Moment Generating Function

Dr. Shafayat Abrar

In probability theory and statistics, the **Poisson distribution** is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant mean rate and independently of the time since the last event.[1] It is named after French mathematician Siméon Denis Poisson (/ˈpwɑːsɒn/; French pronunciation: [pwasɔ̃]). The Poisson distribution can also be used for the number of events in other specified interval types such as distance, area, or volume. It plays an important role for discrete-stable distributions.

For instance, a call center receives an average of 180 calls per hour, 24 hours a day. The calls are independent; receiving one does not change the probability of when the next one will arrive. The number of calls received during any minute has a Poisson probability distribution with mean 3: the most likely numbers are 2 and 3 but 1 and 4 are also likely and there is a small probability of it being as low as zero and a very small probability it could be 10.

Another example is the number of decay events that occur from a radioactive source during a defined observation period.

A discrete random variable X is said to have a Poisson distribution, with parameter $\lambda > 0$, if it has a probability mass function given by:[11]:60

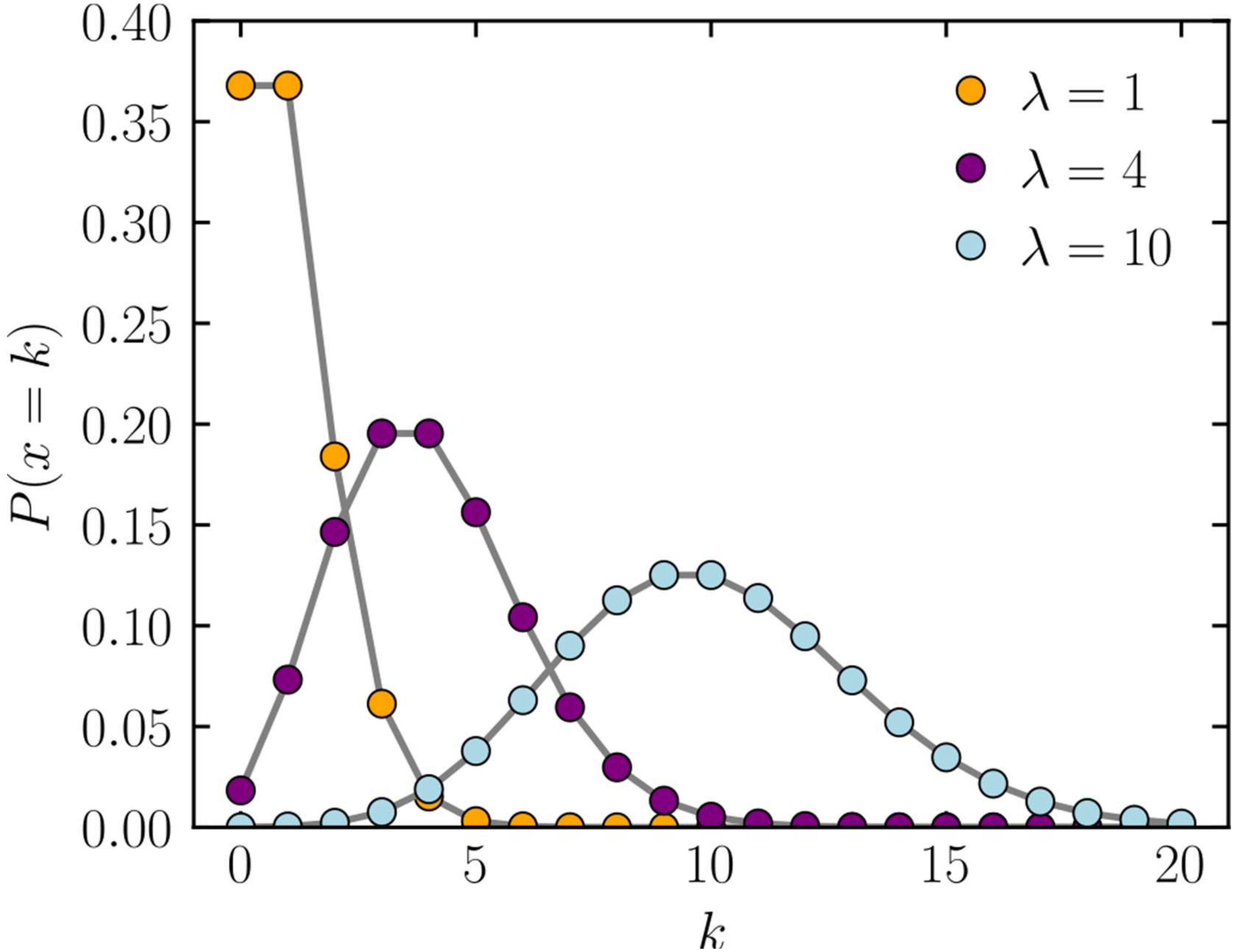
$$f(k;\lambda) = \Pr(X{=}k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

where

- k is the number of occurrences ($k = 0, 1, 2, \ldots$)
- e is Euler's number ($e = 2.71828\ldots$)
- $!$ is the factorial function.

The positive real number λ is equal to the expected value of X and also to its variance.

$$\lambda = \operatorname{E}(X) = \operatorname{Var}(X).$$



0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
$P(x = k)$
0
5
10
15
20
k
$\lambda = 1$
$\lambda = 4$
$\lambda = 10$

Example 01:

On a particular river, overflow floods occur once every 100 years on average. Calculate the probability of k = 0, 1, 2, 3, 4, 5, or 6 overflow floods in a 100 year interval, assuming the Poisson model is appropriate.

Because the average event rate is one overflow flood per 100 years, λ = 1

$$P(k \text{ overflow floods in 100 years}) = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{1^k e^{-1}}{k!}$$

$$P(k = 0 \text{ overflow floods in 100 years}) = \frac{1^0 e^{-1}}{0!} = \frac{e^{-1}}{1} \approx 0.368$$

$$P(k = 1 \text{ overflow flood in 100 years}) = \frac{1^1 e^{-1}}{1!} = \frac{e^{-1}}{1} \approx 0.368$$

$$P(k = 2 \text{ overflow floods in 100 years}) = \frac{1^2 e^{-1}}{2!} = \frac{e^{-1}}{2} \approx 0.184$$

The Poisson distribution can be applied to systems with a large number of possible events, each of which is rare. The number of such events that occur during a fixed time interval is, under the right circumstances, a random number with a Poisson distribution.

The equation can be adapted if, instead of the average number of events λ, we are given the average rate r at which events occur. Then $\lambda = rt$, and:[13]

$$P(k \text{ events in interval } t) = \frac{(rt)^k e^{-rt}}{k!}.$$

SEL 01 (Lecture 05):

On a particular river, overflow floods occur once every 100 years on average. Calculate the probability of $k = 0, 1, 2, 3, 4, 5$, or 6 overflow floods in a 10 year interval, assuming the Poisson model is appropriate.

$$\mathbf{E}[X] = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = ??$$

$$\mathbf{E}[X^2] = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = ??$$

$$\begin{aligned}
\mathbf{E}[X] &= \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} \\
&= \sum_{k=1}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} \\
&= \lambda \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda \sum_{m=0}^{\infty} e^{-\lambda} \frac{\lambda^m}{m!} \\
&= \lambda.
\end{aligned}$$

Not shown here, but we discussed two different methods to evaluate $E[X^2]$ on board.

Discuss with each other if you need help.

Moment Generating Function

The moment-generating function (mgf) of a random variable X is

$$M(t) = E\left(e^{tX}\right)$$

if the expectation is defined. In the discrete case,

$$M(t) = \sum_k e^{tk} p(k)$$

and in the continuous case,

$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

Moment Generating Function

The derivative of $M(t)$ is

$$M'(t) = \frac{d}{dt} \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

It can be shown that differentiation and integration can be interchanged, so that

$$M'(t) = \int_{-\infty}^{\infty} x e^{tx} f(x) dx$$

and

$$M'(0) = \int_{-\infty}^{\infty} x f(x) dx = E(X)$$

Differentiating r times, we find

$$M^{(r)}(0) = E(X^r)$$

Using MGF to obtain moments of Poisson dist.

$$
\begin{aligned}
M(t) &= \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k}{k!} e^{-\lambda} \\
&= \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} e^{-\lambda} \\
&= e^{-\lambda} e^{\lambda e^t} \\
&= e^{\lambda(e^t - 1)}
\end{aligned}
$$

The sum converges for all t. Differentiating, we have

$$
\begin{aligned}
M'(t) &= \lambda e^t e^{\lambda(e^t - 1)} \\
M''(t) &= \lambda e^t e^{\lambda(e^t - 1)} + \lambda^2 e^{2t} e^{\lambda(e^t - 1)}
\end{aligned}
$$

Using MGF to obtain moments of Poisson dist.

$$M'(t) = \lambda e^t e^{\lambda(e^t - 1)}$$
$$M''(t) = \lambda e^t e^{\lambda(e^t - 1)} + \lambda^2 e^{2t} e^{\lambda(e^t - 1)}$$

Evaluating these derivatives at $t = 0$, we find

$$E(X) = M'(t)\,|_{t=0} \;\; = \lambda$$
$$E(X^2) = M''(t)\,|_{t=0} \;\; = \lambda^2 + \lambda$$

from which it follows that

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \lambda$$

We have found that the mean and the variance of a Poisson distribution are equal.

In probability theory and statistics, the gamma distribution is a two-parameter family of continuous probability distributions. The exponential distribution, Erlang distribution, and chi-squared distribution are special cases of the gamma distribution. There are two equivalent parameterizations in common use:

1. With a **shape** parameter k and a **scale** parameter θ.
2. With a **shape** parameter $\alpha = k$ and an **inverse scale** parameter $\beta = 1/\theta$, called a **rate** parameter.

$$\boxed{\begin{aligned} E(X) &= \frac{\alpha}{\beta} \\ E(X^2) &= \frac{\alpha^2}{\beta^2} + \frac{\alpha}{\beta^2} \end{aligned}}$$

In each of these forms, both parameters are positive real numbers.

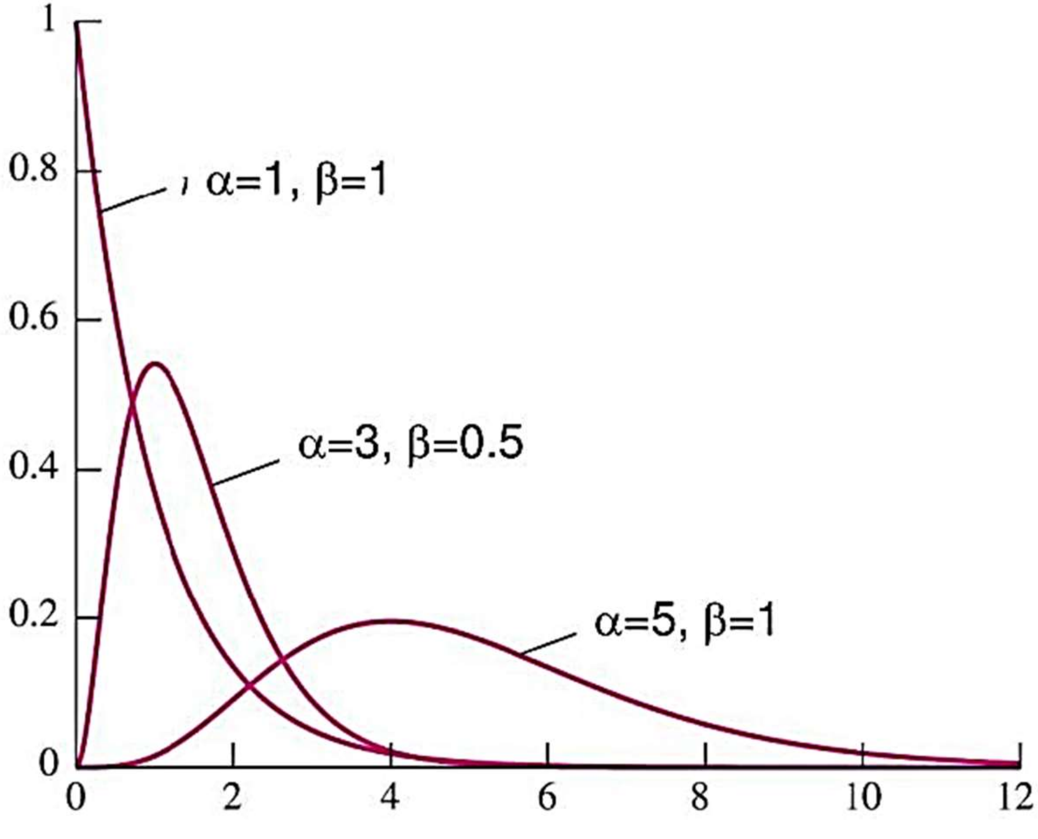
The gamma distribution can be parameterized in terms of a shape parameter $\alpha = k$ and an inverse scale parameter $\beta = 1/\theta$, called a rate parameter. A random variable X that is gamma-distributed with shape α and rate β is denoted

$$X \sim \Gamma(\alpha, \beta) \equiv \mathrm{Gamma}(\alpha, \beta)$$

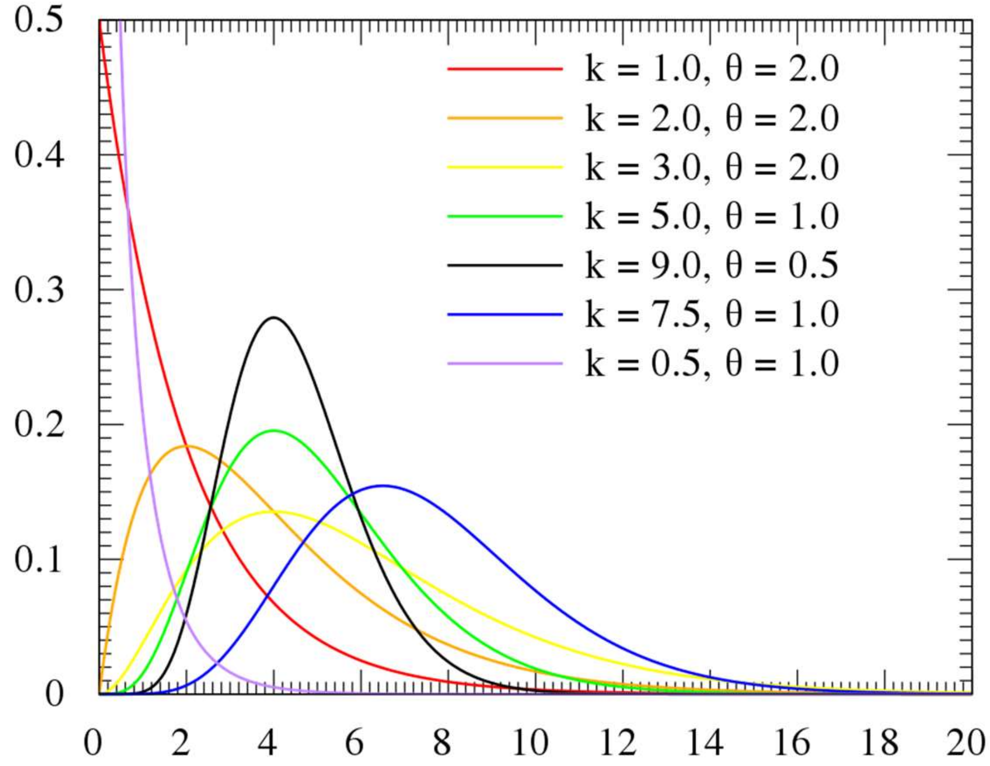
The corresponding probability density function in the shape-rate parameterization is

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1} e^{-\beta x} \beta^{\alpha}}{\Gamma(\alpha)} \quad \text{for } x > 0 \quad \alpha, \beta > 0,$$

where $\Gamma(\alpha)$ is the gamma function. For all positive integers, $\Gamma(\alpha) = (\alpha - 1)!$



1
0.8
0.6
0.4
0.2
0
0
2
4
6
8
10
12
α=1, β=1
α=3, β=0.5
α=5, β=1



0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12
14
16
18
20
k = 1.0, θ = 2.0
k = 2.0, θ = 2.0
k = 3.0, θ = 2.0
k = 5.0, θ = 1.0
k = 9.0, θ = 0.5
k = 7.5, θ = 1.0
k = 0.5, θ = 1.0

In wireless communication, the gamma distribution is used to model the multi-path fading of signal power.

The gamma distribution has been used to model the size of insurance claims and rainfalls. This means that aggregate insurance claims and the amount of rainfall accumulated in a reservoir are modelled by a gamma process.

In oncology, the age distribution of cancer incidence often follows the gamma distribution, wherein the shape and scale parameters predict, respectively, the number of driver events and the time interval between them.

In neuroscience, the gamma distribution is often used to describe the distribution of inter-spike intervals.

In bacterial gene expression, the copy number of a constitutively expressed protein often follows the gamma distribution, where the scale and shape parameter are, respectively, the mean number of bursts per cell cycle and the mean number of protein molecules produced by a single mRNA during its lifetime.

In Bayesian statistics, the gamma distribution is widely used as a conjugate prior. It is the conjugate prior for the precision (i.e. inverse of the variance) of a normal distribution. It is also the conjugate prior for the exponential distribution.

Using MGF to obtain moments of Gamma dist.

Gamma Distribution

The mgf of a gamma distribution is

$$
\begin{aligned}
M(t) &= \int_0^\infty e^{tx} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{x(t-\beta)} dx
\end{aligned}
$$

The latter integral converges for $t < \beta$ and can be evaluated by relating it to the gamma density having parameters α and $\beta - t$. We thus obtain

$$
M(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{\Gamma(\alpha)}{(\beta - t)^\alpha} \right) = \left(\frac{\beta}{\beta - t} \right)^\alpha
$$

Differentiating, we find

$$M'(0) = E(X) = \frac{\alpha}{\beta}$$

$$M''(0) = E\left(X^2\right) = \frac{\alpha(\alpha + 1)}{\beta^2}$$

From these equations, we find that

$$\begin{aligned} \operatorname{Var}(X) &= E\left(X^2\right) - [E(X)]^2 \\ &= \frac{\alpha(\alpha + 1)}{\beta^2} - \frac{\alpha^2}{\beta^2} \\ &= \frac{\alpha}{\beta^2} \end{aligned}$$

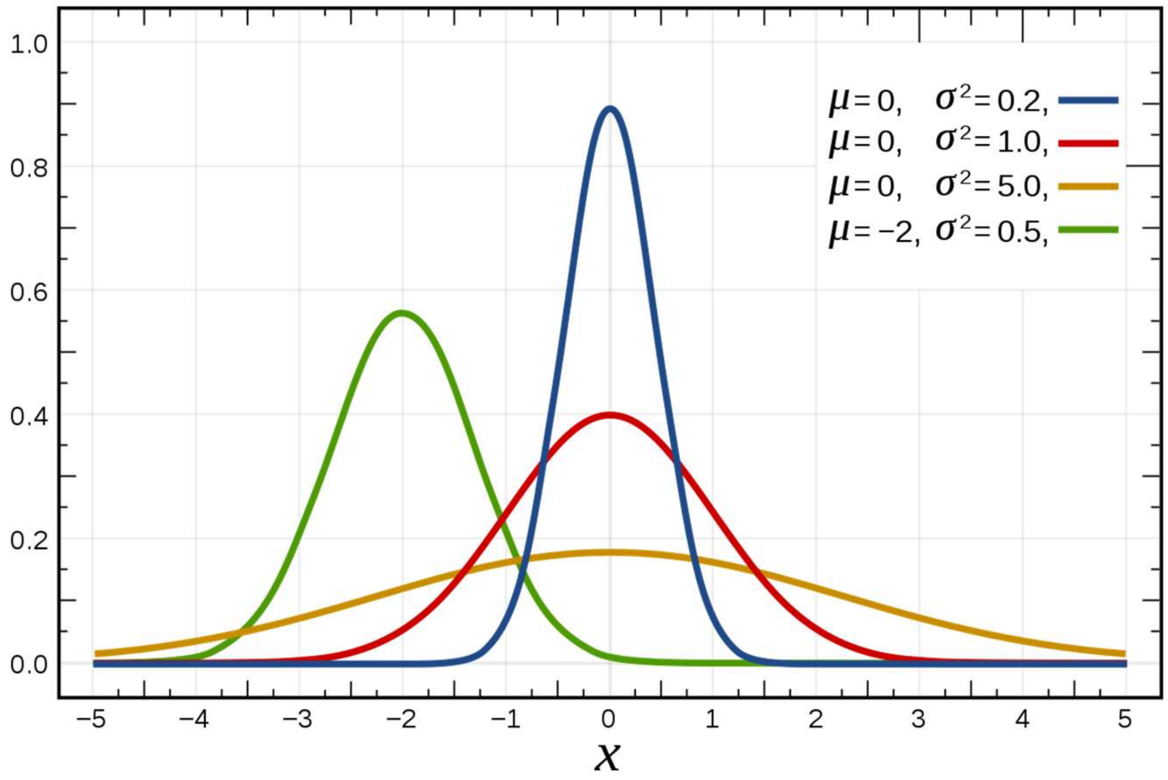
In statistics, a **normal distribution** or **Gaussian distribution** is a type of continuous probability distribution for a real-valued random variable. The general form of its probability density function is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

The parameter μ is the mean or expectation of the distribution (and also its median and mode), while the parameter σ is its standard deviation. The variance of the distribution is σ^2. A random variable with a Gaussian distribution is said to be **normally distributed**, and is called a **normal deviate**.

Normal distributions are important in statistics and are often used in the natural and social sciences to represent real-valued random variables whose distributions are not known.[2][3] Their importance is partly due to the central limit theorem. It states that, under some conditions, the average of many samples (observations) of a random variable with finite mean and variance is itself a random variable—whose distribution converges to a normal distribution as the number of samples increases. Therefore, physical quantities that are expected to be the sum of many independent processes, such as measurement errors, often have distributions that are nearly normal.[4]

Order	moment
1	μ
2	$\mu^2 + \sigma^2$
3	$\mu^3 + 3\mu\sigma^2$
4	$\mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$
5	$\mu^5 + 10\mu^3\sigma^2 + 15\mu\sigma^4$
6	$\mu^6 + 15\mu^4\sigma^2 + 45\mu^2\sigma^4 + 15\sigma^6$
7	$\mu^7 + 21\mu^5\sigma^2 + 105\mu^3\sigma^4 + 105\mu\sigma^6$
8	$\mu^8 + 28\mu^6\sigma^2 + 210\mu^4\sigma^4 + 420\mu^2\sigma^6 + 105\sigma^8$



Theorem: Let X be a random variable following a normal distribution:

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

Then, the moment-generating function of X is

$$M(t) = \exp\left[\mu t + \frac{1}{2}\sigma^2 t^2\right].$$

We discussed the proof on board:

$$\int_{-\infty}^{+\infty} \exp\left[-x^2\right] \mathrm{d}x = \sqrt{\pi}$$

Proof of the theorem is available at the following link:

https://statproofbook.github.io/P/norm-mgf.html

SEL 02 (Lecture 05):

Using MGF, show that

$$E[X] = \mu$$
$$E[X^2] = \mu^2 + \sigma^2$$
$$E[X^3] = \mu^3 + 3\mu\sigma^2$$

Conclusions

- We discussed Poisson, Gamma, and Gaussian distributions.
- We discussed moment generating function to compute the moments of addressed distributions.