Probability

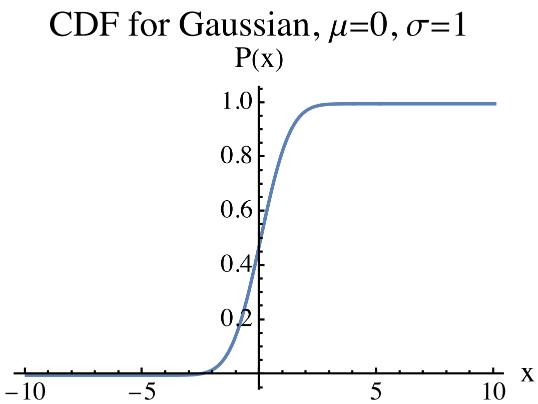
Ruihua Fan

Date

Some Useful Concepts

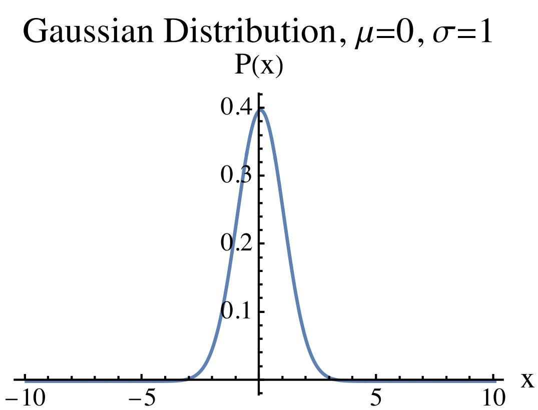
■ *Cumulative distribution function* (CDF) $C(x)$

The probability for the outcome lying in the interval $(-\infty, x]$. It is a monotonically increasing function with limit $C(x \to \infty) = 1$. E.g. ,



■ *Probability distribution function* (PDF) $P(x)$

$P(x) = d\,C(x) / d\,x$. Therefore it is normalized $\int P(x)\,d\,x = 1$. And the probability for the outcome lying in the interval $[x, x + d\,x]$ can be written as $P(x)\,d\,x$. E.g. ,



- For many random variables $\{x_1, x_2, \ldots x_N\}$, we can define the *joint probability distribution function* $P_N(x_1, \ldots, x_N)$. If they are independent, we have, $P_N(x_1, \ldots, x_N) = P_1(x_1) \ldots P_N(x_N)$.

Mean, Variance & Moments

- *Expectation value* or *mean*

$$\mu = \int x\, P(x)\, d\, x$$

- *Variance* and *standard deviation*

$$\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2 = \int x^2\, P(x)\, d\, x - \left(\int x\, P(x)\, d\, x \right)^2$$

- *nth Moment* , depending on the distribution function, it may or may not exist.

$$m_n = \int x^m\, P(x)\, d\, x$$

PDF: $\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma}$			
m_1	m_2	m_3	m_4
μ	$\mu^2 + \sigma^2$	$\mu^3 + 3\,\mu\,\sigma^2$	$\mu^4 + 6\,\mu^2\,\sigma^2 + 3\,\sigma^4$

Generating functions

- *Characteristic function* or *generating function of moments*. It is simply the Fourier transformation of the probability distribution

$$\tilde{p}(k) = \left\langle e^{-i\,k\,x} \right\rangle = \int e^{-i\,k\,x}\, P(x)\, d\, x\, .$$

Its Taylor expansion gives all the moments,

$$\tilde{p}(k) = \sum_{n=0}^{\infty} \frac{(-i\,k)^n}{n!} \langle x^n \rangle$$

- *The generating function of cumulants*. It is simply the logarithm the the Fourier transformation

$$\log \tilde{p}(k) = \log\left(1 + \sum_{n=1}^{\infty} \frac{(-i\,k)^n}{n!} \langle x^n \rangle\right) = \sum_{n=1}^{\infty} \frac{(-i\,k)^n}{n!} \langle x^n \rangle_C,$$

where $\langle x^n \rangle_C$ is called the nth cumulant, which is analogous to the connected n-point function in QFT. The first three cumulants are,

$\langle x \rangle_C = \langle x \rangle$, mean;

$\langle x^2 \rangle_C = \langle x^2 \rangle - \langle x \rangle_C{}^2$, variance;

$\langle x^3 \rangle_C = \langle x^3 \rangle - 3\,\langle x^2 \rangle_C\,\langle x \rangle - \langle x \rangle_C{}^3$, skewness.

For example, in Gaussian distribution, all higher order cumulants vanishes; in Poisson distribution, all cumulants are the same.

PDF: $\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma}$			
C_1	C_2	C_3	C_4
μ	σ^2		

PDF: $e^{-\lambda}\lambda^x/x!, x\in N$				
C_1	C_2	C_3	C_4	C_5
λ	λ	λ	λ	λ

Central Limit Theorem

Motivation

For a generic many-body (macroscopic) system, experiments can only access those macroscopic quantities, such as energy, pressure, temperature and so on. Those macroscopic quantities usually can be expressed as a summation over microscopic degrees of freedom, i.e.

$$Q = \frac{1}{N}\left(q_1 + q_2 + \ldots + q_N\right), \tag{1}$$

where q denotes the physical quantity and $N >> 1$ is the number of particles.

There can be complicated interactions among particles such that q_i's may develop complicated correlation and follow a complicated joint probability distribution. However, due to $N >> 1$, the average may make Q follow a relative simple rule.

The central limit theorem partially gives a positive answer to this question. If q_i's are all independent and satisfy an identical distribution (the second assumption is not necessary but will make the discussion much simpler), as N goes large, Q will satisfy a Gaussian distribution, which is determined by the probability distribution of q_i .

Uniform Distribution

We first consider the *uniform distribution*,

$$P(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & x < 0 \,\&\, x > 1 \end{cases}. \tag{2}$$

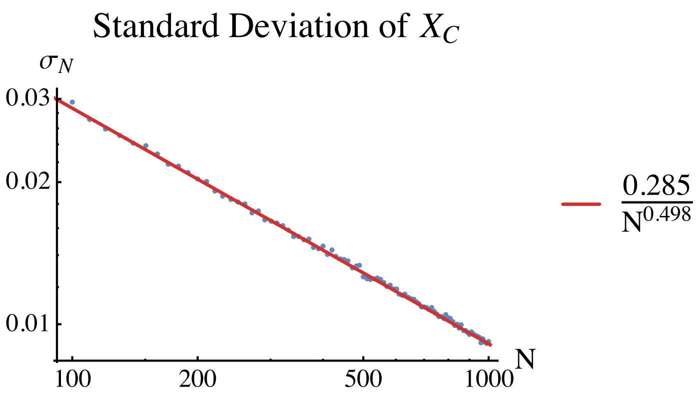
Its mean is $\langle x\rangle = 1/2$. Its standard deviation is $\sigma = \sqrt{\langle x^2\rangle - \langle x\rangle^2} = \frac{1}{2\sqrt{3}} \simeq 0.289$.

Now we study the properties of the probability distribution of

$$X_C = \frac{x_1 + x_2 + \ldots + x_N}{N},$$

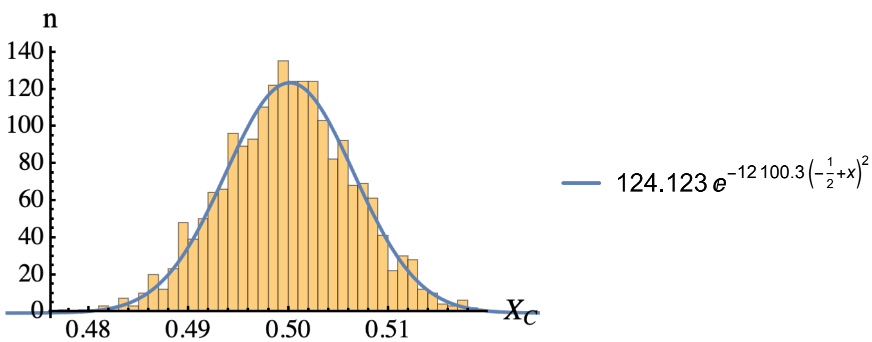
where x_i's are all independent variables drawn from the uniform distribution $P(x)$. X_C indicates that it is like a center-of-mass position.

It is easy to see that $\langle X_C \rangle = \langle x \rangle = 1/2$. But how about its standard deviation? We can first do a numerics.



One can see that σ_N decreases in a fashion as $1/\sqrt{N}$

We can also check its profile, which looks close to a Gaussian distribution.



■ Binomial Distribution

Here we consider the *Binary distribution* with success probability $0 < p < 1$,

$$P(x) = \begin{cases} p & x \in \text{success} \\ 1-p & x \in \text{fail} \end{cases} . \tag{3}$$

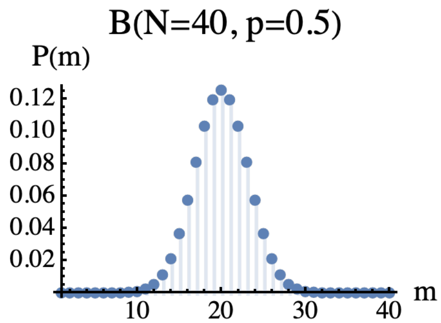
We can use 1 to represent success and 0 to represent fail. If we do the experiment (like tossing a coin) N times, and denote the total outcome as

$$\text{outcome} = x_1 + x_2 + \ldots x_N$$

then the probability for having outcome $= m \in Z$ is

$$\binom{N}{m} p^m (1-p)^{N-m} \tag{4}$$

m satisfies the *Binomial distribution* $B(N,\, p)$, e. g.



For N times repeated experiments and success probability p, the Binomial distribution $B(N, p)$ has,

$$\langle m \rangle = N\,p, \quad {\sigma_m}^2 = N\,p(1-p) \tag{5}$$

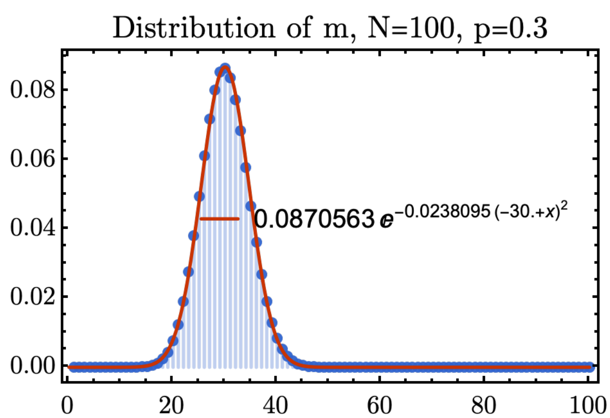
We can also normalize the outcome by N, i.e. define the averaged outcome,

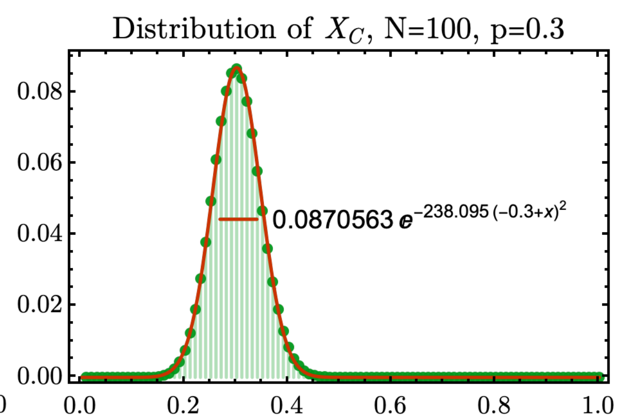
$$X_C = \frac{x_1 + x_2 + \ldots + x_N}{N},$$

which will have,

$$\langle X_C \rangle = p\,, \quad {\sigma_{X_C}}^2 = \frac{p(1-p)}{N} \tag{6}$$

Again, the standard deviation decreases in a $1/\sqrt{N}$ fashion. And The probability distribution for X_C or m will approach a Gaussian distribution as N increases,





The fact that $B(N, p)$ as N increases approaches a Gaussian distribution can also be shown analytically.

$$B(N,\,p;\;m) = \binom{N}{m} p^m (1-p)^{N-m} = \frac{N\,!}{m\,!\,(N-m)\,!}\,p^m (1-p)^{N-m}$$

Now we use Stirling approximation (equivalent to one kind of steepest descent method): $\log N! \simeq N \log N - N + \frac{1}{2}\log 2\pi N$ then we have,

$$B(N,\,p;\;m) \simeq \left(\frac{p}{m}\right)^m \left(\frac{1-p}{N-m}\right)^{N-m} N^N \sqrt{\frac{N}{2\pi\,m(N-m)}}$$

Introducing $x = \frac{m - N p}{N}$, we have,

$$P(x) \simeq \left(\frac{p}{p+x}\right)^{N(x+p)} \left(\frac{1-p}{1-p-x}\right)^{N(1-p-x)} \sqrt{\frac{N}{2\pi\,(x+p)(1-p-x)}}$$

If we only focus on the region near the center of the distribution, i.e. $x << 1$, then we can approximate,

$$\log P(x) = -N(x+p)\log\left(1+\frac{x}{p}\right) - N(1-p-x)\log\left(1-\frac{x}{1-p}\right) + \frac{1}{2}\log(2\pi N(x+p)(1-p-x))$$

$$\simeq -N(x+p)\left(\frac{x}{p} - \frac{x^2}{2p^2}\right) + N(1-p-x)\left(\frac{x}{1-p} + \frac{x^2}{2(1-p)^2}\right) + \frac{1}{2}\log(2\pi N(x+p)(1-p-x))$$

So we have,

$$P(x) \simeq \sqrt{\frac{N}{2\pi\, p(1-p)}}\,(1+O(x))\exp\left(-\frac{x^2}{2p(1-p)/N} + O\left(N x^3\right)\right) \tag{7}$$

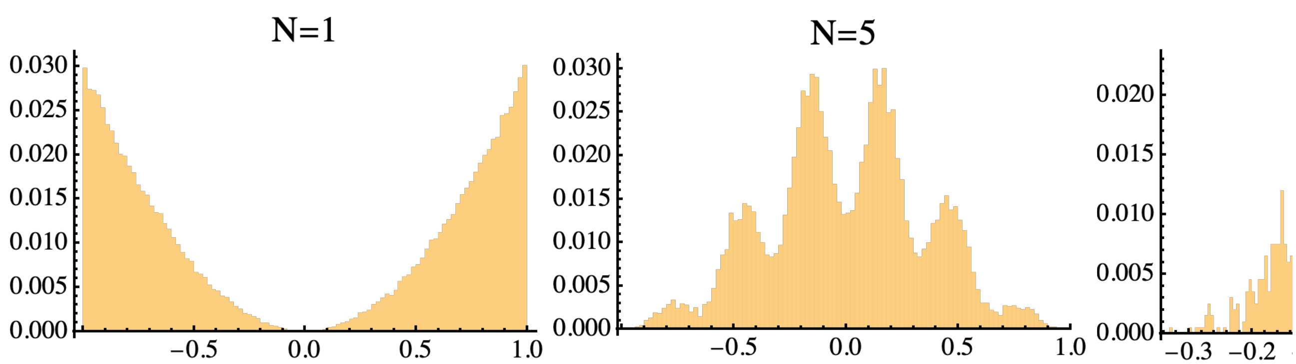
or if we use $\delta = N x$, we have

$$P(\delta) \simeq \sqrt{\frac{1}{2\pi N p(1-p)}}\left(1+O\left(\frac{\delta}{N}\right)\right)\exp\left(-\frac{\delta^2}{2p(1-p)N} + O\left(\frac{\delta^3}{N^2}\right)\right), \tag{8}$$

which as N increases indeed approaches a normal distribution.

Other random examples

Here let's randomly take some distribution function $P(x)$, $0 < x < 1$ and see what the result look like.



The General Theorem

In the examples above, we see that as we draw more from the same probability distribution, the average quantity will satisfy a Gaussian distribution. As we said in the first section, this is not accident but reveals a very deep mathematical fact: the (classical) *central limit theorem (CLT)*.

CLT: Let's take an arbitrary probability distribution $P(x)$ with finite mean μ and variance σ^2. $\{x_1, \ldots, x_N\}$ are N *i.d.d. random variables* drawn from $P(x)$. If we look at the average

$X_C = \frac{1}{N} \sum_{i=1}^{N} x_i$, its probability distribution will approach a normal distribution with mean μ and variance σ^2 / N .

Proof:

By definition, the probability distribution of X_C is,

$$P_N(X_C) = \int d\, x_1 \ldots d\, x_N\, \delta\left(X_C - \frac{1}{N} \sum_{i=1}^{N} x_i\right) P(x_1) \ldots P(x_N)\,,$$

where we have used the i.d.d. condition. Therefore, its Fourier transformation (the so called *characteristic function*) is,

$$\int_{-\infty}^{\infty} e^{-i\,k\,X_C}\, P(X_C)\, d\, X_C = \int d\, x_1 \ldots d\, x_N\, e^{-i\,k\,\frac{1}{N}\sum_{i=1}^{N} x_i}\, P(x_1) \ldots P(x_N) \;= \left(\int d\, x\, e^{-i\,\frac{k}{N}\,x}\, P(x)\right)^N$$

$$= \left[\int d\, x \left(1 - i\,\frac{k}{N}\,x + \frac{1}{2}\left(i\,\frac{k}{N}\,x\right)^2 + O\left(\frac{1}{N^3}\right)\right) P(x)\right]^N$$

$$= \left[1 - i\,\frac{k}{N}\,\mu - \frac{1}{2}\left(\frac{k}{N}\right)^2 \langle x^2\rangle + O\left(\frac{1}{N^3}\right)\right]^N$$

$$= \exp\left[N \log\left(1 - i\,\frac{k}{N}\,\mu - \frac{1}{2}\left(\frac{k}{N}\right)^2 \langle x^2\rangle + O\left(\frac{1}{N^3}\right)\right)\right]$$

In the large-N limit, we can expand the log function and get,

$$\int_{-\infty}^{\infty} e^{-i\,k\,X_C}\, P(X_C)\, d\, X_C = \exp\left[N\left(-i\,\frac{k}{N}\,\mu - \frac{1}{2}\left(\frac{k}{N}\right)^2 \sigma^2 + O\left(\frac{1}{N^3}\right)\right)\right]$$

(I) If $P(x)$ decays fast enough for large $|x - \mu|$, then $\langle x^m\rangle$, $m \geq 2$ are all finite and we can drop the $O(N^{-3})$ terms at large N. Finally, by doing an inverse Fourier transformation, we get,

$$P_N(X_C) = \frac{1}{\sqrt{2\,\pi\,\sigma^2 / N}} \exp\left(-\frac{(X_C - \mu)^2}{2\,\sigma^2 / N}\right), \;\; N \to \infty$$

(II) If $P(x)$ decays too slow, variance and higher order moments may not exist. For example, $P(x) \sim x^{-1-\alpha}$, $0<\alpha<2$ for large x. Then $P_N(X_C)$ still converge to a certain distribution but not Gaussian in general.

QED

Poisson Distribution

■ Derivation

Poisson distribution is also a very common PDF. It appears when:

(1) observing some event over a time period T

(2) the probabilities for the event happening at different times are independent

Classical examples are nuclei decay, receiving phone calls and so on.

Assume the probability for one event in a small time interval dt is λdt, where $T = M\,\mathrm{dt}$. The probability that no event happening in time T is,

$$P_0(\lambda,\ T) = (1 - \lambda\,\mathrm{dt})^M = \left(1 - \lambda\,\frac{T}{M}\right)^M = e^{-\lambda\,T}.$$

Now we want to know after time T, the probability for m events.

We divide T into $M >> 1$ small intervals. In each interval, the probability for one event is $\lambda\,\mathrm{dt}$, the probability for no event is $1 - \lambda\,\mathrm{dt}$ and we omit the probability for multi-event since it is too small.

Therefore, in each interval we have a independent binary distribution with random variable m_i.The total outcome $m = m_1 + \ldots + m_M$. The characteristic function for each binary distribution is

$$p\,e^{-i\,k} + q,\ \ p = \lambda\,\mathrm{dt},\ \ q = 1 - p$$

Those binary distribution for different intervals are independent. Thus the characteristic function for the total distribution is,

$$\left(\lambda\,\mathrm{dt}\,e^{-i\,k} + 1 - \lambda\,\mathrm{dt}\right)^M = \left(1 + \left(e^{-i\,k} - 1\right)\lambda\,\frac{T}{M}\right)^M = \exp\!\left(\left(e^{-i\,k} - 1\right)\lambda\,T\right) = e^{-\lambda\,T}\sum_{m=0}^{\infty}\frac{(\lambda\,T)^m}{m!}\,e^{-i\,k\,m}.$$

We can perform an inverse Fourier transformation to get

$$P(\lambda\,T;\,x) = e^{-\lambda\,T}\sum_{m=0}^{\infty}\delta(x - m)\,\frac{(\lambda\,T)^m}{m!} \qquad (9)$$

Properties

Fix T

The physical meaning of $P(\lambda\,T;\,m)$ is fix a time, the probability for m events happening.

- The self-consistent condition,

```
Sum[Exp[-λT] (λT)^m/m!, {m, 0, ∞}]
```

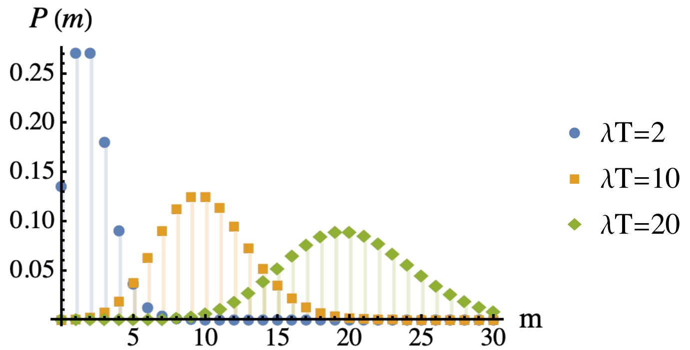
```
1
```

- Mean value of m,

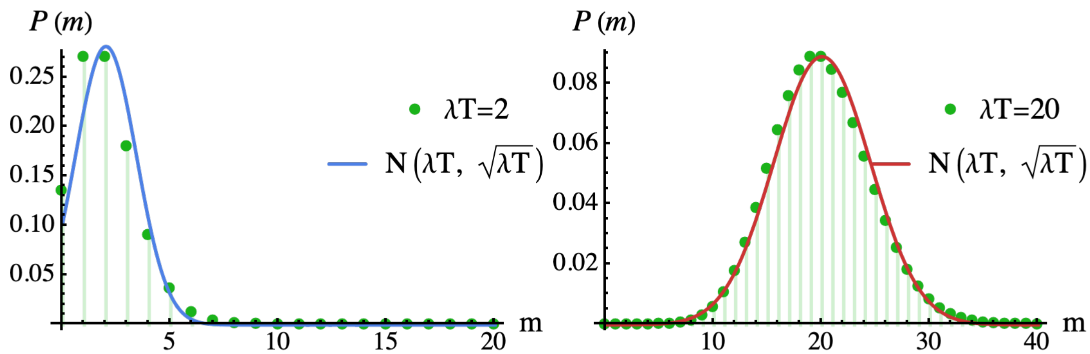
```
Sum[m Exp[-λT] (λT)^m/m!, {m, 0, ∞}]
```

```
λT
```

- Demonstration of the CLT: From the derivation, we see that m can be interpreted as a summation over many i.d.d. random variables drawn from a binary distribution. Therefore, it is not surprising that as T is large, the Poisson distribution will approach a Gaussian distribution.



One can expand $\log P(\lambda\, T;\, m)$ near $m = \langle m \rangle = \lambda\, T$. Notice when $T >> 1$, $\langle m \rangle = \lambda\, T$ is also large thus we can use Stirling formula. And one can prove the Gaussian form. Instead of giving a detailed proof, we simply give a numerical check:



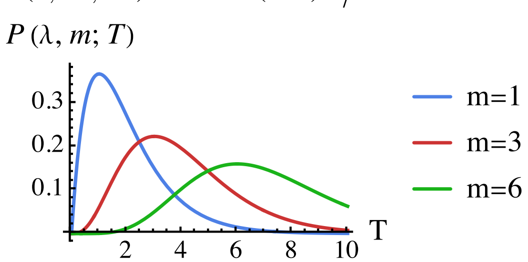
▪ Fix m

We first check the normalization condition

```
Integrate[Exp[-λ T] (λ T)^m / m!, {T, 0, ∞}, Assumptions → m ≥ 0 && λ > 0]
```

$$\frac{1}{\lambda}$$

$P(\lambda,\, m;\; T) = \lambda\, e^{-\lambda\, T} (\lambda\, T)^m \big/ m!$ is a PDF with respect to T.



Because $P(\lambda,\, m;\, T)\, \mathrm{dT} = P(\lambda\, T;\, m)\, \lambda\, d\, T$, where $P(\lambda\, T;\, m)$ is the probability for m events to the time T and $\lambda\, d\, T$ is the probability for one event happening during the interval $[T,\, T + d\, T]$, we conclude the following interpretation:

The physical meaning of $P(\lambda,\, m;\, T)\,\mathrm{dT}$ is the probability for $(m+1)$th event happening at the interval $[T,\ T+d\ T]$.

- Fix m, mean value of T,

```
Integrate[T λ Exp[-λ T] (λ T)^m/m!, {T, 0, ∞}, Assumptions → m ≥ 0 && λ > 0]
```

$$\frac{1+m}{\lambda}$$