \exp\left(t\mu_Z + \tfrac{1}{2}t^2\sigma_Z^2\right)} \tag{139}$$

This is the MGF for a Gaussian distribution with mean μ_Z and variance σ_Z^2.

The above argument justifies the *The Central Limit Theorem*, which states:

"The distribution of the mean values of a set of independent random values tends towards a Gaussian Distribution if the number of samples is large enough."

5 Expectation and Variance of the Poisson and Binomial distributions via MGF

5.1 Deriving the expectation and variance of the Poisson Distribution via the MGF

For the Poisson distribution the MGF is

$$M(x,t) = E[e^{tx}] = \sum_{x=0}^{\infty} e^{tx}\frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\left(\lambda e^t\right)^x}{x!} \tag{140}$$

Now

$$\sum_{x=0}^{\infty}\frac{k^x}{x!} = e^k \tag{141}$$

Therefore

$$\sum_{x=0}^{\infty}\frac{\left(\lambda e^t\right)^x}{x!} = e^{\lambda e^t} \tag{142}$$

Hence

$$\boxed{M(x,t) = e^{\lambda\left(e^t-1\right)}} \tag{143}$$

Therefore

$$\begin{aligned}
E[x] &= \left.\frac{\partial M(x,t)}{\partial t}\right|_{t=0} && (144)\\
&= \left.\frac{\partial}{\partial t}e^{\lambda\left(e^t-1\right)}\right|_{t=0} && (145)\\
&= \left.\lambda e^t e^{\lambda\left(e^t-1\right)})\right|_{t=0} && (146)\\
&= \lambda && (147)
\end{aligned}$$

and

$$\begin{aligned}
V[x] &= \left.\frac{\partial^2 M(x,t)}{\partial t^2}\right|_{t=0} - \left(\left.\frac{\partial M(x,t)}{\partial t}\right|_{t=0}\right)^2 && (148)\\
&= \left.\frac{\partial}{\partial t}\lambda e^t e^{\lambda\left(e^t-1\right)}\right|_{t=0} - \lambda^2 && (149)\\
&= \left.\lambda\left(e^t e^{\lambda\left(e^t-1\right)}) + \lambda e^t e^{\lambda\left(e^t-1\right)}\right)\right|_{t=0} - \lambda^2 && (150)\\
&= \lambda + \lambda^2 - \lambda^2 && (151)\\
&= \lambda && (152)
\end{aligned}$$

5.2 Deriving the expectation and variance of the Binomial Distribution via the MGF

The Binomial Distribution has the form

$$P(x) = \frac{n!}{(n-x)!x!}p^x(1-p)^{n-x} \tag{153}$$

where x is the number of successes in n independent binary-outcome trials, in which the probability of sucess is always p.

Let x be the sum of a set of random variables $\{y_i\}$

$$x = \sum_{i=1}^{n} y_i \tag{154}$$

where

$$y_i = \begin{cases} 1 & \text{Success in trial } i \\ 0 & \text{Failure in trial } i \end{cases} \tag{155}$$

The Moment Generating Function (MGF) for the Binomial distribution is

$$M(x,t) = E[e^{tx}] = E\left[\exp\left\{t\sum_{i=1}^{n} y_i\right\}\right] = E\left[\prod_{i=1}^{n} e^{ty_i}\right] \tag{156}$$

Now if two random variables X and Y are independent

$$E[xy] = E[x]E[y] \tag{157}$$

Therefore

$$M(x,t) = E\left[\prod_{i=1}^{n} e^{ty_i}\right] = \prod_{i=1}^{n} E[e^{ty_i}] \tag{158}$$

Now since in trial i the probability of success $(y_i = 1)$ is p , and failure $(y_i = 0)$ is $1-p$

$$E[e^{ty_i}] = pe^t + 1 - p \tag{159}$$

i.e. it is the same for each trial (as one would intuitively expect given p is trial independent).

Therefore the MGF for the Binomial distribution is

$$M(x,t) = \prod_{i=1}^{n} E[e^{ty_i}] = \left(pe^t + 1 - p\right)^n \tag{160}$$

The MGF can then be used to efficiently derive the expectation and variance of the Binomial Distribution

$$\left.\frac{\partial M(x,t)}{\partial t}\right|_{t=0} = npe^t\left(pe^t + 1 - p\right)^{n-1}\Big|_{t=0} = np \tag{161}$$

Therefore

$$E[x] = np \tag{162}$$

$$\begin{aligned} \frac{\partial^2 M(x,t)}{\partial t^2} &= \frac{\partial}{\partial t}\left(npe^t\left(pe^t + 1 - p\right)^{n-1}\right) && (163) \\ &= npe^t(n-1)pe^t + \left(pe^t + 1 - p\right)^{n-1} npe^t && (164) \end{aligned}$$

$$\therefore \left.\frac{\partial^2 M(x,t)}{\partial t^2}\right|_{t=0} = np^2(n-1) + np = n^2p^2 - np^2 + np \tag{165}$$

The variance of the Binomial distribution is therefore

$$\begin{aligned} V[x] &= \left.\frac{\partial^2 M(x,t)}{\partial t^2}\right|_{t=0} - \left(\left.\frac{\partial M(x,t)}{\partial t}\right|_{t=0}\right)^2 && (166) \\ &= n^2p^2 - np^2 + np - n^2p^2 && (167) \\ &= np(1-p) && (168) \end{aligned}$$