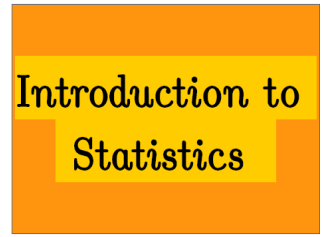


Introduction to Statistics

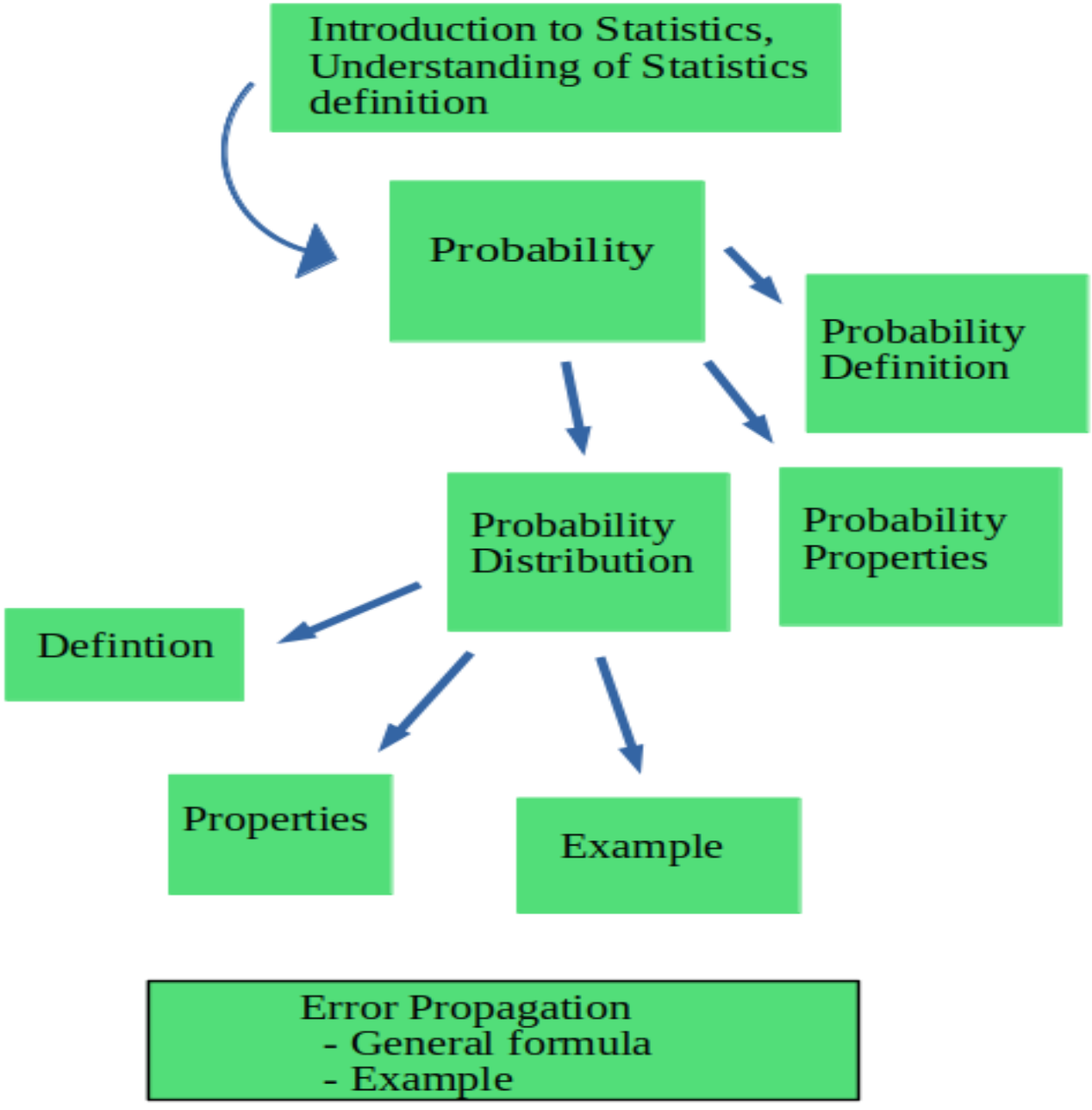
Master Class Winter 2021

Binish Batool

batool@hep.physik.uni-siegen.de

November 5, 2021

Outline


Introduction to Statistics, Understanding of Statistics definition
Probability
Probability Definition
Probability Distribution
Probability Properties
Defintion
Properties
Example
Error Propagation
- General formula
- Example

Introduction

- What is Statistics: Collection of data, organizing, analyzing and conclusion of data.
- What is data: A set of (S) repeated results ($s_1, s_2 ....$) of an experiment E after performing it certain times (n).

  e.g

  "Experiment" (E) = Throwing a dice

  n = 4

  S = $\{s_1 = 1, s_2 = 3, s_3 = 4, s_4 = 6\}$

  S = $\{s_1, s_2, s_3, s_4\}$ Can be repeating

  S = $\{1, 3, 4, 6\}$

  Or

  S = $\{1, 1, 1, 1\}$ etc

Introduction (Cont...)

Understanding the definition of Statistics:

- Collection of data: You throw a dice 60 times and count number of times each face of the dice has appeared.
- Organize: (sorting) What is the total number of appearance of each face.
- Analyzing: If each face appeared with the same amount of times ? (Ideal case: Every face appears same number of times)
- Conclusion/ Inference: If one face appears alot than others. This is baised dice.

Introduction (Cont ...)

Probability: Chance of a certain result to happen

- In the experiment E individual result has (P_r = 1/6) so that all possibilities add to 1 i.e
  ($P_{r1} + P_{r2} + P_{r3} + P_{r4} + P_{r5} + P_{r6} = 1$)
  where r has the maximum number of the results
  :- r has total 6 values
- (P_r > 1/6) if the dice is biased i.e it's one of the faces has the metalic ball/impurity inside it
  e.g
  Throwing such dice for e.g 2000 times will give probablity distribution

Introduction Cont ....

Probability Distribution

- The "rule" or probability distribution that governs the random result of the experiment is not arbitrary. It follows some pattern. It's a function of the result with some parameters a,b,...

$$P(r) = f(r|a, b, ..)$$

e.g
For Perfect Dice: there is no parameter so it's probability distribution

$$P(r) = f(r) \text{ where f(r) = 1/6 for all r}$$

For impure Dice:

$P(r) = f(r|x, y, z)$ where (x,y,z) are the position coordinates of the centre of mass of the metalic ball.

Introduction Cont...

Probability Distribution Continue....

Model: (x,y,z) is the parameter of the model
Inferring the Model (Position of the ball in the dice):
Compare the simulation with the real world experiment.

Prediction from Simulation:

->Take a random initial value of (x_1, y_1, z_1), simulate the results of throwing say 1000 times the dice.

->Change the (x_1, y_1, z_1) to (x_2, y_2, z_2) redo and then (x_3, y_3, z_3) so on.

->Get the data from simulation

->Perform real experiment to some possible number of times and compare with the simulation

->Best matched value of the simulation to real is the prediction of the position of the ball inside the dice.

Introduction Cont...

Properties of the Probability

1 - Probability of all possible results should summed up to 1.

$$\sum_{r}^{r \in R} P(r) = 1$$

2 - Probability of two(or more) Independent Experiments:

Say there are two experiments E_1 and E_2 each has possible results $r_1 \in R_1$ and $r_2 \in R_2$ and probability distribution $P_1(r_1)$ and $P_2(r_2)$. Then the combined experiments will have possible result of

$$r = r_1 \times r_2 \in R_1 \otimes R_2$$

and the probability distribution is:

$$P(r) = P(r_1 \times r_2) = P_1(r_1) \times P_2(r_2)$$

This is called multiplication rule.

Introduction Cont....

Properties of the Probability Cont...

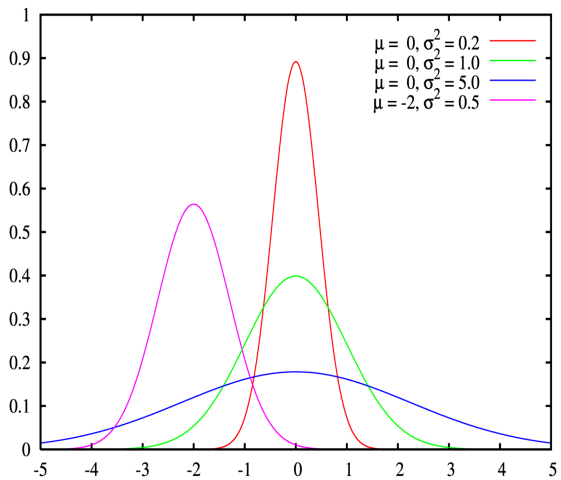
Example: Throwing two dices a and b

$$r = (a, b) \in R = \{(1,1), (1,2), (2,1), ...\}$$
$$P((1,2)) = P_a(1) \times P_b(2) = 1/36$$

3 - Mean $\bar{r}$ or μ

The mean value represents the average value of r if we repeat the experiment for infinite times. It's the location of the centre of the probability distribution and formula:

$$\bar{r} = \sum_r P(r) \times r$$



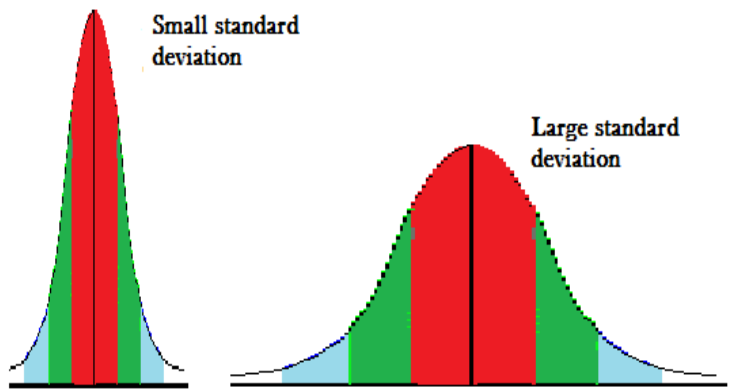
Introduction Cont...

Properties of the Probability Cont...

4 - Standard Deviation σ^2

Standard deviation measures the broadness of the probability distribution and formula:

$$\sigma^2 = \sum_r P(r) \times (r - \bar{r})^2$$



Introduction Cont...

Types of Probability Distribution

1 - Binomial Distribution

Binomial distribution could be thought as easy as **TRUE** or **FALSE**. **SUCCESS** or **FAILURE**.

It is type of distribution which can have two possibilities of the outcomes.

Example:

Let's start from the simplest example; tossing a coin that can bring either **HEAD** or **TAIL** on the top.

If you toss 2 times then $R = \{0head, 1head\}$

Similarly

Tossing n times will result

$R = \{0head, 1head, 2heads, 3heads, ..., nheads\}$

Introduction Cont...

Types of Probability Distribution Cont...

1 - Binomial Distribution Cont...

The probability distribution:

The probability distribution of binomial distribution is

$$P(0) = P(\text{tail}) = 1/2$$

$$P(1) = P(\text{head}) = 1/2$$

$$\sum_{r=1}^{r \in R} P(r) = 1$$

Introduction Cont...

Types of Probability Distribution Cont...

1 - Binomial Distribution Cont...

The probability distribution cont... :

can define $P(r) = f(m|n, p)$, which is the probability to get m heads up out of the n tosses.

m $(0 \leq m \leq n)$ is the index for the random result $r \in R$

p is the chance of head up in 1 toss.

m is the variable

n and p are the parameters.

Introduction Cont...

Types of Probability Distribution Cont...

1 - Binomial Distribution Cont...

The probability distribution cont... :

To calculate $f(m|n,p)$, consider the case of the 1st to the mth tosses giving head up and the following n - m tosses giving tail up. The probability of this result r' to happen is, according to the multiplication rule:

$$P(r') = \underbrace{P(head) \times P(head) \ldots P(head)}_{m \text{ factors}} \times \underbrace{P(tail) \times P(tail) \ldots P(tail)}_{n-m \text{ factors}}$$

$$= P(head)^m P(tail)^{(n-m)}$$

$$= p^m (1-p)^{n-m}$$

Introduction Cont...

Types of Probability Distribution Cont...

1 - Binomial Distribution Cont...

The probability distribution cont... :

However, to get m heads, it is not necessarily that the first m tosses give head. It can be any m tosses out of the n tosses giving head and each of these results have the same probability as above. If there are total N such results:

$$P(r) = N \times P(r')$$

N can be calculated from the knowledge of permutation:

$$N = C_n^m = \frac{n!}{(n-m)!m!}$$

where $n!$ is defined as $n! = 1 \cdot 2 \cdot \ldots \cdot n$.

Introduction Cont...

Types of Probability Distribution Cont...

1 - Binomial Distribution Cont...

The probability distribution cont... :

Final Formula for tossing n times:

$$P(r) \equiv f(m|n,p) = C_n^m p^m (1-p)^{n-m}, \ m = 0, 1, ..., n$$

The distribution $f(m|n)$ is called binomial distribution.

It describes a more general type of question than tossing a coin: the number of successes m in a sequence of n independent experiments with the probability p of one experiment being success.

Introduction Cont...

Types of Probability Distribution Cont...

1 - Binomial Distribution Cont...

The probability distribution cont... :

Mean of Binomial Distribution

$$\bar{m} = np$$

Standard Deviation of Binomial Distribution

$$\sigma^2 = \sum_r P(r) \times (r - \bar{r})^2$$

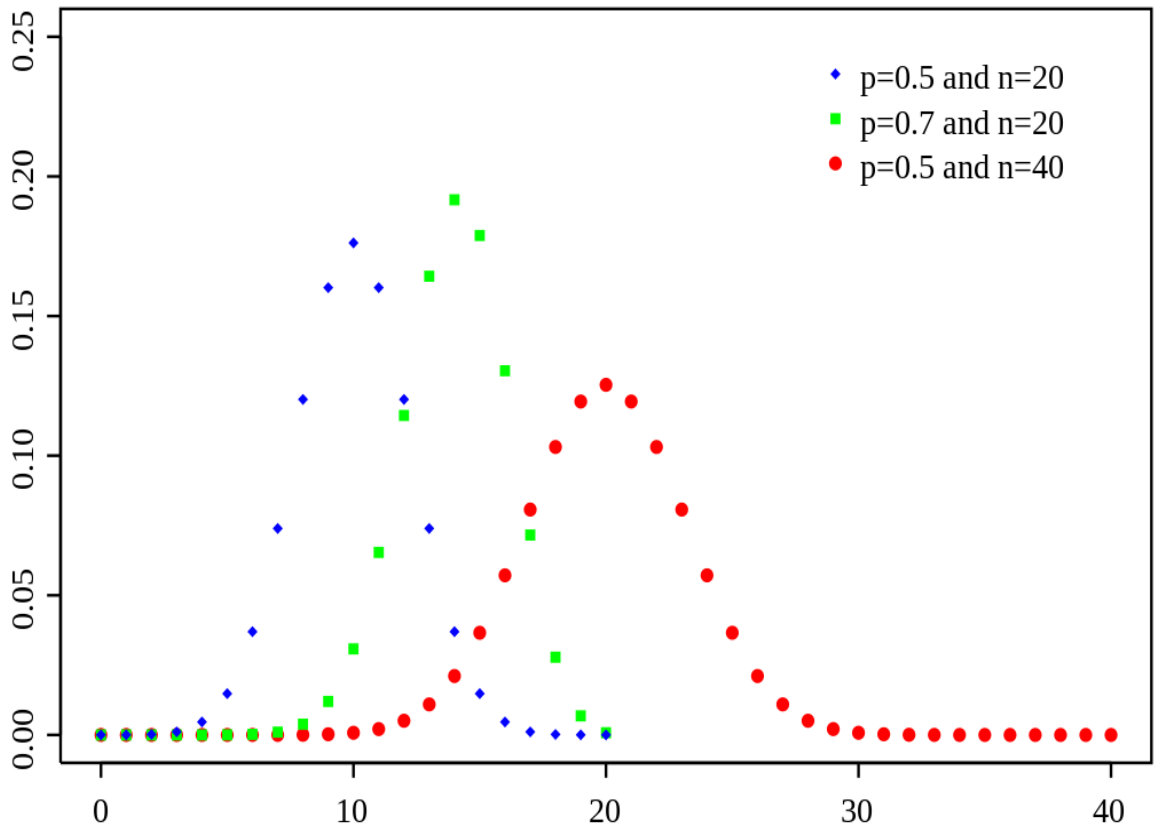
$$\sigma = \sqrt{np(1-p)}$$

Binomial Distribution Example

Example Figure: binomial distribution with different parameter values of *n* and *p*

-> Each mean value represents center of the distribution.

-> Symmetric in mean.



Introduction Cont...

Types of Probability Distribution Cont...

2 - Poisson Distribution

Poisson distribution is actually the extreme case of binomial distribution when the number of trials *n* is infinite large while the success probability *p* is infinite small so that the mean value *np* of the binomial is finite.

Example:

A grosser sells the 4 heads of broccoli per day. It's impractical to say how many heads of lettuce he didn't sell, because we do not know how many customers visited his store or how many they could have bought (and there is really no way to determine the latter). However, we can assume that there were many chances for someone to buy a head of broccoli, so n is very large. The chance of someone buying a head of broccoli at any given moment is very small, so p is small. Finally, the mean, 4 heads of broccoli per day, is known.

Introduction Cont...

Types of Probability Distribution Cont...

2 - Poisson Distribution Cont...

The probability distribution :

Poisson probability distribution will depend on paramenter which is the mean value defined as below:

$$\lambda = np$$

With the above assumption and definition, the Poisson distribution can be derived from the binomial distribution as:

$$P(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Here variable k represents the random result, to be distinguished from m used for binomial distribution. And λ is the parameter of the distribution.

Introduction Cont...

Types of Probability Distribution Cont...

2 - Poisson Distribution Cont...

The probability distribution cont... :

Mean of Poisson Distribution

$$\bar{m} = \lambda$$

Standard Deviation of Poisson Distribution

$$\sigma = \sqrt{\lambda}$$

Poisson Distribution Example

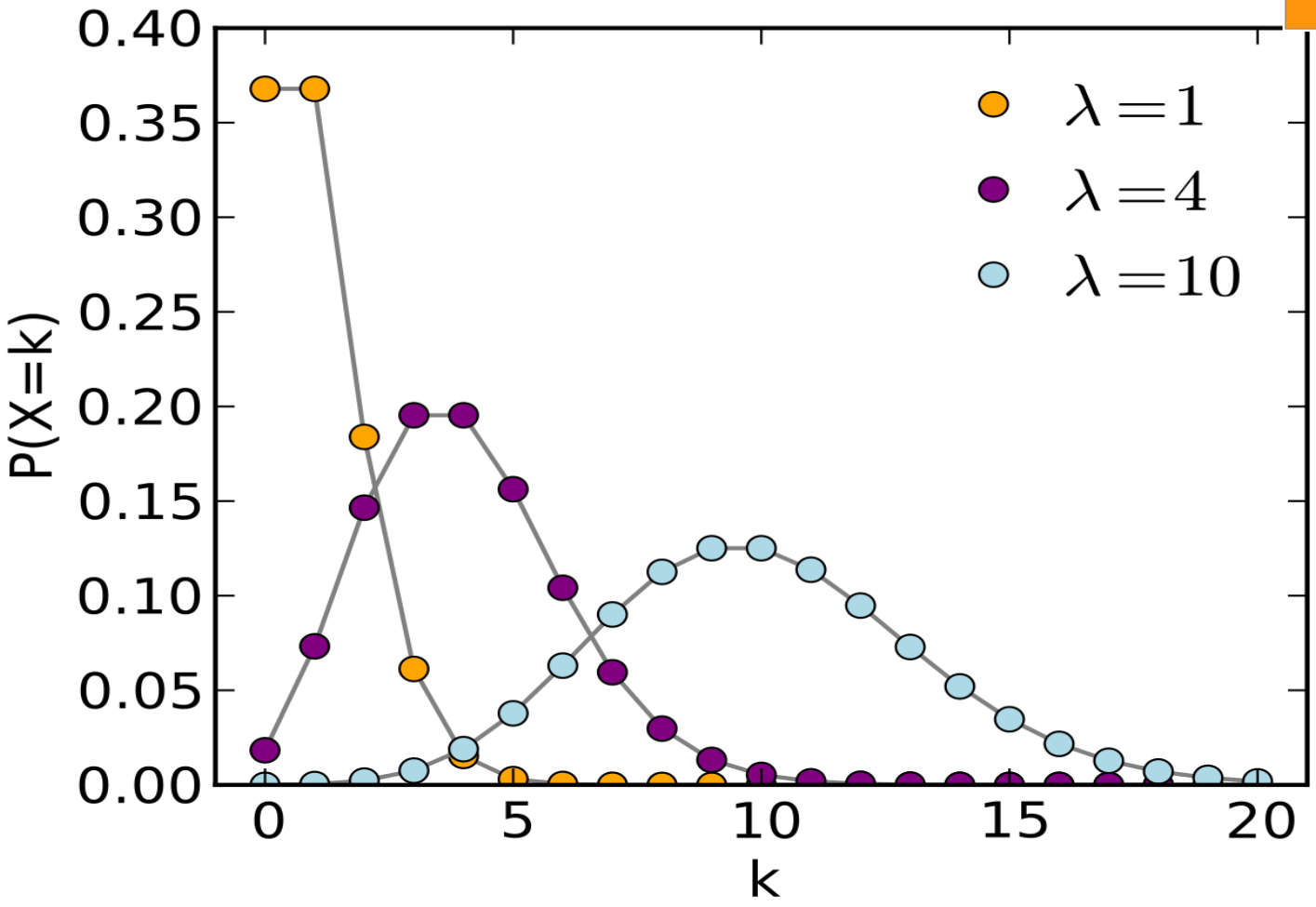


Figure: Poisson distribution with different parameter values of and λ

Introduction Cont...

Types of Probability Distribution Cont...

Poisson Distribution Example: High Energy Physics

Scientists in high energy physics usually collide two beams of particles to create some super rare interaction. In this case, the experiment is the collision of two particles from the two beams. The success is defined as the happening of the rare interaction. If each collision is independent from each other, the number of success in a certain amount of collisions will distribute as a binomial according to the definition of binomial distribution.
However, as the name "super rare" suggests, the chance of having this interaction to happen is very rare, which means p is very tiny.
And again, since the chance is so small, we need to make the two beams so intense and collide them so frequently that the number of trials n gets extremely large. So the two conditions for using the Poisson distribution instead of binomial distribution are satisfied and the probability distribution of the number of rare interactions we observe can be described by Poisson.

Introduction Cont...

Types of Probability Distribution Cont...

3 - Gaussian Distribution Cont...

So when we divide equally the range of the continuous random result $[a, b]$ into n finite pieces, we can talk about the probability of a particular piece $P(i)$ as in the case of discrete random result:

$$P(i) = \int_{x_i}^{x_i+dx} p(x)dx$$

where x_i is the starting point of the piece i an dx the length of the piece.

$p(x)$ is called the probability density distribution or probability density function (p.d.f).

Introduction Cont...

Types of Probability Distribution Cont...

3 - Gaussian Distribution Cont...

Properties:

It has the usual properties of the probability introduced before:

$$\int p(x)dx = 1$$

if z is the combination of x and y, which are independent from each other:

$$p(z) = p(xy) = p(x)p(y)$$

Introduction Cont...

Types of Probability Distribution Cont...

3 - Gaussian Distribution Cont...

Gaussain Probability Distribution :

Now we can introduce the Gaussian distribution, which is used to describe the probability density distribution of continuous random result:

$$p(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\sigma^2\pi}} \exp^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where μ is the mean value and σ the standard deviation (which is easy to validate using the corresponding definitions). The factor before the exponential is to normalize the distribution so that the total probability is one.

Example Gaussian Distribution

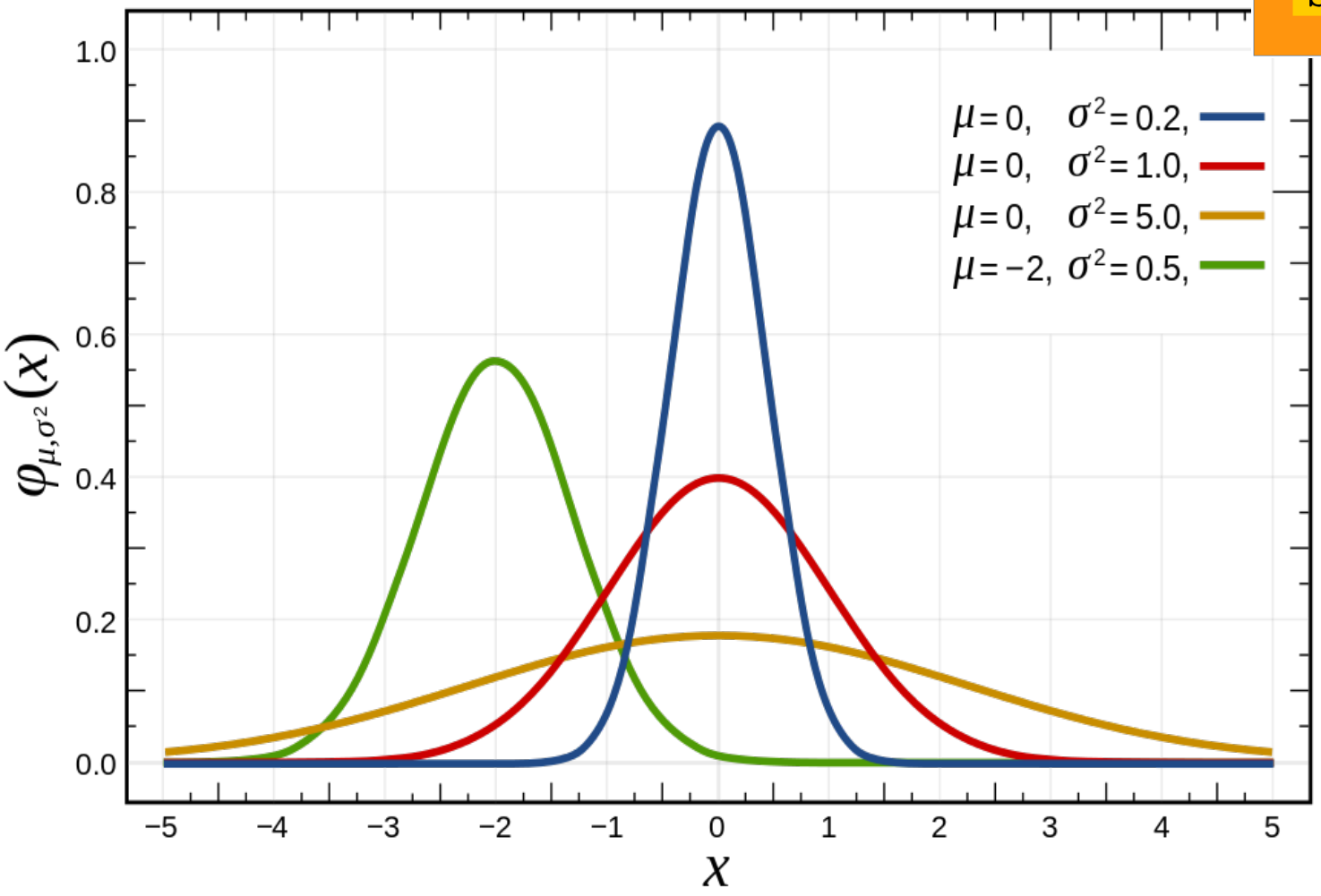


Figure: Gaussian distribution with different parameter values of and λ

Introduction Cont...

Types of Probability Distribution Cont...

Gaussian Distribution Example: High Energy Physics

In high energy physics, the energy of a particle is usually measured by the so-called calorimeter.

The calorimeter is an array of small sub-calorimeters and the energy of the particle is deposited in a bunch of these sub-calorimeters.
So the measured energy of the particle is the sum of the measured energy in each sub-division of calorimeters. However, the measurement in each sub-calorimeter itself is a random process (e.g. due to electronic noise).
So the particle energy is the sum of many independent random processes and it will distribute like a Gaussian.

Introduction Cont...

Measurement and it's error

Measurement

Each time when we measure something, we are actually performing an experiment which gives random result and the result follows some sort of probability distribution $P(r)$ or probability density distribution $p(x)$.

If the measurement result is random in itself, then what are we measuring? It must be something that is certain.
And the answer lies at the purpose of the measurement: usually we perform a measurement (for one time or multiple times) to infer from the measured value(s) about the value of the parameter in $P(r)$ or $p(x)$.
In short, measurement is a kind of parameter inference.

Introduction Cont...

Measurement and it's error

Measurement Error

For example, when we use a ruler to measure the length of a table, we intuitively think that the read from the tick mark tells us the true length of the table.
This is already some kind of inference (although quite simple).
To clearly see the inference we made implicitly, we need to know that the measured length is a random result following a Gaussian distribution whose parameter μ is the true length and σ the resolution of the ruler.

Introduction Cont...

Measurement and it's error

Measurement Error

A nice example is the measuring of the decay rate of some isotope. If in time period T, we observed N decay events, then the measured decay rate is $\Gamma = N/T$.
Is it the true decay rate?
we need to know that the number of observed decay events in time interval T follows a Poisson distribution.
If assuming the true decay rate to be Γ_t, then the number of observed events N follows:

$$P(N|\lambda) = \frac{(\lambda)^N e^{-\lambda}}{N!}$$

where $\lambda = \Gamma_t \cdot T$ is the expected number of decay events that will be observed. Then from the formula $\Gamma = N/T$, we know the measured decay rate Γ is also random.
So We are uncertain about whether we should claim that the measured decay rate is indeed the true decay rate.

Introduction Cont...

Measurement and it's error

Measurement Error

To express this uncertainty about the real position of the true value, we introduce the error of the measurement or error of the inferred parameter σ. And in this case, we say $\Gamma_t = \Gamma \pm \sigma_\Gamma$.

Introduction Cont...

Measurement and it's error

Measurement Error

Error Propagation

In physical measurement, it is often required to derive a physical quantity from the combination of several measured quantities. Consider the simple case below:

$$z = f(x, y)$$

where z is the final quantity we are interested in, x and y are two independently measured quantities, and f is the function relationship between z and x and y.
If x and y are measured to be $x_0 \pm \sigma_x$ and $y_0 \pm \sigma_y$, what is the uncertainty of z: σ_z?
This is a problem of error propagation.

Introduction Cont...

Measurement and it's error

Measurement Error

Error Propagation

adding the uncertainties in quadrature

$$\sigma_z = \sqrt{(\frac{df}{dx}\sigma_x)^2 + (\frac{df}{dy}\sigma_y)^2}$$

Introduction Cont...

Measurement and it's error

Measurement Error

Error Propagation

Take the previous decay rate measurement as example. The expected number of event decay is measured to be $N \pm \sqrt{N}$ and if the total time is also measured with some uncertainty $T \pm \Delta T$, then the decay rate, which is:

$$\Gamma = \frac{N}{T}$$

has the uncertainty of:

$$\sigma_\Gamma = \sqrt{(\frac{1}{T}\sqrt{N})^2 + (-\frac{N}{T^2}\Delta T)^2}$$

Introduction Cont...

Measurement and it's error

Measurement Error

Error Propagation (General)

There is a more efficient formula to calculate the propagated relative uncertainty if the function f is purely the product or ratio of several variables:

$$f(x_1, x_2, x_3, \ldots) = x_1 \cdot x_2 \cdot \ldots$$

then we have:

$$\frac{\sigma_f}{f} = \sqrt{(\frac{\sigma_{x_1}}{x_1})^2 + (\frac{\sigma_{x_2}}{x_2})^2 + (\frac{\sigma_{x_3}}{x_3})^2 + \ldots}$$

Weighted Mean

Let's review the last example
what if we group the first m measurements in one group and the rest $n - m$ the other ?
In this case, the average of the two groups according to the formula are:

$$\bar{x}_m = \frac{x_1 + \ldots + x_m}{m}$$
$$\bar{x}_{n-m} = \frac{x_{m+1} + \ldots + x_n}{n - m}$$

Then combining the results of the two groups again, the final average will be:

$$\begin{aligned}\bar{x} &= \frac{\bar{x}_m + \bar{x}_{n-m}}{2} \\ &= \frac{(n-m)(x_1 + .. + x_m) + m(x_{m+1} + \ldots + x_n)}{2m(n-m)}\end{aligned}$$

which is different from the case of averaging the measurements in one step if $n - m \neq m$!

Introduction Cont...

Weighted Mean

concept of weighted arithmetic mean: which means we assign a smaller weight to the group with less measurements and larger weight to the one with more measurements:

$$\bar{x}_w = w_1 \cdot \bar{x}_m + w_2 \cdot \bar{x}_{n-m}$$

where $w_1 = m/n$ and $w_2 = (n-m)/n$.
In this case, the recombined average is:

$$\begin{aligned} \bar{x} &= \frac{m}{n}\bar{x}_m + \frac{n-m}{n}\bar{x}_{n-m} \\ &= \frac{x_1 + .. + x_m + x_{m+1} + \ldots + x_n}{n} \end{aligned}$$

So we recovered the previous average.

Weighted Mean

Now we can define generally the weighted arithmetic mean as:

$$\bar{x}_w = \sum_i w_i x_i$$

From the error propagation formula, its error is:

$$\sigma_{\bar{x}_w} = \sqrt{\sum_i (w_i dx_i)^2}$$

Usually the weight is chosen as the inverse of squared error of individual measurement entering the averaging:

$$w_i = 1/\sigma_{x_i}^2$$

and then normalized to the total weights:


THANKS
for your
ATTENTION