Statistical analysis of experimental data
Probability distributions and their properties

Aleksander Filip Żarnecki

FACULTY OF PHYSICS
UNIVERSITY OF WARSAW

Lecture 02

October 20, 2022

Probability distributions and their properties

Definition of Probability

Frequentist definition

When repeating the same experiment a large number of times, $N \gg 1$, the probability of A

$$P(A) \;=\; \lim_{N\to\infty} \frac{N(A)}{N}$$

where $N(A)$ is the number of occurrences of the event A

Probability does depends on the definition of the considered sample space!

Frequentist definition

When repeating the same experiment a large number of times, $N \gg 1$, the probability of A

$$P(A) \;=\; \lim_{N\to\infty} \frac{N(A)}{N}$$

where $N(A)$ is the number of occurrences of the event A

Probability does depends on the definition of the considered sample space!

Kolmogorov Axioms

Kolmogorov (1933) formulated the three conditions which have to be fulfilled by probability $P(A)$ of an event $A \subset \Omega$:

1. probability is a non-negative number: $P(A) \geq 0$
2. probability of all possible outcomes (sample space): $P(\Omega) = 1$
3. if A and B are mutually exclusive events: $P(A \cup B) = P(A) + P(B)$

We can derive all properties of the probability from these three axioms...

Statistical Independence

Two events A and B are said to be statistically independent if and only if

$$P(A \cap B) = P(A) \cdot P(B)$$

Two important properties follow:

- mutually exclusive (nonempty) events cannot be independent
- if A is subset of B, $A \subset B$, they cannot be independent, unless $B = \Omega$

Statistical Independence

Two events A and B are said to be statistically independent if and only if

$$P(A \cap B) = P(A) \cdot P(B)$$

Two important properties follow:

- mutually exclusive (nonempty) events cannot be independent
- if A is subset of B, $A \subset B$, they cannot be independent, unless $B = \Omega$

Conditional Probability

When two events are not independent, we can consider probability of event A given that another event B is observed:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{or 0 if } P(B) = 0$$

Statistical Independence

Two events A and B are said to be statistically independent if and only if

$$P(A \cap B) = P(A) \cdot P(B)$$

Two important properties follow:

- mutually exclusive (nonempty) events cannot be independent
- if A is subset of B, $A \subset B$, they cannot be independent, unless $B = \Omega$

Conditional Probability

When two events are not independent, we can consider probability of event A given that another event B is observed:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{or 0 if } P(B) = 0$$

For independent events: $P(A|B) = P(A)$

Total Probability Theorem

Given partition A_i of the sampling space, for any event B we can write

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)$$

Total probability of B can be calculated as a sum over probabilities calculated in separate sub-spaces. Very useful in many cases...

Total Probability Theorem

Given partition A_i of the sampling space, for any event B we can write

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)$$

Total probability of B can be calculated as a sum over probabilities calculated in separate sub-spaces. Very useful in many cases...

Bayes' Theorem

For events A and B the two conditional probabilities are related:

$$P(A|B) = \frac{P(B|A)\, P(A)}{P(B)}$$

This can be also written in a more general form:

$$P(A_i|B) = \frac{P(B|A_i)\, P(A_i)}{\sum_{i=1}^{n} P(A_i)}$$

where A_i is the partition of the sampling space.

Bayesian approach

Bayes theorem can be used to generalize the concept of probability.
In particular, one can consider "probability" of given hypothesis H (theoretical model or model parameter, eg. Hubble constant) when taking into known outcome D (data) of the experiment

$$P(H|D) = \frac{P(D|H)}{P(D)} \cdot P(H)$$

Bayesian approach

Bayes theorem can be used to generalize the concept of probability.
In particular, one can consider "probability" of given hypothesis H (theoretical model or model parameter, eg. Hubble constant) when taking into known outcome D (data) of the experiment

$$P(H|D) = \frac{P(D|H)}{P(D)} \cdot P(H)$$

There are two problems with this approach:

- H can not be considered an event, sampling space can not be defined (no experiment to repeat)
- we need to make a subjective assumption about the "prior" $P(H)$ describing our initial belief in hypothesis H

Bayesian approach

Bayes theorem can be used to generalize the concept of probability.
In particular, one can consider "probability" of given hypothesis H (theoretical model or model parameter, eg. Hubble constant) when taking into known outcome D (data) of the experiment

$$P(H|D) = \frac{P(D|H)}{P(D)} \cdot P(H)$$

There are two problems with this approach:

- H can not be considered an event, sampling space can not be defined (no experiment to repeat)
- we need to make a subjective assumption about the "prior" $P(H)$ describing our initial belief in hypothesis H

For these reasons I rather use term "degree of belief" for the result of the Bayesian procedure applied to non random events

Probability distributions and their properties

1. Random variables
2. Probability distributions
3. Basic probability distributions
   - Binomial distribution
   - Uniform distribution
   - Exponential distribution
   - Poisson distribution
   - Gamma distribution
   - Gaussian distribution

Experiments

So far, we have considered probability as a very general concept.
We only assumed that an experiment delivers data which are a subject to fluctuations. But we did not look at the details of the obtained data.

The outcome of the experiment can be of different nature:

- observation (or non-observation) of given event (true/false)
- observation of an event from given category (classification)
- number of the occurrences of given event (counting)
- value of given observable (measurement)

Experiments

So far, we have considered probability as a very general concept.
We only assumed that an experiment delivers data which are a subject to fluctuations. But we did not look at the details of the obtained data.

The outcome of the experiment can be of different nature:

- observation (or non-observation) of given event (true/false)
- observation of an event from given category (classification)
- number of the occurrences of given event (counting)
- value of given observable (measurement)

We usually present the outcome of the experiment (measurement) in a numerical form. Two general types of variables can be considered:

- discrete (logical, classification and counting)
- continuous (measurement)

Experiments

When repeating the experiment many times, numerical results fluctuate, reflecting fluctuations of the measurement (see previous lecture).

The numerical result of a repeated experiment (measurement) can not be predicted, it is only known when the experiment is made

$\Rightarrow$ that is why we call it a random variable

The true value of the considered physical parameter is usually unknown or known with limited precision only.

By repeating the measurement many times we typically want to increases our knowledge of this parameter.

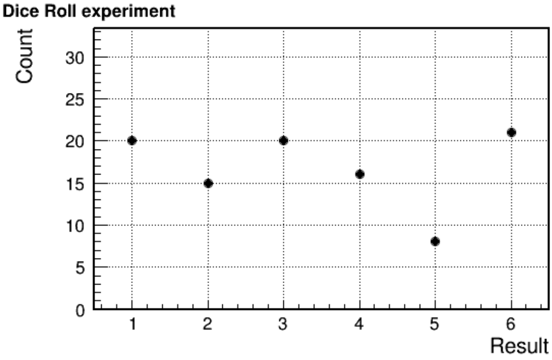
We can also try to understand better the measurement process itself and find the proper description of the observed fluctuations

Distributions

It is very practical to present results of a repeated experiment in a form of a distribution of the considered random variable.

For a discrete variable: plot the count number for each elementary event.

Example of dice roll experiment: 100 rolls, 1st try

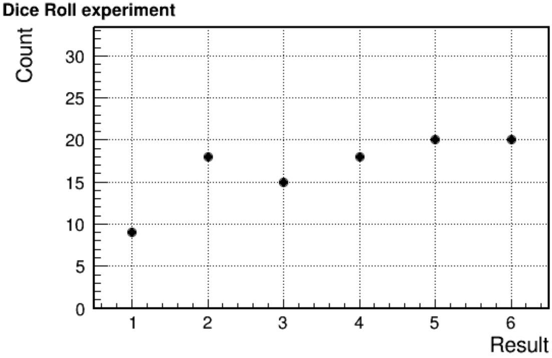


Distributions

It is very practical to present results of a repeated experiment in a form of a distribution of the considered random variable.

For a discrete variable: plot the count number for each elementary event.

Example of dice roll experiment: 100 rolls, 2^{nd} try

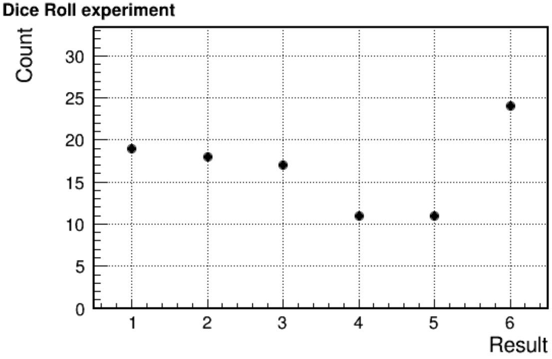


Distributions

It is very practical to present results of a repeated experiment in a form of a distribution of the considered random variable.

For a discrete variable: plot the count number for each elementary event.

Example of dice roll experiment: 100 rolls, 3rd try

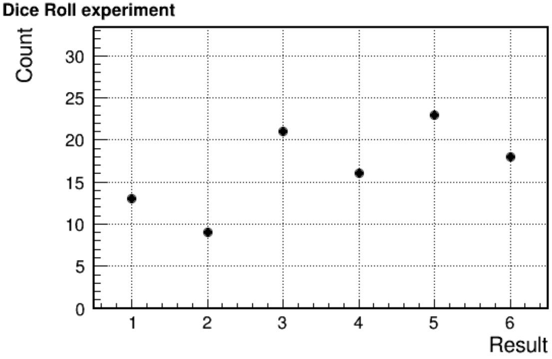


Distributions

It is very practical to present results of a repeated experiment in a form of a distribution of the considered random variable.

For a discrete variable: plot the count number for each elementary event.

Example of dice roll experiment: 100 rolls, 4th try

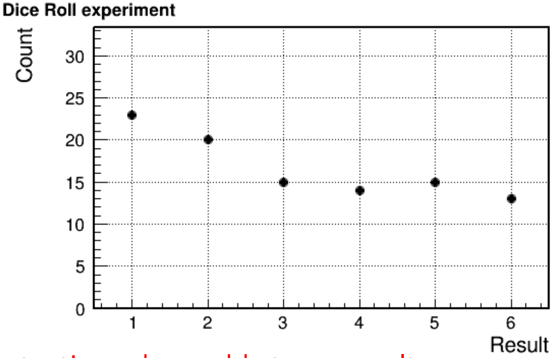


Distributions

It is very practical to present results of a repeated experiment in a form of a distribution of the considered random variable.

For a discrete variable: plot the count number for each elementary event.

Example of dice roll experiment: 100 rolls, 5^{th} try



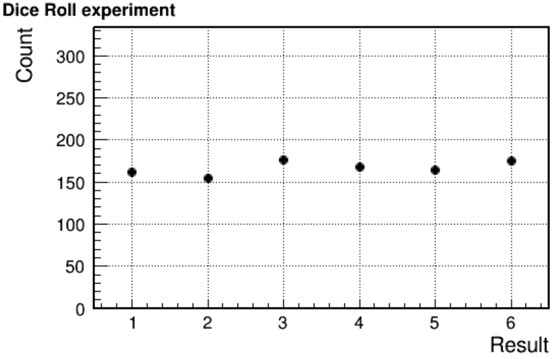
Significant fluctuations observed between results

Distributions

It is very practical to present results of a repeated experiment in a form of a distribution of the considered random variable.

For a discrete variable: plot the count number for each elementary event.

Example of dice roll experiment: 1000 rolls

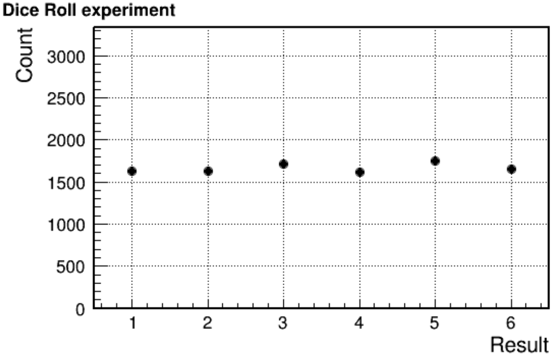


Distributions

It is very practical to present results of a repeated experiment in a form of a distribution of the considered random variable.

For a discrete variable: plot the count number for each elementary event.

Example of dice roll experiment: 10000 rolls

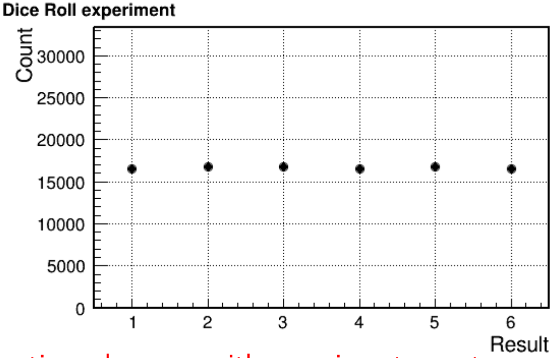


Distributions

It is very practical to present results of a repeated experiment in a form of a distribution of the considered random variable.

For a discrete variable: plot the count number for each elementary event.

Example of dice roll experiment: 100000 rolls



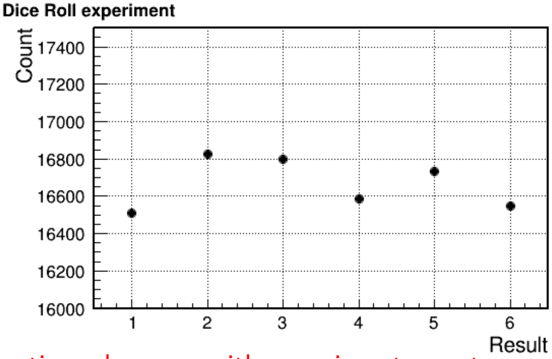
Relative fluctuations decrease with experiment count

Distributions

It is very practical to present results of a repeated experiment in a form of a distribution of the considered random variable.

For a discrete variable: plot the count number for each elementary event.

Example of dice roll experiment: 100000 rolls (zoom)



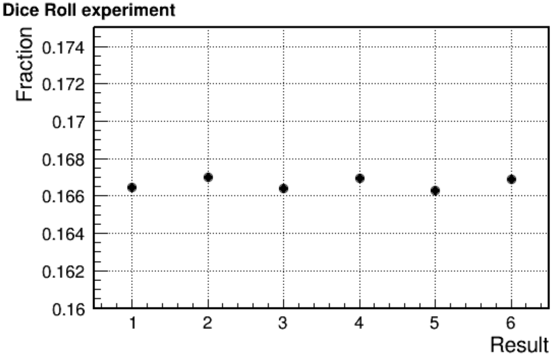
Relative fluctuations decrease with experiment count

Distributions

It is very practical to present results of a repeated experiment in a form of a distribution of the considered random variable.

We can plot also plot the result as the relative fraction.

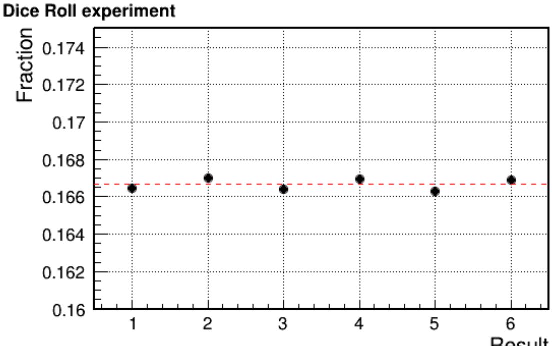
Example of dice roll experiment: 1000000 rolls



Probability distribution function

In the limit of the infinite number of experiments, the relative fraction is given by a probability distribution function (PDF).

For discrete variables, probability distribution function is the probability that a given value of the random variable occurs in a single experiment.

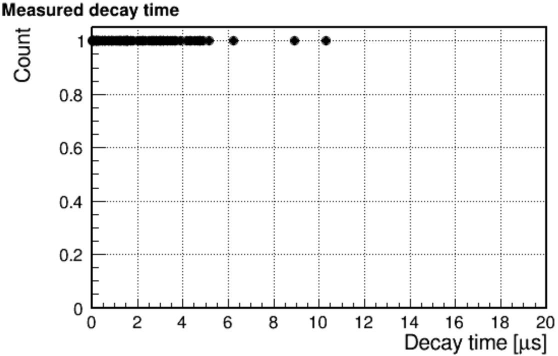


Probability distribution function: $f(n) = P(n) = \frac{1}{6} = const$

Histograms

Graphical presentation becomes more difficult for continuous variable.
With high readout precision, probability of obtaining the same numerical result twice is negligible.

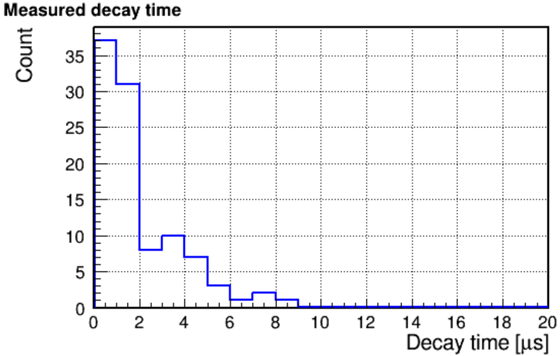
The method of plotting the count number for each result does not work!
Example of decay time measurement: $\tau = 2.2\mu s$, 100 decays



Histograms

Instead of asking for given values, we need to look at defined value ranges.
We usually define a set of value bins covering the whole considered value range and count events when variable value is in given bin.

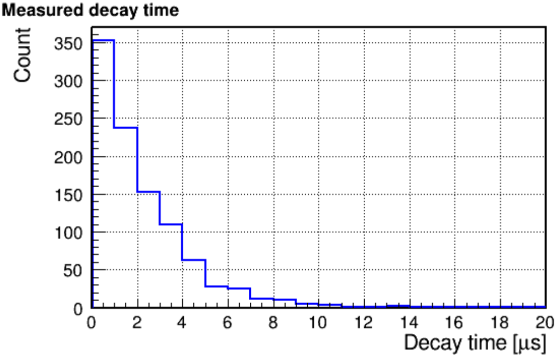
Then we can plot the count number for each bin
Example of decay time measurement: $\tau = 2.2\mu s$, 100 decays



Histograms

Instead of asking for given values, we need to look at defined value ranges.
We usually define a set of value bins covering the whole considered value range and count events when variable value is in given bin.

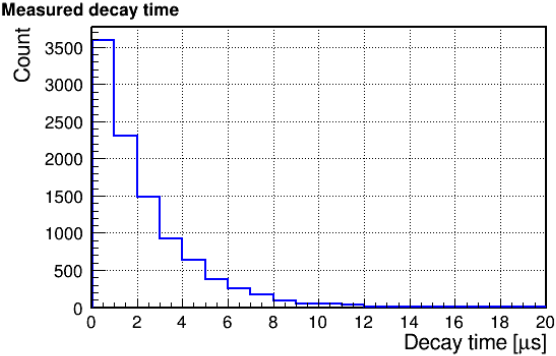
Then we can plot the count number for each bin
Example of decay time measurement: $\tau = 2.2\mu s$, 1000 decays



Histograms

Instead of asking for given values, we need to look at defined value ranges.
We usually define a set of value bins covering the whole considered value range and count events when variable value is in given bin.

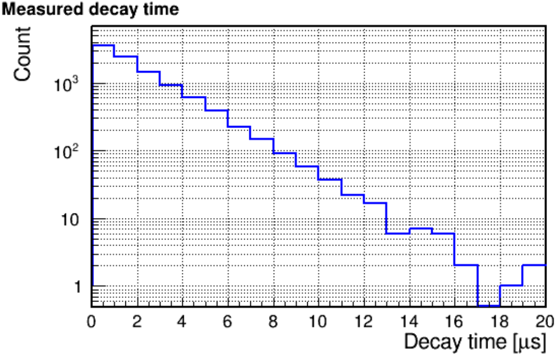
Then we can plot the count number for each bin
Example of decay time measurement: $\tau = 2.2\mu s$, 10000 decays



Histograms

Instead of asking for given values, we need to look at defined value ranges.
We usually define a set of value bins covering the whole considered value range and count events when variable value is in given bin.

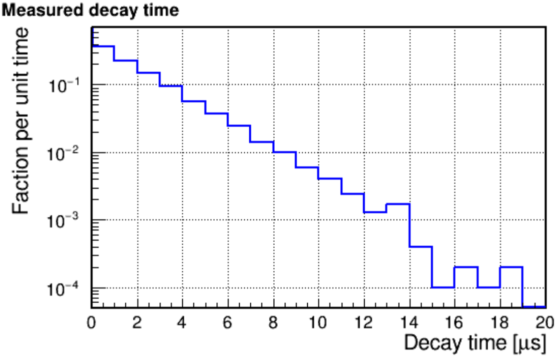
Then we can plot the count number for each bin
Example of decay time measurement: $\tau = 2.2\mu s$, 10000 decays



Probability distribution function

We can calculate the relative fraction of events in each bin.

Example of decay time measurement: $\tau = 2.2\mu s$, 10000 decays

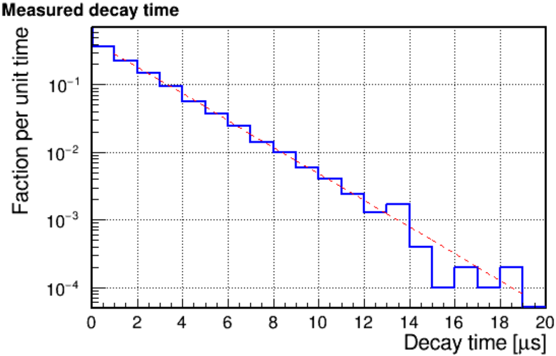


Probability distribution function

We can calculate the relative fraction of events in each bin.

In the limit of the infinite number of experiments (and very narrow bins), the relative fraction is given by a probability distribution function (PDF) of the variable multiplied by the bin width.

Example of decay time measurement: $\tau = 2.2\mu s$, 10000 decays



Probability distributions and their properties

Probability distribution function (PDF)
also called "Probability density function" in some books

For given random variable X, probability distribution function, $f(x)$, describes the probability to obtain given numerical result x (in single experiment). For infinitesimal interval dx:

$$P(x < X < x + dx) \ = \ f(x)\, dx$$

Probability distribution function (PDF)
also called "Probability density function" in some books

For given random variable X, probability distribution function, $f(x)$, describes the probability to obtain given numerical result x (in single experiment). For infinitesimal interval dx:

$$P(x < X < x + dx) \;=\; f(x)\, dx$$

For arbitrary interval $[x_1, x_2]$:

$$P(x_1 < X < x_2) \;=\; \int_{x_1}^{x_2} dx\, f(x)$$

This can be considered an alternative definition of $f(x)$

Cumulative distribution function

We can also define cumulative distribution function $F(x)$, which is the probability that an experiment will result in a value not grater than x:

$$F(x) \;=\; P(X < x)$$

Cumulative distribution function

We can also define cumulative distribution function $F(x)$, which is the probability that an experiment will result in a value not grater than x:

$$F(x) = P(X < x)$$

Probability distribution function can be then written as the derivative:

$$f(x) = \frac{dF(x)}{dx}$$

Cumulative distribution function

We can also define cumulative distribution function $F(x)$, which is the probability that an experiment will result in a value not grater than x:

$$F(x) = P(X < x)$$

Probability distribution function can be then written as the derivative:

$$f(x) = \frac{dF(x)}{dx}$$

Probability that the X value observed is in the range from x_1 to x_2:

$$P(x_1 < X < x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} dx\, f(x)$$

General properties of distribution functions

From the properties of probability

$$f(x) \geq 0 \qquad \int_{-\infty}^{+\infty} dx \, f(x) = 1$$

General properties of distribution functions

From the properties of probability

$$f(x) \geq 0 \qquad \qquad \int_{-\infty}^{+\infty} dx\, f(x) = 1$$

For cumulative distribution:

$$F(x) = \int_{-\infty}^{x} dx'\, f(x')$$

$$\Rightarrow \lim_{x \to -\infty} F(x) = 0 \qquad \lim_{x \to +\infty} F(x) = 1$$

Moments of distribution functions

Expectation value of an arbitrary function $g(x)$ of the random variable X can be defined as

$$\mathbb{E}(g(x)) = \langle\, g(x) \,\rangle = \int_{-\infty}^{+\infty} dx\, g(x)\, f(x)$$

where f(x) is the probability distribution function for X.

Moments of distribution functions

Expectation value of an arbitrary function $g(x)$ of the random variable X can be defined as

$$\mathbb{E}(g(x)) = \langle\, g(x) \,\rangle = \int_{-\infty}^{+\infty} dx\, g(x)\, f(x)$$

where f(x) is the probability distribution function for X.

The expectation value of a random variable itself or **the mean**:

$$\mu = \mathbb{E}(X) = \langle\, x \,\rangle = \bar{x} = \int_{-\infty}^{+\infty} dx\, x\, f(x)$$

Moments of distribution functions

Expectation value of an arbitrary function $g(x)$ of the random variable X can be defined as

$$\mathbb{E}(g(x)) = \langle\, g(x)\,\rangle = \int_{-\infty}^{+\infty} dx\, g(x)\, f(x) \qquad = \sum_i g(x_i)\, f(x_i)$$

where f(x) is the probability distribution function for X.

The expectation value of a random variable itself or **the mean**:

$$\mu = \mathbb{E}(X) = \langle\, x\,\rangle = \bar{x} = \int_{-\infty}^{+\infty} dx\, x\, f(x) \qquad = \sum_i x_i\, f(x_i)$$

For discrete random variables mean is given by the is the sum of all possible values x_i of X multiplied by their corresponding probabilities.

Moment of order n (n^{th} moment) is defined as

$$\mu_n = \mathbb{E}(X^n) = \langle x^n \rangle = \int_{-\infty}^{+\infty} dx\, x^n\, f(x) \qquad = \sum_i x_i^n\, f(x_i)$$

Mean value is, by definition, the first ($n = 1$) moment of the probability distribution, $\mu \equiv \mu_1$.

Moment of order n (n^{th} moment) is defined as

$$\mu_n = \mathbb{E}(X^n) = \langle x^n \rangle = \int_{-\infty}^{+\infty} dx\, x^n\, f(x) \qquad = \sum_i x_i^n\, f(x_i)$$

Mean value is, by definition, the first ($n = 1$) moment of the probability distribution, $\mu \equiv \mu_1$.

Central moment of order n is defined as

$$m_n = \mathbb{E}\left((X-\mu)^n\right) = \langle (x-\mu)^n \rangle = \int_{-\infty}^{+\infty} dx\, (x-\mu)^n\, f(x)$$

$$= \sum_i (x_i - \mu)^n\, f(x_i)$$

By calculating moments of the (unknown) probability distribution we can get information about its shape.

Moments of distribution functions

For the lowest order moments we have:

$\mu_0 \equiv 1$	$m_0 \equiv 1$	normalization of $f(x)$
$\mu_1 \equiv \mu$	$m_1 \equiv 0$	definition of mean value of $f(x)$

Moments of distribution functions

For the lowest order moments we have:

$$\mu_0 \equiv 1 \qquad m_0 \equiv 1 \qquad \text{normalization of } f(x)$$

$$\mu_1 \equiv \mu \qquad m_1 \equiv 0 \qquad \text{definition of mean value of } f(x)$$

The first moment which gives us information about the shape of $f(x)$ is **Variance**, which is the second central moment:

$$\mathbb{V}(X) = m_2 = \langle (x-\mu)^2 \rangle = \int_{-\infty}^{+\infty} dx\, (x-\mu)^2 f(x)$$

$$= \sum_i (x_i - \mu)^2 f(x_i)$$

Moments of distribution functions

For the lowest order moments we have:

$$\mu_0 \equiv 1 \qquad m_0 \equiv 1 \qquad \text{normalization of } f(x)$$

$$\mu_1 \equiv \mu \qquad m_1 \equiv 0 \qquad \text{definition of mean value of } f(x)$$

The first moment which gives us information about the shape of $f(x)$ is **Variance**, which is the second central moment:

$$\sigma^2 = \mathbb{V}(X) = m_2 = \langle (x-\mu)^2 \rangle = \int_{-\infty}^{+\infty} dx\, (x-\mu)^2 f(x)$$

$$= \sum_i (x_i - \mu)^2 f(x_i)$$

The square root of the variance is referred to as the standard deviation σ.
Describes the average difference between measurements x_i and the mean μ

Probability distributions and their properties

Binomial distribution

Consider an experiments with only two possible outcomes
(binary experiment): $\Omega = \{\text{'success'}, \text{'failure'}\}$ (or $\Omega = \{\text{'true'}, \text{'false'}\}$)

This is also the case, when we look for particular event A in wider sampling space.

Consider an experiments with only two possible outcomes
(binary experiment): $\Omega = \{\text{'success'}, \text{'failure'}\}$ (or $\Omega = \{\text{'true'}, \text{'false'}\}$)

This is also the case, when we look for particular event A in wider sampling space.

Assume that the success probability p is known.
What is the probability of having n successes in N tries?
We are not interested in the order in which the successes take place.

Consider an experiments with only two possible outcomes
(binary experiment): $\Omega = \{\text{'success'}, \text{'failure'}\}$ (or $\Omega = \{\text{'true'}, \text{'false'}\}$)

This is also the case, when we look for particular event A in wider sampling space.

Assume that the success probability p is known.
What is the probability of having n successes in N tries?
We are not interested in the order in which the successes take place.

Example

What is the probability to get three 'six' when rolling the die five times?

$$p = \frac{1}{6} \qquad n = 3 \qquad N = 5$$

Consider an experiments with only two possible outcomes
(binary experiment): $\Omega = \{\text{'success'}, \text{'failure'}\}$ (or $\Omega = \{\text{'true'}, \text{'false'}\}$)

This is also the case, when we look for particular event A in wider sampling space.

Assume that the success probability p is known.
What is the probability of having n successes in N tries?
We are not interested in the order in which the successes take place.

Example

What is the probability to get three 'six' when rolling the die five times?

$$p = \frac{1}{6} \qquad n = 3 \qquad N = 5$$

It is important to notice that we ask for a probability for an event from a different, extended sampling space, $\Omega' = \Omega^N$!

Binomial distribution

Describes probability of having n successes in N tries, assuming success probability p in single trial and failure probability $q = 1 - p$

$$P(n) \;=\; \binom{N}{n} p^n \, q^{N-n} \;=\; \frac{N!}{n!(N-n)!} \, p^n \, q^{N-n}$$

The Newton symbol $\binom{N}{n}$ gives the number of possible sequences of N tries giving n successes (regardless of order) and $p^n \, q^{N-n}$ describes the probability for single such sequence (elementary event in Ω').

Binomial distribution

Describes probability of having n successes in N tries, assuming success probability p in single trial and failure probability $q = 1 - p$

$$P(n) = \binom{N}{n} p^n q^{N-n} = \frac{N!}{n!(N-n)!} p^n q^{N-n}$$

The Newton symbol $\binom{N}{n}$ gives the number of possible sequences of N tries giving n successes (regardless of order) and $p^n q^{N-n}$ describes the probability for single such sequence (elementary event in Ω').

Mean (expected value) of the binomial distribution

$$\langle n \rangle = \bar{n} = p N$$

Variance of the distribution important for efficiency uncertainty

$$\sigma^2 = p (1-p) N$$

For given N, distribution is widest for $p = 0.5$

Binomial distribution

Example (1)

What is the probability to get three 'six' when rolling the die five times?

$$p = \frac{1}{6} \qquad n = 3 \qquad N = 5$$

Example (1)

What is the probability to get three 'six' when rolling the die five times?

$$p = \frac{1}{6} \qquad n = 3 \qquad N = 5$$

By putting the numbers directly in the formula we get

$$P(n) \; = \; \frac{N!}{n!(N-n)!} \; p^n \; q^{N-n} \;\; = \;\; \frac{5!}{3! \; 2!} \; \frac{1}{6^3} \; \frac{5^2}{6^2} \; = \; \frac{120 \cdot 1 \cdot 25}{6 \cdot 2 \cdot 216 \cdot 36} \; \approx \; 0.032$$

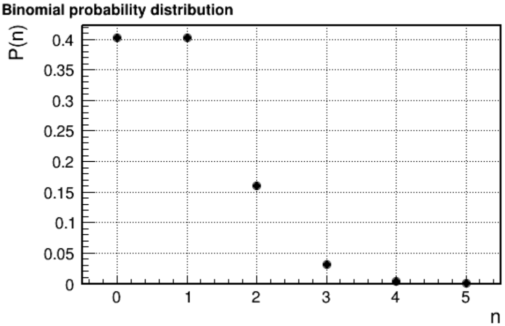
Example (1)

What is the probability to get three 'six' when rolling the die five times?

$$p = \frac{1}{6} \qquad n = 3 \qquad N = 5$$

By putting the numbers directly in the formula we get

$$P(n) \; = \; \frac{N!}{n!(N-n)!} \; p^n \; q^{N-n} \; = \; \frac{5!}{3! \; 2!} \; \frac{1}{6^3} \; \frac{5^2}{6^2} \; = \; \frac{120 \cdot 1 \cdot 25}{6 \cdot 2 \cdot 216 \cdot 36} \; \approx \; 0.032$$



Example (2)

Lecture room has 20 seats and 22 students enrolled for the course.
But students attend 90% of lectures only. Is the room large enough?

Example (2)

Lecture room has 20 seats and 22 students enrolled for the course.
But students attend 90% of lectures only. Is the room large enough?

Simple answer: $\bar{n} = pN = 0.9 \cdot 22 = 19.8 < 20 \Rightarrow$ should be OK...

Example (2)

Lecture room has 20 seats and 22 students enrolled for the course.
But students attend 90% of lectures only. Is the room large enough?

Simple answer: $\bar{n} = pN = 0.9 \cdot 22 = 19.8 < 20 \Rightarrow$ should be OK...

Probability that students will NOT fit in the room:

$$P^{ovfl} = P(22) + P(21) = 0.9^{22} + 22 \cdot 0.9^{21} \cdot 0.1 \approx 0.098 + 0.241 = 0.339$$

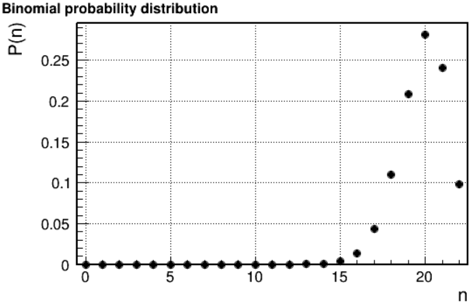
Example (2)

Lecture room has 20 seats and 22 students enrolled for the course.
But students attend 90% of lectures only. Is the room large enough?

Simple answer: $\bar{n} = pN = 0.9 \cdot 22 = 19.8 < 20 \Rightarrow$ should be OK...

Probability that students will NOT fit in the room:

$$P^{ovfl} = P(22) + P(21) = 0.9^{22} + 22 \cdot 0.9^{21} \cdot 0.1 \approx 0.098 + 0.241 = 0.339$$



Example (2)

Lecture room has 20 seats and 22 students enrolled for the course.
But students attend 90% of lectures only. Is the room large enough?

Simple answer: $\bar{n} = pN = 0.9 \cdot 22 = 19.8 < 20 \Rightarrow$ should be OK...

Probability that students will NOT fit in the room:

$$P^{ovfl} = P(22) + P(21) = 0.9^{22} + 22 \cdot 0.9^{21} \cdot 0.1 \approx 0.098 + 0.241 = 0.339$$

There is 66% chance that the room will be large enough for all students.

But this probability applies to single lecture only!
Probability that they will fit in the room for 14 lectures is

$$P^{OK} = (1 - P^{ovfl})^{14} \approx 0.003$$

$\Rightarrow$ we can hardly count on luck in this case, we need larger room !

Uniform probability distribution

Is often used as a model for a “complete randomness” of measurement result in given range. If variable x is restricted to interval $[a, b]$:

$$f(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{for } x > b \end{cases}$$

Uniform distribution

Uniform probability distribution

Is often used as a model for a “complete randomness” of measurement result in given range. If variable x is restricted to interval $[a, b]$:

$$f(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{for } x > b \end{cases}$$

Mean (expected value) of the uniform distribution

$$\bar{x} = \langle x \rangle = \frac{a+b}{2}$$

Uniform probability distribution

Is often used as a model for a "complete randomness" of measurement result in given range. If variable x is restricted to interval $[a, b]$:

$$f(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{for } x > b \end{cases}$$

Mean (expected value) of the uniform distribution

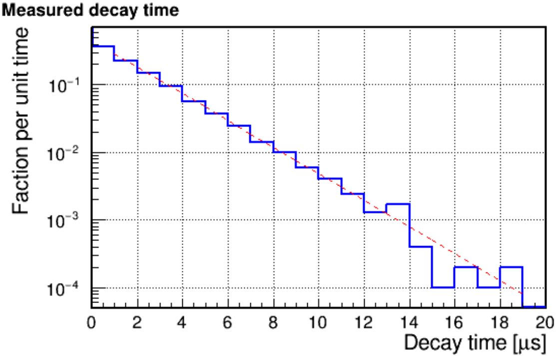
$$\bar{x} = \langle x \rangle = \frac{a+b}{2}$$

Variance of the uniform distribution

$$\mathbb{V}(x) = \sigma^2 = \frac{(b-a)^2}{12}$$

Exponential probability distribution

Describes the probability of waiting time t, when we wait for event A and the probability of A in a small time interval dt is constant: $dp = dt/\tau$.
This is the case for particle and nuclear decays, but also for other phenomena
τ is the only parameter in the problem.



Exponential probability distribution

Describes the probability of waiting time t, when we wait for event A and the probability of A in a small time interval dt is constant: $dp = dt/\tau$.
This is the case for particle and nuclear decays, but also for other phenomena
τ is the only parameter in the problem.

This problem is easily solved when we consider cumulative distribution:

$$F(t + dt) = F(t) + (1 - F(t)) \cdot \frac{dt}{\tau}$$

Exponential probability distribution

Describes the probability of waiting time t, when we wait for event A and the probability of A in a small time interval dt is constant: $dp = dt/\tau$.
This is the case for particle and nuclear decays, but also for other phenomena
τ is the only parameter in the problem.

This problem is easily solved when we consider cumulative distribution:

$$
\begin{aligned}
F(t+dt) &= F(t) + (1 - F(t)) \cdot \frac{dt}{\tau} \\
\frac{d}{dt}\, F(t) &= \frac{1}{\tau} \cdot (1 - F(t)) \\
\frac{d}{dt}\,(1 - F(t)) &= -\frac{1}{\tau} \cdot (1 - F(t))
\end{aligned}
$$

Exponential probability distribution

Describes the probability of waiting time t, when we wait for event A and the probability of A in a small time interval dt is constant: $dp = dt/\tau$.
This is the case for particle and nuclear decays, but also for other phenomena
τ is the only parameter in the problem.

This problem is easily solved when we consider cumulative distribution:

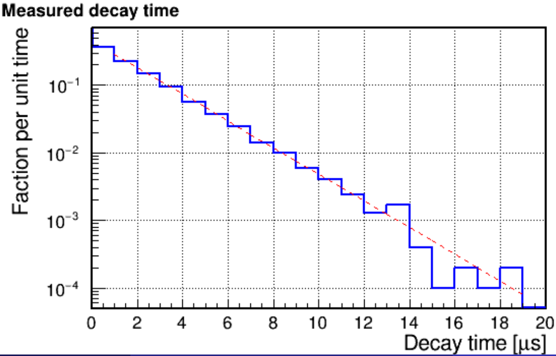
$$\begin{aligned}
F(t+dt) &= F(t) + (1-F(t)) \cdot \frac{dt}{\tau} \\
\frac{d}{dt} F(t) &= \frac{1}{\tau} \cdot (1-F(t)) \\
\frac{d}{dt} (1-F(t)) &= -\frac{1}{\tau} \cdot (1-F(t)) \\
(1-F(t)) &= C \cdot e^{-t/\tau} \\
F(t) &= 1 - e^{-t/\tau} \qquad C = 1 \ \text{from boundary conditions}
\end{aligned}$$

Exponential probability distribution

Resulting formula for the probability distribution is:

$$f(t) = \begin{cases} \frac{1}{\tau} \cdot e^{-t/\tau} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases}$$

indicated by the red dashed line:



Exponential probability distribution

Resulting formula for the probability distribution is:

$$f(t) = \begin{cases} \frac{1}{\tau} \cdot e^{-t/\tau} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases}$$

Mean (expected value) of the exponential distribution

$$\langle\, t \,\rangle = \int_0^{+\infty} dt\ t\ f(t) = \tau \quad \text{integrating by parts}$$

For nuclear/particle decays, parameter τ is the mean lifetime...

Exponential probability distribution

Resulting formula for the probability distribution is:

$$f(t) \; = \; \begin{cases} \frac{1}{\tau} \cdot e^{-t/\tau} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases}$$

Mean (expected value) of the exponential distribution

$$\langle \, t \, \rangle \; = \; \int_0^{+\infty} dt \; t \; f(t) \; = \; \tau \qquad \text{integrating by parts}$$

For nuclear/particle decays, parameter τ is the mean lifetime...

Variance of the exponential distribution

$$\mathbb{V}(t) \; = \; \sigma^2 \; = \; \int_0^{+\infty} dt \; (t - \tau)^2 \; f(t) \; = \; \tau^2$$

Example

What is the probability that the particle does not decay within 10 lifetimes?

Example

What is the probability that the particle does not decay within 10 lifetimes?

We can just look at the cumulative distribution:

$$\begin{aligned} F(t) &= 1 - e^{-t/\tau} \\ 1 - F(t) &= e^{-t/\tau} \\ 1 - F(10\tau) &= e^{-10} \approx 0.0000454 \end{aligned}$$

Probability is very small, but not negligible...

Example

What is the probability that the particle does not decay within 10 lifetimes?

We can just look at the cumulative distribution:

$$\begin{aligned} F(t) &= 1 - e^{-t/\tau} \\ 1 - F(t) &= e^{-t/\tau} \\ 1 - F(10\tau) &= e^{-10} \approx 0.0000454 \end{aligned}$$

Probability is very small, but not negligible...

Half-life

Frequently used in nuclear physics, nuclear medicine etc.
Defined as a time needed for half of the nuclei to decay.

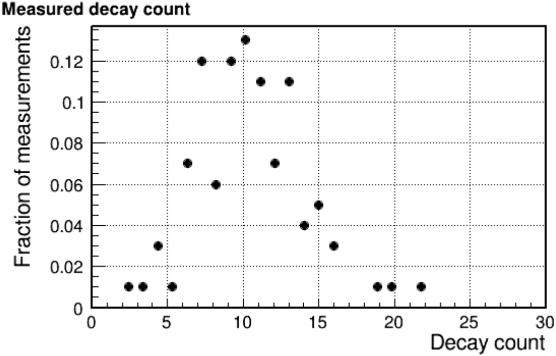
$$\begin{aligned} F(t_{1/2}) &= 0.5 \\ \Rightarrow \quad t_{1/2} &= \ln 2 \cdot \tau \end{aligned}$$

Expected number of decays

Consider radioactive source with 1 decay per second (1 Bq).

What is the expected result of the experiment counting radioactive decays in 10 s time window?

Example of decay count measurement: 100 measurements (100×10 s)

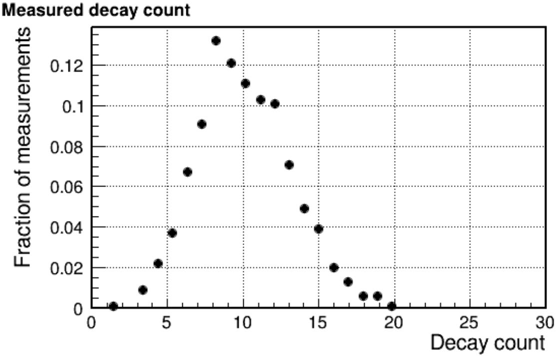


Expected number of decays

Consider radioactive source with 1 decay per second (1 Bq).

What is the expected result of the experiment counting radioactive decays in 10 s time window?

Example of decay count measurement: 1000 measurements

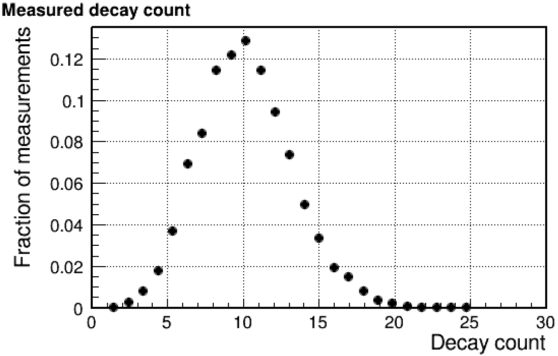


Expected number of decays

Consider radioactive source with 1 decay per second (1 Bq).

What is the expected result of the experiment counting radioactive decays in 10 s time window?

Example of decay count measurement: 10000 measurements

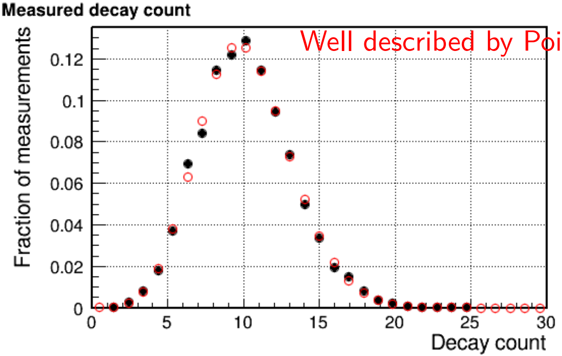


Expected number of decays

Consider radioactive source with 1 decay per second (1 Bq).

What is the expected result of the experiment counting radioactive decays in 10 s time window?

Example of decay count measurement: 10000 measurements



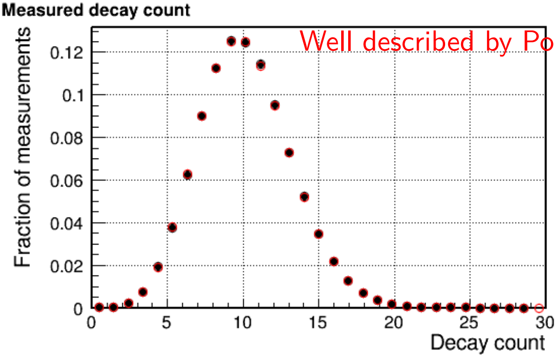
Well described by Poisson distribution

Expected number of decays

Consider radioactive source with 1 decay per second (1 Bq).

What is the expected result of the experiment counting radioactive decays in 10 s time window?

Example of decay count measurement: 1000000 measurements



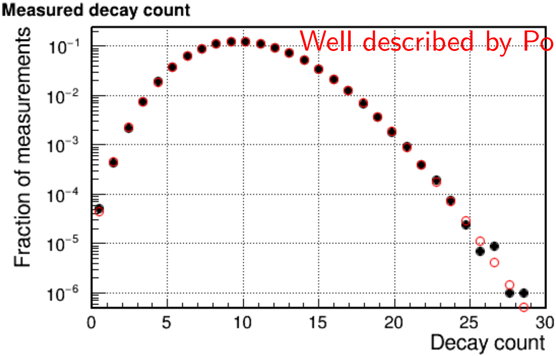
Well described by Poisson distribution

Expected number of decays

Consider radioactive source with 1 decay per second (1 Bq).

What is the expected result of the experiment counting radioactive decays in 10 s time window?

Example of decay count measurement: 1000000 measurements



Well described by Poisson distribution

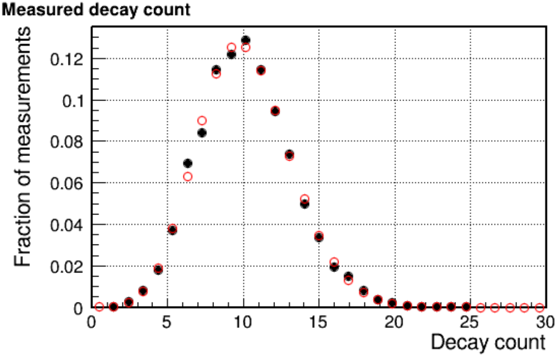
Poisson probability distribution

Formula for the Poisson probability distribution:

$$P(n) \quad = \quad \frac{\mu^n \; e^{-\mu}}{n!} \qquad \text{for } n = 0, 1, 2, \ldots$$

where μ, the expected number of events (mean), is the only parameter (!)

$P(n)$ indicated by the red circles:



Poisson probability distribution

Formula for the Poisson probability distribution:

$$P(n) \;=\; \frac{\mu^n \, e^{-\mu}}{n!} \qquad \text{for } n = 0, 1, 2, \ldots$$

where μ, the expected number of events (mean), is the only parameter (!)

Mean (expected) number of events

$$\bar{n} = \langle \, n \, \rangle \;\equiv\; \mu$$

Poisson probability distribution

Formula for the Poisson probability distribution:

$$P(n) \;=\; \frac{\mu^n \, e^{-\mu}}{n!} \qquad \text{for } n = 0, 1, 2, \ldots$$

where μ, the expected number of events (mean), is the only parameter (!)

Mean (expected) number of events

$$\bar{n} \;=\; \langle\, n \,\rangle \;\equiv\; \mu$$

Variance of the Poisson distribution

$$\mathbb{V}(n) \;=\; \langle\, (n-\mu)^2 \,\rangle \;=\; \sum_n (n-\mu)^2 \; P(n) \;=\; \mu$$

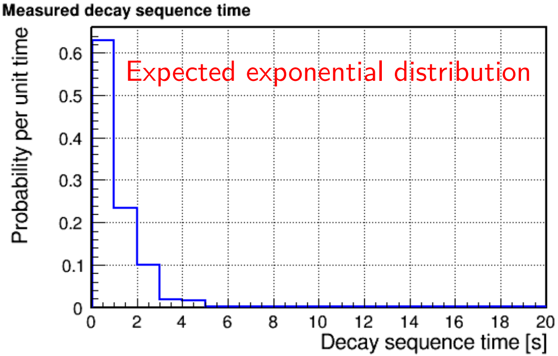
Often defines statistical uncertainty of the measurement...

Expected time of decay sequence

Consider radioactive source with 1 decay per second (1 Bq).

What is the expected result of the experiment measuring time needed for N decays to be observed?

Example of sequence measurements: 1 decay, 1000 measurements

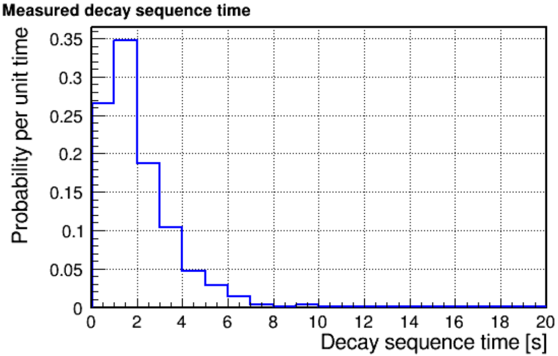


Expected time of decay sequence

Consider radioactive source with 1 decay per second (1 Bq).

What is the expected result of the experiment measuring time needed for N decays to be observed?

Example of sequence measurements: 2 decays, 1000 measurements

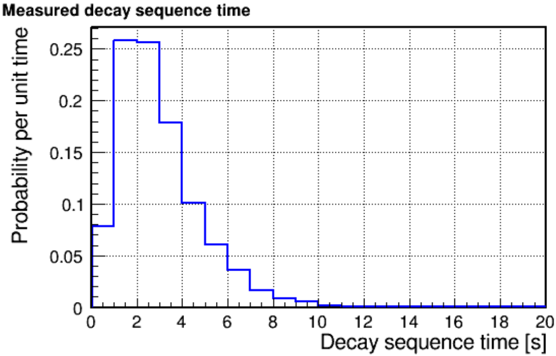


Expected time of decay sequence

Consider radioactive source with 1 decay per second (1 Bq).

What is the expected result of the experiment measuring time needed for N decays to be observed?

Example of sequence measurements: 3 decays, 1000 measurements

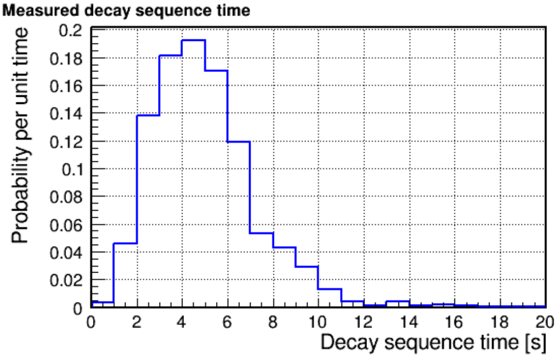


Expected time of decay sequence

Consider radioactive source with 1 decay per second (1 Bq).

What is the expected result of the experiment measuring time needed for N decays to be observed?

Example of sequence measurements: 5 decays, 1000 measurements

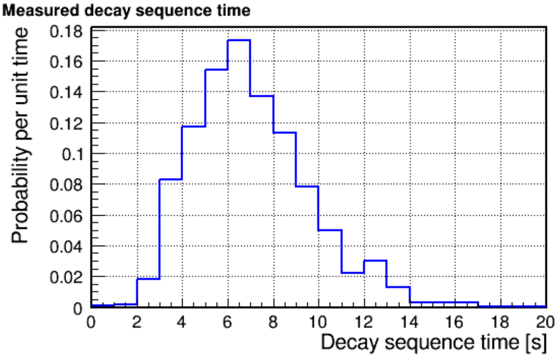


Expected time of decay sequence

Consider radioactive source with 1 decay per second (1 Bq).

What is the expected result of the experiment measuring time needed for N decays to be observed?

Example of sequence measurements: 7 decays, 1000 measurements

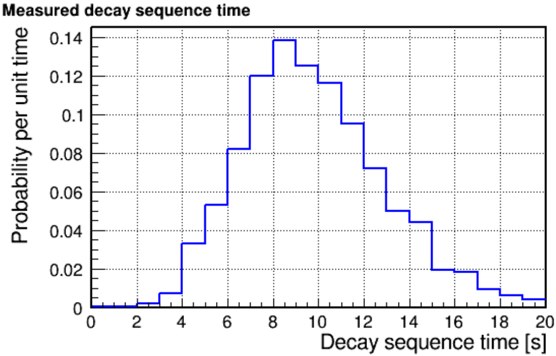


Expected time of decay sequence

Consider radioactive source with 1 decay per second (1 Bq).

What is the expected result of the experiment measuring time needed for N decays to be observed?

Example of sequence measurements: 10 decays, 1000 measurements

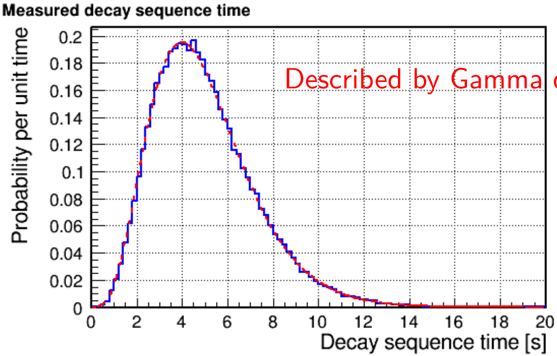


Expected time of decay sequence

Consider radioactive source with 1 decay per second (1 Bq).

What is the expected result of the experiment measuring time needed for N decays to be observed?

Example of sequence measurements: 5 decays, 100000 measurements



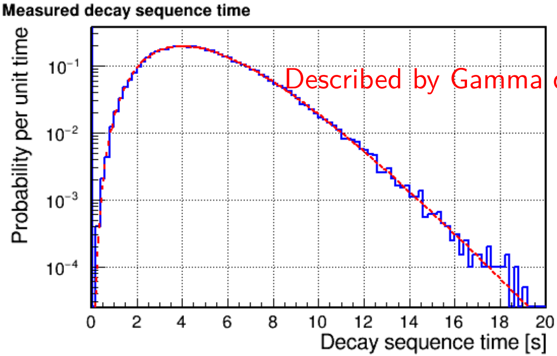
Described by Gamma distribution

Expected time of decay sequence

Consider radioactive source with 1 decay per second (1 Bq).

What is the expected result of the experiment measuring time needed for N decays to be observed?

Example of sequence measurements: 5 decays, 100000 measurements



Described by Gamma distribution

Gamma distribution

Decay sequence time distribution is described by Gamma distribution:

$$f(x) = \begin{cases} \frac{x^{k-1}\,\lambda^k\,e^{-\lambda x}}{\Gamma(k)} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

where $k \geq 0$ and $\lambda \geq 0$ are real parameters of the Gamma distribution.

For decay sequence of n decays: $k = n$ and $\beta = 1/\tau$

Gamma distribution

Decay sequence time distribution is described by Gamma distribution:

$$f(x) = \begin{cases} \frac{x^{k-1} \lambda^k e^{-\lambda x}}{\Gamma(k)} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

where $k \geq 0$ and $\lambda \geq 0$ are real parameters of the Gamma distribution.
For decay sequence of n decays: $k = n$ and $\beta = 1/\tau$

Mean (expected value) of the Gamma distribution

$$\bar{x} = \langle x \rangle = \frac{k}{\lambda}$$

Gamma distribution

Decay sequence time distribution is described by Gamma distribution:

$$f(x) = \begin{cases} \frac{x^{k-1}\,\lambda^k\,e^{-\lambda x}}{\Gamma(k)} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

where $k \geq 0$ and $\lambda \geq 0$ are real parameters of the Gamma distribution.
For decay sequence of n decays: $k = n$ and $\beta = 1/\tau$

Mean (expected value) of the Gamma distribution

$$\bar{x} = \langle x \rangle = \frac{k}{\lambda}$$

Variance of the Gamma distribution

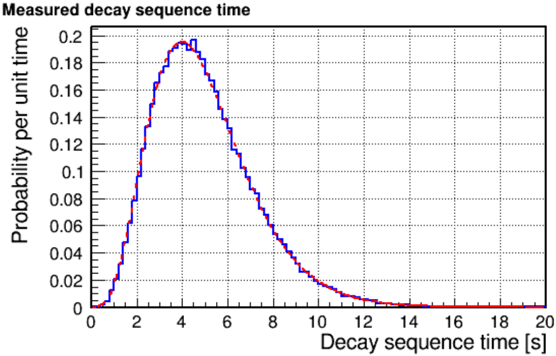
$$\mathbb{V}(x) = \sigma^2 = \frac{k}{\lambda^2}$$

Gamma distribution

For $k > 1$ ($n > 1$) distribution has a maximum at

$$x_0 = \frac{k-1}{\lambda}$$

5 decays, 100000 measurements

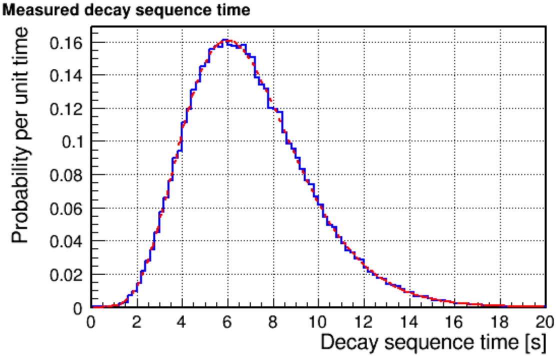


Gamma distribution

For $k > 1$ $(n > 1)$ distribution has a maximum at

$$x_0 = \frac{k-1}{\lambda}$$

7 decays, 100000 measurements

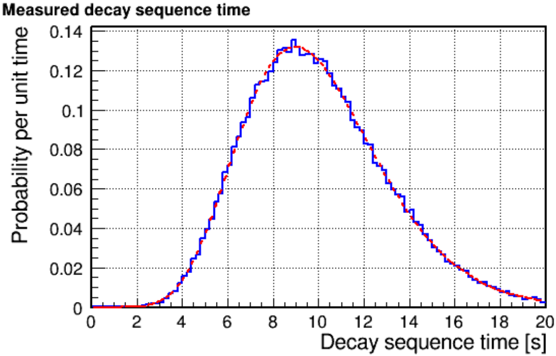


Gamma distribution

For $k > 1$ $(n > 1)$ distribution has a maximum at

$$x_0 = \frac{k-1}{\lambda}$$

10 decays, 100000 measurements



Gamma distribution

For $k > 1$ $(n > 1)$ distribution has a maximum at

$$x_0 = \frac{k-1}{\lambda}$$

Gamma distribution can also be written in equivalent form:

$$f(x) = A \cdot \exp\left[- \left(\frac{x_0}{\sigma_0} \right)^2 \left(\frac{x - x_0}{x_0} - \ln \frac{x}{x_0} \right) \right]$$

where A is normalization factor and σ_0 describes the width of the distribution around maximum:

$$\sigma_0^2 = \frac{k-1}{\lambda^2}$$

Gamma distribution

For $k > 1$ $(n > 1)$ distribution has a maximum at

$$x_0 = \frac{k-1}{\lambda} = \bar{x}\left(1 - \frac{\sigma^2}{\bar{x}^2}\right)$$

Gamma distribution can also be written in equivalent form:

$$f(x) = A \cdot \exp\left[-\left(\frac{x_0}{\sigma_0}\right)^2 \left(\frac{x - x_0}{x_0} - \ln\frac{x}{x_0}\right)\right]$$

where A is normalization factor and σ_0 describes the width of the distribution around maximum:

$$\sigma_0^2 = \frac{k-1}{\lambda^2} = \sigma^2\left(1 - \frac{\sigma^2}{\bar{x}^2}\right)$$

Gamma distribution

Gamma distribution is a "natural" choice for describing many physical processes and measurements.
Its properties are very similar to those of the Gaussian distribution with one additional advantage: negative results are excluded by definition.
One can think of the Gamma distribution as an analogue of the Gaussian distribution restricted to the non-negative results ($\mathbb{R}_+$)

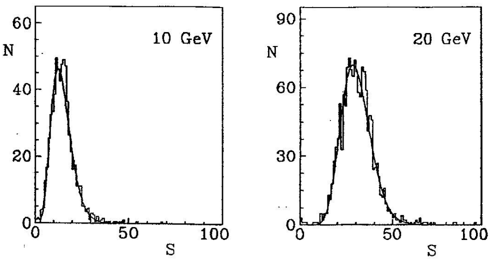
Gamma distribution

Gamma distribution is a "natural" choice for describing many physical processes and measurements.

Its properties are very similar to those of the Gaussian distribution with one additional advantage: negative results are excluded by definition.

One can think of the Gamma distribution as an analogue of the Gaussian distribution restricted to the non-negative results ($\mathbb{R}_+$)

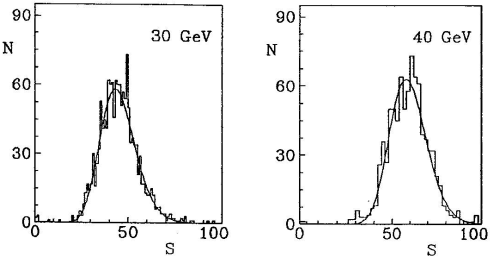
Sampling calorimeter response distribution



Gamma distribution

Gamma distribution is a “natural” choice for describing many physical processes and measurements.
Its properties are very similar to those of the Gaussian distribution with one additional advantage: negative results are excluded by definition.
One can think of the Gamma distribution as an analogue of the Gaussian distribution restricted to the non-negative results ($\mathbb{R}_+$)

Sampling calorimeter response distribution



Gaussian (Normal) distribution

Is most frequently used to describe fluctuations of the measurements and resulting measurement uncertainties

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

where μ and σ are two real parameters of the distribution describing

Mean (expected value) of the Gaussian distribution

$$\mathbb{E}(x) \;=\; \langle\, x \,\rangle \;=\; \mu$$

and **Variance** of the Gaussian distribution

$$\mathbb{V}(x) \;=\; \langle\, (x-\mu)^2 \,\rangle \;=\; \sigma^2$$

Combined distributions

Consider two independent random variables X and Y:

$$f(x,y) \;=\; f_x(x)\cdot f_y(y)$$

The sum of these variables, $z = x + y$, is also a random variable and

$$\bar{z} \,=\, \mathbb{E}(Z) \;=\; \mathbb{E}(X) + \mathbb{E}(Y) \,=\, \bar{x} + \bar{y}$$

Combined distributions

Consider two independent random variables X and Y:

$$f(x, y) = f_x(x) \cdot f_y(y)$$

The sum of these variables, $z = x + y$, is also a random variable and

$$\bar{z} = \mathbb{E}(Z) = \mathbb{E}(X) + \mathbb{E}(Y) = \bar{x} + \bar{y}$$

$$\sigma_z^2 = \mathbb{V}(Z) = \mathbb{V}(X) + \mathbb{V}(Y) = \sigma_x^2 + \sigma_y^2$$

Combined distributions

Consider two independent random variables X and Y:

$$f(x,y) \;=\; f_x(x)\cdot f_y(y)$$

The sum of these variables, $z = x + y$, is also a random variable and

$$\bar{z} \;=\; \mathbb{E}(Z) \;=\; \mathbb{E}(X) + \mathbb{E}(Y) \;=\; \bar{x} + \bar{y}$$

$$\sigma_z^2 \;=\; \mathbb{V}(Z) \;=\; \mathbb{V}(X) + \mathbb{V}(Y) \;=\; \sigma_x^2 + \sigma_y^2$$

If X and Y are described by Gaussian distribution function,
then Z is also described by Gaussian probability distribution !
This is widely used when eg. describing measurement uncertainties.

Combined distributions

Consider two independent random variables X and Y:

$$f(x,y) = f_x(x) \cdot f_y(y)$$

The sum of these variables, $z = x + y$, is also a random variable and

$$\bar{z} = \mathbb{E}(Z) = \mathbb{E}(X) + \mathbb{E}(Y) = \bar{x} + \bar{y}$$

$$\sigma_z^2 = \mathbb{V}(Z) = \mathbb{V}(X) + \mathbb{V}(Y) = \sigma_x^2 + \sigma_y^2$$

If X and Y are described by Gaussian distribution function,
then Z is also described by Gaussian probability distribution !
This is widely used when eg. describing measurement uncertainties.

However, this is also the case for the Gamma distribution !!!

Instead of homework

In the following lectures I will try to show different statistical and numerical methods from more practical side as well.

I will use the environment prepared last year for Modern Particle Physics Experiments course, based on Jupyter-lab and Python.

It is implemented as a docker container.

Instructions on how to install and run docker container on private computer were put in GitHub.

Please try!

Example Python scripts from this lecture will be put on the lecture page.