Statistical analysis of experimental data
Parameter Inference

Aleksander Filip Żarnecki

FACULTY OF PHYSICS
UNIVERSITY OF WARSAW

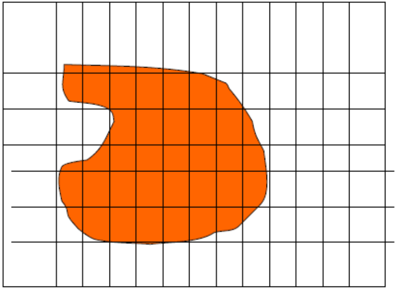
Lecture 06

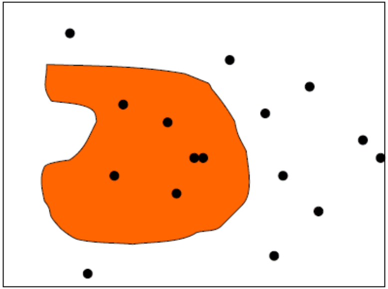
November 16, 2023

Parameter Inference

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

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 according to $f(\mathbf{x})$. 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

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

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:

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

where the second term corrects for the modified pdf.

Weighted Monte Carlo

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

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

Variance of the sum of weights:

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

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

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

For Poisson distributed random number $\mathbb{V}(N) = N$

Weighted mean

$\mathbb{I} = (1, \ldots, 1)^\intercal$

What about averaging measurements which are not independent?

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

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

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:

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

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 $-2\,\ell = -2\ln L$

Parameter Inference

Mean estimate

modified from lecture 05

Let us consider N independent measurements of variable X with given (uniform) uncertainty. Assuming measurement fluctuations are described by Gaussian pdf, the likelihood function is:

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

Log-likelihood:

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

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2}\sum(x_i-\mu) = 0$$

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

Variance estimate

The same method can be also used to estimate the variance of the Gaussian distribution.
Consider partial derivative with respect to σ^2 (we can not neglect normalization now):

$$\ell \;=\; -\frac{1}{2\sigma^2}\sum(x_i-\mu)^2 \;-\frac{N}{2}\;\ln(2\pi\sigma^2)$$

Variance estimate

The same method can be also used to estimate the variance of the Gaussian distribution. Consider partial derivative with respect to σ^2 (we can not neglect normalization now):

$$\ell = -\frac{1}{2\sigma^2}\sum(x_i-\mu)^2 - \frac{N}{2}\ln(2\pi\sigma^2)$$

$$\frac{\partial\ell}{\partial\sigma^2} = \frac{1}{2\sigma^4}\sum(x_i-\mu)^2 - \frac{N}{2\sigma^2}$$

Variance estimate

The same method can be also used to estimate the variance of the Gaussian distribution.
Consider partial derivative with respect to σ^2 (we can not neglect normalization now):

$$\ell = -\frac{1}{2\sigma^2}\sum(x_i-\mu)^2 - \frac{N}{2}\ln(2\pi\sigma^2)$$

$$\frac{\partial\ell}{\partial\sigma^2} = \frac{1}{2\sigma^4}\sum(x_i-\mu)^2 - \frac{N}{2\sigma^2} = 0$$

$$\Rightarrow \quad \sigma^2 = \frac{1}{N}\sum(x_i-\mu)^2$$

Variance estimate

The same method can be also used to estimate the variance of the Gaussian distribution. Consider partial derivative with respect to σ^2 (we can not neglect normalization now):

$$\ell = -\frac{1}{2\sigma^2}\sum(x_i-\mu)^2 - \frac{N}{2}\ln(2\pi\sigma^2)$$

$$\frac{\partial\ell}{\partial\sigma^2} = \frac{1}{2\sigma^4}\sum(x_i-\mu)^2 - \frac{N}{2\sigma^2} = 0$$

$$\Rightarrow \quad \sigma^2 = \frac{1}{N}\sum(x_i-\mu)^2$$

If we extract both μ and σ^2 from the same set of measurements:

$$\sigma^2 = \frac{1}{N}\sum(x_i-\bar{x})^2$$

$\Rightarrow$ ML variance estimator is biased! Bessel's correction missing - lecture 04.

Multiple parameter estimate

Likelihood function (and log-likelihood) can depend on multiple parameters:

$$\boldsymbol{\lambda} = (\lambda_1 \ldots \lambda_p) \qquad L = \prod_{j=1}^{N} f(\mathbf{x}^{(j)}; \boldsymbol{\lambda}) \qquad \ell = \sum_{j=1}^{N} \ln f(\mathbf{x}^{(j)}; \boldsymbol{\lambda})$$

Multiple parameter estimate

Likelihood function (and log-likelihood) can depend on multiple parameters:

$$\boldsymbol{\lambda} = (\lambda_1 \dots \lambda_p) \qquad L = \prod_{j=1}^{N} f(\mathbf{x}^{(j)}; \boldsymbol{\lambda}) \qquad \ell = \sum_{j=1}^{N} \ln f(\mathbf{x}^{(j)}; \boldsymbol{\lambda})$$

Best estimate of $\boldsymbol{\lambda}$, for given set of experimental results $\mathbf{x}^{(j)}$, corresponds to maximum of the likelihood function, which can be found by solving a system of equations:

$$\left.\frac{\partial \ell}{\partial \lambda_i}\right|_{i=1\dots p} = 0$$

Multiple parameter estimate

Likelihood function (and log-likelihood) can depend on multiple parameters:

$$\boldsymbol{\lambda} = (\lambda_1 \ldots \lambda_p) \qquad L = \prod_{j=1}^{N} f(\mathbf{x}^{(j)}; \boldsymbol{\lambda}) \qquad \ell = \sum_{j=1}^{N} \ln f(\mathbf{x}^{(j)}; \boldsymbol{\lambda})$$

Best estimate of $\boldsymbol{\lambda}$, for given set of experimental results $\mathbf{x}^{(j)}$, corresponds to maximum of the likelihood function, which can be found by solving a system of equations:

$$\left.\frac{\partial \ell}{\partial \lambda_i}\right|_{i=1\ldots p} = 0$$

The Likelihood Principle

G. Bohm and G. Zech

Given a p.d.f. $f(\mathbf{x}; \boldsymbol{\lambda})$ containing an unknown parameters of interest $\boldsymbol{\lambda}$ and observations $\mathbf{x}^{(j)}$, all information relevant for the estimation of the parameters $\boldsymbol{\lambda}$ is contained in the likelihood function $L(\boldsymbol{\lambda}; \mathbf{x}) = \prod f(\mathbf{x}^{(j)}; \boldsymbol{\lambda})$.

Multivariate Normal Distribution

Consider experiment resulting in a measurement $\mathbf{x} = (x_1, \ldots, x_n)$.
If we assume each variable follows Gaussian p.d.f, the most general form of the joint probability distribution is:

$$f(\mathbf{x}; \boldsymbol{\lambda}) \;=\; A \,\exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\lambda})^{\mathsf{T}}\, \mathbb{B}\, (\mathbf{x}-\boldsymbol{\lambda})\right]$$

where $\boldsymbol{\lambda}$ is parameter vector and $\mathbb{B}$ is an $n \times n$ matrix.

Since p.d.f. is symmetric about the point $\mathbf{x} = \boldsymbol{\lambda}$:

$$\mathbb{E}(\mathbf{x}-\boldsymbol{\lambda}) \;=\; \int d\mathbf{x}\,(\mathbf{x}-\boldsymbol{\lambda})\, f(\mathbf{x}; \boldsymbol{\lambda}) \;=\; 0$$

$$\Rightarrow \quad \mathbb{E}(\mathbf{x}) \;=\; \boldsymbol{\lambda}$$

Multivariate Normal Distribution

We should note that the derivative of the probability distribution:

$$\frac{\partial f}{\partial \lambda_i} = [(\mathbf{x} - \boldsymbol{\lambda})^{\mathsf{T}} \, \mathbb{B}]_i \cdot f(\mathbf{x}; \boldsymbol{\lambda})$$

Multivariate Normal Distribution

We should note that the derivative of the probability distribution:

$$\frac{\partial f}{\partial \lambda_i} \;=\; \left[(\mathbf{x}-\boldsymbol{\lambda})^{\mathsf{T}}\,\mathbb{B}\right]_i \cdot f(\mathbf{x};\boldsymbol{\lambda})$$

We can now differentiate the formula for $\mathbb{E}(\mathbf{x}-\boldsymbol{\lambda})$ with respect to $\boldsymbol{\lambda}$:

$$\frac{\partial}{\partial \boldsymbol{\lambda}} \int d\mathbf{x}\,(\mathbf{x}-\boldsymbol{\lambda})\,f(\mathbf{x};\boldsymbol{\lambda}) \;=\; \int d\mathbf{x}\,\left[(\mathbf{x}-\boldsymbol{\lambda})\,(\mathbf{x}-\boldsymbol{\lambda})^{\mathsf{T}}\,\mathbb{B}-\mathbb{I}\right]\cdot f(\mathbf{x};\boldsymbol{\lambda}) \;=\; 0$$

Multivariate Normal Distribution

We should note that the derivative of the probability distribution:

$$\frac{\partial f}{\partial \lambda_i} = [(\mathbf{x} - \boldsymbol{\lambda})^\mathsf{T} \, \mathbb{B}]_i \cdot f(\mathbf{x}; \boldsymbol{\lambda})$$

We can now differentiate the formula for $\mathbb{E}(\mathbf{x} - \boldsymbol{\lambda})$ with respect to $\boldsymbol{\lambda}$:

$$\frac{\partial}{\partial \boldsymbol{\lambda}} \int d\mathbf{x} \, (\mathbf{x} - \boldsymbol{\lambda}) \, f(\mathbf{x}; \boldsymbol{\lambda}) = \int d\mathbf{x} \, [(\mathbf{x} - \boldsymbol{\lambda}) \, (\mathbf{x} - \boldsymbol{\lambda})^\mathsf{T} \, \mathbb{B} - \mathbb{I}] \cdot f(\mathbf{x}; \boldsymbol{\lambda}) = 0$$

and realizing that $\mathbb{B}$ and $\mathbb{I}$ are constant we get

$$\left[\int d\mathbf{x} \, (\mathbf{x} - \boldsymbol{\lambda}) \, (\mathbf{x} - \boldsymbol{\lambda})^\mathsf{T} \cdot f(\mathbf{x}; \boldsymbol{\lambda}) \right] \mathbb{B} = \mathbb{I}$$

Multivariate Normal Distribution

We should note that the derivative of the probability distribution:

$$\frac{\partial f}{\partial \lambda_i} = [(\mathbf{x}-\boldsymbol{\lambda})^{\mathsf{T}} \, \mathbb{B}]_i \cdot f(\mathbf{x};\boldsymbol{\lambda})$$

We can now differentiate the formula for $\mathbb{E}(\mathbf{x}-\boldsymbol{\lambda})$ with respect to $\boldsymbol{\lambda}$:

$$\frac{\partial}{\partial \boldsymbol{\lambda}} \int d\mathbf{x} \, (\mathbf{x}-\boldsymbol{\lambda}) \, f(\mathbf{x};\boldsymbol{\lambda}) = \int d\mathbf{x} \, [(\mathbf{x}-\boldsymbol{\lambda}) \, (\mathbf{x}-\boldsymbol{\lambda})^{\mathsf{T}} \, \mathbb{B} - \mathbb{I}] \cdot f(\mathbf{x};\boldsymbol{\lambda}) = 0$$

and realizing that $\mathbb{B}$ and $\mathbb{I}$ are constant we get

$$\left[\int d\mathbf{x} \, (\mathbf{x}-\boldsymbol{\lambda}) \, (\mathbf{x}-\boldsymbol{\lambda})^{\mathsf{T}} \cdot f(\mathbf{x};\boldsymbol{\lambda}) \right] \mathbb{B} = \mathbb{I}$$

$$\Rightarrow \qquad \mathbb{C}_{\mathbf{x}} \, \mathbb{B} = \mathbb{I} \qquad \Rightarrow \qquad \mathbb{C}_{\mathbf{x}} = \mathbb{B}^{-1}$$

Multivariate Normal Distribution

We can now write the joint probability distribution as:

$$f(\mathbf{x};\boldsymbol{\lambda}) \;=\; A\,\exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\lambda})^{\intercal}\,\mathbb{C}^{-1}\,(\mathbf{x}-\boldsymbol{\lambda})\right]$$

where $\mathbb{C}$ is the covariance matrix of variables $\mathbf{x}$.

Multivariate Normal Distribution

We can now write the joint probability distribution as:

$$f(\mathbf{x};\boldsymbol{\lambda}) \quad = \quad A \, \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\lambda})^{\mathsf{T}} \, \mathbb{C}^{-1} \, (\mathbf{x}-\boldsymbol{\lambda})\right]$$

where $\mathbb{C}$ is the covariance matrix of variables $\mathbf{x}$. Log-likelihood:

$$\ell(\boldsymbol{\lambda};\mathbf{x}) \quad = \quad -\frac{1}{2}(\mathbf{x}-\boldsymbol{\lambda})^{\mathsf{T}} \, \mathbb{C}^{-1} \, (\mathbf{x}-\boldsymbol{\lambda}) \;\; + \; \text{const}$$

$$\frac{\partial \ell}{\partial \lambda_i} \quad = \quad \left[(\mathbf{x}-\boldsymbol{\lambda})^{\mathsf{T}} \, \mathbb{C}^{-1}\right]_i$$

similar to p.d.f. derivative

Multivariate Normal Distribution

We can now write the joint probability distribution as:

$$f(\mathbf{x};\boldsymbol{\lambda}) = A \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\lambda})^{\mathsf{T}}\,\mathbb{C}^{-1}\,(\mathbf{x}-\boldsymbol{\lambda})\right]$$

where $\mathbb{C}$ is the covariance matrix of variables $\mathbf{x}$. Log-likelihood:

$$\ell(\boldsymbol{\lambda};\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\lambda})^{\mathsf{T}}\,\mathbb{C}^{-1}\,(\mathbf{x}-\boldsymbol{\lambda}) + \text{const}$$

$$\frac{\partial\ell}{\partial\lambda_i} = \left[(\mathbf{x}-\boldsymbol{\lambda})^{\mathsf{T}}\,\mathbb{C}^{-1}\right]_i \qquad \text{similar to p.d.f. derivative}$$

$$\frac{\partial^2\ell}{\partial\lambda_i\,\partial\lambda_j} = -\left[\mathbb{C}^{-1}\right]_{ij} \qquad \Rightarrow \qquad \mathbb{C} = \left(-\frac{\partial^2\ell}{\partial\lambda_i\,\partial\lambda_j}\right)^{-1}$$

Parameter covariance matrix

For the considered case of multivariate normal distribution, best parameter estimates $\hat{\boldsymbol{\lambda}}$ are given by the measured variable values $\mathbf{x}$ (single measurement!)

Unlike parameters $\boldsymbol{\lambda}$, parameter estimates $\hat{\boldsymbol{\lambda}}$ are random variables (functions of $\mathbf{x}$ in general) and so we can consider covariance matrix for $\hat{\boldsymbol{\lambda}}$:

$$\mathbb{C}_{\mathbf{x}} = \mathbb{C}_{\hat{\boldsymbol{\lambda}}} = \left(- \frac{\partial^2 \ell}{\partial \lambda_i \, \partial \lambda_j} \right)^{-1}$$

Parameter covariance matrix

For the considered case of multivariate normal distribution, best parameter estimates $\hat{\boldsymbol{\lambda}}$ are given by the measured variable values $\mathbf{x}$ (single measurement!)

Unlike parameters $\boldsymbol{\lambda}$, parameter estimates $\hat{\boldsymbol{\lambda}}$ are random variables (functions of $\mathbf{x}$ in general) and so we can consider covariance matrix for $\hat{\boldsymbol{\lambda}}$:

$$\mathbb{C}_{\mathbf{x}} = \mathbb{C}_{\hat{\boldsymbol{\lambda}}} = \left(- \frac{\partial^2 \ell}{\partial \lambda_i \, \partial \lambda_j} \right)^{-1}$$

Knowing the likelihood function, we can not only estimate parameter values, but also extract uncertainties and correlations of these estimates!

Parameter covariance matrix

For the considered case of multivariate normal distribution, best parameter estimates $\hat{\boldsymbol{\lambda}}$ are given by the measured variable values $\mathbf{x}$ (single measurement!)

Unlike parameters $\boldsymbol{\lambda}$, parameter estimates $\hat{\boldsymbol{\lambda}}$ are random variables (functions of $\mathbf{x}$ in general) and so we can consider covariance matrix for $\hat{\boldsymbol{\lambda}}$:

$$\mathbb{C}_{\mathbf{x}} = \mathbb{C}_{\hat{\boldsymbol{\lambda}}} = \left(- \frac{\partial^2 \ell}{\partial \lambda_i \, \partial \lambda_j} \right)^{-1}$$

Knowing the likelihood function, we can not only estimate parameter values, but also extract uncertainties and correlations of these estimates!

For the uncorrelated parameters (diagonal covariance matrix):

$$\sigma_{\hat{\boldsymbol{\lambda}}_i} = \left(- \frac{\partial^2 \ell}{\partial \lambda_i^2} \right)^{-1/2}$$

Parameter covariance matrix

Considered example was based on the Gaussian distribution.
Standard deviation is one of the parameters of the p.d.f., can be easily extracted from log-likelihood:

$$\sigma_i \quad = \quad \sqrt{\mathbb{C}_{ii}}$$

However, this procedure works only in the Gaussian approximation.
How to define parameter uncertainty in the general case?

Parameter covariance matrix

Considered example was based on the Gaussian distribution.
Standard deviation is one of the parameters of the p.d.f., can be easily extracted from log-likelihood:

$$\sigma_i = \sqrt{\mathbb{C}_{ii}}$$

However, this procedure works only in the Gaussian approximation.
How to define parameter uncertainty in the general case?

For 1-D case we have:

$$\ell(\hat{\lambda} + \sigma; x) = \ln f(x; x + \sigma) = \ell(\hat{\lambda}; x) - \frac{1}{2}$$

Parameter covariance matrix

Considered example was based on the Gaussian distribution.
Standard deviation is one of the parameters of the p.d.f., can be easily extracted from log-likelihood:

$$\sigma_i \quad = \quad \sqrt{\mathbb{C}_{ii}}$$

However, this procedure works only in the Gaussian approximation.
How to define parameter uncertainty in the general case?

For 1-D case we have:

$$\ell(\hat{\lambda} + 2\sigma; x) \quad = \quad \ln f(x; x + 2\sigma) \; = \; \ell(\hat{\lambda}; x) - \frac{4}{2}$$

Parameter covariance matrix

Considered example was based on the Gaussian distribution.
Standard deviation is one of the parameters of the p.d.f., can be easily extracted from log-likelihood:

$$\sigma_i \quad = \quad \sqrt{\mathbb{C}_{ii}}$$

However, this procedure works only in the Gaussian approximation.
How to define parameter uncertainty in the general case?

For 1-D case we have:

$$\ell(\hat{\lambda} + 3\sigma; x) \quad = \quad \ln f(x; x + 3\sigma) \ = \ \ell(\hat{\lambda}; x) - \frac{9}{2}$$

Parameter covariance matrix

Considered example was based on the Gaussian distribution.
Standard deviation is one of the parameters of the p.d.f., can be easily extracted from log-likelihood:

$$\sigma_i = \sqrt{\mathbb{C}_{ii}}$$

However, this procedure works only in the Gaussian approximation.
How to define parameter uncertainty in the general case?

Recipe for a parameter uncertainty

G. Bohm and G. Zech

Standard error intervals of the extracted parameter are defined by the decrease of the log-likelihood function by 0.5 for one, by 2 for two and by 4.5 for three standard deviations.

This definition works for arbitrary p.d.f. shape, also for multiple parameters

Multiple parameter estimate

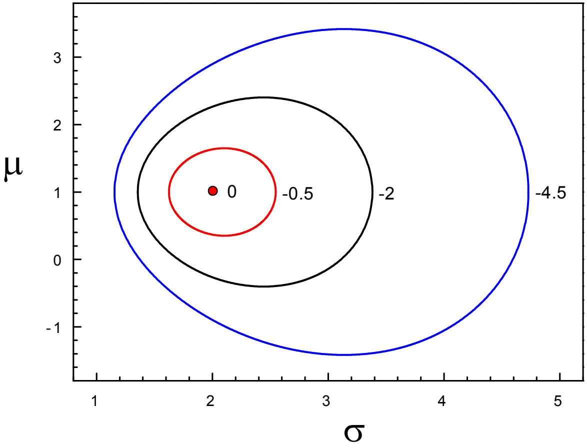
Example log-likelihood function contours

$$\ell(\mu, \sigma; \hat{\mu}, \hat{\sigma})$$

for a sample of 10 events from normal distribution, with extracted parameter values $\hat{\mu} = 1$ and $\hat{\sigma} = 2$.

Figure from:
G. Bohm and G. Zech, Introduction to Statistics and Data Analysis for Physicsts, Verlag Deutsches Elektronen-Synchrotron, 3rd edition

Do we understand the shape of ℓ?



Multiple parameter estimate

For the set **x** of N measurements we can write:

$$
\begin{aligned}
\sum (x_i - \mu)^2 &= \sum (x_i^2 - 2\mu x_i + \mu^2) \\
&= N \left(\langle x^2 \rangle - 2\mu \langle x \rangle + \mu^2 \right)
\end{aligned}
$$

Multiple parameter estimate

For the set **x** of N measurements we can write:

$$\sum (x_i - \mu)^2 = \sum (x_i^2 - 2\mu x_i + \mu^2)$$
$$= N\left(\langle x^2\rangle - 2\mu\langle x\rangle + \mu^2\right)$$
$$= N\left(\hat{\sigma}^2 + \hat{\mu}^2 - 2\mu\hat{\mu} + \mu^2\right) \qquad \hat{\sigma}^2 = \langle x^2\rangle - \langle x\rangle^2$$

Multiple parameter estimate

For the set **x** of N measurements we can write:

$$\begin{aligned}
\sum (x_i - \mu)^2 &= \sum (x_i^2 - 2\mu x_i + \mu^2) \\
&= N\left(\langle x^2 \rangle - 2\mu \langle x \rangle + \mu^2\right) \\
&= N\left(\hat{\sigma}^2 + \hat{\mu}^2 - 2\mu\hat{\mu} + \mu^2\right) \qquad \hat{\sigma}^2 = \langle x^2 \rangle - \langle x \rangle^2
\end{aligned}$$

Log-likelihood function for the example is then:

$$\ell(\mu, \sigma; \hat{\mu}, \hat{\sigma}) = -\frac{1}{2\sigma^2} \sum (x_i - \mu)^2 - \frac{N}{2} \ln(2\pi\sigma^2)$$

Multiple parameter estimate

For the set **x** of N measurements we can write:

$$\begin{aligned} \sum (x_i - \mu)^2 &= \sum (x_i^2 - 2\mu x_i + \mu^2) \\ &= N\left(\langle x^2 \rangle - 2\mu\langle x \rangle + \mu^2\right) \\ &= N\left(\hat{\sigma}^2 + \hat{\mu}^2 - 2\mu\hat{\mu} + \mu^2\right) \qquad \hat{\sigma}^2 = \langle x^2 \rangle - \langle x \rangle^2 \end{aligned}$$

Log-likelihood function for the example is then:

$$\ell(\mu, \sigma; \hat{\mu}, \hat{\sigma}) = -\frac{N}{2\sigma^2}\left(\hat{\sigma}^2 + (\mu - \hat{\mu})^2\right) - \frac{N}{2} \ln(2\pi\sigma^2)$$

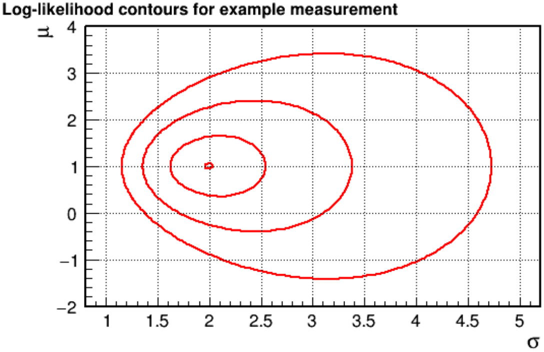
This corresponds to the Gaussian shape for μ, but very asymmetric for σ...

Also, the two parameters are not independent!

Multiple parameter estimate

06_mlm_func.ipynb Open in Colab

Example log-likelihood function contours for a sample of 10 events from normal distribution, with extracted parameter values $\hat{\mu} = 1$ and $\hat{\sigma} = 2$.



Result from G. Bohm and G. Zech is nicely reproduced...

Multiple parameter estimate

As mentioned before, $\hat{\mu}$ and $\hat{\sigma}$ (extracted from measurements $\mathbf{x}$) are random variables! We can calculate their joint probability distribution:

$$f(\hat{\mu}, \hat{\sigma}^2; \mu, \sigma) \quad = \quad A \cdot \exp\left(-\frac{N\,(\hat{\mu}-\mu)^2}{2\,\sigma^2}\right) \cdot \left(\frac{N\hat{\sigma}^2}{\sigma^2}\right)^{k-1} \exp\left(-\lambda \frac{N\hat{\sigma}^2}{\sigma^2}\right)$$

where $\frac{N\hat{\sigma}^2}{\sigma^2}$ is distributed according to the Gamma distribution with $k = (N-1)/2$ and $\lambda = 1/2$ (particular case referred to as χ^2 distribution; we will discuss it at the next lectures).

Multiple parameter estimate

As mentioned before, $\hat{\mu}$ and $\hat{\sigma}$ (extracted from measurements $\mathbf{x}$) are random variables! We can calculate their joint probability distribution:

$$f(\hat{\mu}, \hat{\sigma}^2; \mu, \sigma) \quad = \quad A \cdot \exp\left(-\frac{N\,(\hat{\mu}-\mu)^2}{2\,\sigma^2}\right) \cdot \left(\frac{N\hat{\sigma}^2}{\sigma^2}\right)^{k-1} \exp\left(-\lambda \frac{N\hat{\sigma}^2}{\sigma^2}\right)$$

where $\frac{N\hat{\sigma}^2}{\sigma^2}$ is distributed according to the Gamma distribution with $k = (N-1)/2$ and $\lambda = 1/2$ (particular case referred to as χ^2 distribution; we will discuss it at the next lectures).

The two variables, $\hat{\mu}$ and $\hat{\sigma}$, are independent!

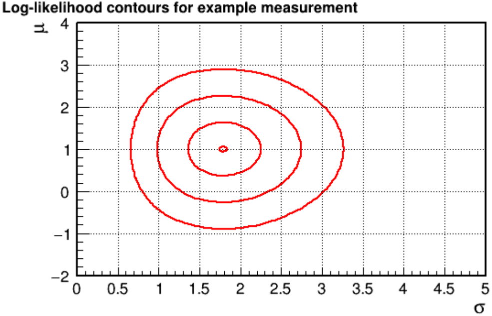
PDF for $\hat{\sigma}$ is asymmetric, but much less than the likelihood function !!!

Multiple parameter estimate

06_mlm_func2.ipynb Open in Colab

One needs to stress that likelihood function for p.d.f. parameters are not equivalent to probability distribution of parameter estimators!

Contours of the joint probability distribution function $f(\hat{\mu}, \hat{\sigma}; \mu, \sigma)$ for $\mu = 1$ and $\sigma = 2$



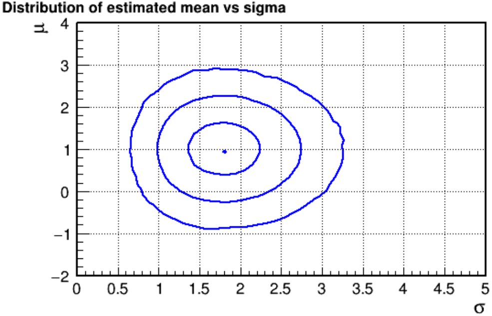
Contours corresponding to log f decrease by 0.5, 2 and 4.5 from maximum...

Multiple parameter estimate

06_mlm_mc_2d.ipynb Open in Colab

One needs to stress that likelihood function for p.d.f. parameters are not equivalent to probability distribution of parameter estimators!

Results of Monte Carlo simulation (contours from 1 000 000 experiments)



In each experiment, mean and sigma are calculated from 10 generated numbers

Parameter Inference

Presenting measurement results

When doing the measurement, we usually quote the final result as numerical value (with units) and estimated uncertainty:

$$x \pm \sigma_x$$

We can often calculate the uncertainty from the data itself (eg. when result is obtained by averaging a large number of independent measurements) or from the variation of the log-likelihood function.

Attributing proper uncertainty to the result is crucial!

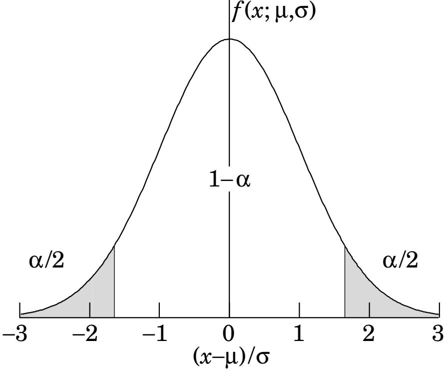
But what does it tell as after all?!

Normal distribution

Meaning of σ is well defined for Gaussian distribution.

Probability for the experimental result to be consistent with the true value within $\pm N\,\sigma$:

		$1-\alpha$	
$\pm 1\,\sigma$	$\Rightarrow$	68.27	%
$\pm 2\,\sigma$	$\Rightarrow$	95.45	%
$\pm 3\,\sigma$	$\Rightarrow$	99.73	%
$\pm 4\,\sigma$	$\Rightarrow$	99.9937	%
$\pm 5\,\sigma$	$\Rightarrow$	99.999943	%



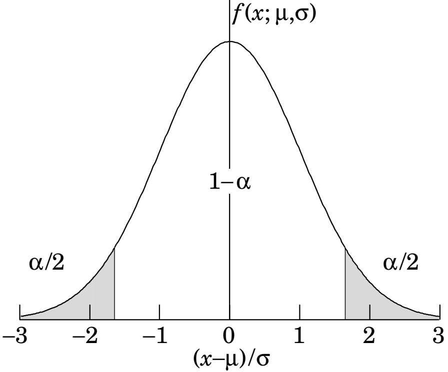
There is a non-zero chance for deviation grater than 5σ, but it is extremely small

Normal distribution

Meaning of σ is well defined for Gaussian distribution.

Probability for the experimental result to differ from the true value by more than $N\sigma$:

		α	
$\pm 1\ \sigma$	$\Rightarrow$	31.73	%
$\pm 2\ \sigma$	$\Rightarrow$	4.55	%
$\pm 3\ \sigma$	$\Rightarrow$	0.27	%
$\pm 4\ \sigma$	$\Rightarrow$	0.0063	%
$\pm 5\ \sigma$	$\Rightarrow$	0.000057	%

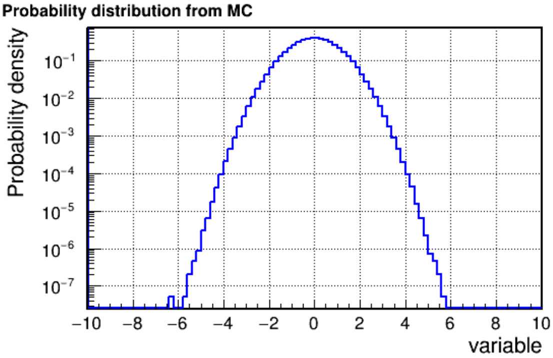


Fluctuations up and down are observed with equal probability...

Normal distribution

06_interval.ipynb   Open in Colab

Results of the Monte Carlo test
($\mu = 0$, $\sigma = 1$, 100 000 000 generations)



Down fluctuations:

$< -1\,\sigma$ p = 0.15864225
$< -2\,\sigma$ p = 0.0227686
$< -3\,\sigma$ p = 0.00134872
$< -4\,\sigma$ p = 3.204E-05
$< -5\,\sigma$ p = 3.2E-07

Fluctuations up:

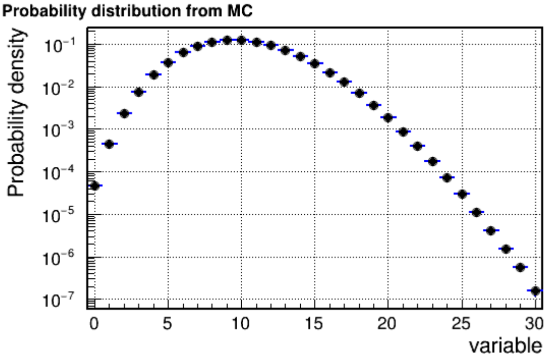
$> 1\,\sigma$ p = 0.15861607
$> 2\,\sigma$ p = 0.02275479
$> 3\,\sigma$ p = 0.00135162
$> 4\,\sigma$ p = 3.263E-05
$> 5\,\sigma$ p = 2.8E-07

Good agreement with expectations

Poisson distribution

06_interval_2.ipynb Open in Colab

Results of the Monte Carlo test
($\mu = 10$, 100 000 000 generations)



Down fluctuations:

$< -1\,\sigma$	p = 0.1301325
$< -2\,\sigma$	p = 0.01033324
$< -3\,\sigma$	p = 4.55E-05
$< -4\,\sigma$	p = 0
$< -5\,\sigma$	p = 0

Fluctuations up:

$> 1\,\sigma$	p = 0.13553469
$> 2\,\sigma$	p = 0.02702754
$> 3\,\sigma$	p = 0.00344483
$> 4\,\sigma$	p = 0.00029364
$> 5\,\sigma$	p = 1.774e-05

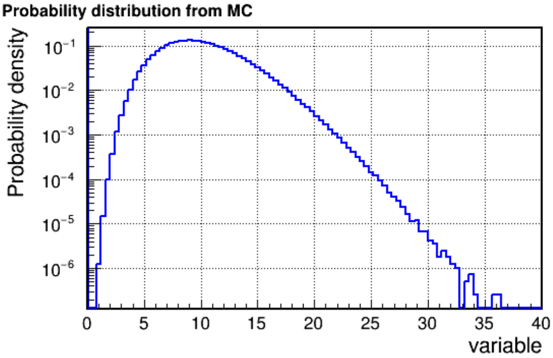
Much longer tail of positive fluctuations!

Gamma distribution

06_interval_3.ipynb Open in Colab

Results of the Monte Carlo test

($\mu = \sigma^2 = 10$, 10 000 000 generations)



Down fluctuations:

$< -1\,\sigma$	p = 0.1533757
$< -2\,\sigma$	p = 0.0046445
$< -3\,\sigma$	p = 0
$< -4\,\sigma$	p = 0
$< -5\,\sigma$	p = 0

Fluctuations up:

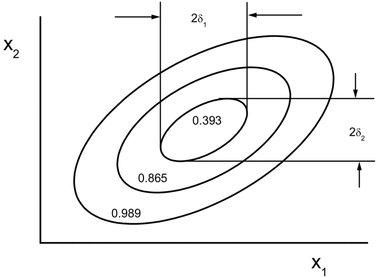
$> 1\,\sigma$	p = 0.1554444
$> 2\,\sigma$	p = 0.0368358
$> 3\,\sigma$	p = 0.0067406
$> 4\,\sigma$	p = 0.0010105
$> 5\,\sigma$	p = 0.0001305

Even longer, 5σ fluctuations not excluded

Normal distribution in N-D

It is also important to notice that the fractions presented previously (eg. 68% within $\pm 1\sigma$) refer to one-dimensional normal distribution only!

If we consider 2-D distribution



Fractions within $N\sigma$ contours:

Deviation	Dimension			
	1	2	3	4
1 σ	0.683	0.393	0.199	0.090
2 σ	0.954	0.865	0.739	0.594
3 σ	0.997	0.989	0.971	0.939
4 σ	1.	1.	0.999	0.997

1σ fraction above 50% only for N=1 !

G. Bohm and G. Zech

Less than 40% is contained inside 1σ contour...

Interpreting results

As demonstrated above, quoting numerical result with uncertainties gives only partial information on the measurement...

It is sufficient, if we can assume normal distribution of the variable.

Also, we need to assume that the width of the distribution does not depend on the measured parameter. Only then the likelihood function will be Gaussian as well...

Interpreting results

As demonstrated above, quoting numerical result with uncertainties gives only partial information on the measurement...

It is sufficient, if we can assume normal distribution of the variable.

Also, we need to assume that the width of the distribution does not depend on the measured parameter. Only then the likelihood function will be Gaussian as well...

In the general case, parameter uncertainty does not give us full information on the shape (in particular the tails) of the distribution...

How should we present results of the experiment, if we are more concerned about the probability of (large) result fluctuations?...

Interpreting results

There is also another problem, which has to be noticed!

So far we have only considered distribution of experimental results for given probability distribution, $f(\mathbf{x};\boldsymbol{\lambda})$, when the parameter values $\boldsymbol{\lambda}$ are known.

Interpreting results

There is also another problem, which has to be noticed!

So far we have only considered distribution of experimental results for given probability distribution, $f(\mathbf{x}; \boldsymbol{\lambda})$, when the parameter values $\boldsymbol{\lambda}$ are known.

The actual situation is usually different: for given set of measurements $\mathbf{x}$ we extract estimates of the parameter values $\hat{\boldsymbol{\lambda}}$.

Uncertainties estimated from log-likelihood variation indicate the expected level of agreement (in Gaussian approximation) between our estimate $\hat{\boldsymbol{\lambda}}$ and the true parameter values $\boldsymbol{\lambda}$.

Can we present measurement results in a way which gives us more precise information about the possible fluctuations in the estimate $\hat{\boldsymbol{\lambda}}$?

Interpreting results

There is also another problem, which has to be noticed!

So far we have only considered distribution of experimental results for given probability distribution, $f(\mathbf{x}; \boldsymbol{\lambda})$, when the parameter values $\boldsymbol{\lambda}$ are known.

The actual situation is usually different: for given set of measurements $\mathbf{x}$ we extract estimates of the parameter values $\hat{\boldsymbol{\lambda}}$.

Uncertainties estimated from log-likelihood variation indicate the expected level of agreement (in Gaussian approximation) between our estimate $\hat{\boldsymbol{\lambda}}$ and the true parameter values $\boldsymbol{\lambda}$.

Can we present measurement results in a way which gives us more precise information about the possible fluctuations in the estimate $\hat{\boldsymbol{\lambda}}$?

Yes, but we need to define the problem differently... We should not consider probability of $\hat{\boldsymbol{\lambda}}$...

Frequentist confidence intervals

Classical (frequentist) definition of the confidence interval refers directly to the probability distribution of the experimental results, $f(\mathbf{x}; \boldsymbol{\lambda})$.

For given outcome of the experiment x_m, $1-\alpha$ confidence level (C.L.) interval for parameter λ is $[\lambda_1, \lambda_2]$, if for all values $\lambda' \in [\lambda_1, \lambda_2]$, our result x_m is inside the corresponding $1-\alpha$ probability interval for $f(x; \lambda')$.

Frequentist confidence intervals

Classical (frequentist) definition of the confidence interval refers directly to the probability distribution of the experimental results, $f(\mathbf{x};\boldsymbol{\lambda})$.

For given outcome of the experiment x_m, $1-\alpha$ confidence level (C.L.) interval for parameter λ is $[\lambda_1,\lambda_2]$, if for all values $\lambda' \in [\lambda_1,\lambda_2]$, our result x_m is inside the corresponding $1-\alpha$ probability interval for $f(x;\lambda')$.

This definition clearly depends on the way we define probability intervals for $f(x;\lambda')$ - it is rather a concept, more assumptions are needed.

We always refer to probability distribution for x!

Frequentist confidence intervals

It is interesting to note that the modern concept of confidence intervals was proposed by Polish statistician Jerzy Neyman only in 1937

Jerzy Neyman obtained his PhD (1924) and habilitation (1928) at UW

X—Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability

By J. Neyman

Reader in Statistics, University College, London

(Communicated by H. Jeffreys, F.R.S.—Received 20 November, 1936—Read 17 June, 1937)

Contents

J.Neyman, Phil. Trans. Royal Soc. London, Series A, 236 333-80 (1937).

Frequentist confidence intervals

As mentioned above, to define confidence interval for parameter, we need to define how the probability interval for our measurement is defined.

There are three “natural” choices:

- We constrain the measurement from above:

$$\int_{x_{ul}}^{+\infty} dx \; f(x;\lambda) \;\; = \;\; \alpha$$

- We constrain the measurement from below:

$$\int_{-\infty}^{x_{ll}} dx \; f(x;\lambda) \;\; = \;\; \alpha$$

- We use central probability interval: as presented for Gaussian pdf

$$\int_{-\infty}^{x_1} dx \; f(x;\lambda) \; = \; \alpha/2 \quad \text{and} \quad \int_{x_2}^{+\infty} dx \; f(x;\lambda) \; = \; \alpha/2$$

Frequentist confidence intervals

Let us consider the simplest possible example, Gaussian pdf:

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

assuming σ is known (and fixed). What is the central 90% C.L. interval for λ?

Frequentist confidence intervals

Let us consider the simplest possible example, Gaussian pdf:

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

assuming σ is known (and fixed). What is the central 90% C.L. interval for λ?

From the Gaussian pdf properties we can directly obtain:

$$x_1 = \lambda - 1.64\sigma \qquad x_2 = \lambda + 1.64\sigma$$

Frequentist confidence intervals

Let us consider the simplest possible example, Gaussian pdf:

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

assuming σ is known (and fixed). What is the central 90% C.L. interval for λ?

From the Gaussian pdf properties we can directly obtain:

$$x_1 = \lambda - 1.64\sigma \qquad x_2 = \lambda + 1.64\sigma$$

Definition of the confidence interval for λ is based on the condition: we measure $x = x_m$

$$x_1 \;<\; x_m \;<\; x_2$$

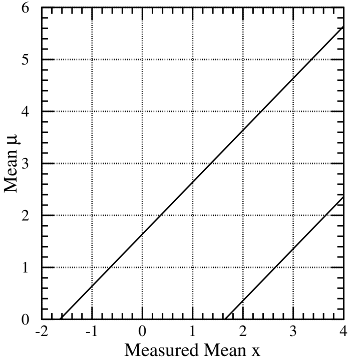
Which is fulfilled for all λ in range:

$$\lambda_1 \;=\; x_m - 1.64\sigma \;<\; \lambda \;<\; x_m + 1.64\sigma \;=\; \lambda_2$$

So the central 90% C.L. interval for λ is $[x_m - 1.64\sigma, x_m + 1.64\sigma]$...

Frequentist confidence intervals

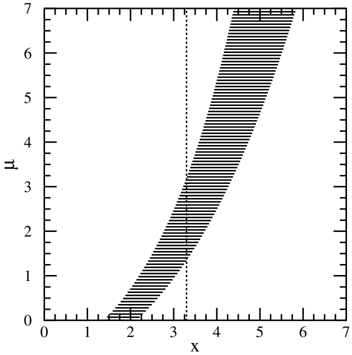
Graphical presentation of the procedure for $\sigma = 1$



G.J.Feldman, R.D.Cousins,
A Unified Approach to the Classical Statistical Analysis of Small Signals,
Phys.Rev.D57:3873-3889,1998; arXiv:physics/9711021

Frequentist confidence intervals

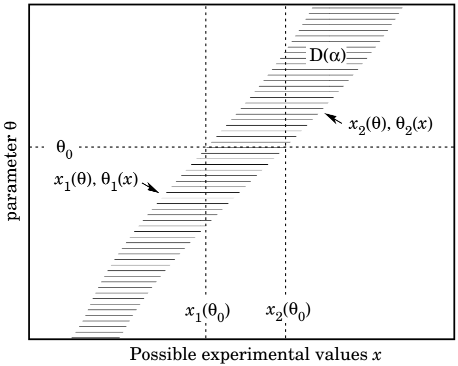
Graphical presentation of the general procedure



G.J.Feldman, R.D.Cousins,
A Unified Approach to the Classical Statistical Analysis of Small Signals,
Phys.Rev.D57:3873-3889,1998; arXiv:physics/9711021

Frequentist confidence intervals

General procedure

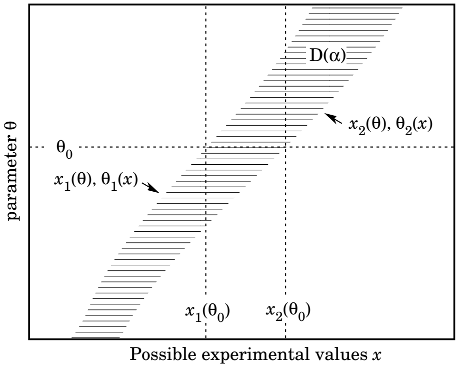


- calculate limits of probability intervals for x, $x_1(\theta)$ and $x_2(\theta)$, for different values of θ

R.L. Workman et al. (Particle Data Group), Prog. Theor. Exp. Phys. 2022, 083C01 (2022), PDG web page

Frequentist confidence intervals

General procedure

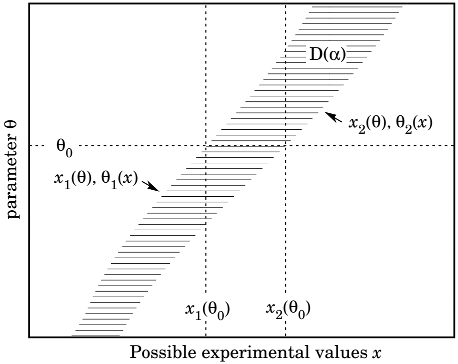


- calculate limits of probability intervals for x, $x_1(\theta)$ and $x_2(\theta)$, for different values of θ
- calculated intervals define the "accepted region" in (θ, x)

R.L. Workman et al. (Particle Data Group), Prog. Theor. Exp. Phys. 2022, 083C01 (2022), PDG web page

Frequentist confidence intervals

General procedure

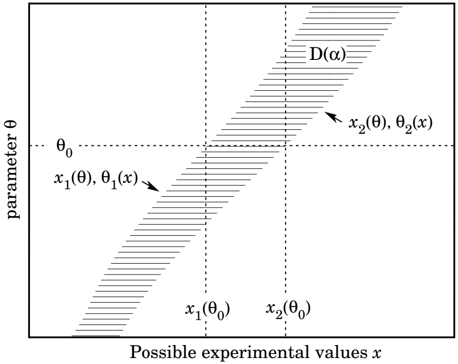


- calculate limits of probability intervals for x, $x_1(\theta)$ and $x_2(\theta)$, for different values of θ
- calculated intervals define the "accepted region" in (θ, x)
- confidence interval for θ is defined by drawing line $x = x_m$ in the accepted region

R.L. Workman et al. (Particle Data Group), Prog. Theor. Exp. Phys. 2022, 083C01 (2022), PDG web page

Frequentist confidence intervals

General procedure



- calculate limits of probability intervals for x, $x_1(\theta)$ and $x_2(\theta)$, for different values of θ
- calculated intervals define the "accepted region" in (θ, x)
- confidence interval for θ is defined by drawing line $x = x_m$ in the accepted region

$\Rightarrow$ limit on θ for given x_m, $\theta_1(x_m)$, corresponds to limit on x for given θ: $x_m = x_1(\theta_1)$.

R.L. Workman et al. (Particle Data Group), Prog. Theor. Exp. Phys. 2022, 083C01 (2022), PDG web page

Frequentist limits

The procedure is more unique, if we want to constrain the parameter from above or below. We first define an upper or lower limit for the measured value x, for given λ (at $1-\alpha$ CL):

$$\int_{x_{ul}(\lambda)}^{+\infty} dx\ f(x;\lambda) \ = \ \alpha \quad \text{or} \quad \int_{-\infty}^{x_{ll}(\lambda)} dx\ f(x;\lambda) \ = \ \alpha$$

Frequentist limits

The procedure is more unique, if we want to constrain the parameter from above or below. We first define an upper or lower limit for the measured value x, for given λ (at $1-\alpha$ CL):

$$\int_{x_{ul}(\lambda)}^{+\infty} dx\, f(x;\lambda) \;=\; \alpha \quad \text{or} \quad \int_{-\infty}^{x_{ll}(\lambda)} dx\, f(x;\lambda) \;=\; \alpha$$

Assuming that $\langle \mathbf{x} \rangle$ increases with $\boldsymbol{\lambda}$, and the measurement resulted in value x_m:

- Upper limit on parameter λ can be defined by the condition: $x_{ll}(\lambda_{ul}) = x_m$

For $\lambda > \lambda_{ul}$, the probability that the experiment result is not larger than x_m is less than α.

$\Rightarrow$ We state that values $\lambda > \lambda_{ul}$ are excluded at $(1-\alpha)$ confidence level (CL)

Frequentist limits

The procedure is more unique, if we want to constrain the parameter from above or below. We first define an upper or lower limit for the measured value x, for given λ (at $1-\alpha$ CL):

$$\int_{x_{ul}(\lambda)}^{+\infty} dx\, f(x;\lambda) = \alpha \quad \text{or} \quad \int_{-\infty}^{x_{ll}(\lambda)} dx\, f(x;\lambda) = \alpha$$

Assuming that $\langle \mathbf{x} \rangle$ increases with $\boldsymbol{\lambda}$, and the measurement resulted in value x_m:

- Upper limit on parameter λ can be defined by the condition: $x_{ll}(\lambda_{ul}) = x_m$

For $\lambda > \lambda_{ul}$, the probability that the experiment result is not larger than x_m is less than α.
$\Rightarrow$ We state that values $\lambda > \lambda_{ul}$ are excluded at $(1-\alpha)$ confidence level (CL)

- Lower limit on parameter λ can be defined by the condition: $x_{ul}(\lambda_{ll}) = x_m$

Parameter Inference

Frequentist limits on the quark radius

Physics Letters B 757 (2016) 468–472

Contents lists available at ScienceDirect

Physics Letters B

www.elsevier.com/locate/physletb

Limits on the effective quark radius from inclusive *ep* scattering at HERA

ZEUS Collaboration

H. Abramowicz [y,31], I. Abt [t], L. Adamczyk [h], M. Adamus [ae], S. Antonelli [b], V. Aushev [q], O. Behnke [j], U. Behrens [j], A. Bertolin [v], S. Bhadra [ag], I. Bloch [k], E.G. Boos [o], I. Brock [c], N.H. Brook [ac], R. Brugnera [w], A. Bruni [a], P.J. Bussey [l], A. Caldwell [t], M. Capua [e], C.D. Catterall [ag], J. Chwastowski [g], J. Ciborowski [ad,33], R. Ciesielski [j,16], A.M. Cooper-Sarkar [u], M. Corradi [a,11], R.K. Dementiev [s], R.C.E. Devenish [u], S. Dusini [v], B. Foster [m,23], G. Gach [h], E. Gallo [m,24], A. Garfagnini [w], A. Geiser [j], A. Gizhko [j], L.K. Gladilin [s], Yu.A. Golubkov [s], G. Grzelak [ad], M. Guzik [h], C. Gwenlan [u], W. Hain [j], O. Hlushchenko [q], D. Hochman [af], R. Hori [n], Z.A. Ibrahim [f], Y. Iga [x], M. Ishitsuka [z], F. Januschek [j,17], N.Z. Jomhari [f], I. Kadenko [q], S. Kananov [y], U. Karshon [af], P. Kaur [d,12], D. Kisielewska [h], R. Klanner [m], U. Klein [j,18],

HERA electron(positron)-proton collider at DESY

e^+/e^- p

27.5 GeV

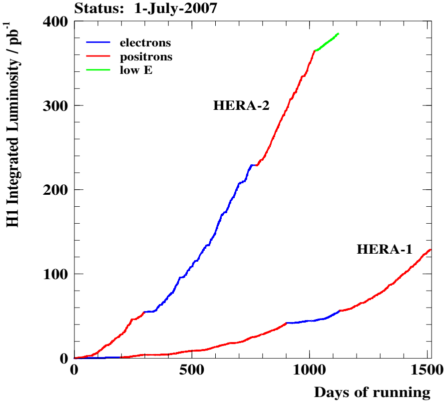
820 GeV 1994–1997
920 GeV 1998–2007

HERA I 1994-2000

about $100 pb^{-1}$ collected per experiment
mainly e^+p data, unpolarised

HERA II 2002-2007

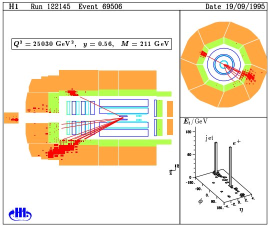
about $400 pb^{-1}$ per experiment
similar amount of e^-p and e^+p data
with longitudinal polarization of $e^\pm$ beams (30-40%)
and small samples collected at reduced proton beam energy



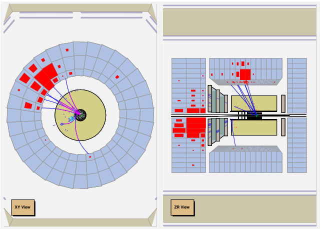
Deep Inelastic $e^{\pm}p$ Scattering

Main process studied by H1 and ZEUS

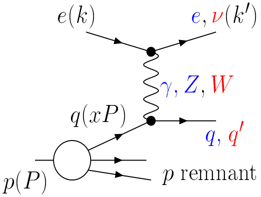
NC DIS



CC DIS



Kinematic variables:



$$Q^2 = -(k - k')^2$$

|virtuality| of the exchanged boson

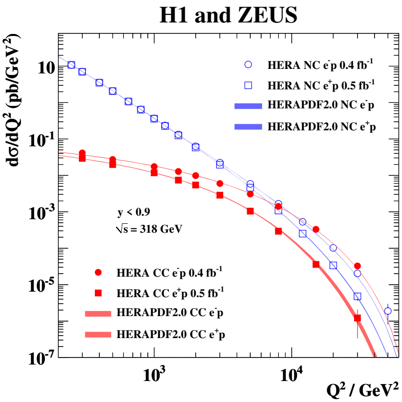
$$x = \frac{Q^2}{2P \cdot (k - k')}$$

fraction of proton momenta carried by stuck quark

$$y = \frac{P \cdot (k - k')}{P \cdot k}$$

fraction of lepton energy transfered in the proton rest frame

SM predictions vs measurements from HERA



NC and CC DIS cross sections comparable for the highest Q^2 values

$$Q^2 \sim M_Z^2, M_W^2$$

Combined QCD+EW analysis shows good agreement with SM predictions

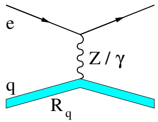
Phys. Rev. D 93 (2016) 092002, arXiv:1603.09628

High precision data could also be used to look for possible BSM effects...

Quark form factor

"classical" method to look for possible fermion (sub)structure.

If a quark has finite size, the standard model cross-section is expected to decrease at high momentum transfer:



$$\frac{d\sigma}{dQ^2} = \frac{d\sigma^{SM}}{dQ^2} \cdot \left[1 - \frac{R_q^2}{6} Q^2\right]^2 \cdot \left[1 - \frac{R_e^2}{6} Q^2\right]^2$$

where R_q is the root mean-square radius of the electroweak charge distribution in the quark.

We do not consider the possibility of finite electron size...

same dependence expected for e^+p and e^-p !

Data model

Model description extended to take into account the possible BSM contributions

$$-2\,\ell\,(\boldsymbol{p},\boldsymbol{s},\eta) = \sum_i \frac{\left[m^i(p,\eta) + \sum_j s_j \gamma^i_j m^i(p,\eta) - \mu^i_0\right]^2}{\left(\delta^2_{i,\mathrm{stat}} + \delta^2_{i,\mathrm{uncor}}\right)\,(\mu^i_0)^2} + \sum_j s_j^2$$

$\boldsymbol{p}$ and $\boldsymbol{s}$ are vectors of PDF parameters p_k (describe proton structure) and systematic shifts s_j, η is the parameter describing BSM contribution (here: $\eta = R^2_q$)

$\Rightarrow$ we look for global likelihood maximum for the combined HERA data

$$R^{2\ \mathrm{Data}}_q \quad = \quad -0.2 \cdot 10^{-33}\ \mathrm{cm}^2$$

This one number summarizes the results of all our measurements...

μ^i_0 and $m^i(\boldsymbol{p},\eta)$ are measured and predicted (SM+BSM) cross sections,
γ^i_j, $\delta_{i,\mathrm{stat}}$ and $\delta_{i,\mathrm{uncor}}$ are the relative correlated systematic, relative statistical and relative uncorrelated systematic uncertainties of the input data point i

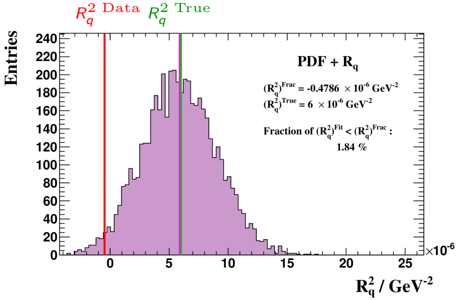
Limit setting

Limits derived using the technique of "MC replicas" (a must for frequentist approach).

Replicas are generated sets of cross-section values that are calculated for given $R_q^{2\ \mathrm{True}}$ and varied randomly according to the statistical and systematic uncertainties (including correlations) of the input data.

Each replica is then used as an input to SM+BSM fit $\Rightarrow R_q^{2\ \mathrm{Fit}}$

Many replicas for each considered $R_q^{2\ \mathrm{True}}$ value $\Rightarrow$ distribution of $R_q^{2\ \mathrm{Fit}}$



$R_q^{2\ \mathrm{True}}$ is tested by comparing $R_q^{2\ \mathrm{Fit}}$ distribution with the value of $R_q^{2\ \mathrm{Data}}$

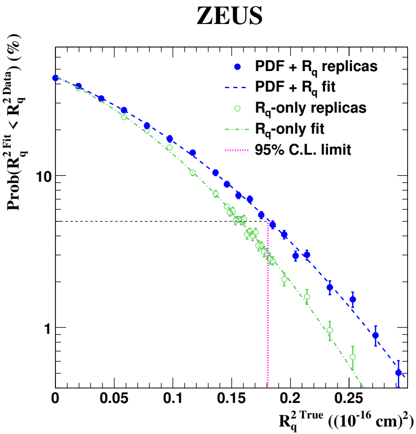
Limit setting

The probability of obtaining a $R_q^{2\,\mathrm{Fit}}$ value smaller than that obtained for the actual data

$$\mathrm{Prob}(R_q^{2\,\mathrm{Fit}} < R_q^{2\,\mathrm{Data}})$$

is studied as a function of $R_q^{2\,\mathrm{True}}$

$R_q^{2\,\mathrm{True}}$ values corresponding to the probability smaller than 5% are excluded at the 95% C.L.

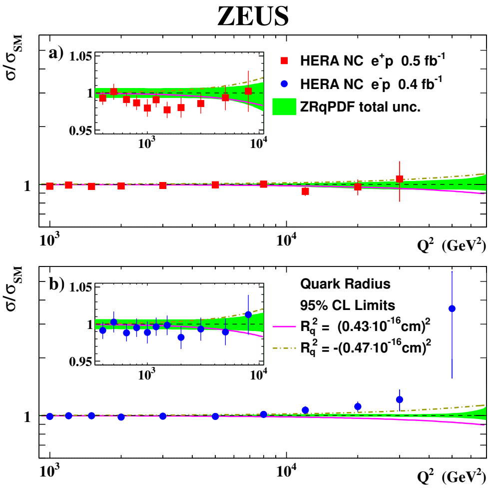


$$R_q < 0.43 \cdot 10^{-16}\mathrm{cm}$$

Quark radius limits

Results

Phys. Lett. B757 (2016) 468, arXiv:1604.01280



$$-(0.47 \cdot 10^{-16}\mathrm{cm})^2 < R_q^2 < (0.43 \cdot 10^{-16}\mathrm{cm})^2$$

Parameter Inference

Homework

Solutions to be uploaded by November 30.

Consider the example discussed today:

sample of 10 measurements from the normal distributions $G(\mu, \sigma)$.

Assuming that the measured variance of the sample is $\hat{\sigma} = 2$:

- calculate the 95% CL upper limit on the true variance of the distribution, σ, based on the properties of the Gamma distribution (SciPy library).
- verify the result by the means of the Monte Carlo simulation.