Statistical analysis of experimental data
Variable distributions

Aleksander Filip Żarnecki

FACULTY OF PHYSICS
UNIVERSITY OF WARSAW

Lecture 11
December 21, 2023

Variable distributions

Iterative procedure (Brandt)

We start from some "initial guess" of parameter values $\mathbf{a_0}$.

Assuming small variations of the model parameters, $\mathbf{a} = \mathbf{a_0} + \delta\mathbf{a}$, we can expand χ^2 in a series:

$$\chi^2(\mathbf{a}) \;=\; \chi^2(\mathbf{a_0}) - 2\,\mathbf{b}\cdot(\mathbf{a}-\mathbf{a_0}) + \ldots$$

where $\mathbf{b}$ is the negative gradient of χ^2:

$$\mathbf{b} \;=\; -\frac{1}{2}\,\nabla\,\chi^2(\mathbf{a_0}) \qquad\qquad b_j \;=\; -\frac{1}{2}\frac{\partial\chi^2}{\partial a_j} \;=\; \sum_{i=1}^{N}\frac{y_i-\mu_i}{\sigma_i^2}\cdot\frac{\partial\mu_i}{\partial a_j}$$

Vector $\mathbf{b}$ defines the direction of steepest χ^2 descent.
One of the possible procedures is to make a step in this direction:

$$\mathbf{a_1} \;=\; \mathbf{a_0} \;+\; \varepsilon\,\mathbf{b}$$

with small $\varepsilon > 0$ and then repeat the whole procedure...

Iterative procedure

(Brandt)

We can try to be "smarter". Expanding χ^2 to quadratic term:

$$\chi^2(\mathbf{a}) = \chi^2(\mathbf{a_0}) - 2\,\mathbf{b}\cdot(\mathbf{a}-\mathbf{a_0}) + (\mathbf{a}-\mathbf{a_0})^{\intercal}\mathbb{A}(\mathbf{a}-\mathbf{a_0}) + \ldots$$

where $\mathbb{A}$ is the so called Hessian matrix of second derivatives:

$$\mathbb{A}_{jk} = \frac{1}{2}\,\frac{\partial^2\chi^2}{\partial a_j\,\partial a_k}\bigg|_{\mathbf{a}=\mathbf{a_0}} \approx \sum_{i=1}^{N}\frac{1}{\sigma_i^2}\cdot\frac{\partial\mu_i}{\partial a_j}\cdot\frac{\partial\mu_i}{\partial a_k} \qquad \left(\text{neglecting } \frac{\partial^2\mu_i}{\partial a_j\,\partial a_k}\right)$$

In this approximation, we can calculate the expected position of the χ^2 minimum:

$$\nabla\,\chi^2(\mathbf{a}) = -2\,\mathbf{b} + 2\,\mathbb{A}\,(\mathbf{a}-\mathbf{a_0}) = 0$$

$$\Rightarrow \quad \mathbf{a_1} = \mathbf{a_0} + \mathbb{A}^{-1}\mathbf{b}$$

and we can try to "jump" directly to the minimum...

Iterative procedure

(Brandt)

We can try to be "smarter". Expanding χ^2 to quadratic term:

$$\chi^2(\mathbf{a}) \;=\; \chi^2(\mathbf{a_0}) - 2\,\mathbf{b}\cdot(\mathbf{a}-\mathbf{a_0}) + (\mathbf{a}-\mathbf{a_0})^\intercal \mathbb{A} (\mathbf{a}-\mathbf{a_0}) + \ldots$$

where $\mathbb{A}$ is the so called Hessian matrix of second derivatives:

$$\mathbb{A}_{jk} \;=\; \frac{1}{2}\,\frac{\partial^2 \chi^2}{\partial a_j\,\partial a_k}\bigg|_{\mathbf{a}=\mathbf{a_0}} \approx \sum_{i=1}^{N} \frac{1}{\sigma_i^2}\cdot\frac{\partial \mu_i}{\partial a_j}\cdot\frac{\partial \mu_i}{\partial a_k} \qquad \left(\text{neglecting } \frac{\partial^2 \mu_i}{\partial a_j\,\partial a_k}\right)$$

In this approximation, we can calculate the expected position of the χ^2 minimum:

$$\nabla\,\chi^2(\mathbf{a}) \;=\; -2\,\mathbf{b} + 2\,\mathbb{A}\,(\mathbf{a}-\mathbf{a_0}) \;=\; 0$$

$$\Rightarrow\quad \mathbf{a_1} \;=\; \mathbf{a_0} + \left(\mathbb{A} + \lambda\cdot\mathbb{I}\right)^{-1}\mathbf{b}$$

or interpolate between "safe" and "fast" approach... Marquardt Minimization

Method of Lagrange Multipliers (Behnke)

The method, invented by J.L.Lagrange in 1788, applies to general minimization problem with additional constraints imposed.

Problem of finding minimum of χ^2 $(\mathbf{a})$ with constraints $w_k(\mathbf{a}) = 0$ is equivalent to finding a stationary point (point with all first derivatives at zero) of the Lagrange function:

$$\mathcal{L}(\mathbf{a}, \boldsymbol{\lambda}) \;=\; \chi^2(\mathbf{a}) + \sum_k 2\lambda_k w_k(\mathbf{a})$$

where we introduce additional K parameters λ_k - Lagrange multipliers

Our problem is now reduced to finding parameters $\mathbf{a}$ and $\boldsymbol{\lambda}$ fulfilling

$$\frac{\partial \mathcal{L}}{\partial a_j} \;=\; 0 \qquad \text{and} \qquad \frac{\partial \mathcal{L}}{\partial \lambda_k} \;=\; 0$$

(without any additional constraints)

Method of Lagrange Multipliers

We can write these equations the matrix form:

$$\underbrace{\left(\begin{array}{c|c} \mathbb{A} & \mathbb{D} \\ \hline \mathbb{D}^T & 0 \end{array}\right)}_{\tilde{\mathbb{A}}} \cdot \left(\begin{array}{c} \mathbf{a} \\ \hline \boldsymbol{\lambda} \end{array}\right) = \left(\begin{array}{c} \mathbf{b} \\ \hline \mathbf{c} \end{array}\right)$$

where: $\mathbb{A}_{jk} = \sum_{i=1}^{N} \frac{f_j(x_i)\; f_k(x_i)}{\sigma_i^2}$, $\mathbb{D}_{jk} = d_{k,\,j}$ and $b_j = \sum_{i=1}^{N} \frac{f_j(x_i)\; y_i}{\sigma_i^2}$

and the problem can be solved by inverting matrix $\tilde{\mathbb{A}}$.

Covariance matrix for **a** can be extracted as: (seems to work for linear problems)

$$(\mathbb{C}_{\mathbf{a}})_{ij} = (\tilde{\mathbb{A}}^{-1})_{ij} \qquad i,j = 1 \ldots M$$

General procedure

General procedure for including systematic uncertainties in the analysis is to consider corresponding systematic shifts as additional model parameters

$$\mu_i = \mu(x_i; \mathbf{a}, \mathbf{s}) = \mu(x_i; \mathbf{a}')$$

$$\chi^2(\mathbf{a}, \mathbf{s}) = \sum_{i=1}^{N} \frac{(y_i - \mu(x_i, \mathbf{a}, \mathbf{s}))^2}{\sigma_i^2} + \sum_{k=1}^{K} \frac{(s_k - s_{0,\,k})^2}{\sigma_{s_k}^2}$$

$$\chi^2(\mathbf{a}') = \sum_{i=1}^{N} \frac{(y_i - \mu(x_i, \mathbf{a}'))^2}{\sigma_i^2} + \sum_{k=1}^{K} \delta_k^2 \qquad \delta_k = \frac{s_k - s_{0,\,k}}{\sigma_{s_k}}$$

If systematic parameters are not independent (are correlated)

$$\chi^2(\mathbf{a}') = \sum_{i=1}^{N} \frac{(y_i - \mu(x_i, \mathbf{a}'))^2}{\sigma_i^2} + \sum_{k,j} (s_k - s_{0,\,k})(s_j - s_{0,\,j}) \left(\mathbb{C}_{\mathbf{s}}\right)^{-1}_{j,k}$$

General procedure

χ^2 minimization procedure is basically unchanged, only the additional terms (systematic constrains) need to be included in calculations (as for the parameter constraints).

Negative gradient of χ^2 uncorrelated systematics

$$b_j = -\frac{1}{2}\frac{\partial \chi^2}{\partial a'_j} = \sum_{i=1}^{N} \frac{y_i - \mu_i}{\sigma_i^2} \cdot \frac{\partial \mu_i}{\partial a'_j} - \frac{s_j - s_{0,\,j}}{\sigma_{s_j}^2}$$

Hessian matrix of second derivatives:

$$\mathbb{A}_{jk} = \frac{1}{2}\frac{\partial^2 \chi^2}{\partial a'_j\, \partial a'_k} = \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \cdot \frac{\partial \mu_i}{\partial a'_j} \cdot \frac{\partial \mu_i}{\partial a'_k} + \frac{\delta_{jk}}{\sigma_{s_k}^2}$$

where systematic shifts **s** are assumed to go first in **a**' (for proper indexing)

General procedure

χ^2 minimization procedure is basically unchanged, only the additional terms (systematic constrains) need to be included in calculations (as for the parameter constraints).

Negative gradient of χ^2 general case

$$b_j = -\frac{1}{2}\frac{\partial \chi^2}{\partial a'_j} = \sum_{i=1}^{N} \frac{y_i - \mu_i}{\sigma_i^2} \cdot \frac{\partial \mu_i}{\partial a'_j} - \sum_k (s_k - s_{0,\,k})\,(\mathbb{C}_{\mathbf{s}})^{-1}_{j,k}$$

Hessian matrix of second derivatives:

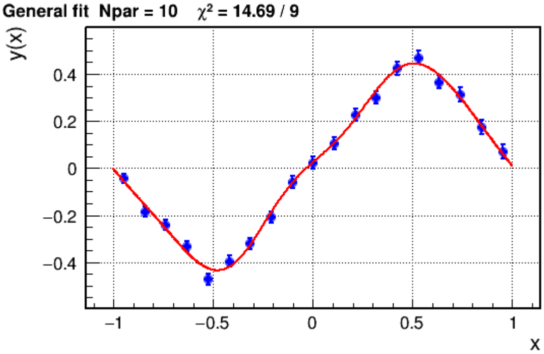
$$\mathbb{A}_{jk} = \frac{1}{2}\frac{\partial^2 \chi^2}{\partial a'_j\,\partial a'_k} = \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \cdot \frac{\partial \mu_i}{\partial a'_j} \cdot \frac{\partial \mu_i}{\partial a'_k} + (\mathbb{C}_{\mathbf{s}})^{-1}_{j,k}$$

where systematic shifts **s** are assumed to go first in **a**' (for proper indexing)

Variable distributions

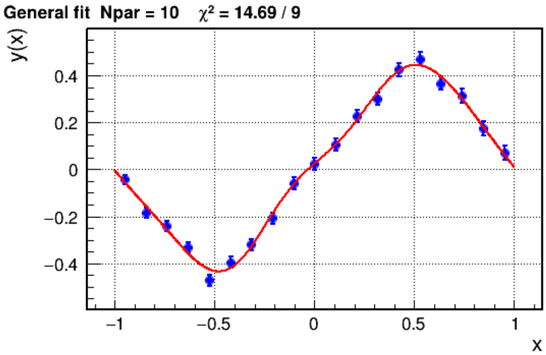
Problem

So far, we have considered sets of independent measurements of a random variable Y, which could depend on some controlled variable X (and a number of model parameters $\mathbf{a}$), assuming measurement fluctuations are described by Gaussian pdf.



Problem

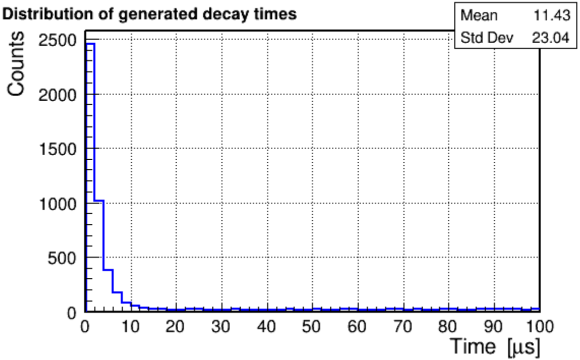
So far, we have considered sets of independent measurements of a random variable Y, which could depend on some controlled variable X (and a number of model parameters $\mathbf{a}$), assuming measurement fluctuations are described by Gaussian pdf.



Problem

11_example_data.ipynb Open in Colab

However, the problem frequently met (in high energy physics experiments in particular) is that we want to extract model parameters not from the fit of $Y(X)$ dependence, but just from the distribution of the measured X values. Results are often presented in a form of a histogram:

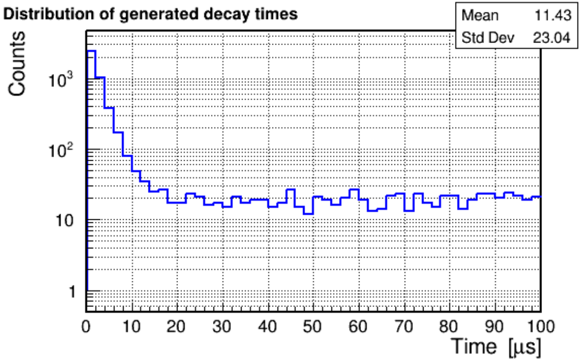


Example simulation of P3 experiment results (Advanced Physics Laboratory)

Problem

11_example_data.ipynb   Open in Colab

However, the problem frequently met (in high energy physics experiments in particular) is that we want to extract model parameters not from the fit of $Y(X)$ dependence, but just from the distribution of the measured X values. Results are often presented in a form of a histogram:



Example simulation of P3 experiment results (Advanced Physics Laboratory)

Problem

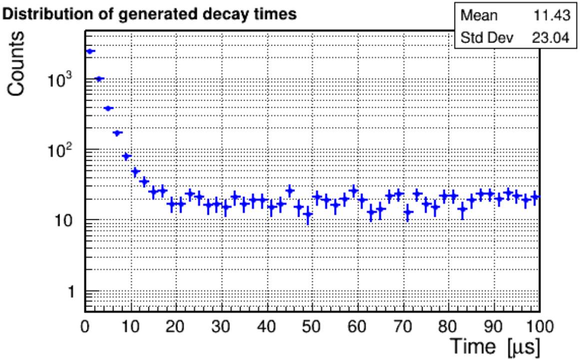
11_example_data2.ipynb Open in Colab

We can consider number of events in each bin of the histogram, n_i, as an independent measurement depending on the controlled variable x_i. The only problem we have, to use the χ^2 minimization procedure, is to attribute measurement uncertainties to measured numbers of events.

Simple guess:

assume $\sigma_{n_i} = \sqrt{n_i}$

works very well for large n_i



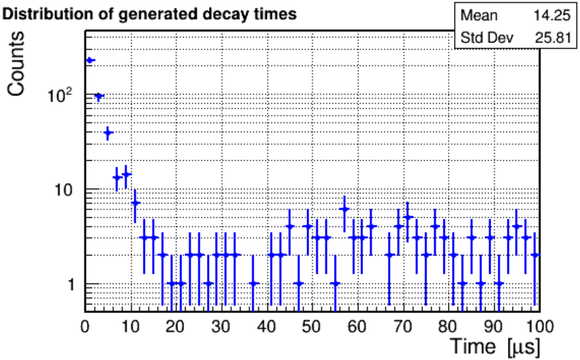
Problem

11_example_data2.ipynb Open in Colab

We can consider number of events in each bin of the histogram, n_i, as an independent measurement depending on the controlled variable x_i. The only problem we have, to use the χ^2 minimization procedure, is to attribute measurement uncertainties to measured numbers of events.

Simple guess:
assume $\sigma_{n_i} = \sqrt{n_i}$

results become biased when n_i small, problem with $n_i = 0$



Least-squares fit

(Behnke)

Considered approach was proposed by K.Pearson in 1900:

$$\chi^2 = \sum_{i=1}^{N_{bin}} \frac{(n_i - \mu_i)^2}{n_i}$$

when we use the property of the Poisson distribution $\mathbb{V}(n_i) = \mathbb{E}(n_i) = \mu_i$

Least-squares fit

(Behnke)

Considered approach was proposed by K.Pearson in 1900:

$$\chi^2 = \sum_{i=1}^{N_{bin}} \frac{(n_i - \mu_i)^2}{n_i}$$

when we use the property of the Poisson distribution $\mathbb{V}(n_i) = \mathbb{E}(n_i) = \mu_i$

Unfortunately, the result of the minimization turns out to be biased!

Let us assume that the model expectations can be presented as

$$\mu_i = N\, f_i$$

where f_i represents the probability density (properly normalized)

From minimization condition, $\frac{\partial\chi^2}{\partial\mathbf{a}} = 0$, **biased estimate** of N is obtained:

$$\hat{N} = N - \chi^2_{\min}$$

Example

It can be easily demonstrated for the flat distribution

$$\mu_i \equiv \mu \quad \Rightarrow \quad \chi^2 = \sum_{i=1}^{N_{bin}} \frac{(n_i - \mu)^2}{n_i} \; = \sum_{i=1}^{N_{bin}} \left(n_i - 2\mu + \frac{\mu^2}{n_i} \right)$$

Example

It can be easily demonstrated for the flat distribution

$$\mu_i \equiv \mu \quad \Rightarrow \quad \chi^2 = \sum_{i=1}^{N_{bin}} \frac{(n_i - \mu)^2}{n_i} = \sum_{i=1}^{N_{bin}} \left(n_i - 2\mu + \frac{\mu^2}{n_i} \right)$$

we then obtain:

$$0 = \frac{\partial \chi^2}{\partial \mu} = -2N_{bin} + 2\mu \sum_{i=1}^{N_{bin}} \frac{1}{n_i}$$

Example

It can be easily demonstrated for the flat distribution

$$\mu_i \equiv \mu \quad \Rightarrow \quad \chi^2 = \sum_{i=1}^{N_{bin}} \frac{(n_i - \mu)^2}{n_i} \; = \sum_{i=1}^{N_{bin}} \left(n_i - 2\mu + \frac{\mu^2}{n_i} \right)$$

we then obtain:

$$0 \; = \; \frac{\partial \chi^2}{\partial \mu} \; = \; -2N_{bin} + 2\mu \sum_{i=1}^{N_{bin}} \frac{1}{n_i}$$

$$\Rightarrow \quad \hat{\mu} \; = \; \left(\frac{1}{N_{bin}} \sum_{i=1}^{N_{bin}} \frac{1}{n_i} \right)^{-1}$$

which is harmonic mean of n_i (not equal to the expected arithmetic mean)

Least-squares fit (Behnke)

Alternative approach was proposed by J.Neyman in 1949:

$$\chi^2 = \sum_{i=1}^{N_{bin}} \frac{(n_i - \mu_i)^2}{\mu_i}$$

when we try to use “true” uncertainty in the denominator...

Least-squares fit

(Behnke)

Alternative approach was proposed by J.Neyman in 1949:

$$\chi^2 = \sum_{i=1}^{N_{bin}} \frac{(n_i - \mu_i)^2}{\mu_i}$$

when we try to use "true" uncertainty in the denominator...

Unfortunately, this approach also turns out to be biased!

Higher μ_i values are "preferred", as they result in smaller χ^2 contribution (for given difference).

Estimate of N obtained from minimization condition, $\frac{\partial \chi^2}{\partial \mathbf{a}} = 0$:

$$\hat{N} = N + \frac{1}{2}\chi^2_{\min}$$

Example

Let as consider flat probability distribution again

$$\mu_i \equiv \mu \quad \Rightarrow \quad \chi^2 = \sum_{i=1}^{N_{bin}} \frac{(n_i - \mu)^2}{\mu} = \sum_{i=1}^{N_{bin}} \left(\frac{n_i^2}{\mu} - 2n_i + \mu \right)$$

Example

Let as consider flat probability distribution again

$$\mu_i \equiv \mu \quad \Rightarrow \quad \chi^2 = \sum_{i=1}^{N_{bin}} \frac{(n_i - \mu)^2}{\mu} = \sum_{i=1}^{N_{bin}} \left(\frac{n_i^2}{\mu} - 2n_i + \mu \right)$$

we then obtain:

$$0 = \frac{\partial \chi^2}{\partial \mu} = -\sum_{i=1}^{N_{bin}} \frac{n_i^2}{\mu^2} + N_{bin}$$

Example

Let as consider flat probability distribution again

$$\mu_i \equiv \mu \quad \Rightarrow \quad \chi^2 = \sum_{i=1}^{N_{bin}} \frac{(n_i - \mu)^2}{\mu} = \sum_{i=1}^{N_{bin}} \left(\frac{n_i^2}{\mu} - 2n_i + \mu \right)$$

we then obtain:

$$0 = \frac{\partial \chi^2}{\partial \mu} = -\sum_{i=1}^{N_{bin}} \frac{n_i^2}{\mu^2} + N_{bin}$$

$$\Rightarrow \quad \hat{\mu}^2 = \frac{1}{N_{bin}} \sum_{i=1}^{N_{bin}} n_i^2$$

Example

Let as consider flat probability distribution again

$$\mu_i \equiv \mu \quad \Rightarrow \quad \chi^2 = \sum_{i=1}^{N_{bin}} \frac{(n_i - \mu)^2}{\mu} = \sum_{i=1}^{N_{bin}} \left(\frac{n_i^2}{\mu} - 2n_i + \mu \right)$$

we then obtain:

$$0 = \frac{\partial \chi^2}{\partial \mu} = -\sum_{i=1}^{N_{bin}} \frac{n_i^2}{\mu^2} + N_{bin}$$

$$\Rightarrow \quad \hat{\mu}^2 = \frac{1}{N_{bin}} \sum_{i=1}^{N_{bin}} n_i^2$$

$$\langle \hat{\mu}^2 \rangle = \langle n^2 \rangle = \langle n \rangle^2 + \langle (n - \langle n \rangle)^2 \rangle = \mu^2 + \mu$$

which shows that the method results in biased (too high) value of $\hat{\mu}$

Maximum Likelihood

The χ^2 method, while giving (almost) correct results for large n_i is clearly not suitable to fit variable distributions when n_i can be small.

Solution is to use general Maximum Likelihood Method, look for parameter values for which the likelihood function has a (global) maximum.

Maximum Likelihood

The χ^2 method, while giving (almost) correct results for large n_i is clearly not suitable to fit variable distributions when n_i can be small.

Solution is to use general Maximum Likelihood Method, look for parameter values for which the likelihood function has a (global) maximum.

We have N_{bin} independent measurements and each is described by the Poisson probability distribution. So the likelihood function is:

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

Maximum Likelihood

The χ^2 method, while giving (almost) correct results for large n_i is clearly not suitable to fit variable distributions when n_i can be small.

Solution is to use general Maximum Likelihood Method, look for parameter values for which the likelihood function has a (global) maximum.

We have N_{bin} independent measurements and each is described by the Poisson probability distribution. So the likelihood function is:

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

Log-likelihood:

$$\ell = \sum (n_i \ln \mu_i - \mu_i) - \sum \ln n_i!$$

were the last term can be neglected in minimization (constant)

Example

For our example of flat distribution, $\mu_i \equiv \mu$

$$L = \prod_{i=1}^{N_{bin}} \frac{\mu^{n_i} e^{-\mu}}{n!}$$

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

Example

For our example of flat distribution, $\mu_i \equiv \mu$

$$L = \prod_{i=1}^{N_{bin}} \frac{\mu^{n_i} e^{-\mu}}{n!}$$

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

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

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

we obtain an unbiased estimate of the expected value.

Maximum Likelihood fit

In the Maximum Likelihood approach, we can use all methods introduced for χ^2 minimization, one only needs to make substitution

$$\chi^2(\mathbf{x};\mathbf{a}) \qquad \longrightarrow \qquad -2\ell(\mathbf{x};\mathbf{a})$$

Maximum Likelihood fit

In the Maximum Likelihood approach, we can use all methods introduced for χ^2 minimization, one only needs to make substitution

$$\chi^2(\mathbf{x};\mathbf{a}) \quad \longrightarrow \quad -2\ell(\mathbf{x};\mathbf{a})$$

In the general fit approach:

$$\mathbf{b} = -\frac{1}{2}\nabla\chi^2(\mathbf{a}) \quad \longrightarrow \quad \mathbf{b} = \nabla\ell(\mathbf{a}) \quad \text{or} \quad b_j = \frac{\partial\ell}{\partial a_j}$$

Maximum Likelihood fit

In the Maximum Likelihood approach, we can use all methods introduced for χ^2 minimization, one only needs to make substitution

$$\chi^2(\mathbf{x};\mathbf{a}) \quad \longrightarrow \quad -2\ell(\mathbf{x};\mathbf{a})$$

In the general fit approach:

$$\mathbf{b} = -\frac{1}{2}\nabla\chi^2(\mathbf{a}) \quad \longrightarrow \quad \mathbf{b} = \nabla\ell(\mathbf{a}) \quad \text{or} \quad b_j = \frac{\partial\ell}{\partial a_j}$$

$$\mathbb{A}_{jk} = \frac{1}{2}\frac{\partial^2\chi^2}{\partial a_j\,\partial a_k} \quad \longrightarrow \quad \mathbb{A}_{jk} = -\frac{\partial^2\ell}{\partial a_j\,\partial a_k}$$

Maximum Likelihood fit

In the Maximum Likelihood approach, we can use all methods introduced for χ^2 minimization, one only needs to make substitution

$$\chi^2(\mathbf{x};\mathbf{a}) \quad \longrightarrow \quad -2\ell(\mathbf{x};\mathbf{a})$$

In the general fit approach:

$$\mathbf{b} = -\frac{1}{2}\nabla\chi^2(\mathbf{a}) \quad \longrightarrow \quad \mathbf{b} = \nabla\ell(\mathbf{a}) \quad \text{or} \quad b_j = \frac{\partial\ell}{\partial a_j}$$

$$\mathbb{A}_{jk} = \frac{1}{2}\frac{\partial^2\chi^2}{\partial a_j\,\partial a_k} \quad \longrightarrow \quad \mathbb{A}_{jk} = -\frac{\partial^2\ell}{\partial a_j\,\partial a_k}$$

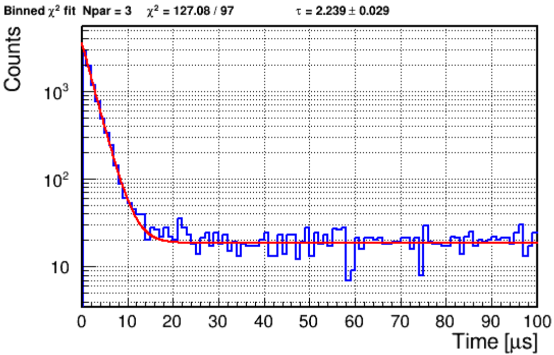
These derivatives can be directly calculated for the Poisson distribution:

$$b_j = \sum_{i=1}^{N_{bin}}\left(\frac{n_i}{\mu_i}-1\right)\frac{\partial\mu_i}{\partial a_j} \quad \text{and} \quad \mathbb{A}_{jk} = \sum_{i=1}^{N_{bin}}\frac{n_i}{\mu_i^2}\frac{\partial\mu_i}{\partial a_j}\frac{\partial\mu_i}{\partial a_k}$$

Maximum Likelihood fit example

11_binned_fit.ipynb Open in Colab

Iterative χ^2 fit, using (modified) Pearson approach $(\sigma_{n_i} = \sqrt{n_i + 1})$

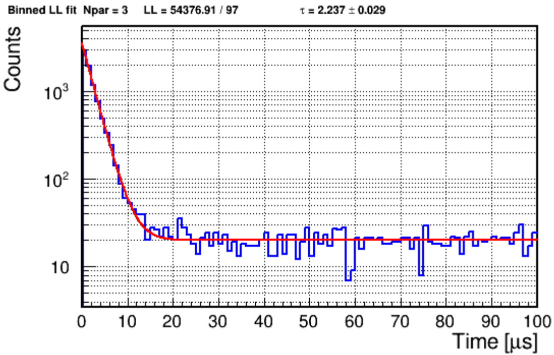


N=10000

Maximum Likelihood fit example

11_binned_fit2.ipynb Open in Colab

Iterative Maximum Log-Likelihood fit

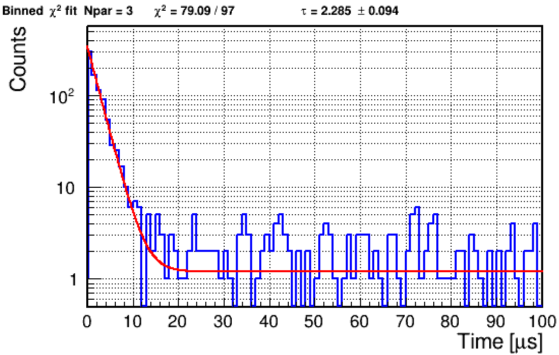


N=10000

Higher background level estimate in likelihood fit (underestimated in χ^2 fit)

Maximum Likelihood fit example

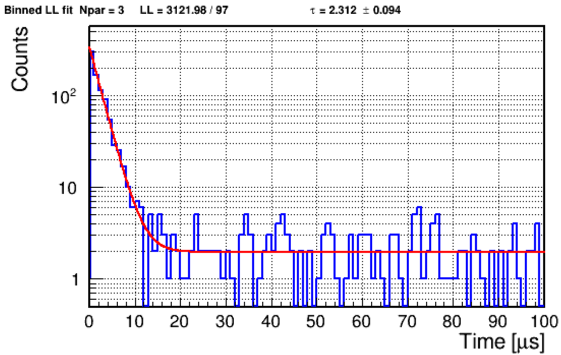
Iterative χ^2 fit, using (modified) Pearson approach $(\sigma_{n_i} = \sqrt{n_i + 1})$



N=1000

Maximum Likelihood fit example

Iterative Maximum Log-Likelihood fit

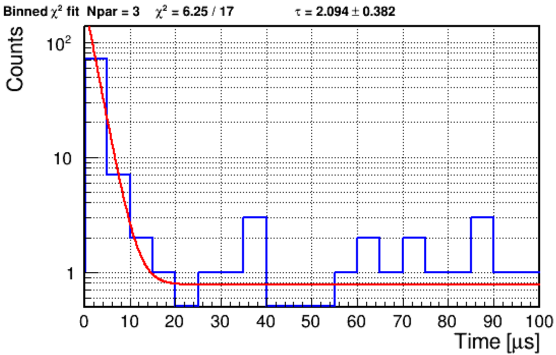


N=1000

Difference between two methods becomes significant for small n_i

Maximum Likelihood fit example

Iterative χ^2 fit, using (modified) Pearson approach $(\sigma_{n_i} = \sqrt{n_i + 1})$

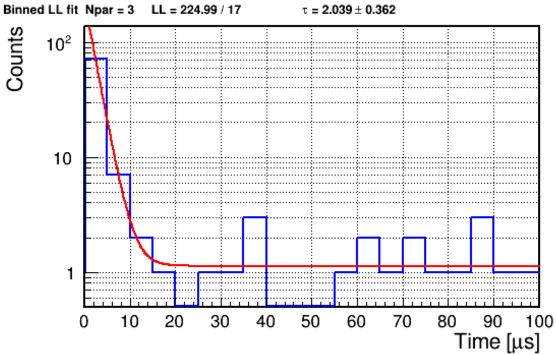


N=100

Difference between two methods becomes significant for small n_i

Maximum Likelihood fit example

Iterative Maximum Log-Likelihood fit



N=100

Difference between two methods becomes significant for small n_i

Variable distributions

Free normalization

It is quite often the case that normalization of our data sample (total number of registered events) is one of the (unknown) parameters of our model. We can then present model expectations as

$$\mu_i(\mathbf{a}') = A\, f_i(\mathbf{a})$$

where $f_i(\mathbf{a})$ is the probability density depending on model parameters $\mathbf{a}$

Free normalization

It is quite often the case that normalization of our data sample (total number of registered events) is one of the (unknown) parameters of our model. We can then present model expectations as

$$\mu_i(\mathbf{a}') = A\, f_i(\mathbf{a})$$

where $f_i(\mathbf{a})$ is the probability density depending on model parameters $\mathbf{a}$

We can try to maximize log-likelihood with respect to A:

$$\ell = \sum (n_i \ln A + n_i \ln f_i - A f_i) - \sum \ln n_i!$$

Free normalization

It is quite often the case that normalization of our data sample (total number of registered events) is one of the (unknown) parameters of our model. We can then present model expectations as

$$\mu_i(\mathbf{a}') = A\, f_i(\mathbf{a})$$

where $f_i(\mathbf{a})$ is the probability density depending on model parameters $\mathbf{a}$

We can try to maximize log-likelihood with respect to A:

$$\ell = \sum (n_i \ln A + n_i \ln f_i - A f_i) - \sum \ln n_i!$$

$$\frac{\partial \ell}{\partial A} = \sum \left(\frac{n_i}{A} - f_i \right) \quad \Rightarrow \quad A = \frac{\sum n_i}{\sum f_i}$$

which corresponds to the normalization condition: $\sum \mu_i = \sum n_i$

Normalization fit

If we do not care for normalization (and its uncertainty) and only want to consider shape of the distribution, we can reduce number of model parameters by including normalization condition in the definition of the model function:

$$\mu_i(\mathbf{a}) = \frac{\sum_{j=1}^{N_{bin}} n_j}{\sum_{k=1}^{N_{bin}} f_k} \; f_i(\mathbf{a})$$

Normalization of the distribution

Normalization fit

If we do not care for normalization (and its uncertainty) and only want to consider shape of the distribution, we can reduce number of model parameters by including normalization condition in the definition of the model function:

$$\mu_i(\mathbf{a}) \quad = \quad \frac{\sum_{j=1}^{N_{bin}} n_j}{\sum_{k=1}^{N_{bin}} f_k} \; f_i(\mathbf{a})$$

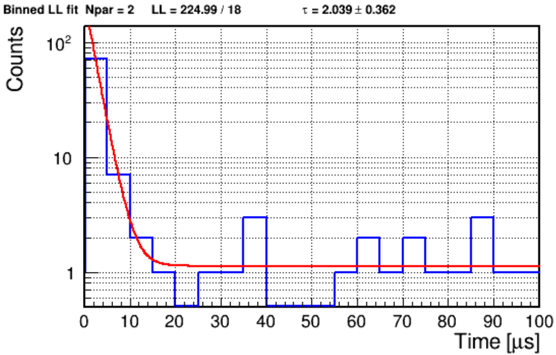
If normalization is not correlated with other model parameters, derivatives of the normalization term can be neglected in the fit.

If there are correlations, uncertainties on model parameters can be underestimated...

Normalization fit example

11_shape_fit.ipynb Open in Colab

Example of the likelihood fit including normalization condition



Results agree perfectly with the previous fit (with normalization as parameter)

Normalization constrain

It can also be the case that the normalization of data is known from theory or independent measurement (eg. luminosity). Let us assume it is known with relative uncertainty Δ. We can write the log-likelihood as:

$$\ell \;=\; \sum (n_i \ln s + n_i \ln \mu_i \;-\; s\mu_i) \;-\; \frac{1}{2}\frac{(s-1)^2}{\Delta^2}$$

where s is the factor scaling the nominal model expectations $(s_0 = 1)$

Normalization constrain

It can also be the case that the normalization of data is known from theory or independent measurement (eg. luminosity). Let us assume it is known with relative uncertainty Δ. We can write the log-likelihood as:

$$\ell = \sum (n_i \ln s + n_i \ln \mu_i - s\mu_i) - \frac{1}{2}\frac{(s-1)^2}{\Delta^2}$$

where s is the factor scaling the nominal model expectations ($s_0 = 1$)

We could use general approach to systematic effects, as described before. However, in case of the normalization systematics, the problem factorizes. We can extract s from derivative:

$$\frac{\partial \ell}{\partial s} = \frac{1}{s}\sum n_i - \sum \mu_i - \frac{s-1}{\Delta^2} = 0$$

Normalization constrain

To simplify this formula, let us introduce normalization shift, $s = 1 + \delta$.

$$\sum n_i \; - \; (1+\delta)\sum \mu_i \; - \; \delta(1+\delta)\frac{1}{\Delta^2} \;\; = \;\; 0$$

If we now assume that normalization variation is small, $\delta \ll 1$, $\delta^2 \ll \delta$

$$\sum n_i \; - \; \sum \mu_i \;\; = \;\; \delta \left(\sum \mu_i + \frac{1}{\Delta^2} \right)$$

Normalization constrain

To simplify this formula, let us introduce normalization shift, $s = 1 + \delta$.

$$\sum n_i \; - \; (1+\delta) \sum \mu_i \; - \; \delta(1+\delta)\frac{1}{\Delta^2} \;\; = \;\; 0$$

If we now assume that normalization variation is small, $\delta \ll 1$, $\delta^2 \ll \delta$

$$\sum n_i \; - \; \sum \mu_i \;\; = \;\; \delta \left(\sum \mu_i + \frac{1}{\Delta^2} \right)$$

$$\delta \; = \; \frac{\sum n_i \; - \; \sum \mu_i}{\sum \mu_i + \frac{1}{\Delta^2}}$$

Normalization constrain

To simplify this formula, let us introduce normalization shift, $s = 1 + \delta$.

$$\sum n_i \; - \; (1+\delta)\sum \mu_i \; - \; \delta(1+\delta)\frac{1}{\Delta^2} \;\; = \;\; 0$$

If we now assume that normalization variation is small, $\delta \ll 1$, $\delta^2 \ll \delta$

$$\sum n_i \; - \; \sum \mu_i \;\; = \;\; \delta \left(\sum \mu_i + \frac{1}{\Delta^2} \right)$$

$$\delta \; = \; \frac{\sum n_i \; - \; \sum \mu_i}{\sum \mu_i + \frac{1}{\Delta^2}} \quad \Rightarrow \quad s \; = \; 1 + \delta \; = \; \frac{\sum n_i \; + \; \frac{1}{\Delta^2}}{\sum \mu_i \; + \; \frac{1}{\Delta^2}}$$

This constraint can be included in the model definition as well!

For $\Delta \to 0$ normalization becomes fixed ($s \equiv 1$)

For $\Delta \to \infty$ we reproduce "free normalization" result...

Normalization constrain

Normalization constrain can also be considered in the χ^2 minimization:

$$\chi^2 = \sum_{i=1}^{N_{bin}} \frac{(n_i - s\mu_i)^2}{\sigma_{n_i}^2} + \frac{(s-1)^2}{\Delta^2}$$

Normalization constrain

Normalization constrain can also be considered in the χ^2 minimization:

$$\chi^2 = \sum_{i=1}^{N_{bin}} \frac{(n_i - s\mu_i)^2}{\sigma_{n_i}^2} + \frac{(s-1)^2}{\Delta^2}$$

$$\frac{\partial \chi^2}{\partial s} = 2s \sum \frac{\mu_i^2}{\sigma_{n_i}^2} - 2 \sum \frac{n_i \mu_i}{\sigma_{n_i}^2} + \frac{2(s-1)}{\Delta^2} = 0$$

Normalization constrain

Normalization constrain can also be considered in the χ^2 minimization:

$$\chi^2 = \sum_{i=1}^{N_{bin}} \frac{(n_i - s\mu_i)^2}{\sigma_{n_i}^2} + \frac{(s-1)^2}{\Delta^2}$$

$$\frac{\partial \chi^2}{\partial s} = 2s \sum \frac{\mu_i^2}{\sigma_{n_i}^2} - 2 \sum \frac{n_i \mu_i}{\sigma_{n_i}^2} + \frac{2(s-1)}{\Delta^2} = 0$$

$$s \left(\sum \frac{\mu_i^2}{\sigma_{n_i}^2} + \frac{1}{\Delta^2} \right) = \sum \frac{n_i \mu_i}{\sigma_{n_i}^2} + \frac{1}{\Delta^2}$$

Normalization constrain

Normalization constrain can also be considered in the χ^2 minimization:

$$\chi^2 = \sum_{i=1}^{N_{bin}} \frac{(n_i - s\mu_i)^2}{\sigma_{n_i}^2} + \frac{(s-1)^2}{\Delta^2}$$

$$\frac{\partial\chi^2}{\partial s} = 2s\sum \frac{\mu_i^2}{\sigma_{n_i}^2} - 2\sum \frac{n_i\mu_i}{\sigma_{n_i}^2} + \frac{2(s-1)}{\Delta^2} = 0$$

$$s\left(\sum \frac{\mu_i^2}{\sigma_{n_i}^2} + \frac{1}{\Delta^2}\right) = \sum \frac{n_i\mu_i}{\sigma_{n_i}^2} + \frac{1}{\Delta^2}$$

$$s = \frac{\sum \frac{n_i\mu_i}{\sigma_{n_i}^2} + \frac{1}{\Delta^2}}{\sum \frac{\mu_i^2}{\sigma_{n_i}^2} + \frac{1}{\Delta^2}}$$

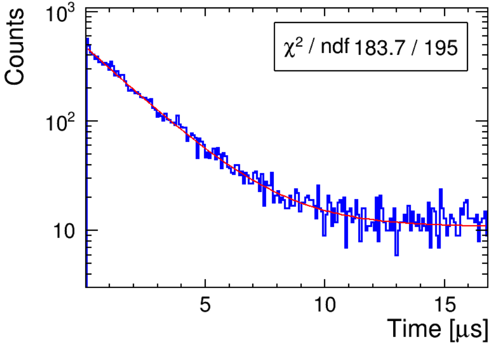
which reduces to the previous result, if we assume $\sigma_{n_i}^2 = \mu_i$

Variable distributions

Problem

When defining parameters of the histogram, which will be used to extract parameters of the variable distribution, we have to be very careful!

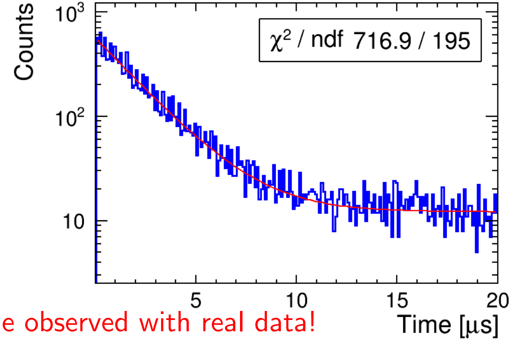
Example histogram from P3 exercise (real data), time bin $\Delta t = 84$ ns



Problem

When defining parameters of the histogram, which will be used to extract parameters of the variable distribution, we have to be very careful!

Example histogram from P3 exercise (real data), time bin $\Delta t = 100$ ns



Unexpected effects can be observed with real data!

Unbinned likelihood

We do need the histogram to visualize our data. But we do not need it to extract parameters of the distribution. We can do it directly from the data.

Likelihood of our data set can be calculated from single events:

$$L = \prod_{i=1}^{N} f(x_i; \mathbf{a})$$

$$\text{or} \quad \ell = \sum_{i=1}^{N} \ln f(x_i; \mathbf{a})$$

when the sum runs over all collected events.

We can then fit parameters by looking for maximum of (log-)likelihood...
We look at the shape of the distribution only (normalization fixed)!

Example

Very simple example is the decay time measurement.
Let us assume that we measured decay times, t_i, of N identical particles.
We know the probability distribution function:

$$f(t) \;=\; \frac{1}{\tau}\, e^{-\frac{t}{\tau}}$$

where the mean lifetime, τ, is the only parameter.
We can write the formula for log-likelihood

$$\ell \;=\; \sum_{i=1}^{N} \ln f(t_i;\tau) \;=\; \sum_{i=1}^{N} \left(-\ln\tau - \frac{t_i}{\tau} \right)$$

Example

Very simple example is the decay time measurement.
Let us assume that we measured decay times, t_i, of N identical particles.
We know the probability distribution function:

$$f(t) = \frac{1}{\tau} e^{-\frac{t}{\tau}}$$

where the mean lifetime, τ, is the only parameter.
We can write the formula for log-likelihood

$$\ell = \sum_{i=1}^{N} \ln f(t_i;\tau) = \sum_{i=1}^{N} \left(-\ln\tau - \frac{t_i}{\tau} \right)$$

$$\frac{\partial \ell}{\partial \tau} = -\frac{N}{\tau} + \frac{1}{\tau^2} \sum_{i=1}^{N} t_i = 0$$

Example

Very simple example is the decay time measurement.
Let us assume that we measured decay times, t_i, of N identical particles.
We know the probability distribution function:

$$f(t) = \frac{1}{\tau} e^{-\frac{t}{\tau}}$$

where the mean lifetime, τ, is the only parameter.
We can write the formula for log-likelihood

$$\ell = \sum_{i=1}^{N} \ln f(t_i;\tau) = \sum_{i=1}^{N} \left(-\ln\tau - \frac{t_i}{\tau} \right)$$

$$\frac{\partial \ell}{\partial \tau} = -\frac{N}{\tau} + \frac{1}{\tau^2} \sum_{i=1}^{N} t_i = 0 \qquad \Rightarrow \qquad \tau = \frac{1}{N} \sum_{i=1}^{N} t_i$$

General case

As before, we can use all methods introduced for χ^2 minimization with proper substitution

$$\chi^2(\mathbf{x};\mathbf{a}) \quad \longrightarrow \quad -2\ell(\mathbf{x};\mathbf{a})$$

General case

As before, we can use all methods introduced for χ^2 minimization with proper substitution

$$\chi^2(\mathbf{x};\mathbf{a}) \quad \longrightarrow \quad -2\ell(\mathbf{x};\mathbf{a})$$

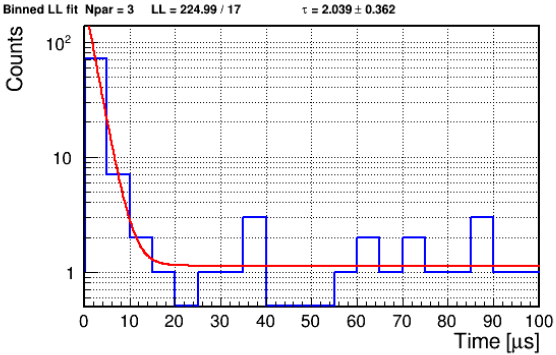
In the general unbinned log-likelihood fit: $f_i = f(x_i)$

$$b_j = \frac{\partial \ell}{\partial a_j} = \sum_{i=1}^{N} \frac{1}{f_i} \frac{\partial f_i}{\partial a_j}$$

$$\mathbb{A}_{jk} = -\frac{\partial^2 \ell}{\partial a_j\, \partial a_k} = \sum_{i=1}^{N} \frac{1}{f_i^2} \frac{\partial \mu_i}{\partial a_j} \frac{\partial \mu_i}{\partial a_k}$$

Unbinned likelihood fit example

Iterative binned log-likelihood fit for comparison

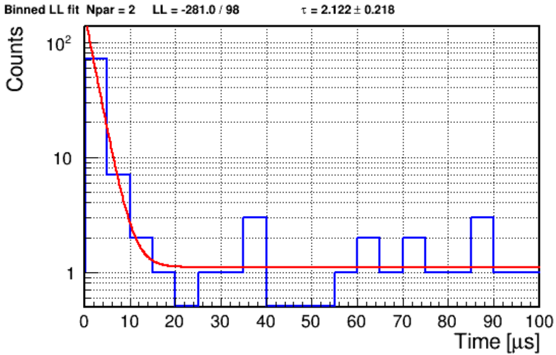


N=100

Unbinned likelihood fit example

11_unbinned_fit.ipynb Open in Colab

Iterative unbinned log-likelihood fit



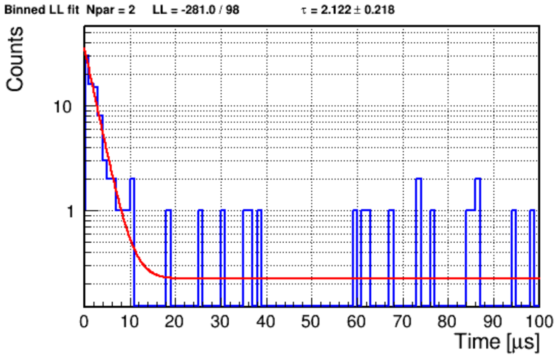
N=100

Higher precision of the lifetime estimate!

Unbinned likelihood fit example

11_unbinned_fit.ipynb Open in Colab

Iterative unbinned log-likelihood fit



N=100

Higher precision of the lifetime estimate! More details "visible"…

Extended Maximum Likelihood

(Behnke)

"Standard" (unbinned) maximum likelihood fit is sensitive only to the shape of the probability distribution (normalized by definition).

However, we can extend the likelihood definition to take possible normalization fluctuations into account:

$$L(x_i;\mu,\mathbf{a}) \;=\; \frac{\mu^N e^{-\mu}}{N!}\prod_{i=1}^{N} f(x_i;\mathbf{a})$$

where μ is the now the expected total number of observed events.

$$\ell(x_i;\mu,\mathbf{a}) \;=\; N\ln\mu \;-\; \mu \;+\; \sum_{i=1}^{N}\ln f(x_i;\mathbf{a}) \;+\; \mathrm{const}$$

Extended Maximum Likelihood (Behnke)

When μ is independent of **a**, maximum of the (extended) likelihood corresponds to

$$\mu = N$$

and we reproduce our previous result.

Extended Maximum Likelihood

(Behnke)

When μ is independent of $\mathbf{a}$, maximum of the (extended) likelihood corresponds to

$$\mu = N$$

and we reproduce our previous result.

However, we need to use the extended approach, if the total expected number of events depends on the model parameters

$$\mu \rightarrow \mu(\mathbf{a})$$

and so it is related to the shape of the distribution.

If this is the case, extended approach is also required to get a correct estimate of the parameter uncertainties.

Variable distributions

Problem

So far, we focused on the problem of extracting model parameters from the collected data sample. We used maximum likelihood approach (or χ^2 minimization, which is a special case).

Problem

So far, we focused on the problem of extracting model parameters from the collected data sample. We used maximum likelihood approach (or χ^2 minimization, which is a special case).

However, what we often want to do is to "make choice", discriminate between two (or more) hypothesis based on the collected data.

We already addressed this problem (partially) when discussing limits (lecture 07) and consistency of the fit (lecture 08).

Problem

So far, we focused on the problem of extracting model parameters from the collected data sample. We used maximum likelihood approach (or χ^2 minimization, which is a special case).

However, what we often want to do is to "make choice", discriminate between two (or more) hypothesis based on the collected data.

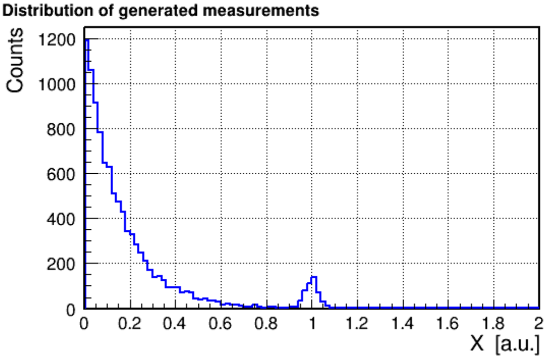
We already addressed this problem (partially) when discussing limits (lecture 07) and consistency of the fit (lecture 08).

The general formulation of the problem:
how to discriminate between two model hypothesis H_0 and H_1 based on the collected data D?

Common case:
H_0 - Standard Model is valid, H_1 - SM + additional BSM contribution
D - the whole collected data sample, subset, or a single measurement

Example

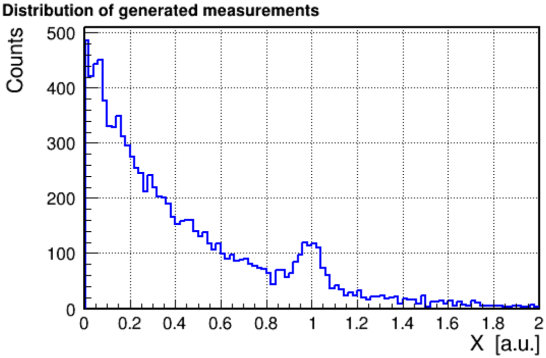
We can consider measurement of X, where exponential decrease is expected in the SM and BSM signal is expected to be visible as a peak



This is a case with very clear separation...

Example

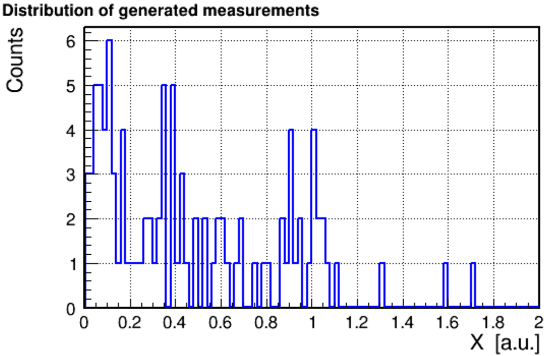
We can consider measurement of X, where exponential decrease is expected in the SM and BSM signal is expected to be visible as a peak



More difficult when the two distributions overlap

Example

We can consider measurement of X, where exponential decrease is expected in the SM and BSM signal is expected to be visible as a peak



More difficult when the two distributions overlap and statistics is small...

Neyman–Pearson Lemma

According to Neymann and Pearson, the optimal, "most powerful" method to discriminate between the two hypothesis is to look at likelihood ratio

$$Q(D) = \frac{L(D|H_1)}{L(D|H_0)}$$

When considering single measurements, making a cut on Q(x) is the optimal way to classify events. By using likelihood ratio, multi-dimensional measurements (whole events) are also presented as single number...

Neyman–Pearson Lemma

According to Neymann and Pearson, the optimal, “most powerful” method to discriminate between the two hypothesis is to look at likelihood ratio

$$Q(D) = \frac{L(D|H_1)}{L(D|H_0)}$$

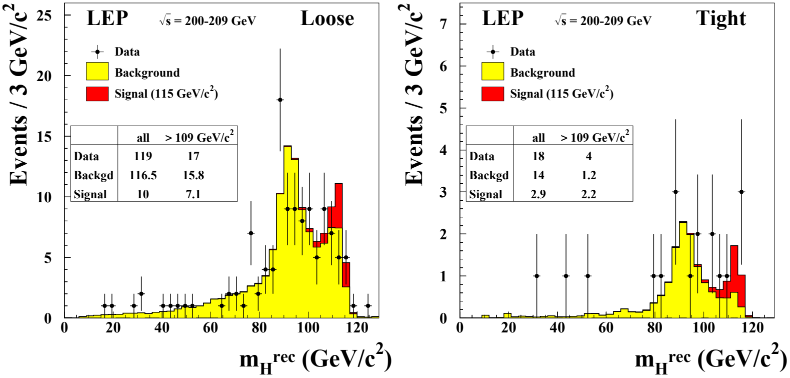
When considering single measurements, making a cut on Q(x) is the optimal way to classify events. By using likelihood ratio, multi-dimensional measurements (whole events) are also presented as single number...

When we consider the whole sample of collected data, value of Q(D) is the best discriminant between the two hypothesis.

Still, one needs to compare the value of $Q(D)$ resulting from the measurement, with the expected Q distributions for the two hypothesis.

CL_s method

This method was introduced at the end of LEP running, when some hints for Higgs boson production were observed



$\mathbf{CL_s}$ method

The two hypothesis we consider in this case:

H_0 - Standard Model without Higgs contribution - "background" only (b)

H_1 - SM with Higgs contribution - "signal+background" (s+b)

where we can consider different masses of the Higgs, m_H

CL_s method

The two hypothesis we consider in this case:
H_0 - Standard Model without Higgs contribution - “background” only (b)
H_1 - SM with Higgs contribution - “signal+background” (s+b)
where we can consider different masses of the Higgs, m_H

Instead of using Q, it is more convenient to use

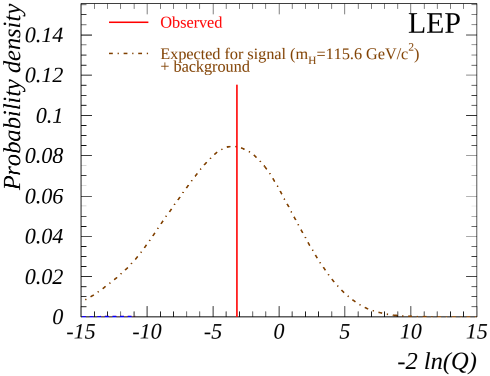
$$q = -2\ln Q = -2\ell(D|H_1) + 2\ell(D|H_0) = \chi^2(D|H_1) - \chi^2(D|H_0)$$

where:

- positive q values are expected for data looking more like background only (H_0)
- negative q values indicate that data are better described by signal+background (H_1)

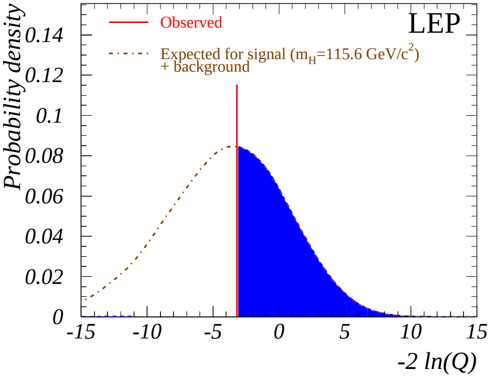
CL_s method

Value of q from LEP, q_{dat}, was compared with distribution obtained with multiple Monte Carlo experiments for $m_H = 115.6$ GeV.



CL_s method

Value of q from LEP, q_{dat}, was compared with distribution obtained with multiple Monte Carlo experiments for $m_H = 115.6$ GeV.



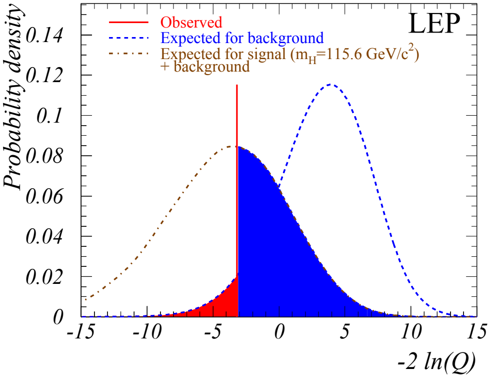
We can define

$$CL_{s+b} = \int_{q_{dat}}^{+\infty} dq\, f^{H1}(q)$$

$\Leftarrow$ indicated as blue area

CL_s method

Value of q from LEP, q_{dat}, was compared with distribution obtained with multiple Monte Carlo experiments for $m_H = 115.6$ GeV.



We can define

$$CL_{s+b} = \int_{q_{dat}}^{+\infty} dq\, f^{H1}(q)$$

$\Leftarrow$ indicated as blue area and compare it with

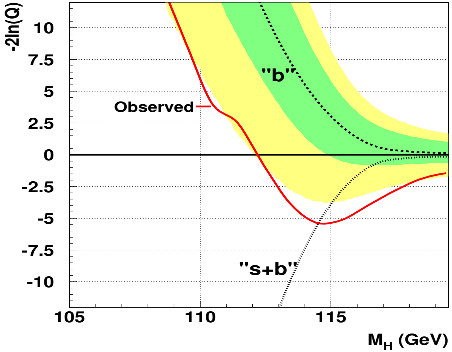
$$CL_b = \int_{q_{dat}}^{+\infty} dq\, f^{H0}(q)$$

$\Leftarrow$ indicated as red is $1 - CL_b$

CL_s method

Measured and expected q (for two hypothesis) as a function of the assumed Higgs mass.

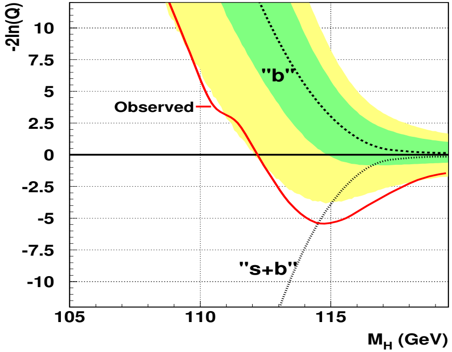
The green and yellow bands represent the 68% and 95% probability bands about the median background expectation



Looks like we exclude H_0 up to mass ~ 118 GeV (Frequentist 97.5%CL)

CL_s method

Measured and expected q (for two hypothesis) as a function of the assumed Higgs mass.

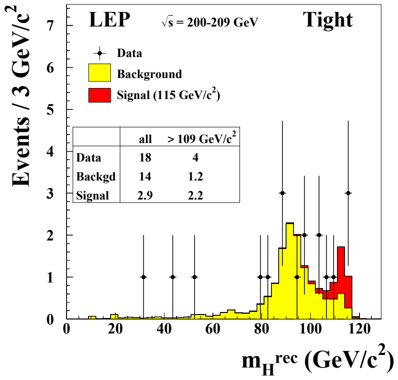


The green and yellow bands represent the 68% and 95% probability bands about the median background expectation

Looks like we exclude H_0 up to mass $\sim$ 118 GeV (Frequentist 97.5%CL)

But there is almost no difference between expectations for H_1 and H_0 ?!

CL_s method



With tight event selection, LEP experiments observed 4 candidate events with reconstructed mass, $m_H^{rec} > 109\,GeV$.

$\mathrm{CL_s}$ method

	$1-\alpha$ =90%		$1-\alpha$ =95%	
n	μ_1	μ_2	μ_1	μ_2
0	0.00	2.44	0.00	3.09
1	0.11	4.36	0.05	5.14
2	0.53	5.91	0.36	6.72
3	1.10	7.42	0.82	8.25
4	1.47	8.60	1.37	9.76
5	1.84	9.99	1.84	11.26
6	2.21	11.47	2.21	12.75
7	3.56	12.53	2.58	13.81
8	3.96	13.99	2.94	15.29
9	4.36	15.30	4.36	16.77
10	5.50	16.50	4.75	17.82

Unified intervals (RPP)

With tight event selection, LEP experiments observed 4 candidate events with reconstructed mass, $m_H^{rec} > 109\,GeV$.

Expectations of the background only hyposthesis, $b = 1.2$ is below 95% CL limit (both Cental and Unified, refer lecture 07)

CL_s method

	$1-\alpha$ =90%		$1-\alpha$ =95%	
n	μ_1	μ_2	μ_1	μ_2
0	0.00	2.44	0.00	3.09
1	0.11	4.36	0.05	5.14
2	0.53	5.91	0.36	6.72
3	1.10	7.42	0.82	8.25
4	1.47	8.60	1.37	9.76
5	1.84	9.99	1.84	11.26
6	2.21	11.47	2.21	12.75
7	3.56	12.53	2.58	13.81
8	3.96	13.99	2.94	15.29
9	4.36	15.30	4.36	16.77
10	5.50	16.50	4.75	17.82

Unified intervals (RPP)

With tight event selection, LEP experiments observed 4 candidate events with reconstructed mass, $m_H^{rec} > 109\,GeV$.

Expectations of the background only hyposthesis, $b = 1.2$ is below 95% CL limit (both Cental and Unified, refer lecture 07)

Does it mean that we can exclude H_0 hypothesis (Standard Model predictions)?

All we can say is that
"probability of SM reproducing this data is below 5%"...

But we know that it can still be due to fluctuations...

$\mathbf{CL_s}$ method

Experiments at LEP, running with energy up to $\sqrt{s} = 210$ GeV, could only observe Higgs bosons with mass of up to about 118 GeV (produced together with Z boson: $e^+e^- \to Z\ H$)

For higher masses, signal+background hypothesis (H_1) becomes indistinguishable from background only one (H_0)

In strictly frequentiest approach we could exclude (on 95%CL) not only the SM, but also all Higgs scenarios (H_1) with $m_H > 118 GeV$!..

Frequentist approach gives us result which is correct (from statistical point of view) but not very useful for discriminating the two hypothesis! Too sensitive to background fluctuations...

$\mathbf{CL_s}$ method

Experiments at LEP, running with energy up to $\sqrt{s} = 210$ GeV, could only observe Higgs bosons with mass of up to about 118 GeV (produced together with Z boson: $e^+e^- \to Z\ H$)

For higher masses, signal+background hypothesis (H_1) becomes indistinguishable from background only one (H_0)

In strictly frequentiest approach we could exclude (on 95%CL) not only the SM, but also all Higgs scenarios (H_1) with $m_H > 118 GeV$!..

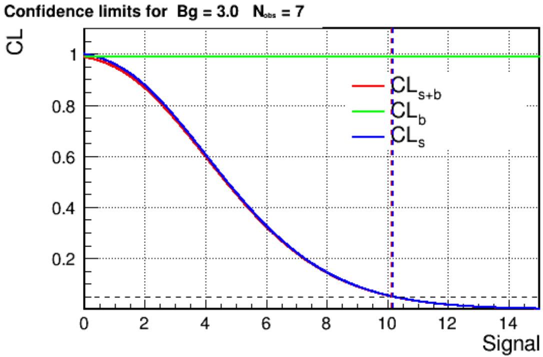
Frequentist approach gives us result which is correct (from statistical point of view) but not very useful for discriminating the two hypothesis! Too sensitive to background fluctuations...

Solution is to look for confidence level of H_1 relative to H_0:

$$CL_s = \frac{CL_{s+b}}{CL_b}$$

$\mathbf{CL_s}$ method example

Counting experiment with expected background $\mu_{bg} = 3$ and $N_{obs} = 7$

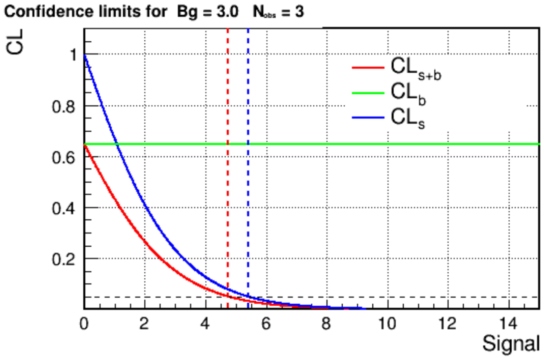


Probability of background hypothesis to result in $N_{obs} \leq 7$ is 98.8%

$\Rightarrow$ CL_s limit on number of signal events is 10.17 (95% CL)

CL_s method example

Counting experiment with expected background $\mu_{bg} = 3$ and $N_{obs} = 3$

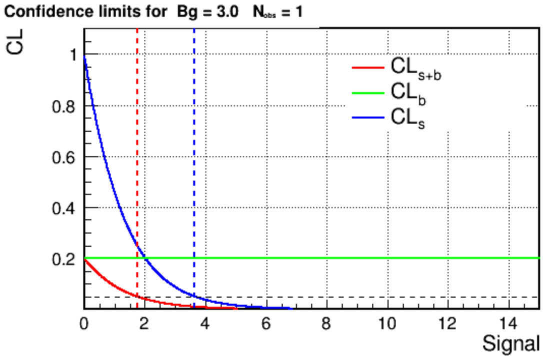


Probability of background hypothesis to result in $N_{obs} \leq 3$ is 64.7%

$\Rightarrow$ CL_s limit on number of signal events is 5.40 (95% CL)

CL_s method example

Counting experiment with expected background $\mu_{bg} = 3$ and $N_{obs} = 1$

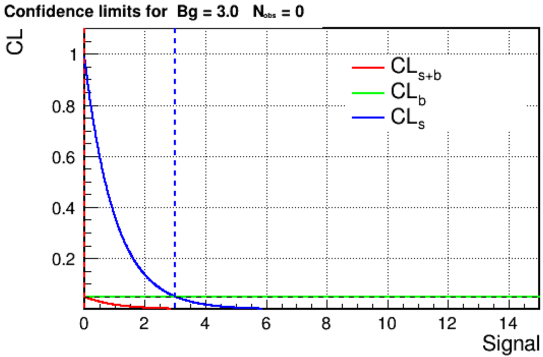


Probability of background hypothesis to result in $N_{obs} \leq 1$ is 19.9%

$\Rightarrow$ CL_s limit on number of signal events is 3.64 (95% CL)

CL_s method example

Counting experiment with expected background $\mu_{bg} = 3$ and $N_{obs} = 0$



Probability of background hypothesis to result in $N_{obs} = 0$ is 4.98%

$\Rightarrow$ CL_s limit on number of signal events is 3.00 (95% CL)

CL_s method

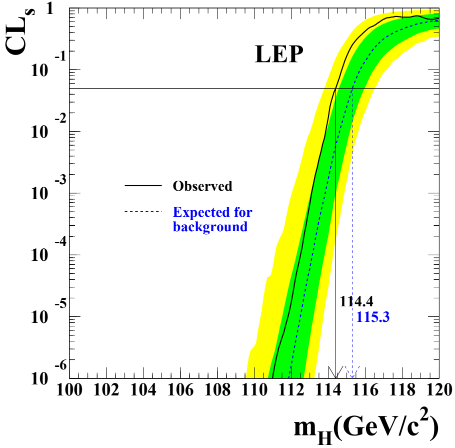
In the modified approach, we exclude (at 95% CL) all scenarios with

$$CL_s \quad < 0.05$$

This means that the probability of H_1 to reproduce the collected data is less than 5% of the SM probability:

$$P(q > q_{dat}|H_1) \; < \; 0.05 \; P(q > q_{dat}|H_0)$$

Final Higgs limits from LEP



Variable distributions

Homework

Solutions to be uploaded by January 18.

1000 events were collected in the muon lifetime measurement.
Distribution can be described by the formula:

$$\frac{dN}{dt} = \frac{N_{sig}}{\tau} e^{-\frac{t}{\tau}} + \frac{dN_{bg}}{dt} \qquad 0 \leq t \leq 15\ \mu s$$

with flat background level known to be $\frac{dN_{bg}}{dt} = 10 \pm \Delta\ \mu s^{-1}$

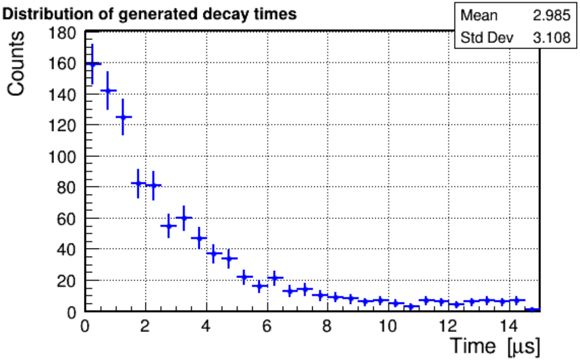
Estimate the dependence of the uncertainty on the muon lifetime, σ_τ, obtained from the fit on the assumed background uncertainty Δ.

Hint: consider the Hessian matrix of the fit

Homework

Solutions to be uploaded by January 18.

Example data



Homework

Solutions to be uploaded by January 18.

Example data