Statistical analysis of experimental data
Monte Carlo methods

Aleksander Filip Żarnecki

FACULTY OF PHYSICS
UNIVERSITY OF WARSAW

Lecture 05

October 31, 2024

Monte Carlo methods

Multiple variable case

Definition of covariance can be generalized to the set of N variables X_i:

$$c_{ij} \;=\; Cov(X_i, X_j) \;=\; \mathbb{E}((X_i - \mu_i)(X_j - \mu_j))$$

We can present it in a form of the covariance matrix:

$$\mathbb{C} \;=\; \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1N} \\ c_{21} & c_{22} & \dots & c_{2N} \\ \vdots & & & \vdots \\ c_{N1} & c_{N2} & \dots & c_{NN} \end{pmatrix}$$

where diagonal elements correspond to variances of the variables

$$c_{ii} \;=\; Cov(X_i, X_i) \;\equiv\; \mathbb{V}(X_i)$$

Functions of Random Variables

Change of Variables can be also considered in multi-dimensional case

$$\mathbf{y} = \mathbf{y}(\mathbf{x})$$

where components of vector $\mathbf{y}$ are given by functions $y_i(\mathbf{x})$, $i = 1, \ldots, N$

Probability density function for dependent variables $\mathbf{y}$ is given by

$$g(\mathbf{y}) = f(\mathbf{x}(\mathbf{y}))\,|J|$$

assuming that function $\mathbf{y}(\mathbf{x})$ is one-to-one and can be inverted, with

$$J = \left(\frac{\partial x_i}{\partial y_j}\right)$$

being the Jacobian of the variable transformation (square matrix)

General form

Covariance matrix for $\mathbf{y} = \mathbf{y}(\mathbf{x})$ can be approximate as:

$$Cov(Y_k, Y_l) = \sum_{i,j} \frac{\partial y_k}{\partial x_i} \frac{\partial y_l}{\partial x_j} Cov(X_i, X_j)$$

In matrix notation:

$$\mathbb{C}_{\mathbf{Y}} = A\, \mathbb{C}_{\mathbf{X}}\, A^{\mathsf{T}}$$

where A is a matrix of partial derivatives:

$$A_{i,j} = \left. \frac{\partial y_i}{\partial x_j} \right|_{\hat{\mu}_{\mathbf{x}}}$$

Direct Method

Let us assume we have a large number of uniformly distributed random numbers corresponding to the uniform probability distribution:

How to generate numbers from arbitrary probability distribution $f(x)$?

We need to consider cumulative distribution function for X, $F(x)$:

$$P(X \leq x) \quad = \quad F(x)$$

If cumulative distribution function $F(x)$ can be inverted, we can generate random numbers from $f(x)$ by using relation:

$$x \quad = \quad F^{-1}(r)$$

where r is uniformly distributed random number (from $u(r)$)

Direct Method example

Nice example is Cauchy distribution:

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} \quad \Rightarrow \quad F(x) = \frac{1}{\pi} \cdot \arctan(x) + \frac{1}{2}$$

which can be easily inverted, resulting in:

$$x \quad = \quad \tan\left(\pi\left(r - \frac{1}{2}\right)\right)$$

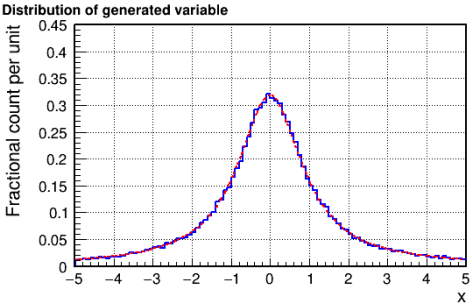
where r is uniformly distributed random number

Direct Method example

Nice example is Cauchy distribution:

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} \quad \Rightarrow \quad F(x) = \frac{1}{\pi} \cdot \arctan(x) + \frac{1}{2}$$

Example generation, N = 100000



von Neumann method

When cumulative distribution function can not be inverted we can define

$$f_{\max} = \max_x f(x) = N$$

Assume $f(x)$ is non-zero only for $a \leq x \leq b$.

We can then apply the following procedure:

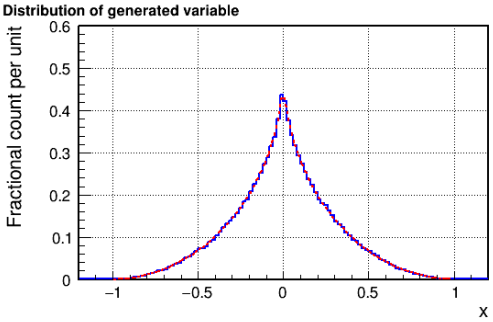
- generate value x uniformly distributed in $[a, b]$
- generate test variable r from uniform distribution ($[0, 1[$)
- accept generated value of x, if $r \cdot f_{\max} < f(x)$
  otherwise repeat from the beginning

This procedure is called von Neumann Acceptance–Rejection Technique

von Neumann method example

Acceptance–Rejection Technique applied to the example problem:

Test generation, N = 1000000



However, this procedure can not be directly applied if a or $b \to \pm\infty$

General method

For arbitrary probability distribution $f(x)$ (also with infinite domain).

We need to find "similar" distribution $g(x)$, such that its cumulative distribution exists and can be inverted, so we know how to generate random numbers from $g(x)$ distribution. We define

$$\max_x \frac{f(x)}{g(x)} = f_{max}$$

The following procedure can then be used:

- generate value x distributed according to $g(x)$
- generate test variable r from uniform distribution $([0, 1[)$
- accept generated value of x, if $r \cdot f_{max} < f(x)/g(x)$
  otherwise repeat from the beginning

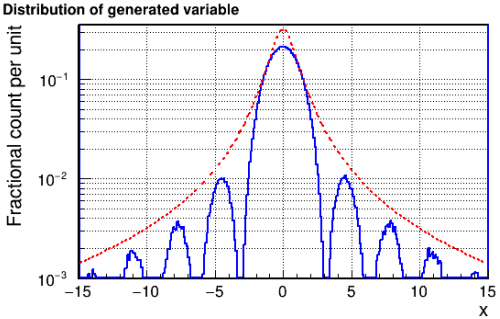
The closer $g(x)$ is to $f(x)$, the more efficient is the procedure...

General method example

Photon scattering angles for diffractive scattering

$$f(x) \quad = \quad N \cdot \frac{\sin^2 x}{x^2}$$

Test generation, N = 1000000



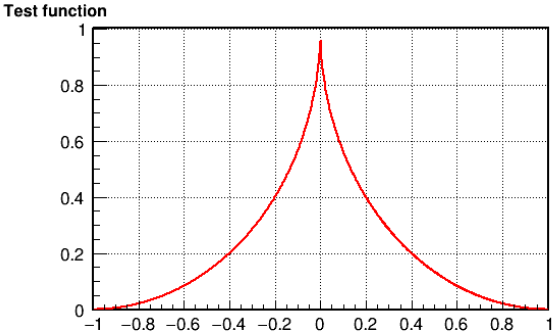
Monte Carlo methods

von Neumann method for random number generation

Let us considered test function:

$$f(x) = \begin{cases} 0 & \text{for } x < -1 \\ 1 - \sqrt{1 - (1 - |x|)^2} & \text{for } -1 \le x \\ 0 & \text{for } x > 1 \end{cases}$$

not normalized now
(contrary to the example in lecture 4)

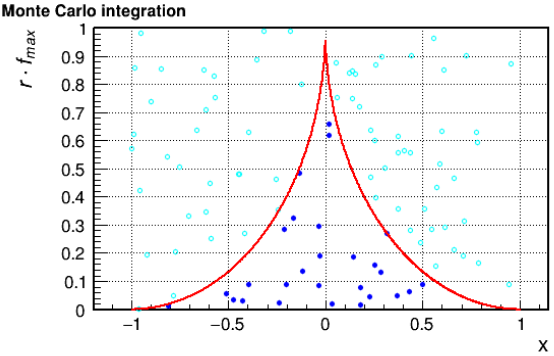


von Neumann method for random number generation 05_generation.ipynb Open in Colab

At each step, we generate pair of numbers $(x,\ r \cdot f_{max})$ from uniform distributions

$\Rightarrow$ random point inside $2 \times f_{max}$ rectangular

Test generation, N $= 100 \Rightarrow$ 27 points accepted

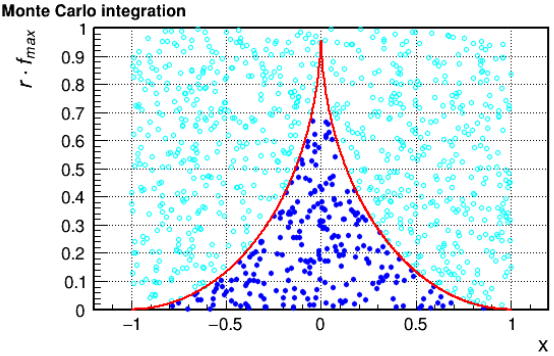


von Neumann method for random number generation 05_generation.ipynb Open in Colab

At each step, we generate pair of numbers $(x,\ r \cdot f_{max})$ from uniform distributions

$\Rightarrow$ random point inside $2 \times f_{max}$ rectangular

Test generation, N = 1000 $\Rightarrow$ 236 points accepted

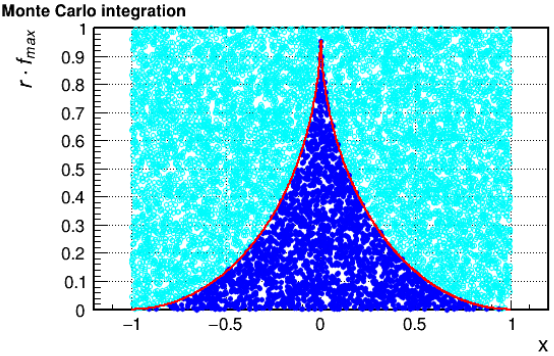


von Neumann method for random number generation 05_generation.ipynb Open in Colab

At each step, we generate pair of numbers $(x,\ r \cdot f_{max})$ from uniform distributions

$\Rightarrow$ random point inside $2 \times f_{max}$ rectangular

Test generation, N = 10000 $\Rightarrow$ 2146 points accepted

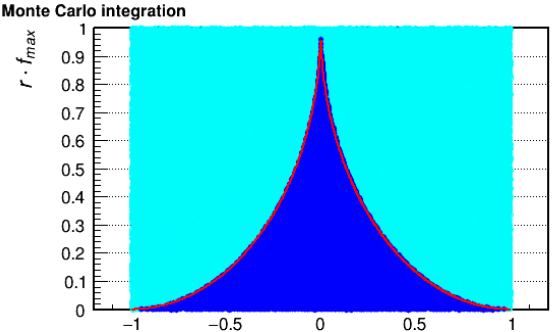


von Neumann method for random number generation 05_generation.ipynb Open in Colab

At each step, we generate pair of numbers $(x,\ r \cdot f_{max})$ from uniform distributions

$\Rightarrow$ random point inside $2 \times f_{max}$ rectangular

Test generation, N = 100000 $\Rightarrow$ 21388 points accepted



$\Rightarrow$ Fraction of accepted events is proportional to the to the integral of $f(x)$

General approach

Assuming number of events accepted in N tries is N_f, probability of accepting an event:

$$p = \lim_{N\to\infty} \frac{N_f}{N} = \frac{\int dx\ f(x)}{f_{max} \cdot \int dx\ g(x)} = \frac{\int dx\ f(x)}{I_0}$$

where I_0 is the total surface covered by generated $(x,\ r \cdot f_{max})$ pairs.

General approach

Assuming number of events accepted in N tries is N_f, probability of accepting an event:

$$p = \lim_{N\to\infty} \frac{N_f}{N} = \frac{\int dx\ f(x)}{f_{max} \cdot \int dx\ g(x)} = \frac{\int dx\ f(x)}{I_0}$$

where I_0 is the total surface covered by generated $(x,\ r \cdot f_{max})$ pairs.
We can thus write for integral of f:

$$I_f = \int dx\ f(x) = \frac{N_f}{N} \cdot f_{max} \cdot \int dx\ g(x) = p \cdot I_0$$

General approach

Assuming number of events accepted in N tries is N_f, probability of accepting an event:

$$p = \lim_{N\to\infty} \frac{N_f}{N} = \frac{\int dx\ f(x)}{f_{max} \cdot \int dx\ g(x)} = \frac{\int dx\ f(x)}{I_0}$$

where I_0 is the total surface covered by generated $(x,\ r \cdot f_{max})$ pairs.
We can thus write for integral of f:

$$I_f = \int dx\ f(x) = \frac{N_f}{N} \cdot f_{max} \cdot \int dx\ g(x) = p \cdot I_0$$

with uncertainty, which can be estimated from binomial distribution assuming $N \gg 1$

$$\sigma_I = \sqrt{\frac{p(1-p)}{N}} \cdot I_0$$

Example

Our test problem: uniform distribution was used for $g(x)$

$$f(x) = 1 - \sqrt{1 - (1 - |x|)^2} \quad \text{and} \quad g(x) = 0.5 \quad \text{for} \quad -1 \leq x \leq 1$$

Example

Our test problem: uniform distribution was used for $g(x)$

$$f(x) = 1 - \sqrt{1 - (1 - |x|)^2} \quad \text{and} \quad g(x) = 0.5 \quad \text{for} \quad -1 \leq x \leq 1$$

We find that:

$$\begin{aligned} f_{max} &= 2 \\ I_0 &= f_{max} \cdot \int dx\, g(x) = f_{max} = 2 \end{aligned}$$

Example

Our test problem: uniform distribution was used for $g(x)$

$$f(x) = 1 - \sqrt{1 - (1 - |x|)^2} \quad \text{and} \quad g(x) = 0.5 \quad \text{for} \quad -1 \leq x \leq 1$$

We find that:

$$\begin{aligned} f_{max} &= 2 \\ I_0 &= f_{max} \cdot \int dx\, g(x) = f_{max} = 2 \end{aligned}$$

From the last (largest) generation we have:

$$\begin{aligned} N = 100000 \quad N_f = 21388 \quad \Rightarrow p &= 0.21388 \\ I_f &= 0.42776 \\ \sigma_I &= 0.00259 \end{aligned}$$

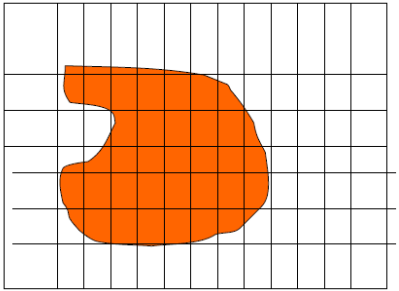
where expected values is $\langle I_f \rangle = 2 - \frac{\pi}{2} = 0.429203673...$

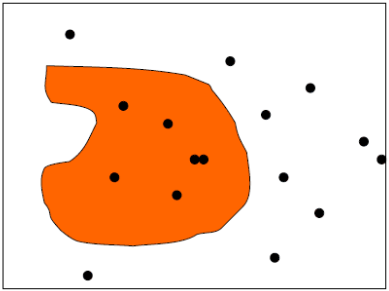
Applications

Described procedure can be used not only to calculate integrals of one-dimensional functions, it is much more general... It can be easily extended to multiple dimensions...

Applications

Described procedure can be used not only to calculate integrals of one-dimensional functions, it is much more general... How to calculate volume of a given shape?





Standard procedure:
scan all dimensions using dense point grid and sum cells with centers inside the volume

Monte Carlo integration:
Generate random points in the considered space and count points inside the volume

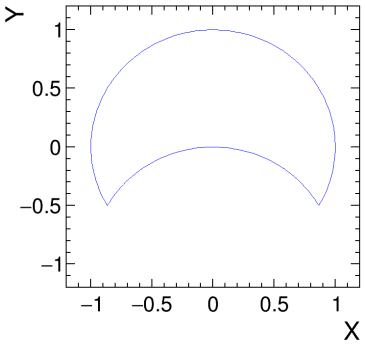
2-D example

05_generation_2.ipynb Open in Colab

Consider surface calculation for the partially eclipsed sun.

Assume $R_S = R_M = 1$ and distance between centers $\Delta = 1$

What is the surface of this shape?

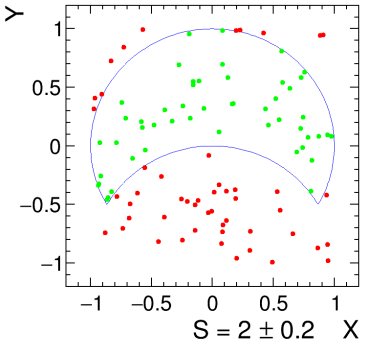


2-D example

05_generation_2.ipynb Open in Colab

Consider surface calculation for the partially eclipsed sun.

Generation results for $N = 100$



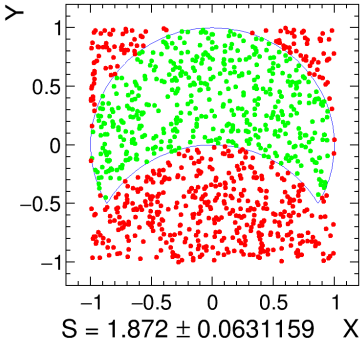
2-D example

05_generation_2.ipynb

Open in Colab

Consider surface calculation for the partially eclipsed sun.

Generation results for $N = 1000$

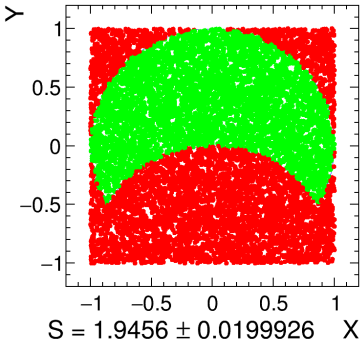


2-D example

05_generation_2.ipynb

Consider surface calculation for the partially eclipsed sun.

Generation results for N = 10'000



2-D example

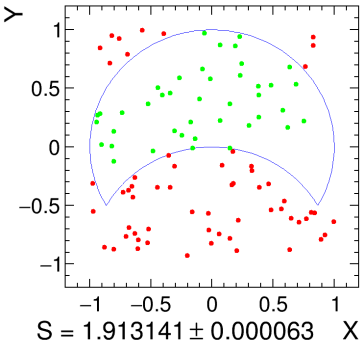
05_generation_2.ipynb Open in Colab

Consider surface calculation for the partially eclipsed sun.

Generation results for N = 1'000'000'000

only 100 points shown

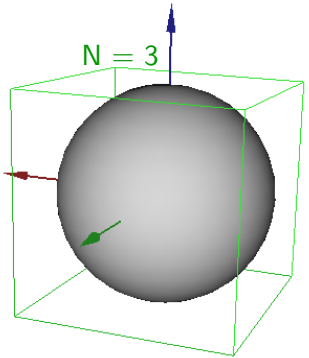
Exact calculation: $S = \frac{\pi}{3} + \frac{\sqrt{3}}{2} \approx 1.91322296$



Example (3)

Simplest possible case: calculate volume of a sphere in N dimensions
Unit sphere volume can be defined as:

$$V^N = \{\mathbf{x} \in \mathbb{R}^N : |\mathbf{x}| \leq 1\}$$

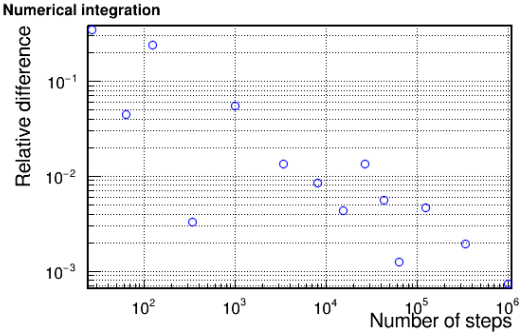


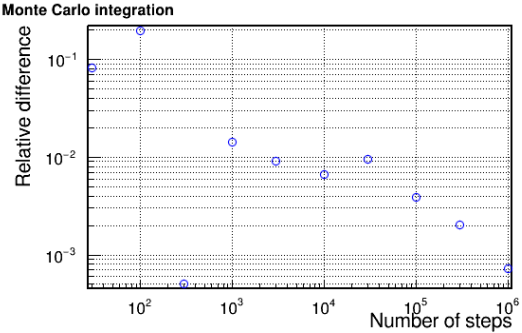
Example (3)

05_numerical_integration.ipynb 05_mc_integration.ipynb

Compare two approaches: standard variable scan and MC integration

Precision of the result $\frac{V_{num}-V_{true}}{V_{true}}$ for $N = 3$, as a function of the total number of steps





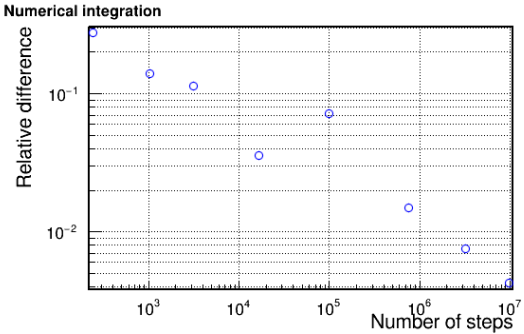
Both methods give comparable precision...

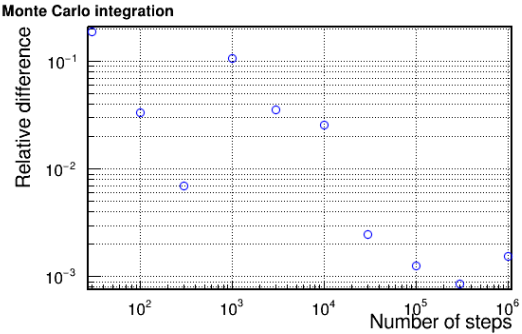
Example (3)

05_numerical_integration_2.ipynb 05_mc_integration_2.ipynb

Compare two approaches: standard variable scan and MC integration

Precision of the result $\frac{V_{num}-V_{true}}{V_{true}}$ for $N = 5$, as a function of the total number of steps

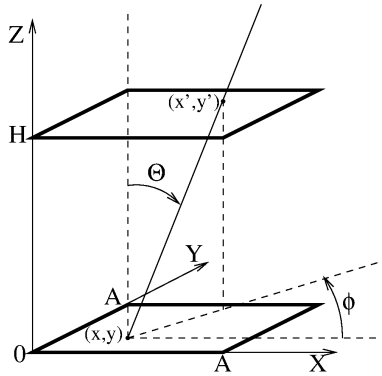




MC method converges much faster... (note different step number scale)

Example (4)

In one of the exercises at the Physics Laboratory (P8): total flux of the secondary cosmic rays. To extract the flux, we need the effective solid angle corresponding to the detector acceptance.



Assume the flux is measured by a coincidence of two parallel scintillating detector planes, separated by distance H.

Square detector ($A \times A$) are placed directly one above the other and have negligible thickness.

We need to know what is the average solid angle "observed" be a small element of the lower counter?

Example (4)

Solid angle "visible" from the particular point (x, y) on the lower detector surface can be defined by the integral:

$$d\Omega(x,y) = \int_0^1 d(cos\Theta) \int_0^{2\pi} d\phi \cdot F(x', y')$$

where Θ and ϕ are polar coordinates defining the particle direction and (x', y') are coordinates of the particle in the upper detector plane:

$$\begin{aligned} x' &= x + H \cdot \tan(\theta) \cdot \cos(\phi) \\ y' &= y + H \cdot \tan(\theta) \cdot \sin(\phi) \end{aligned}$$

Function $F(x', y')$ defines the condition that the particle crosses the upper counter:

$$\begin{aligned} F(x', y') &= 1 \quad \text{for } 0 < x' < A \text{ and } 0 < y' < A \\ &= 0 \quad \text{otherwise} \end{aligned}$$

Example (4)

To obtain the effective solid angle we need to average $d\Omega(x, y)$ over the lower surface:

$$\begin{aligned} \Omega &= \frac{1}{A^2} \int_0^A dx \int_0^A dy \cdot d\Omega(x, y) \\ &= \frac{1}{A^2} \int_0^A dx \int_0^A dy \int_0^1 d(cos\Theta) \int_0^{2\pi} d\phi \cdot F(x', y') \end{aligned}$$

Even in this relatively simple problem, we get a multidimensional integral which is very difficult to calculate analytically

On the other hand, it can be very easily integrated with MC approach

Uniform distribution on the sphere corresponds to uniform distribution in $\cos\Theta$ and ϕ.

Example (4)

Integration procedure:

- generate random point (x, y) on the lower detector surface
  uniform distributions for $x \in [0, A[$ and $y \in [0, A[$

Example (4)

Integration procedure:

- generate random point (x, y) on the lower detector surface
  uniform distributions for $x \in [0, A[$ and $y \in [0, A[$
- generate random direction in space
  uniform distribution for $\cos\Theta \in [0, 1[$ and $\phi \in [0, 2\pi[$

Example (4)

Integration procedure:

- generate random point (x, y) on the lower detector surface
  uniform distributions for $x \in [0, A[$ and $y \in [0, A[$
- generate random direction in space
  uniform distribution for $\cos\Theta \in [0, 1[$ and $\phi \in [0, 2\pi[$
- extrapolate particle track to the upper detector plane
  calculate coordinates (x', y') in the upper plane

Example (4)

Integration procedure:

- generate random point (x, y) on the lower detector surface
  uniform distributions for $x \in [0, A[$ and $y \in [0, A[$
- generate random direction in space
  uniform distribution for $\cos\Theta \in [0, 1[$ and $\phi \in [0, 2\pi[$
- extrapolate particle track to the upper detector plane
  calculate coordinates (x', y') in the upper plane
- count particles which pass the active detector surface (N_p)

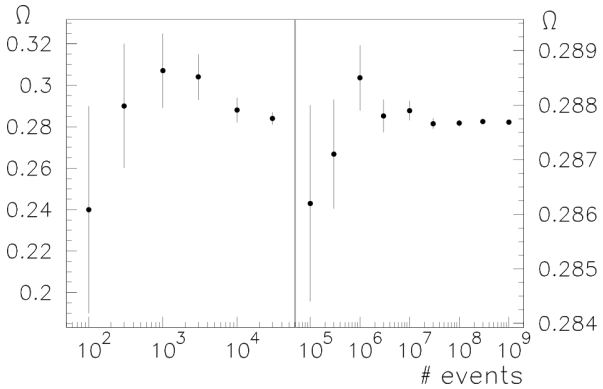
The result of the integration will be given by the fraction of the events passing the second detector, multiplied by the solid angle corresponding to the whole hemisphere (2π)

$$\Omega \;=\; 2\pi \cdot \frac{N_p}{N}$$

where N is the total number of generated events

Example (4)

Example integration results for $A = 15\,\text{cm}$ and $H = 26\,\text{cm}$



We can easily get precision much higher than precision of input parameters

General case

Examples presented considered the special case: input random variables had uniform distribution and "test function" was binary (returning 0 or 1).

In the general case we want to determine an expectation value of a function $h(\mathbf{x})$ of random variable vector $\mathbf{x}$ described by $f(\mathbf{x})$ pdf:

$$\mu_h \equiv \mathbb{E}_f[h(\mathbf{x})] = \int d\mathbf{x}\, h(\mathbf{x})\, f(\mathbf{x})$$

General case

Examples presented considered the special case: input random variables had uniform distribution and "test function" was binary (returning 0 or 1).

In the general case we want to determine an expectation value of a function $h(\mathbf{x})$ of random variable vector $\mathbf{x}$ described by $f(\mathbf{x})$ pdf:

$$\mu_h \equiv \mathbb{E}_f[h(\mathbf{x})] = \int d\mathbf{x}\, h(\mathbf{x})\, f(\mathbf{x})$$

Monte Carlo determination of μ_h assumes we can generate random variables from $f(\mathbf{x})$. We can then calculate:

$$\mu_{MC} = \lim_{N\to\infty} \frac{1}{N} \sum_i h(\mathbf{x}_i)$$

where $\mathbf{x}_i$, $i = 1, \ldots, N$ are random (input) variables generated from $f(\mathbf{x})$

Importance sampling

When $h(\mathbf{x})$ varies strongly in the considered variable range, statistical precision on the mean can be poor. Can it be improved?

Possible solution is to generate $\mathbf{x}$ using probability density more "focused" on the areas where $h(\mathbf{x})$ is large. Optimal choice turns out to be

$$g(\mathbf{x}) \quad \sim \quad h(\mathbf{x})\, f(\mathbf{x})$$

but approximate descriptions also work well.
When generating input variables from $g(\mathbf{x})$, the mean value of $h(\mathbf{x})$ can be now calculated as:

$$\mu_{IS} \quad = \quad \frac{1}{N} \sum_i h(\mathbf{x}_i) \cdot \frac{f(\mathbf{x}_i)}{g(\mathbf{x}_i)}$$

where the second term corrects for the modified pdf.

Weighted Monte Carlo

General method for generating random points in multi-dimensional space using acceptance–rejection technique can have very low efficiency, if probability distribution function $f(\mathbf{x})$ varies a lot, eg. has sharp peaks.

Weighted Monte Carlo

General method for generating random points in multi-dimensional space using acceptance–rejection technique can have very low efficiency, if probability distribution function $f(\mathbf{x})$ varies a lot, eg. has sharp peaks.

Assume we know how to generate random numbers from $g(x)$.
We can then apply the following procedure:

- generate $\mathbf{x}_i$ distributed according to $g(\mathbf{x})$
- accept all generated value $\mathbf{x}_i$,
  but consider them with additional weight: $w_i = f(\mathbf{x})/g(\mathbf{x})$

Weighted Monte Carlo

General method for generating random points in multi-dimensional space using acceptance–rejection technique can have very low efficiency, if probability distribution function $f(\mathbf{x})$ varies a lot, eg. has sharp peaks.

Assume we know how to generate random numbers from $g(x)$.
We can then apply the following procedure:

- generate $\mathbf{x}_i$ distributed according to $g(\mathbf{x})$
- accept all generated value $\mathbf{x}_i$,
  but consider them with additional weight: $w_i = f(\mathbf{x})/g(\mathbf{x})$

For example, when calculating the expectation value of $h(\mathbf{x})$:

$$\mu_{MC} \rightarrow \mu_{wMC} = \frac{\sum_i w_i\, h(\mathbf{x}_i)}{\sum_i w_i}$$

"unweighted" samples considered previously correspond to $w_i \equiv 1$

Weighted Monte Carlo

When using weighted Monte Carlo "events", number of events has to be replaced by sum of weights:

$$N \rightarrow N_w = \sum_i w_i$$

Weighted Monte Carlo

When using weighted Monte Carlo "events", number of events has to be replaced by sum of weights:

$$N \quad \rightarrow \quad N_w \; = \; \sum_i w_i$$

Variance of the sum of weights:

$$\mathbb{V}(N_w) \quad = \quad \sum_i w_i^2$$

Weighted Monte Carlo

When using weighted Monte Carlo "events", number of events has to be replaced by sum of weights:

$$N \quad \rightarrow \quad N_w \; = \; \sum_i w_i$$

Variance of the sum of weights:

$$\mathbb{V}(N_w) \quad = \quad \sum_i w_i^2$$

Statistical power of the weighted Monte Carlo sample is equivalent to unweighted sample of:

$$N_{eq} \quad = \quad \frac{N_w^2}{\mathbb{V}(N_w)} \quad = \quad \frac{(\sum_i w_i)^2}{\sum_i w_i^2}$$

Same relative uncertainty, $\frac{\sigma_{N_w}}{N_w}$, as for N_{eq} events measured from Poisson distribution ($\mathbb{V}(N_{eq}) = N_{eq}$)

General remarks

Monte Carlo techniques are widely used not only for integration but for modeling of multi-dimensional random variable distributions in general.

Simulation of particle collision events or particle interactions with matter are just special cases, widely used in particle physics. It allows to predict the experimental result with high precision.

General remarks

Monte Carlo techniques are widely used not only for integration but for modeling of multi-dimensional random variable distributions in general.

Simulation of particle collision events or particle interactions with matter are just special cases, widely used in particle physics. It allows to predict the experimental result with high precision.

While Monte Carlo methods allow us to perform very complicated computations in an efficient way, **we still need to know all details.**

It is not a "magic box" - if we do not know any of the input distributions or parameters, we can not perform the integration or simulation needed.

Simplifying assumptions are made sometimes to cover our lack of knowledge (eg. assuming uniform distribution for particle decays) but this has to be clearly stated, as can result in systematic bias of results.

Monte Carlo methods

Weighted mean

If we perform N experiments resulting in set of measurements x_i, $i = 1 \ldots N$, then the sample mean (see lecture 4):

$$\bar{x} = \frac{1}{N} \sum_i x_i$$

gives us an unbiased estimator of the true mean μ for random variable X.

However, is it the optimal estimate? (with smallest variance)

Weighted mean

If we perform N experiments resulting in set of measurements x_i, $i = 1 \ldots N$, then the sample mean (see lecture 4):

$$\bar{x} = \frac{1}{N} \sum_i x_i$$

gives us an unbiased estimator of the true mean μ for random variable X.

However, is it the optimal estimate? (with smallest variance)

This is the case, if the same experiment is repeated many times,
so the probability distribution function for all input x_i is the same.

If input measurements have different pdf's (different variances),
"simple mean" is not the best choice, we can do better...

Weighted mean example

Consider measurements of the mean lifetime of X particle performed by N groups. Particle is difficult to produce so the measurement precision is dominated by statistical fluctuations in the measured decay times.

How to combine results τ_i presented by different groups?

Weighted mean example

Consider measurements of the mean lifetime of X particle performed by N groups. Particle is difficult to produce so the measurement precision is dominated by statistical fluctuations in the measured decay times.

How to combine results τ_i presented by different groups?

The best procedure would be to average individual decay times measured by all groups. This is equivalent to:

$$\bar{\tau} = \frac{1}{N} \sum_i N_i \, \tau_i$$

where N_i is the number of decays measured by group i and N is the total number of measured decays

$$N = \sum_i N_i$$

Weighted mean example

What can we do, if groups do not present event numbers, only the mean lifetime uncertainty?

We can use properties of the exponential distribution

$$\mathbb{V}(t) \;=\; \tau^2$$

Weighted mean example

What can we do, if groups do not present event numbers, only the mean lifetime uncertainty?

We can use properties of the exponential distribution

$$\sigma^2 \;=\; \mathbb{V}(\,\langle t \rangle\,) \;\;=\;\; \tau^2/N$$

Weighted mean example

What can we do, if groups do not present event numbers, only the mean lifetime uncertainty?

We can use properties of the exponential distribution

$$\sigma^2 \;=\; \mathbb{V}(\,\langle t \rangle\,) \;=\; \tau^2/N$$

and estimating the number of decays measured, $N_i = \tau^2/\sigma_i^2$, we get:

$$\bar{\tau} \;=\; \sigma^2 \sum_i \frac{\tau_i}{\sigma_i^2}$$

where σ^2 is the expected variance of the weighted mean:

$$\frac{1}{\sigma^2} \;=\; \sum_i \frac{1}{\sigma_i^2}$$

Weighted mean

We can also obtain this formula from minimum variance requirement.
General expression for mean:

$$\bar{x} \;=\; \sum_i a_i\, x_i \;, \qquad \text{where } \sum_i a_i = 1 \;.$$

Weighted mean

We can also obtain this formula from minimum variance requirement.
General expression for mean:

$$\bar{x} \;=\; \sum_i a_i\, x_i \;, \qquad \text{where } \sum_i a_i = 1 \;.$$

Variance of the linear combination (lecture 4)

$$\mathbb{V}(\bar{x}) \;=\; \sum_i a_i^2\, \sigma_i^2$$

Weighted mean

We can also obtain this formula from minimum variance requirement.
General expression for mean:

$$\bar{x} \;=\; \sum_i a_i\, x_i \;, \qquad \text{where } \sum_i a_i = 1 \;.$$

Variance of the linear combination (lecture 4) λ - Lagrange multiplier

$$\mathbb{V}(\bar{x}) \;=\; \sum_i a_i^2\, \sigma_i^2 - 2\,\lambda\,(\sum_i a_i - 1)$$

For minimum variance, partial derivatives should be zero:

$$\frac{\partial \mathbb{V}}{\partial a_i} \;=\; 2\, a_i\, \sigma_i^2 \;-\; 2\,\lambda \;=0 \;\;\Rightarrow\;\; a_i \;=\; \frac{\lambda}{\sigma_i^2}$$

Weighted mean

What about averaging measurements which are not independent?

$\bar{x} = \mathbf{a}^\intercal \mathbf{x}$

In the most general case, variance of the weighted mean is given by

$\mathbb{I}$ - vector of ones

$$\sigma^2_{\bar{x}} = \mathbf{a}^\intercal \mathbb{C}_\mathbf{x} \mathbf{a}$$

Weighted mean

What about averaging measurements which are not independent?

$\bar{x} = \mathbf{a}^\intercal \mathbf{x}$

In the most general case, variance of the weighted mean is given by

$\mathbb{I}$ - vector of ones

$$\sigma^2_{\bar{x}} \;=\; \mathbf{a}^\intercal \, \mathbb{C}_{\mathbf{x}} \, \mathbf{a} \;-\; 2\,\lambda\,(\mathbf{a}^\intercal \mathbb{I} - 1)$$

Weighted mean

What about averaging measurements which are not independent?

$\bar{x} = \mathbf{a}^\intercal \mathbf{x}$

In the most general case, variance of the weighted mean is given by

$\mathbb{I}$ - vector of ones

$$\sigma^2_{\bar{x}} \;=\; \mathbf{a}^\intercal \, \mathbb{C}_\mathbf{x} \, \mathbf{a} \;-\; 2\,\lambda\,(\mathbf{a}^\intercal \mathbb{I} - 1)$$

Minimizing mean variance we compare partial derivatives to zero and get

$$\mathbb{C}_\mathbf{x} \, \mathbf{a} \;=\; \lambda \cdot \mathbb{I}$$

where λ can be constrained from the boundary condition $\mathbf{a}^\intercal \mathbb{I} = 1$.

This is a linear set of equations, which can be solved:

$$\mathbf{a} = \lambda \, \mathbb{C}_\mathbf{x}^{-1} \, \mathbb{I}$$

Weighted mean

What about averaging measurements which are not independent?

$\bar{x} = \mathbf{a}^\intercal \mathbf{x}$

In the most general case, variance of the weighted mean is given by

$\mathbb{I}$ - vector of ones

$$\sigma^2_{\bar{x}} \;=\; \mathbf{a}^\intercal \, \mathbb{C}_\mathbf{x} \, \mathbf{a} \;-\; 2\,\lambda\,(\mathbf{a}^\intercal \mathbb{I} - 1)$$

Minimizing mean variance we compare partial derivatives to zero and get

$$\mathbb{C}_\mathbf{x} \, \mathbf{a} \;=\; \lambda \cdot \mathbb{I}$$

where λ can be constrained from the boundary condition $\mathbf{a}^\intercal \mathbb{I} = 1$.

This is a linear set of equations, which can be solved:

$$\mathbf{a} = \frac{\mathbb{C}_\mathbf{x}^{-1}\,\mathbb{I}}{\mathbb{I}^\intercal \mathbb{C}_\mathbf{x}^{-1}\,\mathbb{I}}$$

Weighted mean example

Let four groups reported average number of Gamma Ray Bursts per year, per unit solid angle.
For simplicity, let us assume each group observes one unit of solid angle.
Number of bursts observed by group i in year j is denoted as n_{ij}. Total number is N_i.

First group wanted to be first, published their result already after one year. They quoted:

$$\mu_1 \;=\; N_1 \pm \sqrt{N_1} \;=\; n_{11} \pm \sqrt{n_{11}}$$

Weighted mean example

Let four groups reported average number of Gamma Ray Bursts per year, per unit solid angle.
For simplicity, let us assume each group observes one unit of solid angle.
Number of bursts observed by group i in year j is denoted as n_{ij}. Total number is N_i.

First group wanted to be first, published their result already after one year. They quoted:

$$\mu_1 \;=\; N_1 \pm \sqrt{N_1} \;=\; n_{11} \pm \sqrt{n_{11}}$$

Second group published results after two years of data taking:

$$\mu_2 \;=\; \frac{1}{2} N_2 \pm \frac{1}{2}\sqrt{N_2} \;=\; \frac{1}{2}(n_{21}+n_{22}) \pm \frac{1}{2}\sqrt{n_{21}+n_{22}}$$

Weighted mean example

Let four groups reported average number of Gamma Ray Bursts per year, per unit solid angle.
For simplicity, let us assume each group observes one unit of solid angle.
Number of bursts observed by group i in year j is denoted as n_{ij}. Total number is N_i.

First group wanted to be first, published their result already after one year. They quoted:

$$\mu_1 \;=\; N_1 \pm \sqrt{N_1} \;=\; n_{11} \pm \sqrt{n_{11}}$$

Second group published results after two years of data taking:

$$\mu_2 \;=\; \frac{1}{2} N_2 \pm \frac{1}{2}\sqrt{N_2} \;=\; \frac{1}{2}(n_{21}+n_{22}) \pm \frac{1}{2}\sqrt{n_{21}+n_{22}}$$

Third and fourth group published results covering three and four years.

$$\mu_3 \;=\; \frac{1}{3}(n_{31}+n_{32}+n_{33}) \qquad \mu_4 \;=\; \frac{1}{4}(n_{41}+n_{42}+n_{43}+n_{44})$$

How should we average these measurements ?!

Weighted mean example

The proper procedure to combine these measurements strongly depend on the actual way the observations were done. In this particular case, the key questions is:

Did they observe the same region of the sky at the same time?
(so they did observe and count the same GRBs)

How much did their acceptance (in time or solid angle coverage) overelap?

Weighted mean example

The proper procedure to combine these measurements strongly depend on the actual way the observations were done. In this particular case, the key questions is:

Did they observe the same region of the sky at the same time?
(so they did observe and count the same GRBs)

How much did their acceptance (in time or solid angle coverage) overelap?

Let us consider two extreme cases:

- four experiments observing different sky regions, unrelated observations
- four experiments observing exactly the same sky region, same bursts

We do assume that the GRB rate is constant in time.

Weighted mean example Case 1 - observing different parts of the sky

We assume that N_1, N_2, N_3 and N_4 are independent random numbers.
The covariance matrix for the variables N_i is thus diagonal:

$$\mathbb{C}_{\mathbf{N}} = \begin{pmatrix} N_1 & & & \\ & N_2 & & \\ & & N_3 & \\ & & & N_4 \end{pmatrix}$$

Weighted mean example Case 1 - observing different parts of the sky

We assume that N_1, N_2, N_3 and N_4 are independent random numbers.
The covariance matrix for the variables N_i is thus diagonal:

$$\mathbb{C}_{\mathbf{N}} = \begin{pmatrix} N_1 & & & \\ & N_2 & & \\ & & N_3 & \\ & & & N_4 \end{pmatrix}$$

We can use it to calculate covariance matrix for set of measurements μ_i, using the partial derivatives:

$$\frac{\partial \mu_k}{\partial n_j} = \begin{cases} \frac{1}{k} & \text{for} \quad k = j \\ 0 & \quad k \neq j \end{cases}$$

see lecture 4

Weighted mean example Case 1 - observing different parts of the sky

The covariance matrix for the set of measurements μ_i is:

$$\mathbb{C}_\mu = \begin{pmatrix} N_1 & & & \\ & \frac{1}{4}N_2 & & \\ & & \frac{1}{9}N_3 & \\ & & & \frac{1}{16}N_4 \end{pmatrix}$$

Weighted mean example Case 1 - observing different parts of the sky

The covariance matrix for the set of measurements μ_i is:

$$\mathbb{C}_\mu = \begin{pmatrix} N_1 & & & \\ & \frac{1}{4}N_2 & & \\ & & \frac{1}{9}N_3 & \\ & & & \frac{1}{16}N_4 \end{pmatrix}$$

By solving the set of equations we get:

$$\mathbf{a}^\mathsf{T} = \lambda \left(\frac{1}{N_1}, \frac{4}{N_2}, \frac{9}{N_3}, \frac{16}{N_4} \right) \approx \frac{\lambda}{\mu}(1,\ 2,\ 3,\ 4)$$

Where we assumed all measurements are consistent expected results $\langle N_i \rangle = i \cdot \mu$

Weighted mean example Case 1 - observing different parts of the sky

Taking normalization condition into account we finally get

$$\mathbf{a}^{\mathsf{T}} = \frac{1}{10}(1,\ 2,\ 3,\ 4)$$

and the final result is

$$\begin{aligned}\bar{\mu} = \mathbf{a}^{\mathsf{T}}\,\mu &= \frac{1}{10}(\mu_1 + 2\mu_2 + 3\mu_3 + 4\mu_4) \\ &= \frac{1}{10}(N_1 + N_2 + N_3 + N_4) = \frac{N_{tot}}{T_{tot}}\end{aligned}$$

As the measurements are independent, we can just take the total event count N_{tot} and divide by the total measurement time (in years) T_{tot}. This gives the most precise estimate.

Weighted mean example Case 2 - observing same region in the sky

We assume n_1, n_2, n_3 and n_4 are numbers of GRB in each year (same for each experiment). They are independent, the covariance matrix for the variables n_i is thus diagonal:

$$\mathbb{C}_{\mathbf{n}} = \begin{pmatrix} n_1 & & & \\ & n_2 & & \\ & & n_3 & \\ & & & n_4 \end{pmatrix}$$

Weighted mean example Case 2 - observing same region in the sky

We assume n_1, n_2, n_3 and n_4 are numbers of GRB in each year (same for each experiment). They are independent, the covariance matrix for the variables n_i is thus diagonal:

$$\mathbb{C}_{\mathbf{n}} = \begin{pmatrix} n_1 & & & \\ & n_2 & & \\ & & n_3 & \\ & & & n_4 \end{pmatrix}$$

We can calculate covariance matrix for set of measurements μ_i, using the partial derivatives in new variables:

$$\frac{\partial \mu_k}{\partial n_j} = \begin{cases} \frac{1}{k} & \text{for} \quad k \geq j \\ 0 & \quad k < j \end{cases}$$

Weighted mean example Case 2 - observing same region in the sky

The covariance matrix for the set of measurements μ_i is more complicated now:

$$\mathbb{C}_\mu = \begin{pmatrix} n_1 & \frac{n_1}{2} & \frac{n_1}{3} & \frac{n_1}{4} \\ \frac{n_1}{2} & \frac{n_1+n_2}{4} & \frac{n_1+n_2}{6} & \frac{n_1+n_2}{8} \\ \frac{n_1}{3} & \frac{n_1+n_2}{6} & \frac{n_1+n_2+n_3}{9} & \frac{n_1+n_2+n_3}{12} \\ \frac{n_1}{4} & \frac{n_1+n_2}{8} & \frac{n_1+n_2+n_3}{12} & \frac{n_1+n_2+n_3+n_4}{16} \end{pmatrix}$$

Weighted mean example Case 2 - observing same region in the sky

The covariance matrix for the set of measurements μ_i is more complicated now:

$$\mathbb{C}_\mu = \begin{pmatrix} n_1 & \frac{n_1}{2} & \frac{n_1}{3} & \frac{n_1}{4} \\ \frac{n_1}{2} & \frac{n_1+n_2}{4} & \frac{n_1+n_2}{6} & \frac{n_1+n_2}{8} \\ \frac{n_1}{3} & \frac{n_1+n_2}{6} & \frac{n_1+n_2+n_3}{9} & \frac{n_1+n_2+n_3}{12} \\ \frac{n_1}{4} & \frac{n_1+n_2}{8} & \frac{n_1+n_2+n_3}{12} & \frac{n_1+n_2+n_3+n_4}{16} \end{pmatrix}$$

By solving the set of equations we get:

$$\mathbf{a}^\mathsf{T} = \lambda \left(\frac{1}{n_1} - \frac{1}{n_2}, \frac{2}{n_2} - \frac{2}{n_3}, \frac{3}{n_3} - \frac{3}{n_4}, \frac{4}{n_4} \right)$$

If GRB rate is constant in time ($n_1 \approx n_2 \approx n_3 \approx n_4$) average is clearly dominated by the fourth measurement (as expected!).

Weighted mean example Case 2 - observing same region in the sky

If we assume that $\langle n_1 \rangle = \langle n_2 \rangle = \langle n_3 \rangle = \langle n_4 \rangle = n$ (constant rate)
then the (true) covariance matrix for the set of measurements μ_i is:

$$\mathbb{C}_{\mathbf{x}} = \begin{pmatrix} n & \frac{n}{2} & \frac{n}{3} & \frac{n}{4} \\ \frac{n}{2} & \frac{n}{2} & \frac{n}{3} & \frac{n}{4} \\ \frac{n}{3} & \frac{n}{3} & \frac{n}{3} & \frac{n}{4} \\ \frac{n}{4} & \frac{n}{4} & \frac{n}{4} & \frac{n}{4} \end{pmatrix}$$

and by solving the set of equations we get:

$$\mathbf{a}^{\mathsf{T}} = \lambda \left(0,\ 0,\ 0,\ \frac{4}{n}\right)$$

Weighted mean example Case 2 - observing same region in the sky

If we assume that $\langle n_1 \rangle = \langle n_2 \rangle = \langle n_3 \rangle = \langle n_4 \rangle = n$ (constant rate)
then the (true) covariance matrix for the set of measurements μ_i is:

$$\mathbb{C}_{\mathbf{x}} = \begin{pmatrix} n & \frac{n}{2} & \frac{n}{3} & \frac{n}{4} \\ \frac{n}{2} & \frac{n}{2} & \frac{n}{3} & \frac{n}{4} \\ \frac{n}{3} & \frac{n}{3} & \frac{n}{3} & \frac{n}{4} \\ \frac{n}{4} & \frac{n}{4} & \frac{n}{4} & \frac{n}{4} \end{pmatrix}$$

and by solving the set of equations we get:

$$\mathbf{a}^{\mathsf{T}} = (0,\ 0,\ 0,\ 1)$$

Only the last measurement, including all observations, is relevant. All earlier can be discarded...

Monte Carlo methods

General problem

Presented above was a simple example of a more general problem: how to estimate parameters of the probability distribution function from the results of the experiment (measurements).

In many cases, parameter value can not be directly extracted from the measurement results

In the general case, shape of the probability density function for measurement result $\mathbf{x}$:

$$\mathbf{x} = (x_1, \ldots, x_n)$$

depends on a set of pdf parameters:

$$\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)$$

so the probability density should be written as:

$$f(\mathbf{x}; \boldsymbol{\lambda})$$

Likelihood Function

Probability density functions describes probability for given outcome of the experiment to be observed:

$$dP = f(\mathbf{x}; \boldsymbol{\lambda})\, d\mathbf{x}$$

Likelihood Function

Probability density functions describes probability for given outcome of the experiment to be observed:

$$dP = f(\mathbf{x}; \boldsymbol{\lambda})\, d\mathbf{x}$$

If experiment is repeated N times, we have N independent measurements
$\Rightarrow$ then the combined probability

$$dP = \prod_{j=1}^{N} f(\mathbf{x}^{(j)}; \boldsymbol{\lambda})\, d\mathbf{x}$$

This probability clearly depends on the parameter values $\boldsymbol{\lambda}$.

Likelihood Function

Probability density functions describes probability for given outcome of the experiment to be observed:

$$dP = f(\mathbf{x};\boldsymbol{\lambda})\, d\mathbf{x}$$

If experiment is repeated N times, we have N independent measurements
$\Rightarrow$ then the combined probability

$$dP = \prod_{j=1}^{N} f(\mathbf{x}^{(j)};\boldsymbol{\lambda})\, d\mathbf{x}$$

We can use this probability to compare different parameter sets. If

$$\prod_{j=1}^{N} f(\mathbf{x}^{(j)};\boldsymbol{\lambda}_1) > \prod_{j=1}^{N} f(\mathbf{x}^{(j)};\boldsymbol{\lambda}_2)$$

we can conclude that $\boldsymbol{\lambda}_1$ describes our experimental results better than $\boldsymbol{\lambda}_2$

Maximum Likelihood Method

The product:

$$L = \prod_{j=1}^{N} f(\mathbf{x}^{(j)}; \boldsymbol{\lambda})$$

is called a likelihood function.

Maximum Likelihood Method

The product:

$$L = \prod_{j=1}^{N} f(\mathbf{x}^{(j)}; \boldsymbol{\lambda})$$

is called a likelihood function.

The most commonly used approach to parameter estimation is the maximum likelihood approach: as the best estimate of the parameter set $\boldsymbol{\lambda}$ we choose the parameter values for which the likelihood function has a (global) maximum.

Maximum Likelihood Method

The product:

$$L = \prod_{j=1}^{N} f(\mathbf{x}^{(j)}; \boldsymbol{\lambda})$$

is called a likelihood function.

The most commonly used approach to parameter estimation is the maximum likelihood approach: as the best estimate of the parameter set $\boldsymbol{\lambda}$ we choose the parameter values for which the likelihood function has a (global) maximum.

Frequently used is also log-likelihood function

$$\ell = \ln L = \sum_{j=1}^{N} \ln f(\mathbf{x}^{(j)}; \boldsymbol{\lambda})$$

we can look for maximum value of ℓ or minimum of $-2\,\ell = -2 \ln L$

Example

Let us consider N measurements of source radioactivity (numbers of decays in given time window). Each measurement is described by the Poisson probability distribution. So the likelihood function is:

$$L = \prod_{i=1}^{N} P(n_i; \mu) = \prod_{i=1}^{N} \frac{\mu^{n_i} e^{-\mu}}{n!}$$

Example

Let us consider N measurements of source radioactivity (numbers of decays in given time window). Each measurement is described by the Poisson probability distribution. So the likelihood function is:

$$L = \prod_{i=1}^{N} P(n_i;\mu) = \prod_{i=1}^{N} \frac{\mu^{n_i} e^{-\mu}}{n!}$$

Log-likelihood:

$$\ell = \ln\mu \sum n_i - N\,\mu - \sum \ln n!$$

Example

Let us consider N measurements of source radioactivity (numbers of decays in given time window). Each measurement is described by the Poisson probability distribution. So the likelihood function is:

$$L = \prod_{i=1}^{N} P(n_i;\mu) = \prod_{i=1}^{N} \frac{\mu^{n_i} e^{-\mu}}{n!}$$

Log-likelihood:

$$\ell = \ln\mu \sum n_i - N\,\mu - \sum \ln n!$$

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\mu} \sum n_i - N$$

Example

Let us consider N measurements of source radioactivity (numbers of decays in given time window). Each measurement is described by the Poisson probability distribution. So the likelihood function is:

$$L = \prod_{i=1}^{N} P(n_i;\mu) = \prod_{i=1}^{N} \frac{\mu^{n_i} e^{-\mu}}{n!}$$

Log-likelihood:

$$\ell = \ln\mu \sum n_i - N\,\mu - \sum \ln n!$$

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\mu}\sum n_i - N = 0$$

$$\Rightarrow \quad \mu = \frac{1}{N}\sum n_i$$

we reproduce previous result (mean of the individual measurements)

Example (2)

Let us consider N independent measurements of variable X with non-uniform uncertainties. Assuming measurement fluctuations are described by Gaussian pdf, the likelihood function is:

$$L = \prod_{i=1}^{N} G(x_i; \mu, \sigma_i) = \prod_{i=1}^{N} \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x_i - \mu)^2}{\sigma_i^2}\right)$$

Example (2)

Let us consider N independent measurements of variable X with non-uniform uncertainties. Assuming measurement fluctuations are described by Gaussian pdf, the likelihood function is:

$$L = \prod_{i=1}^{N} G(x_i; \mu, \sigma_i) = \prod_{i=1}^{N} \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma_i^2}\right)$$

Log-likelihood:

$$\ell = -\frac{1}{2}\sum \frac{(x_i-\mu)^2}{\sigma_i^2} + \text{const}$$

Example (2)

Let us consider N independent measurements of variable X with non-uniform uncertainties. Assuming measurement fluctuations are described by Gaussian pdf, the likelihood function is:

$$L = \prod_{i=1}^{N} G(x_i; \mu, \sigma_i) = \prod_{i=1}^{N} \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma_i^2}\right)$$

Log-likelihood:

$$\ell = -\frac{1}{2}\sum \frac{(x_i-\mu)^2}{\sigma_i^2} + \text{const}$$

$$\frac{\partial \ell}{\partial \mu} = \sum \frac{x_i-\mu}{\sigma_i^2}$$

Example (2)

Let us consider N independent measurements of variable X with non-uniform uncertainties. Assuming measurement fluctuations are described by Gaussian pdf, the likelihood function is:

$$L = \prod_{i=1}^{N} G(x_i; \mu, \sigma_i) = \prod_{i=1}^{N} \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma_i^2}\right)$$

Log-likelihood:

$$\ell = -\frac{1}{2}\sum \frac{(x_i-\mu)^2}{\sigma_i^2} + \text{const}$$

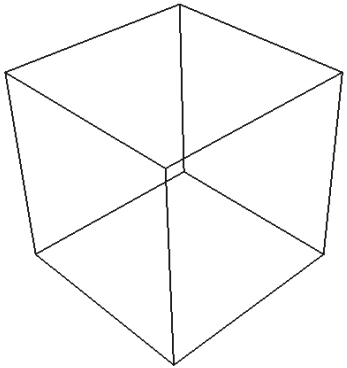
$$\frac{\partial \ell}{\partial \mu} = \sum \frac{x_i-\mu}{\sigma_i^2} = 0$$

$$\Rightarrow \quad \mu = \sigma^2 \sum \frac{x_i}{\sigma_i^2} \qquad \text{with } \frac{1}{\sigma^2} = \sum \frac{1}{\sigma_i^2}$$

Monte Carlo methods

Homework Solutions to be uploaded by November 13.

Use Monte Carlo method to calculate volume of a solid constructed as an intersection of three cylinders, with unit diameter and unit height, and perpendicular axes.

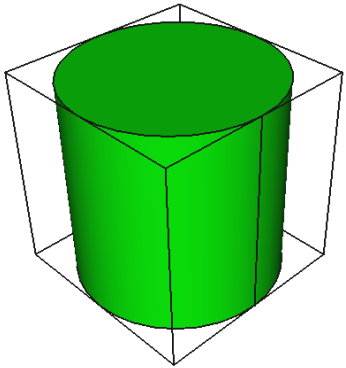


Unit cube

Homework

Solutions to be uploaded by November 13.

Use Monte Carlo method to calculate volume of a solid constructed as an intersection of three cylinders, with unit diameter and unit height, and perpendicular axes.

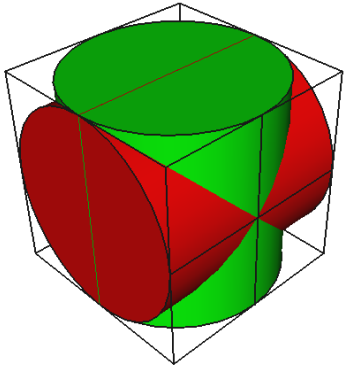


First cylinder

Homework

Solutions to be uploaded by November 13.

Use Monte Carlo method to calculate volume of a solid constructed as an intersection of three cylinders, with unit diameter and unit height, and perpendicular axes.

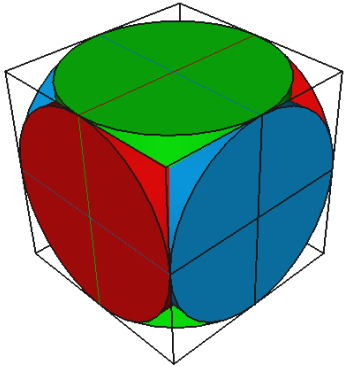


Second cylinder

Homework

Solutions to be uploaded by November 13.

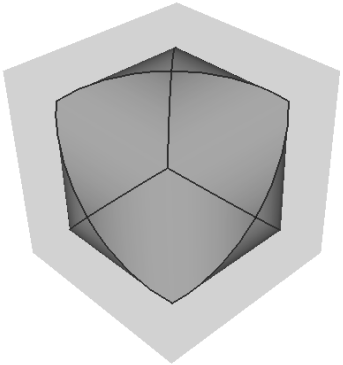
Use Monte Carlo method to calculate volume of a solid constructed as an intersection of three cylinders, with unit diameter and unit height, and perpendicular axes.



Third cylinder

Homework Solutions to be uploaded by November 13.

Use Monte Carlo method to calculate volume of a solid constructed as an intersection of three cylinders, with unit diameter and unit height, and perpendicular axes.



Intersection of three cylinders