Earth Observation - Astrophysics - Satellite Technologies

Lecture 2: Probability distributions

Hoàng Đức Thường
Department of Space and Applications (DSA), USTH

Hanoi, 2021

Probability distributions

In this class I will review several special probability distributions and simple applications.

Probability distributions

At the end of this class, you should be able to do:

- Apply probability distribution: Gaussian, Poisson, exponential

Random variable (RV)

Random variable

- Values of outcomes of a random phenomenon. There are two types of random variables: **discrete** and **continuous**.
- Example: Tossing a coin, number of shooting stars, the temperature in random day in Hanoi, …

Discrete random variable

- Their values is in a discrete set. Discrete random variables are often a countable number.
- Example: The number N of children of a random person, number of stars in a squared sky.

Continuous random variable

- Their values in arbitrary ranges of real numbers. Continuous random variables are often measurements.
- Example: The age, height, weight of Vietnamese population.

Probability distributions

- A **probability distribution** $P(x)$ is a **function** which assigns a probability for range of values / or the law of probability of variable.

- **Normalized (=1)** $\int_{-\infty}^{\infty} P(x)\, dx = 1$ or for discrete variables $\sum_i P(x_i) = 1$

- Continuous random variable: P(x) is also called **probability density function**: PDF

- Probability density in a range [a, b] = $prob(a < x < b) = \int_a^b P(x)dx$

- P(x) is a single, non-negative value for all real x.

- If we know the probability function, it means we determine completely the random variable.

Mean and Variance of Probability Distributions

Mean

- Discrete $\mu = = \bar{x} = \langle x \rangle = \sum_i x_i P(x_i)$

- Continuous $\mu = \int x P(x) dx$

Variance

$$\sigma^2 = \sum_i (x_i - \mu)^2 P(x_i) = \sum_i (x_i)^2 P(x_i) - \mu^2$$

$$\sigma^2 = \int (x-\mu)^2 P(x) dx = \int x^2 P(x) dx - \mu^2$$

$$= \langle x^2 \rangle - \langle x \rangle^2$$

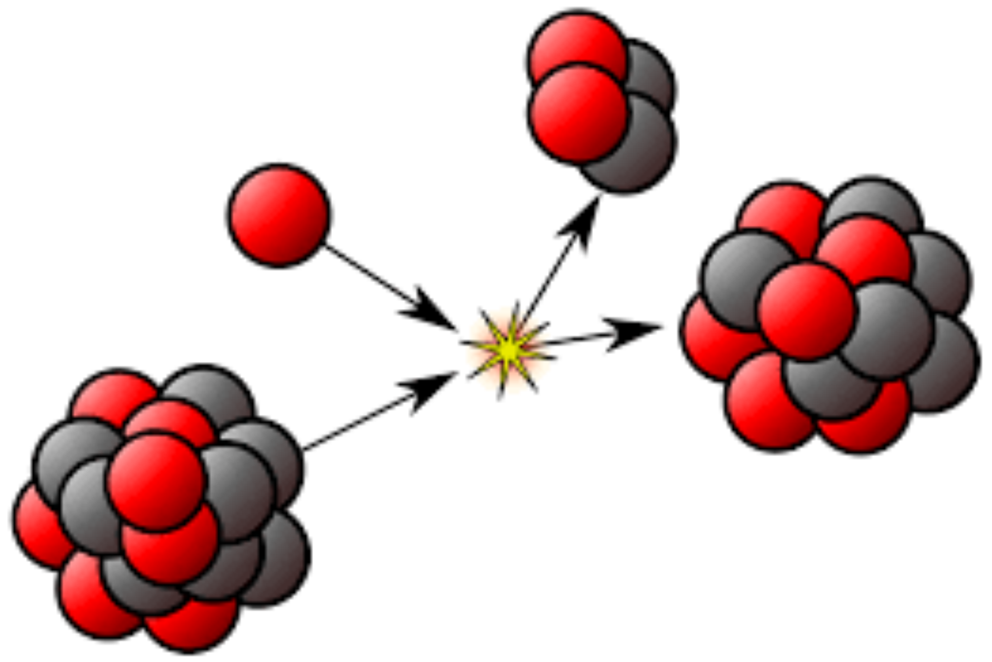
Probability distributions

Discrete variable Distribution

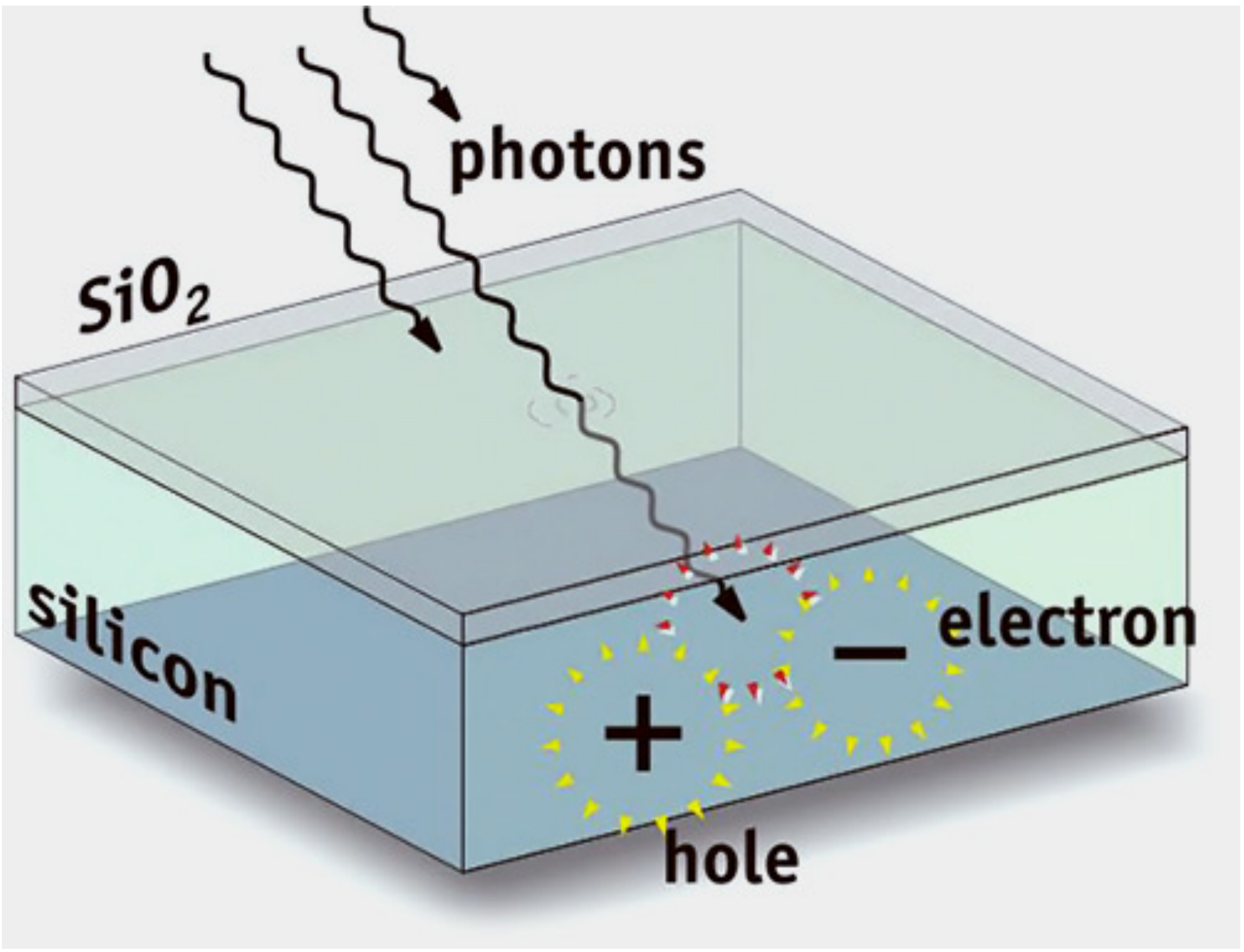
1. Binomial distribution
2. Poisson distribution

Continuous variables Distribution (=Probability Density Function - PDF)

3. *Normal distribution/ Center limit theorem*
4. Uniform distribution
5. Gamma distribution
6. Exponential distribution
7. Student's t-distribution
8. Chi-square distribution



(Photons received on a sensor)



1. Binomial Distribution

- There are a random process with **two possible outcomes** with probabilities ρ (success) and 1-ρ (failure). This common distribution gives the chance of n successes in N trials.

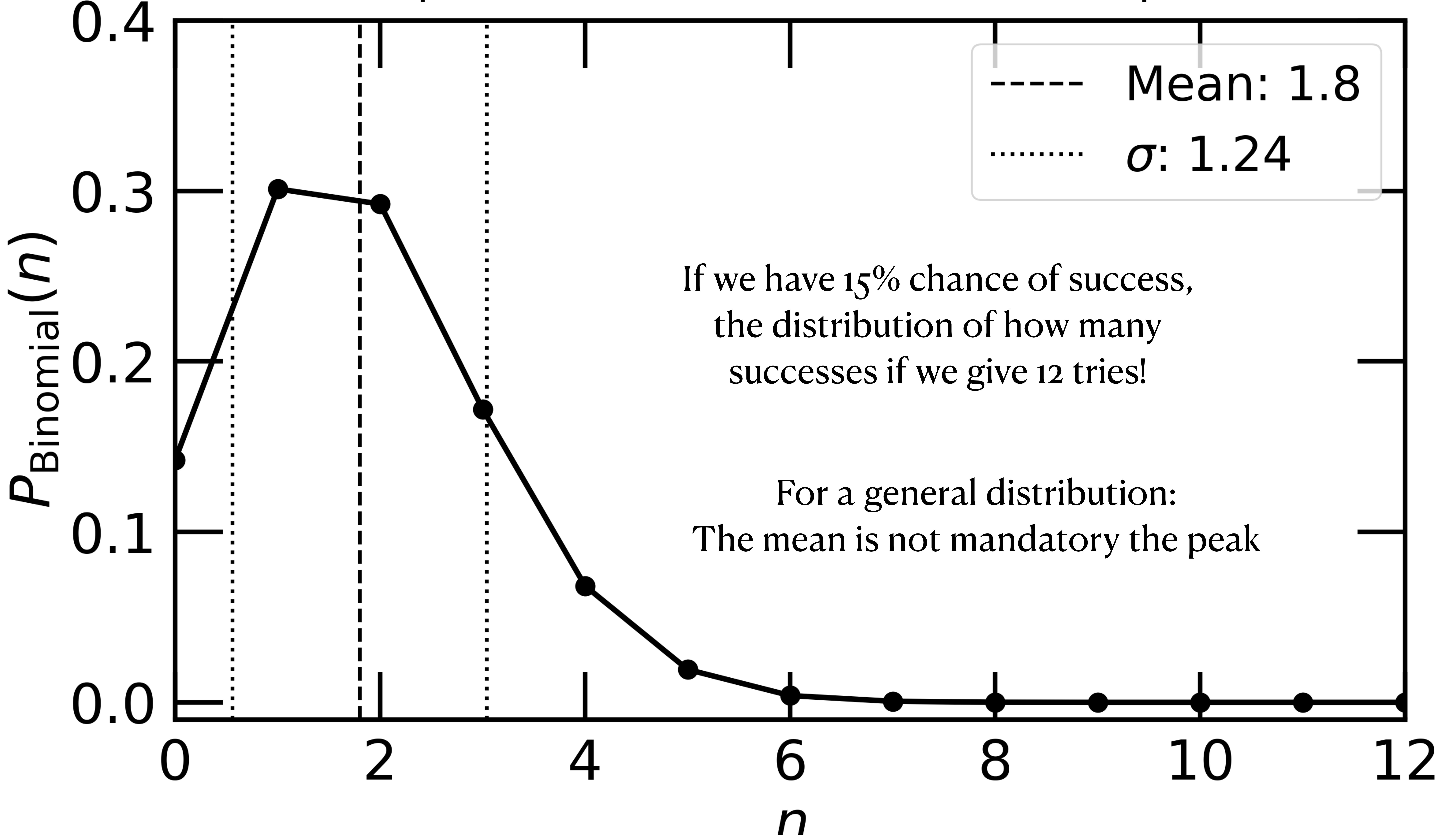
- If we have N trials, and the probability of success in each is ρ, the the probability to obtain n successes is:

$$P_{Binomial}(n) = \frac{N!}{n!(N-n)!}\rho^n(1-\rho)^{N-n}$$

Example: Tossing a coin N times. What is the probability of head (success).

- The mean of this distribution: $\bar{n} = \rho N$
- The variance of this distribution: $Var(n) = N\rho\left(1-\rho\right)$.

Example Binomial distribution: N=12, p=0.15



If we have 15% chance of success, the distribution of how many successes if we give 12 tries!

For a general distribution:
The mean is not mandatory the peak

Binomial Distribution

- **Example**, a satellite observes a data set of 100 galaxies. 30 of them have an elliptical shape. What is the best estimate of the elliptical shape and its error?

Binomial Distribution

- Example, a satellite observes a data set of 100 galaxies. 30 of them have an elliptical shape. What is the best estimate of the elliptical shape and its error?

- The elliptical shape success probability = ρ = 30/100 = 0.3.

- There are 2 possible outcomes: a galaxy has the elliptical shape or not, then the best estimation is the binomial distribution.

- Error in elliptical fraction:

remember: **error in mean** $= \dfrac{\sigma}{\sqrt{N}} = \sqrt{N\rho\left(1-\rho\right)/N} = \sqrt{100 * 0.3(1-0.3)/100} = 0.046$

Answer: $p = 0.3 \pm 0.046$

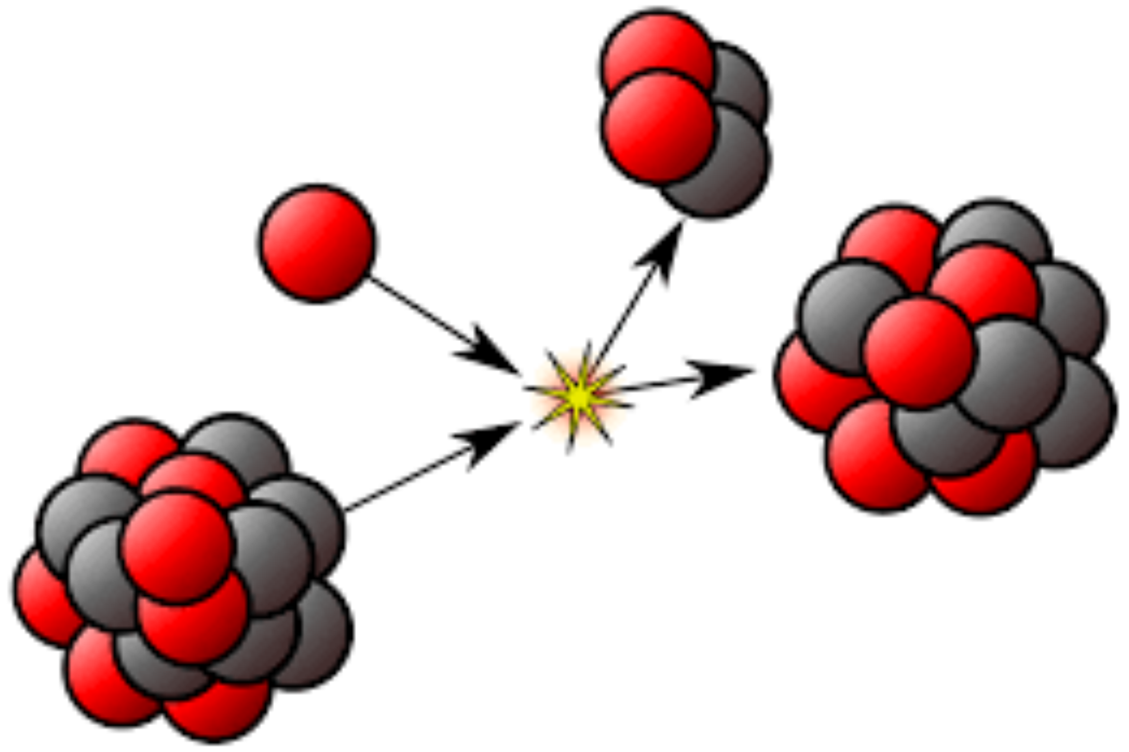
2. Poisson Distribution

- Poisson distribution derives from the **Binomial distribution** ($\rho \rightarrow 0, N \rightarrow \infty$). It applies to a process of counting something in an interval (of time).

  - Number of stars in a random square on the sky.
  - Number of photons received on a bolometer / CCD (charge-coupled device).
  - Radioactive decay.

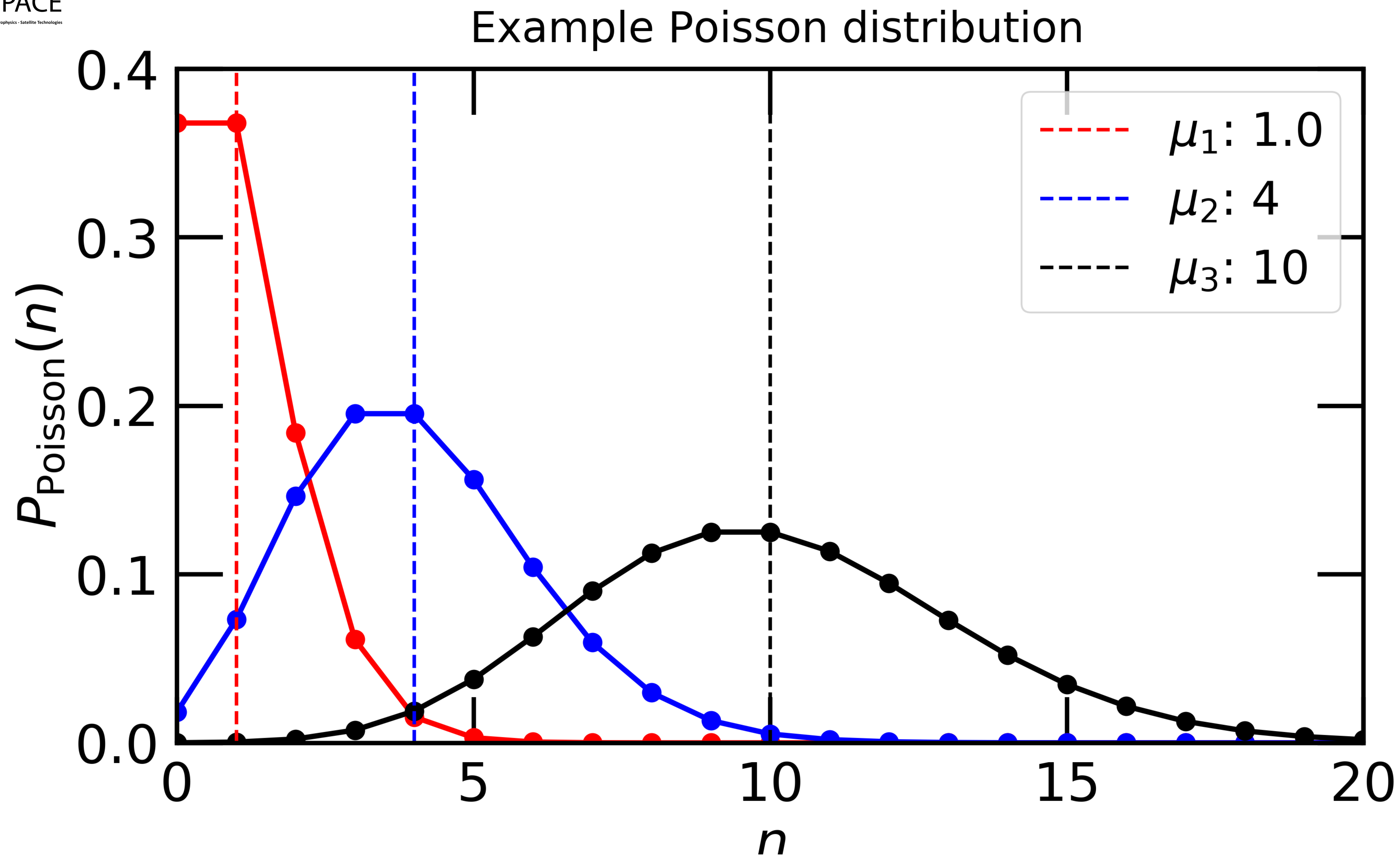
- The probability of observing n events, with the mean number μ:

$$P_{Poisson}(n) = \frac{\mu^n e^{-\mu}}{n!}$$



- The mean and variance of this distribution: $\bar{n} = Var(n) = \mu$.

- e = 2.71828.


Example Poisson distribution
0.4
0.3
0.2
0.1
0.0
0
5
10
15
20
n
$P_{\mathrm{Poisson}}(n)$
μ_1: 1.0
μ_2: 4
μ_3: 10

Poisson Distribution

- The number of photons/particles arriving a sensor (charge-coupled device: CCD) in an integration time is obeyed Poisson statistics.
- For example: The probability of a photon/particle (the rate of photon/particle) λ arriving a CCD is small during an integration time. If we assume during the integration time t. Hence the mean of photons/particles received at CCD is $\mu = \lambda t$, the $\sigma = \sqrt{\lambda t}$.
- If we integrate time more $\sigma \propto \sqrt{t}$, and the Signal $\propto t$. Then the Signal-to-Noise ratio $SNR \propto \sqrt{t}$

- If CCD of telescope observes faint objects, It is *photon-limited* case:

$$SNR \propto \frac{\mu}{\sqrt{\mu}} \; or \propto \sqrt{t}\,.$$

- If CCD of telescope observes bright objects, It is *readout-limited* case:

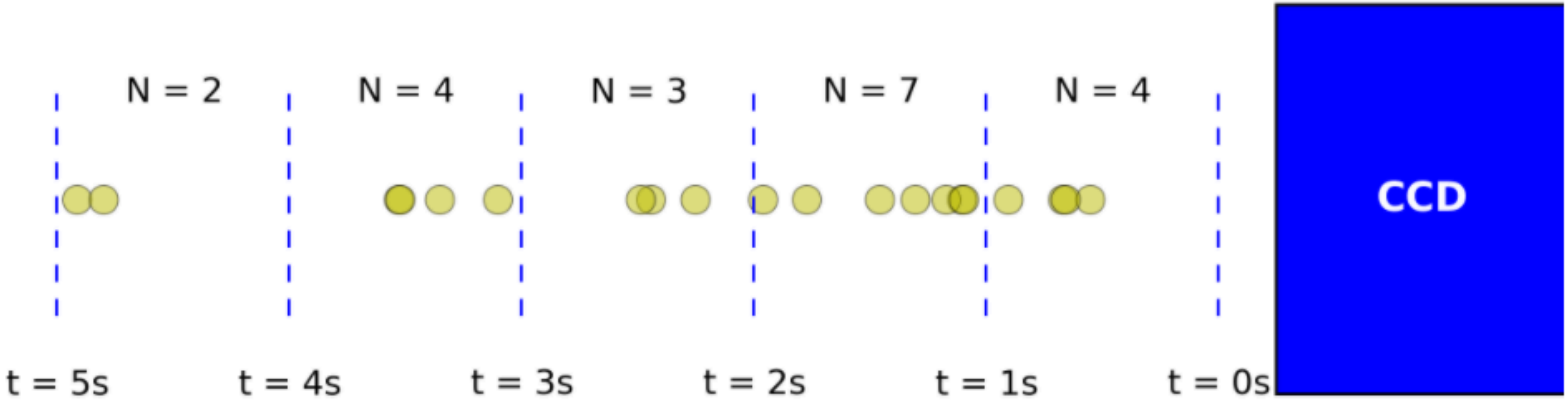
$$SNR \propto \frac{\mu}{\sigma_{CCD}} \; or \propto t\,.$$

- In radio astronomy, we use bolometer. It is *receiver-limited* case:

$$SNR \propto \frac{Signal}{\sigma_{bolo}/\sqrt{t}} \; or \propto \sqrt{t}\,.$$

http://slittlefair.staff.shef.ac.uk/teaching/phy217/lectures/instruments/L14/index.html#photons

Poisson Distribution



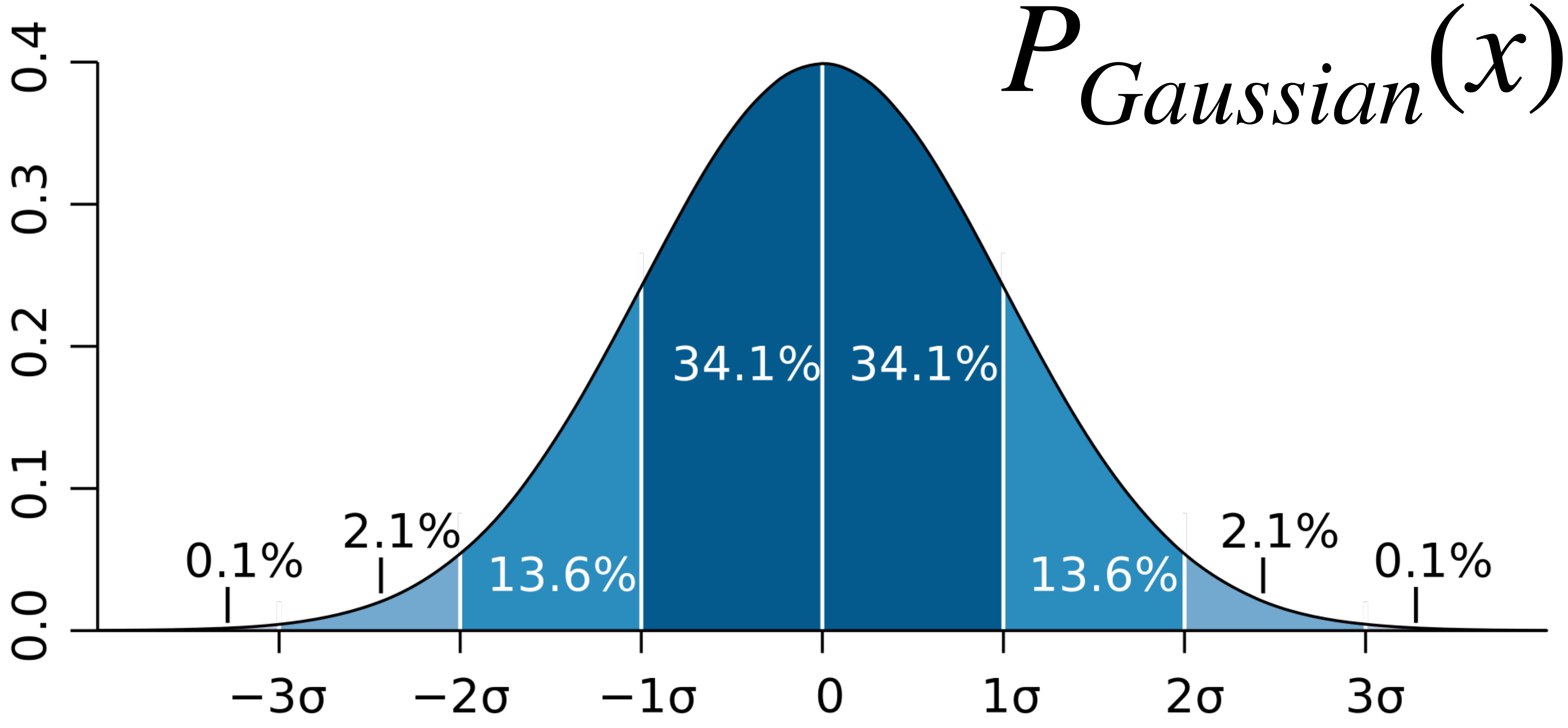
- For example: On average, a source emits 4 photons per second, and they are randomly. Therefore at a time, a telescope can detect 7 or 2 photons. This random bunching of photon is known as shot noise or bunching noise.
- If number of photons received in a fixed interval of time has an average N(=4), and photons are randomly spaced. The probability of detecting *k* photons is following a Poisson distribution.

$$P_{Poisson}(k) = \frac{N^k e^{-N}}{k!}$$

3. Normal (Gaussian) Distribution: $N(\mu, \sigma^2)$

- The **Gaussian** (or **Normal**) distribution is a continuous probability distribution. It is a ubiquitous and important probability distribution. Gaussian distribution is often used in the nature and social science.
- In cases of a large N in Binomial, a large μ in Poisson distributions, both become Gaussian distribution.

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

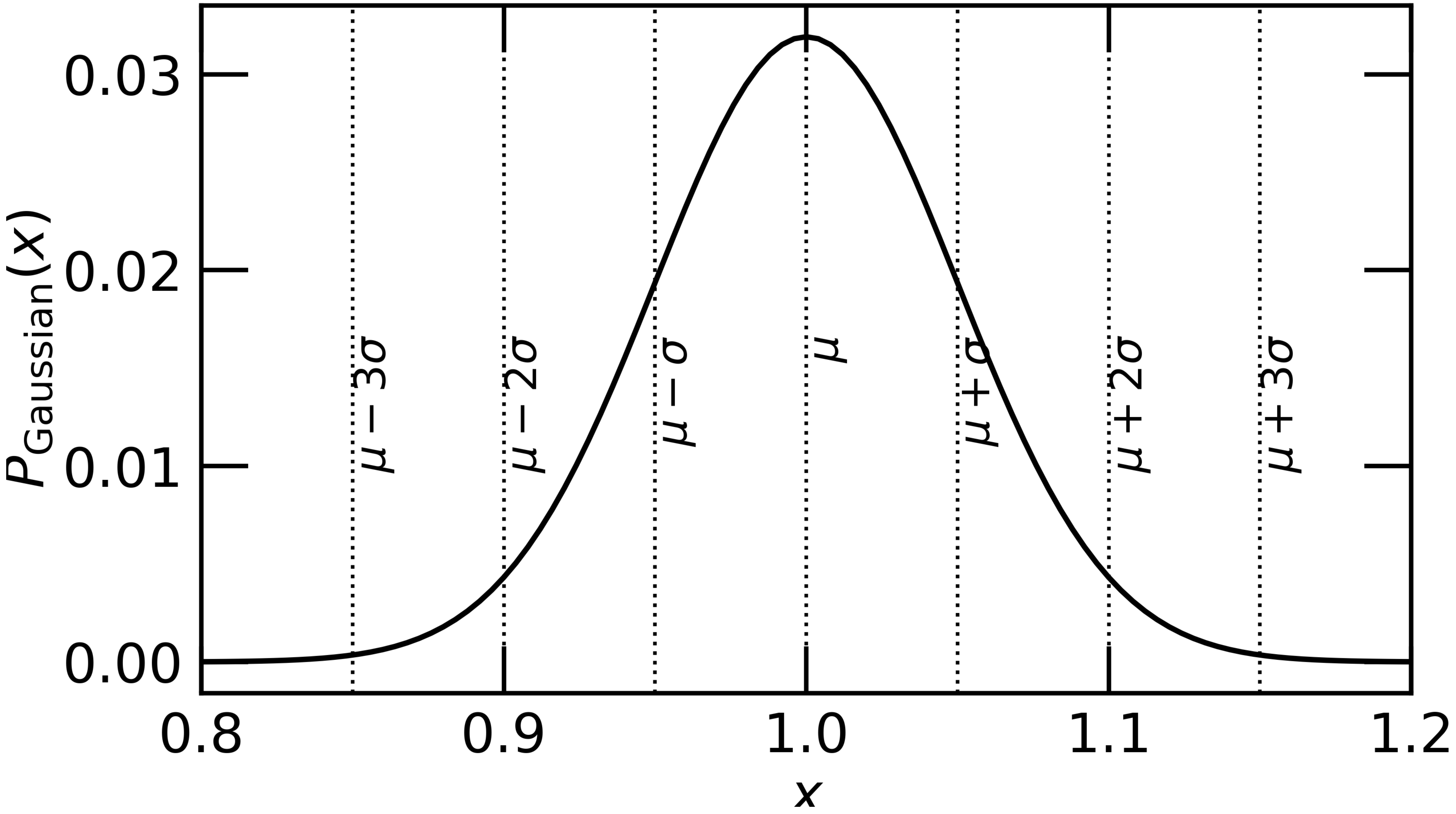


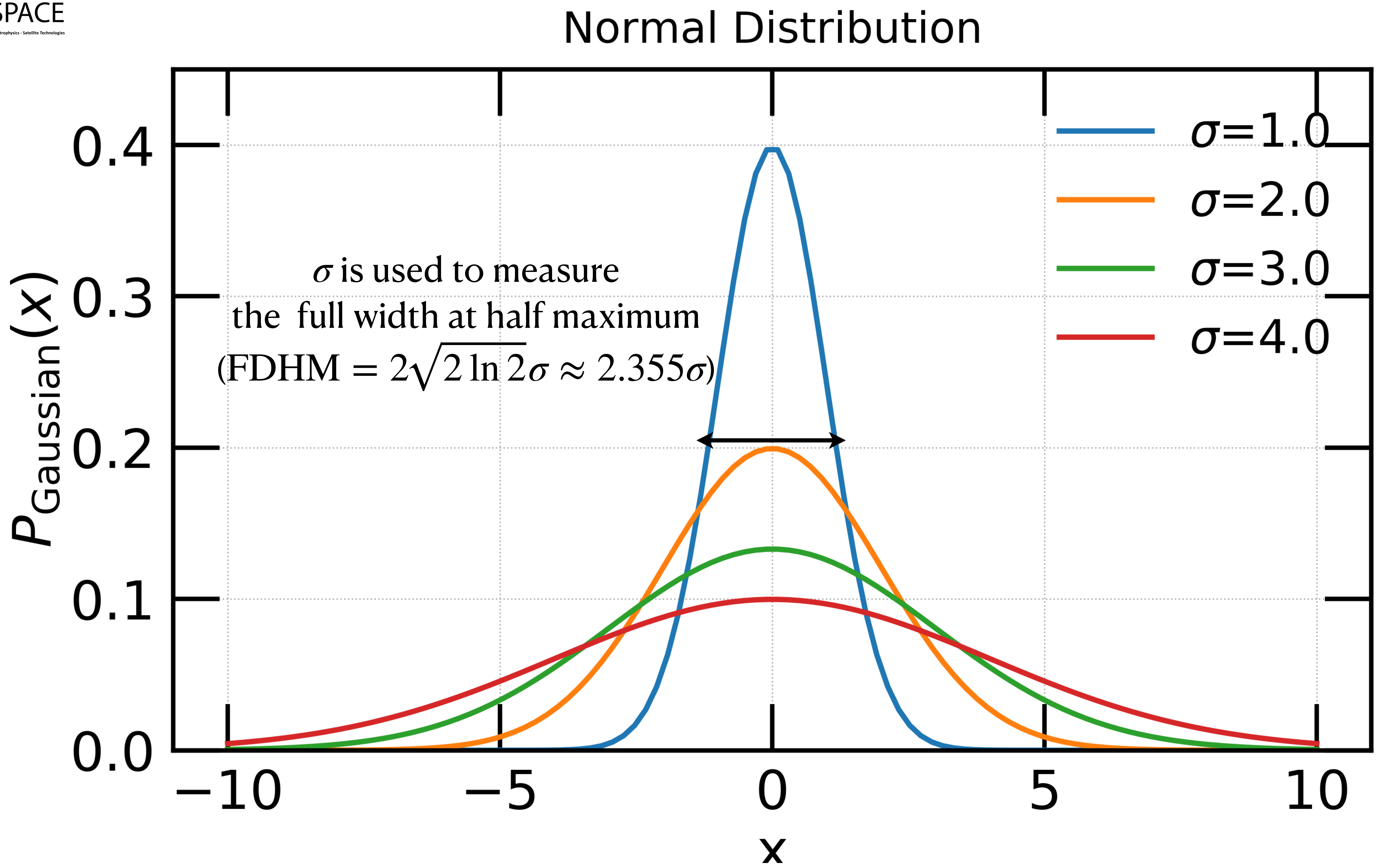
$x =$ *continuous variable*,

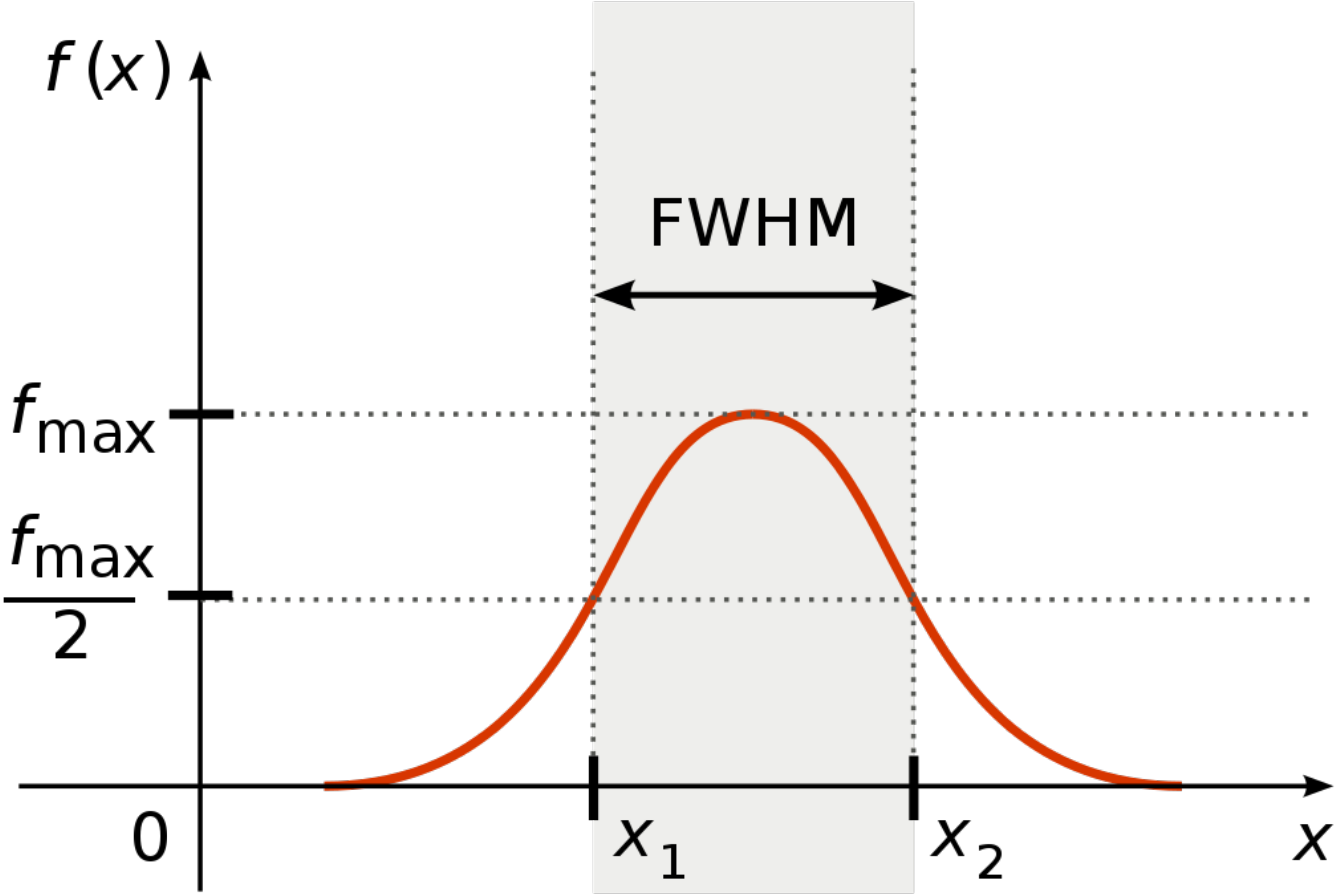
$\mu =$ *mean*

$\sigma =$ *standard deviation*

Example: Normal Distribution $\mu = 1, \sigma = 0.05$

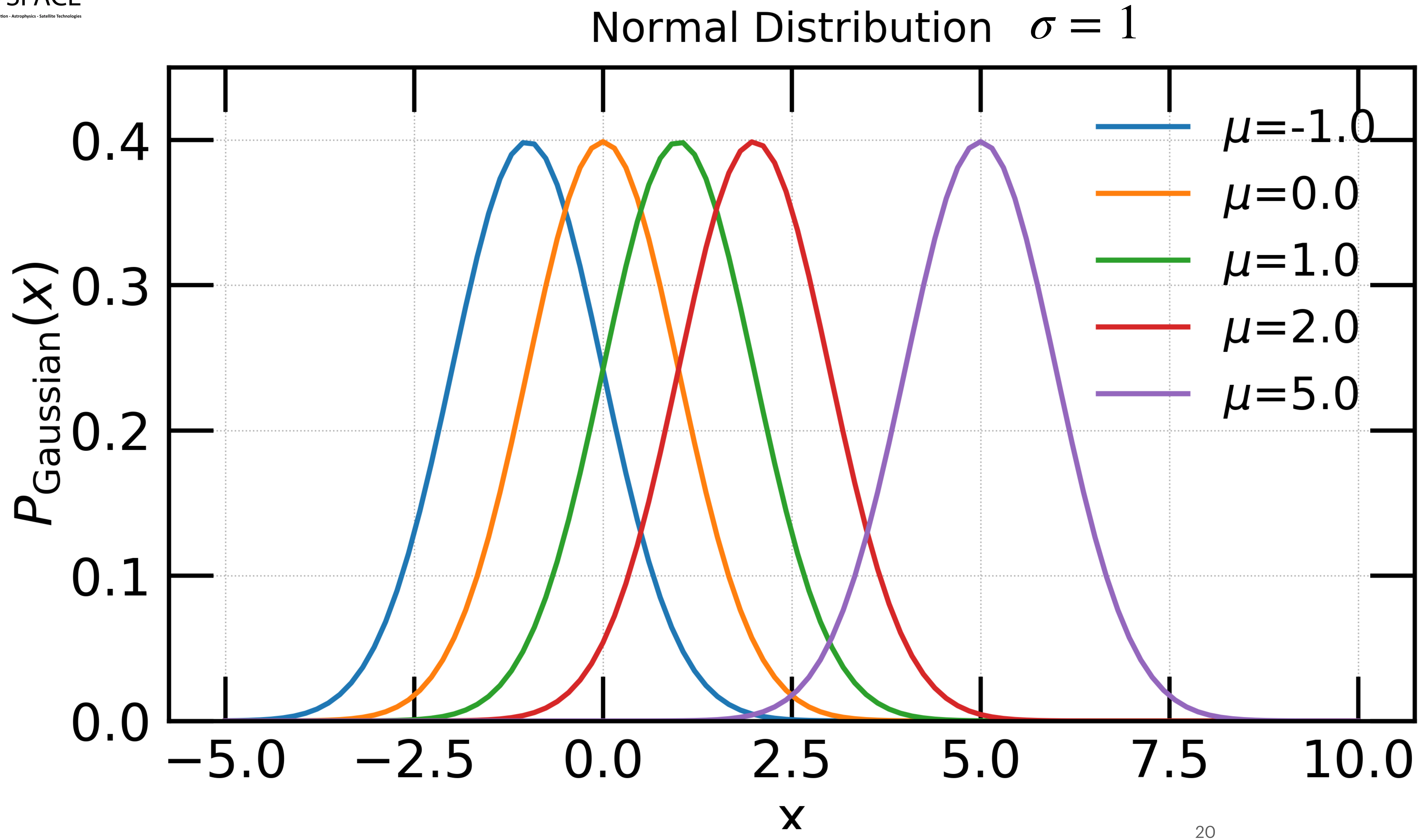



Normal Distribution
σ is used to measure
the full width at half maximum
(FDHM = $2\sqrt{2\ln 2}\sigma \approx 2.355\sigma$)
σ=1.0
σ=2.0
σ=3.0
σ=4.0
$P_{\mathrm{Gaussian}}(x)$
x
0.0
0.1
0.2
0.3
0.4
−10
−5
0
5
10

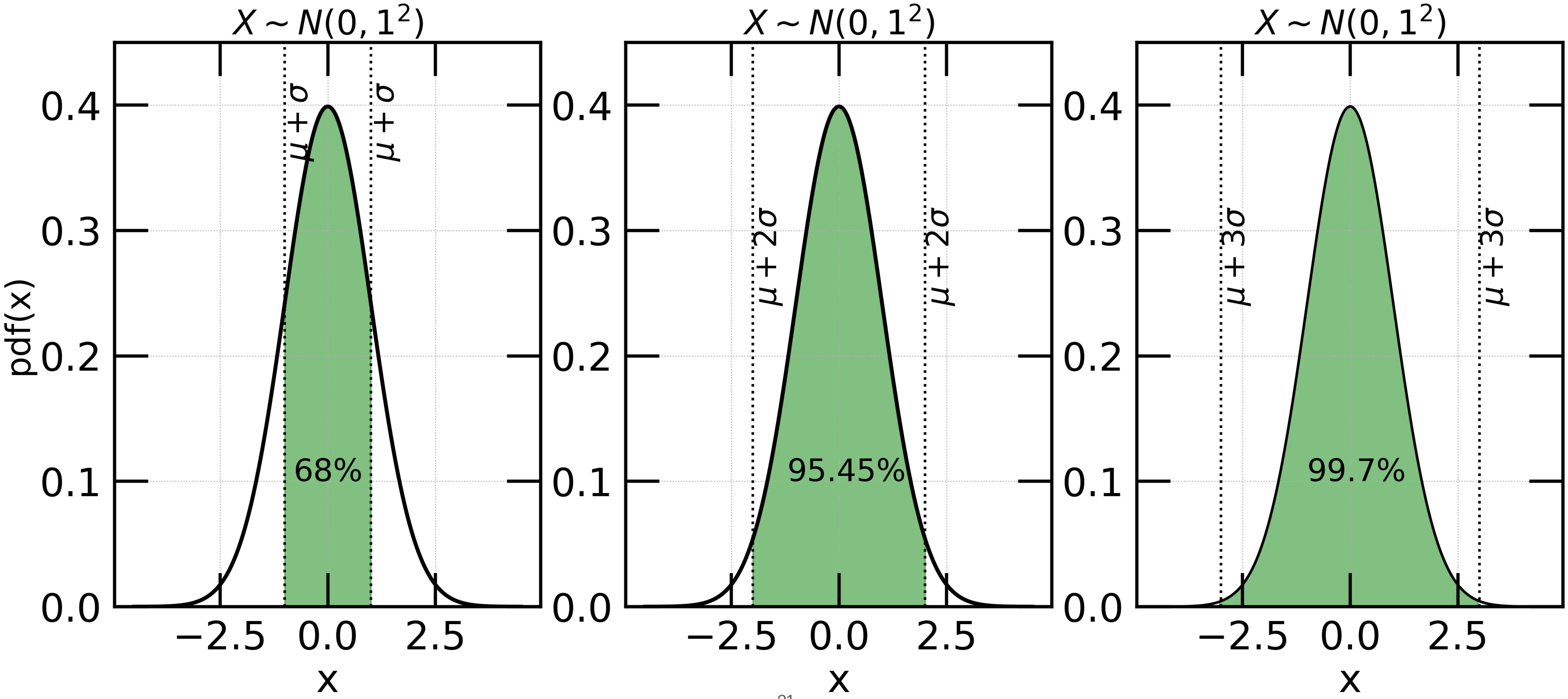

f(x)
FWHM
f max
f max
2
0
x 1
x 2
x

USTH MASTER SPACE
Earth Observation - Astrophysics - Satellite Technologies

USTH
VIETNAM FRANCE UNIVERSITY

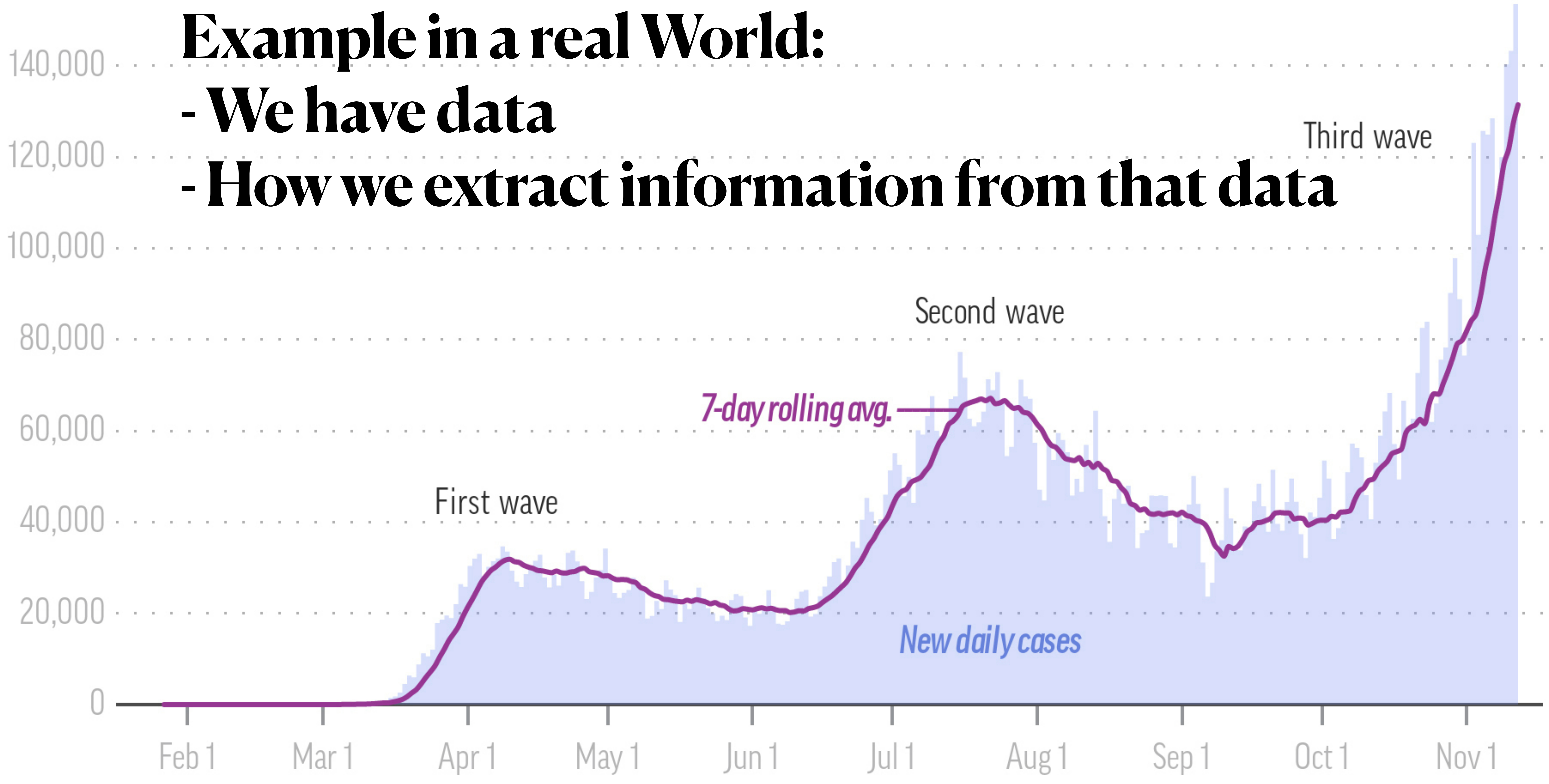


3. Normal Distribution: 68-95-99.7 rule!



The latest case counts of COVID in the US

Example in a real World:
- We have data
- How we extract information from that data



Source: Johns Hopkins University / Graphic: Phil Holm & Nicky Forster

2020

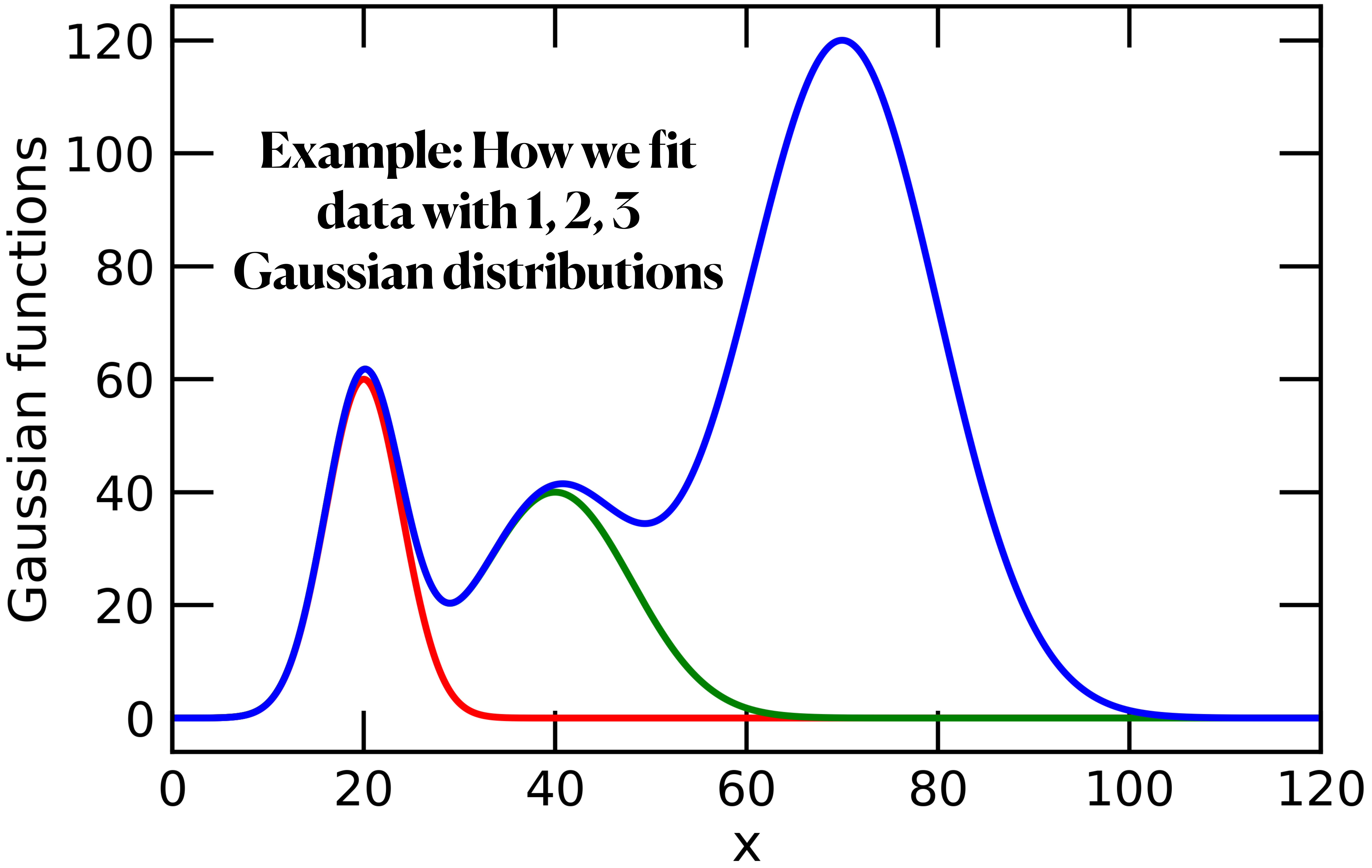
3. Normal Distribution

Exercise:

Draw a normal distribution for the Hubble constant with mean and deviation:

$$\boxed{H_0 = 70 \pm 5\ kms^{-1}Mpc^{-1}.}$$

$$60.0e^{-\frac{(x-20.0)^2}{2*4.0^2}} + 40.0e^{-\frac{(x-40.0)^2}{2*8.0^2}} + 120.0e^{-\frac{(x-70.0)^2}{2*10.0^2}}$$



Multivariate normal distribution

- The multivariate normal distribution is often used to describe correlated real-valued random variables.

- Probability Density Function (PDF):

$$P(X_1, X_2, \ldots X_n) = P(\mathbf{X}) = \frac{1}{\sqrt{(2\pi)^n \det[\mathbf{N}]}} \exp\left[-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})^T \mathbf{N}^{-1}(\mathbf{X}-\boldsymbol{\mu})\right]$$

$$\propto \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n} a_{ij}\left(\frac{X_i-\mu_i}{\sigma_i}\right)\left(\frac{X_j-\mu_j}{\sigma_j}\right)\right]$$

$\boldsymbol{\mu}$ is mean: n-dimensional real vector

N $(some\ books\ use : \Sigma)$

is Covariance matrix: n-dimensional real vector

a_{ij} is amplitude.

$$Cov(X_i X_j) = N_{ij}$$

Multivariate normal distribution

In two dimensions X, Y:

$$P(X,Y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(X-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{(X-\mu_X)(Y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(Y-\mu_Y)^2}{\sigma_Y^2}\right)\right]$$

- This equation is explicitly parametrized in its mean μ_X, μ_Y, standard deviation $\sigma_X,\ \sigma_Y$, and correlation ρ.

- In case the covariance matrix is diagonal, it means also two variables are independent: $\rho = 0$.

$$P(X,Y) = \frac{1}{2\pi\sigma_X}\exp\left[-\frac{1}{2}\frac{(X-\mu_X)^2}{\sigma_X^2}\right]\cdot\frac{1}{2\pi\sigma_Y}\exp\left[-\frac{1}{2}\frac{(Y-\mu_Y)^2}{\sigma_Y^2}\right]$$

Multivariate normal distribution

$\mu_1 = 0; \mu_2 = 1$

Diagonal Covariance Matrix [3, 15]

The central limit theorem - CLT (Law of Large Numbers)

Summations of independent random variables will tend to "look Gaussian."

- Suppose $X_1, X_2, \ldots, X_N$ are series of independent random variables, each variable has the mean μ_i and variance σ_i^2. Then the sum distribution $S = \sum X_i$ will have the mean $\sum \mu_i$, and the variance $\sum \sigma_i^2$.

- When N is a **large number** (the size of sample is increased - N > ~30), In particular $N \rightarrow \infty$:

$$\frac{S - \sum_{i=1}^{N} \mu_i}{\sqrt{\sum_{i=1}^{N} \sigma_i^2}} \rightarrow N(0,1)\,.$$

Standard Normal distribution

The central limit theorem - CLT

- Example:

  For a usual case where $\mu_i = \mu$, and $\sigma_i = \sigma$. Let $\overline{X}_N$ is the **sample mean** of $X_1, X_2, \ldots, X_N$:

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{N} \qquad \mu_{\overline{X}} = \mu \qquad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}}$$

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

$$E(\overline{X}) = \frac{E(X_1 + X_2 + \ldots + X_n)}{n}$$

$$E(\overline{X}) = \frac{E(X_1) + E(X_2) + \ldots + E(X_n)}{n}$$

$$E(\overline{X}) = \frac{nE(X)}{n}$$

$$E(\overline{X}) = E(X) = \mu$$

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

$$Var(\overline{X}) = Var\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right)$$

$$Var(\overline{X}) = \frac{1}{n^2} Var(X_1 + X_2 + \ldots + X_n)$$

$$Var(\overline{X}) = \frac{n}{n^2} Var(X) = \frac{\sigma^2}{n}$$

The central limit theorem - CLT

- Then, the normalized random variable:

$$Z_N = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{N}}} = \frac{X_1 + X_2 + \ldots + X_N - N\mu}{\sqrt{N}\sigma} \quad \rightarrow N(0,1)$$

Standard Normal distribution

- This powerful theorem is very import in theoretical and practical statistics. The **sample mean** $\overline{X}$ will be a normal distribution associated with population mean μ_i.
- When the sample size increases, the variance of the **sample mean** $\overline{X}$ becomes smaller.

The central limit theorem - CLT

- **Exercise (Home Work):**
  - Form groups of 2 students, present about Center limit theorem and Law of Larger Number less than 10 slides. The content should cover:
    - Theory (associated with Normal distribution)
    - Examples/Applications
    - (optional) Jupiter Notebook

4. Uniform Distribution

- This distribution is used to describe an experiment where the output is lie between certain bounds a, b. The uniform can be used in the simulation of perfect filters. The uniform distribution for a real variable X:

$$f(X) = \frac{1}{b-a} \quad , \quad X \in [a, b]$$

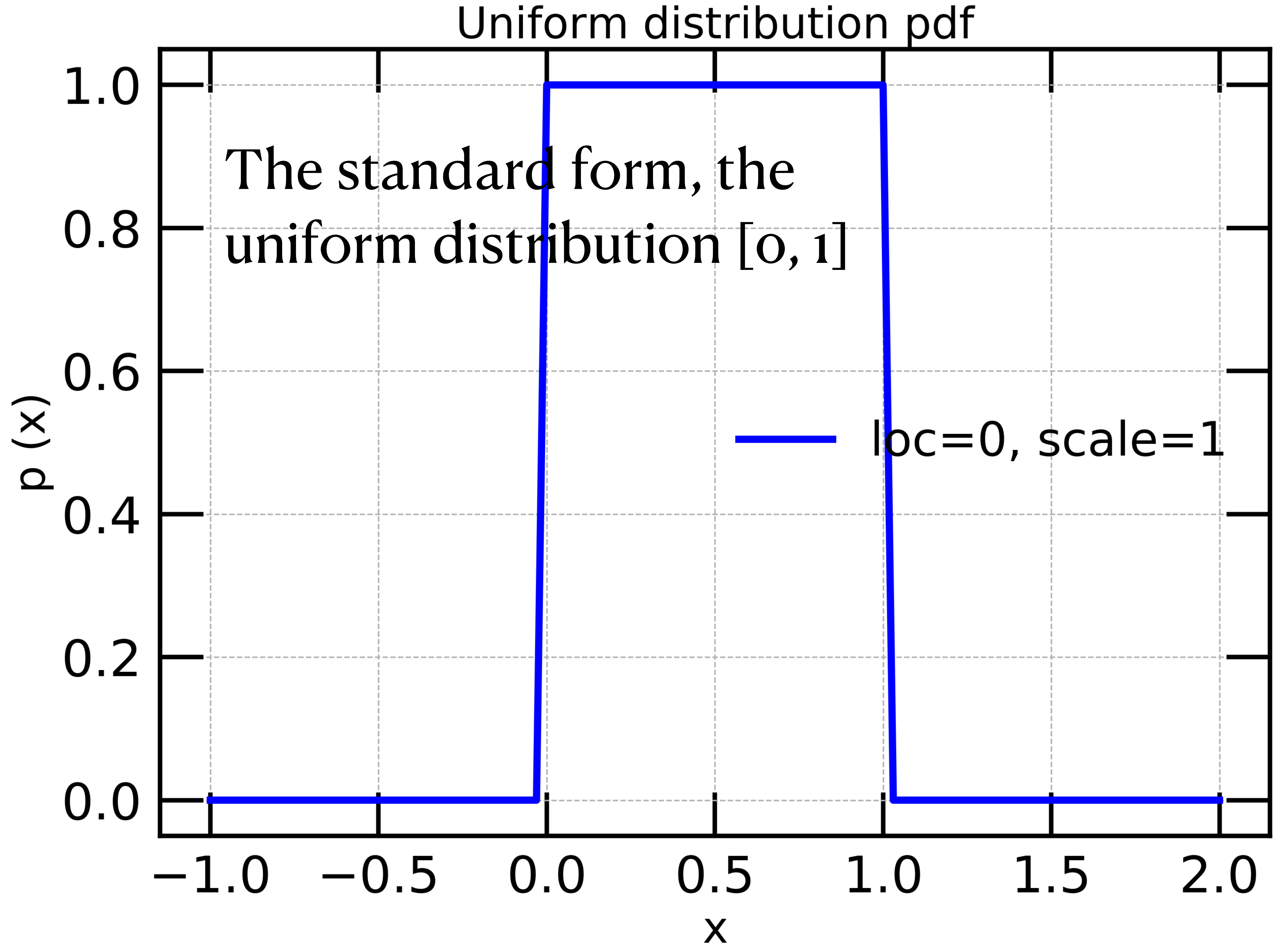
$$f(X) = 0 \quad \text{otherwise}$$

- Expectation (mean):

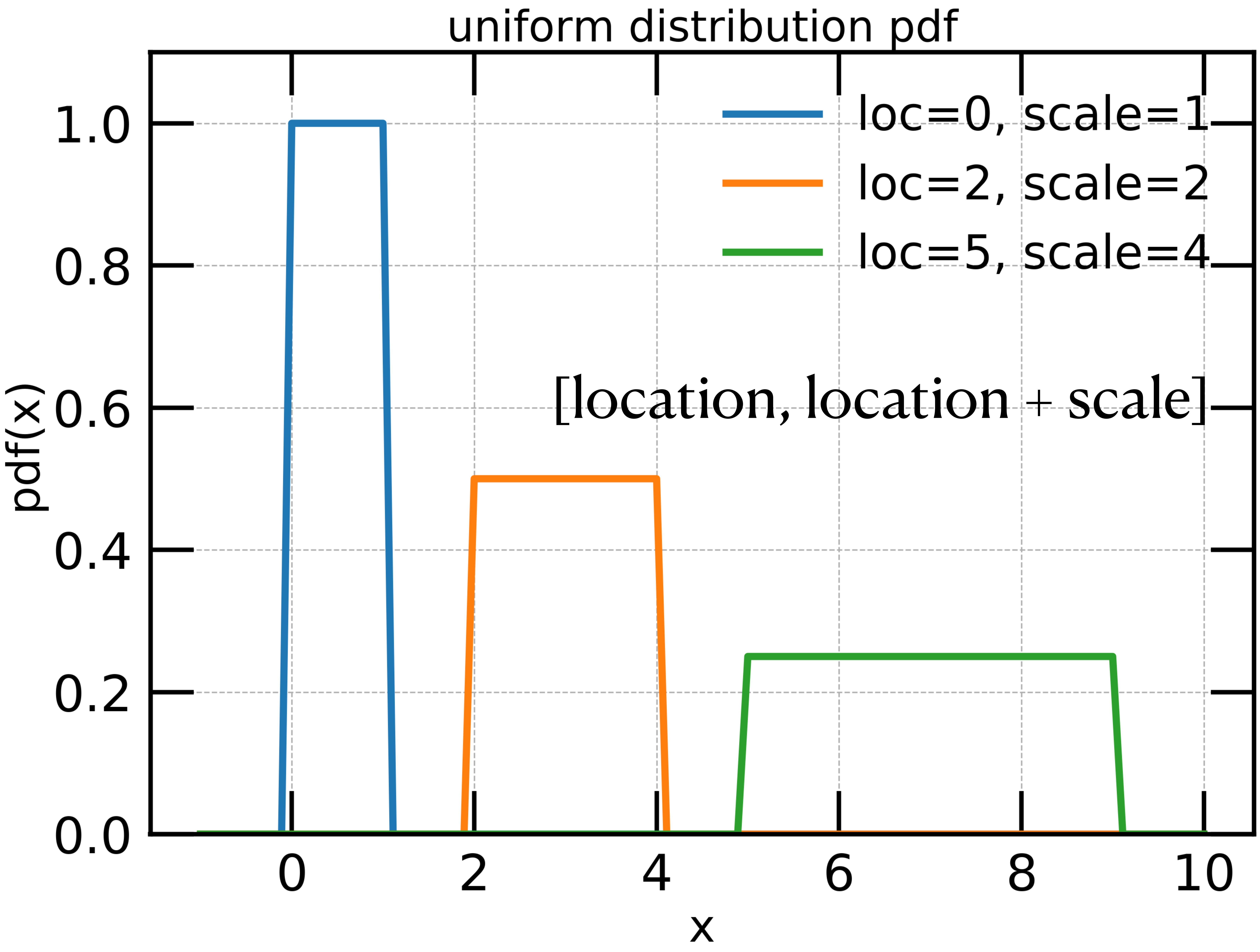
$$E(X) = \frac{a+b}{2}$$

- Variance:

$$Var(X) = \frac{(b-a)^2}{12}$$



The standard form, the uniform distribution [0, 1]

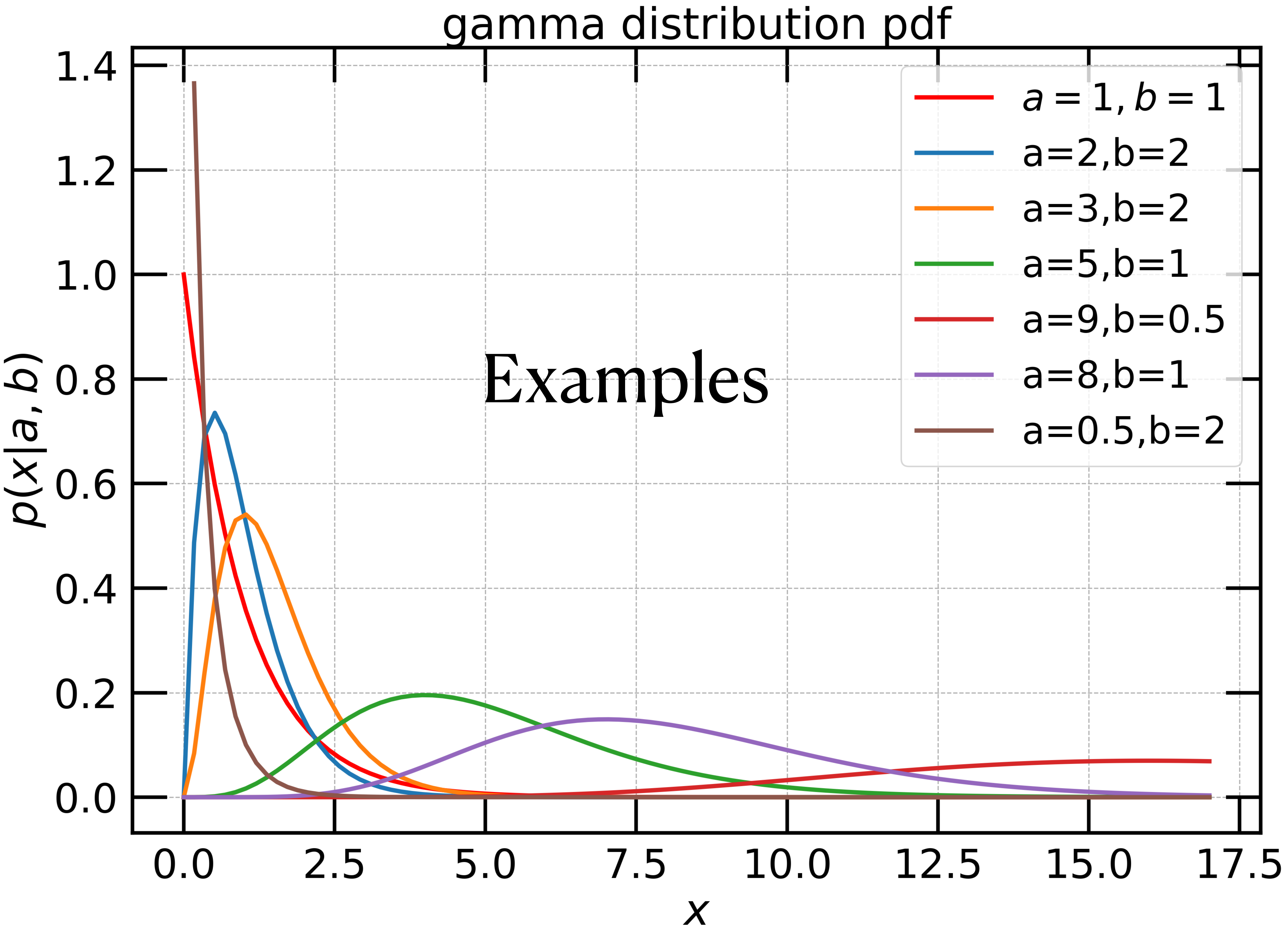

uniform distribution pdf
loc=0, scale=1
loc=2, scale=2
loc=5, scale=4
[location, location + scale]
pdf(x)
x
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10

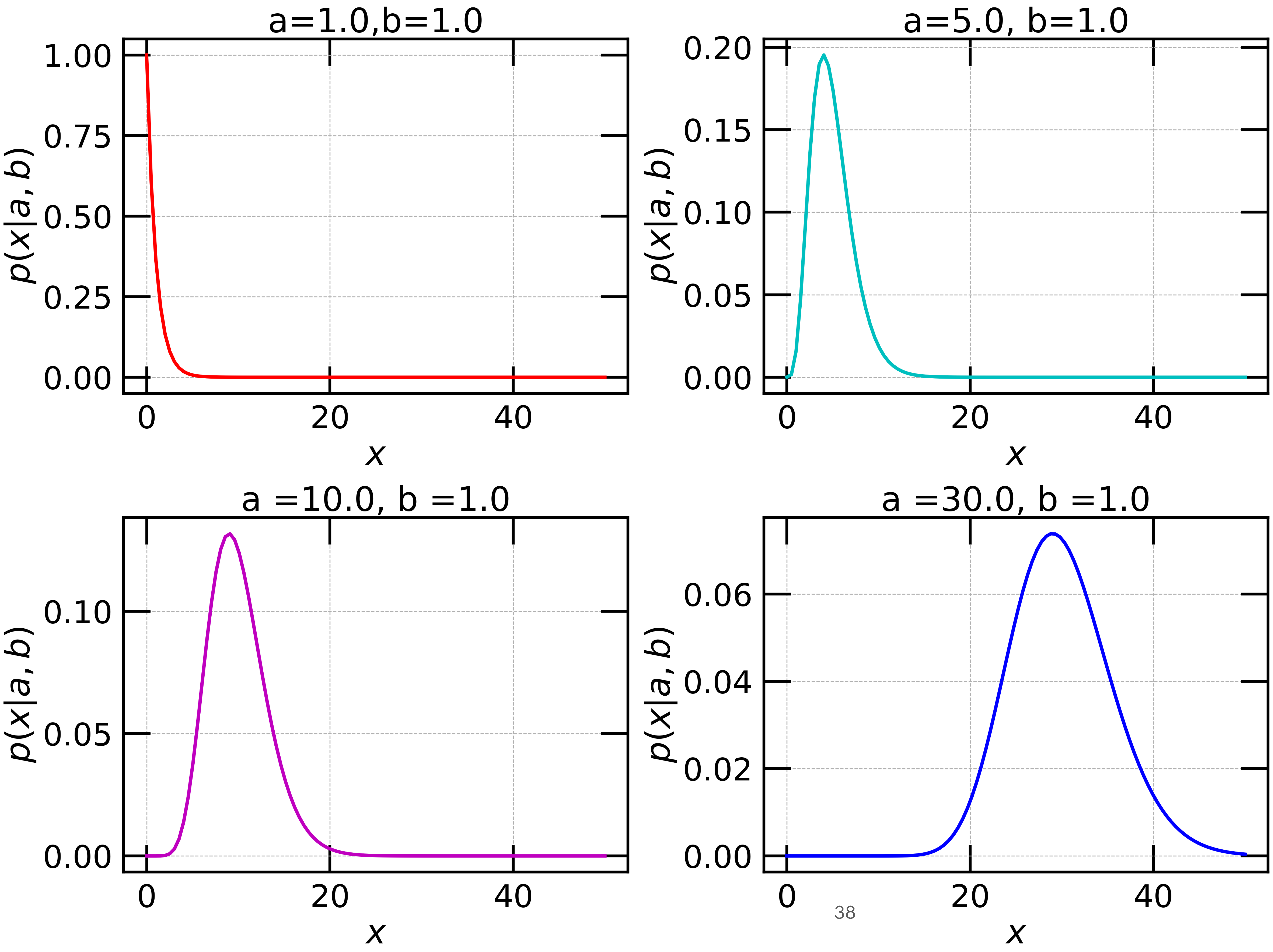
5. Gamma distribution: $Gamma(a, b)$

- The gamma distribution is the maximum entropy, it can be used to model the variables that are always positive $x \in [0,\infty]$.
- The exponential distribution, Erlang distribution, and chi-square distribution are related and special cases of the gamma distribution.
- There are two different parameter: **a shape parameter** a, and an inverse scale parameter b **(rate parameter).**

$$f(x \mid a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$$

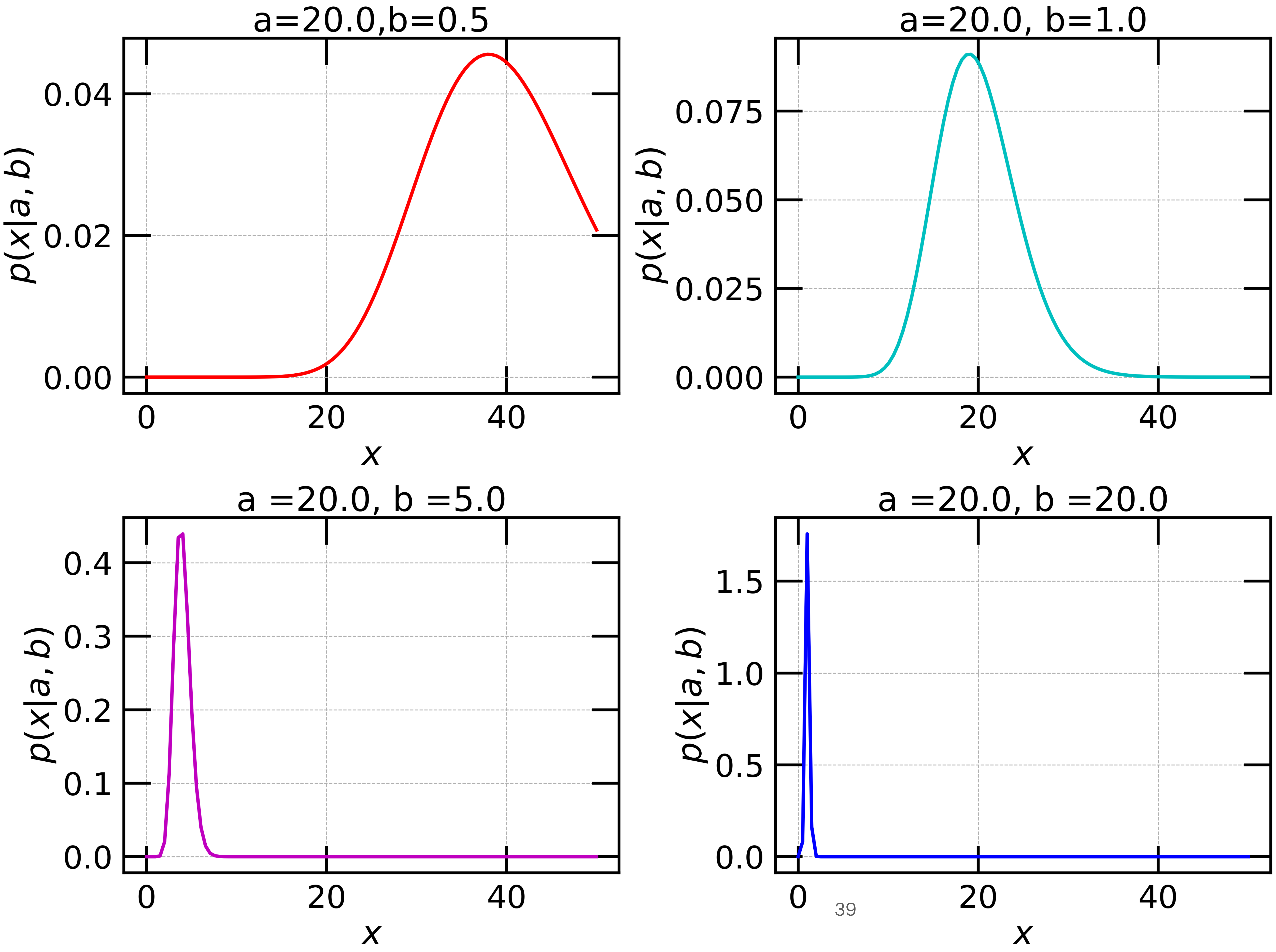
- Expectation (mean): $E(x) = \frac{a}{b}$
- Variance: $Var(x) = \frac{a}{b^2}$


gamma distribution pdf
$a = 1, b = 1$
a=2,b=2
a=3,b=2
a=5,b=1
a=9,b=0.5
a=8,b=1
a=0.5,b=2
Examples
$p(x|a,b)$
x
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5



Vary shape parameter

$Gamma(a, b = 1)$



Vary inverse shape parameter

$Gamma(a = 20, b)$

Gamma function properties

- The gamma function is an extension of the fractional function for real and complex number.
- For a positive real number α:

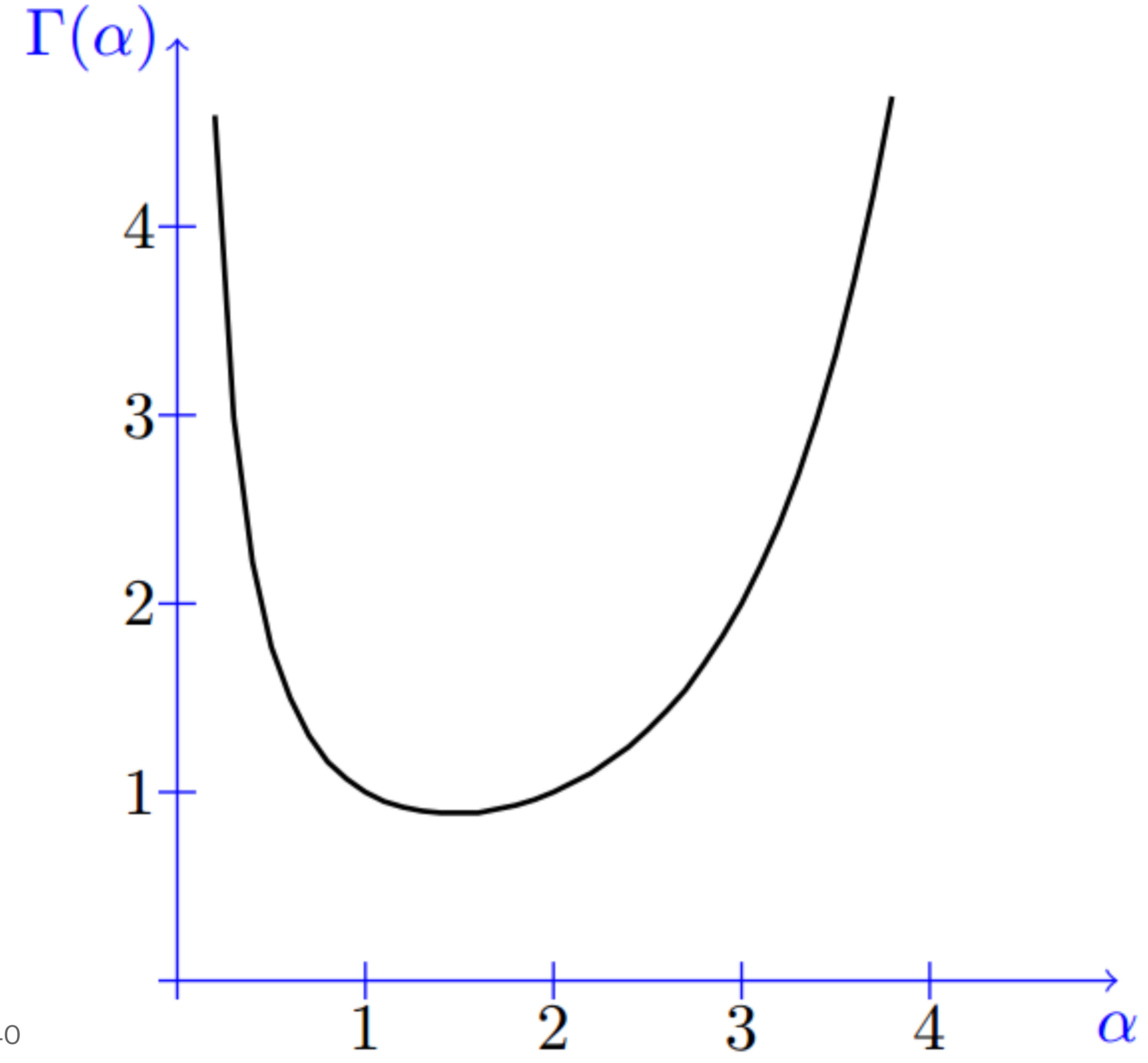
$$(1)\ \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

$$(2)\ \int_0^\infty x^{\alpha-1} e^{-\tau x} dx = \frac{\Gamma(\alpha)}{\tau^\alpha}$$

$$(3)\ \Gamma(\alpha+1) = \alpha\Gamma(\alpha)$$

$$(4)\ \Gamma(n) = (n-1)! \quad n \in [1,2,3,\ldots]$$

$$(5)\ \Gamma(\frac{1}{2}) = \sqrt{\pi}$$



6. Exponential distribution (τ)

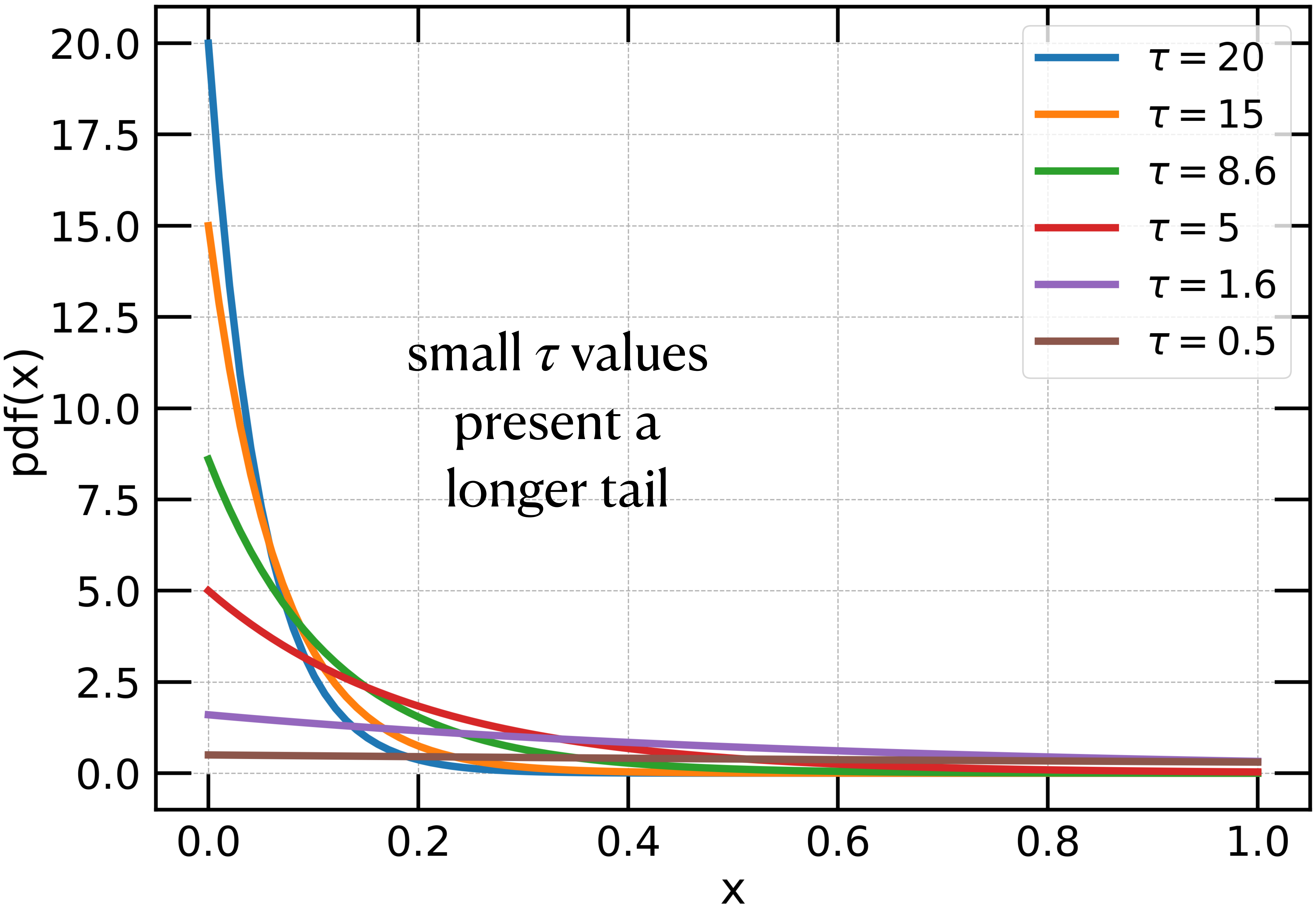
- The exponential distribution is widely used continuous distribution for model the **time decay** events.
- if $a = 1$ in the **gamma distribution:** $Gamma(1,\ b)$, in many experiments, b is the time response τ of events.

$$f(x) = \tau e^{-\tau x}$$

- Expectation (mean): $E(x) = \frac{1}{\tau}$
- Variance: $Var(x) = \frac{1}{\tau^2}$

- Generally, if we sum n independent $exponential(\tau)$ random variables, we will get a $Gamma(n, \tau)$ random variable.

6. Exponential distribution (τ)



6. Exponential distribution (τ)

Example:

Number of Cosmic Microwaves background photons is about 400 photons per second.

https://online.stat.psu.edu/stat414/lesson/15/15.3

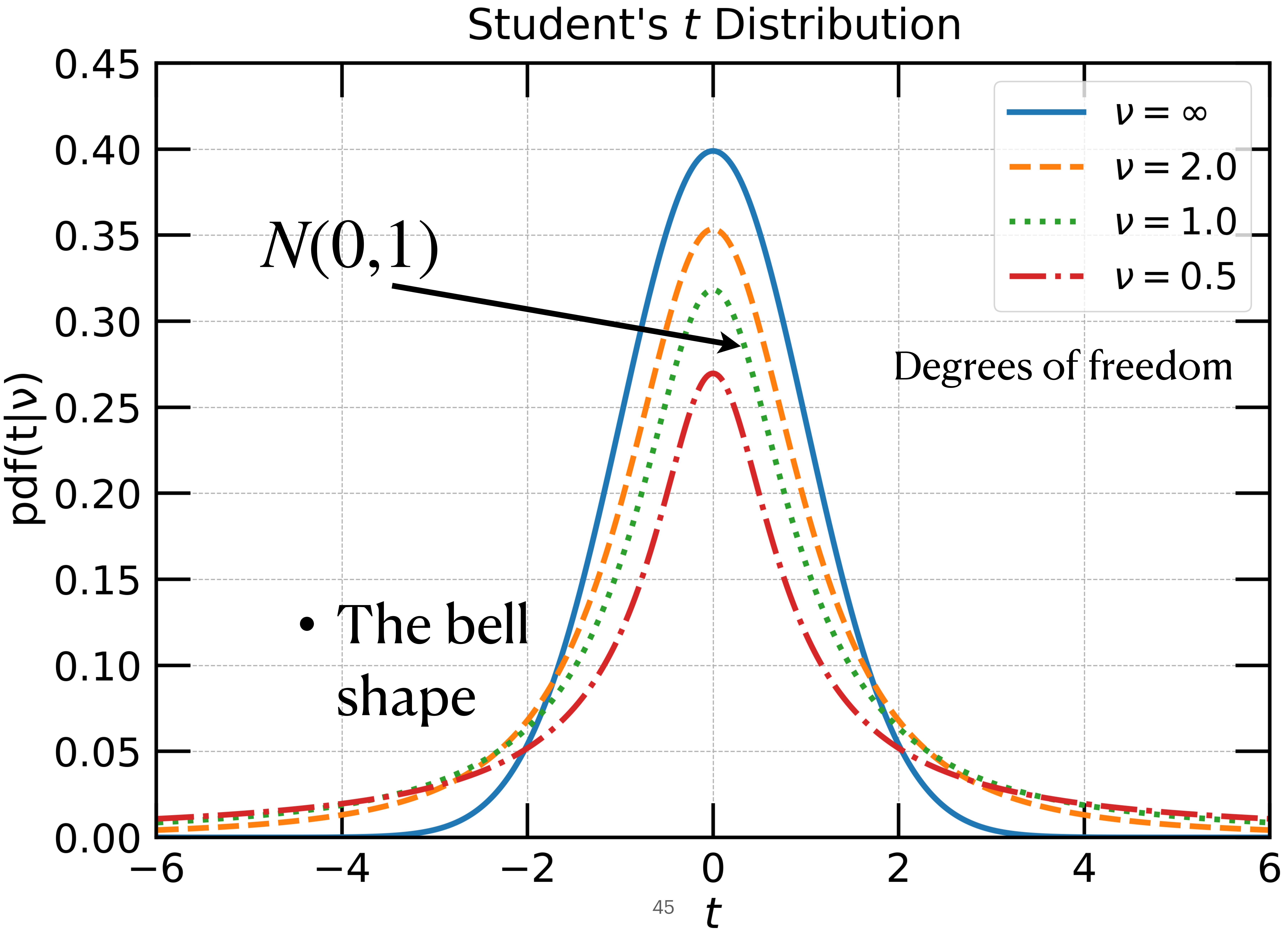
7. Student's t-distribution

- When we estimate the mean of a normal distribution where the sample size is small and the standard deviation is unknown.
- The probability density function is symmetric and has the bell shape of a normal distribution $N(0,1)$.

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

- ν is the number of degrees of freedom.

- Expectation (mean): $E(t) = 0$
- Variance: $Var(t) = \dfrac{\nu}{\nu - 2} \quad ; \nu > 2$


Student's t Distribution
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
pdf(t|ν)
−6
−4
−2
0
2
4
6
t
$\nu = \infty$
$\nu = 2.0$
$\nu = 1.0$
$\nu = 0.5$
$N(0,1)$
Degrees of freedom
• The bell shape

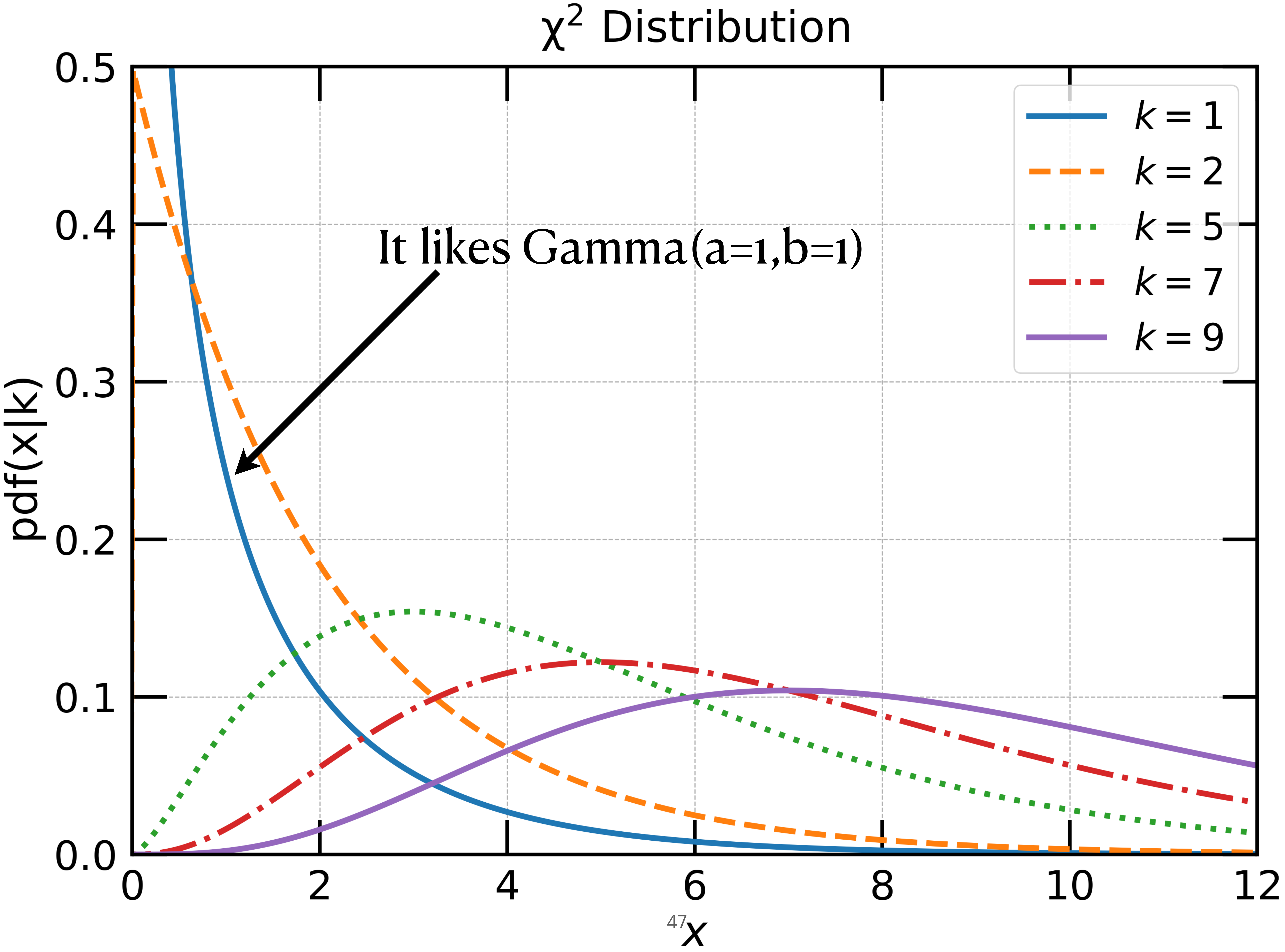
8. Chi-Square Distribution: $\chi^2(k)$

- Chi-square distribution is **sum** of the **square** of k independent standard **normal distributions** $N(0,1)$. Chi-square distribution is a particular case of the gamma distribution $Gamma(a, b = 1)$.
- Chi-square distribution is mostly used in **hypothesis testing** and **confidence level** construction, **goodness of fit.**
-

$$f(x) = \frac{x^{\frac{k}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})}$$

k: degrees of freedom

- Expectation (mean): $E(x) = k$
- Variance: $Var(x) = 2k$



χ² Distribution
It likes Gamma(a=1,b=1)
k = 1
k = 2
k = 5
k = 7
k = 9
pdf(x|k)
x
0.0
0.1
0.2
0.3
0.4
0.5
0
2
4
6
8
10
12

Relationship among distributions