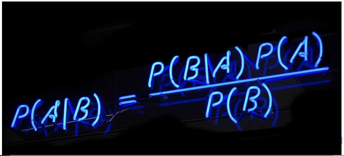
Statistics in Data Analysis

All you ever wanted to know about statistics but never dared to ask

part 5

Pawel Brückman de Renstrom
(`pawel.bruckman@ifj.edu.pl`)

April 10, 2024



Question from the previous lecture

Suppose a beam of particles is known to consist of charged pions and muons. For each particle in the beam we measure a variable t, whose distribution for pions (π) and muons (μ) is

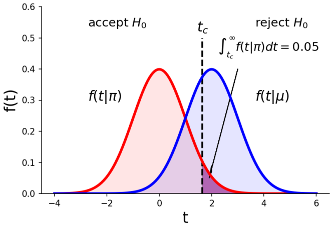
$$f(t;\pi) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(t-\mu_\pi)^2/2\sigma^2}, \qquad f(t;\mu) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(t-\mu_\mu)^2/2\sigma^2},$$

where $\mu_\pi = 0$, $\mu_\mu = 2$ and $\sigma = 1$. For each particle we want to test the hypothesis H_0 that it is a pion against the alternative H_1 that it is a muon. The critical region of the test is given by $t > t_c$ where t_c is a given constant.

1. Suppose we want to have a test of size $\alpha = 0.05$. Illustrate where the critical region lies and what α means on a sketch of the p.d.f.s $f(t|\pi)$ and $f(t|\mu)$ and show that t_c is numerically about 1.64.
2. Suppose a sample of particles is known to consist of 99% pions and 1% muons. What is the purity of the muon sample selected by $t > t_c$? Here, purity means the probability to be a muon given that the particle had $t > t_c$ (i.e., it was rejected as a pion and thus selected as a muon candidate).

Solution to be sent to me before the next lecture

Solution



1 $\alpha = 0.05$ (test size) means that:

$$1 - F(t_c; 0.0, 1.0) = \int_{t_c}^{\infty} f(t|\pi) = 0.05$$

$$1 - erf(t_c) = 0.05 \implies t_c = erf^{-1}(0.95) = 1.64485...$$

2 For the second part, we need to calculate:

$$1 - \beta = 1 - F(t_c; 2.0, 1.0) = \int_{t_c}^{\infty} f(t|\mu) = 0.6388...$$

$$N_\pi = 0.99 * N, \qquad N_\mu = 0.01 * N$$

$$\text{purity} = \frac{(1-\beta)N_\mu}{(1-\beta)N_\mu + \alpha N_\pi} \simeq \frac{0.01 * 0.639}{0.01 * 0.639 + 0.99 * 0.05} = 0.114...$$

The choice of the critical point t_c (selection working point) results in muon selection efficiency of 64% with purity (given the beam composition) of 11.4% .

Maximum Likelihood estimator - reminder

Let $f(x;\theta)$ is a p.d.f. of a known form but unknown parameter θ (more generally $\boldsymbol{\theta} = (\theta_1, ...\theta_m)$). Let $x_1, x_2, ..., x_n$ be a sample of n events drawn from the above p.d.f. Generally, $\boldsymbol{x_i}$ may be a multidimensional vector. We define:

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) \tag{1}$$

called the **likelihood function**.

- L is technically a joint p.d.f. of x but, assuming a fixed data sample, represents a function of θ.
- The **maximum likelihood** (ML) estimator $\widehat{\boldsymbol{\theta}}$ is given by:

$$\frac{\partial L}{\partial \theta_i} = 0, \qquad i = 1, ..., m. \tag{2}$$

- **Log-likelihood function** is commonly used:

$$\log L(\theta) = \sum_{i=1}^{n} \ln f(x_i;\theta) \tag{3}$$

Variance of ML estimator

RCF bound

Rao-Cramér-Frechet (RCF) inequality states that:

$$V[\widehat{\theta}] \geq \left(1 + \frac{\partial b}{\partial \theta}\right)^2 \Bigg/ E\left[-\frac{\partial^2 \log L}{\partial \theta^2}\right]. \quad (4)$$

In case of equality (i.e. minimum variance) the estimator is said to be **efficient**.

- E.g., $\widehat{\tau} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is an efficient estimator for the parameter τ. (show!)
- ML are always efficient in the large sample limit!

Assuming efficiency and zero bias, for a general case when $\boldsymbol{\theta} = (\theta_1, ...\theta_m)$ we get:

$$\left(V^{-1}\right)_{ij} = E\left[-\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j}\right] \quad \longrightarrow \quad \left(\widehat{V^{-1}}\right)_{ij} = -\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j}\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \quad (5)$$

For a single parameter θ this reduces to: $\widehat{\sigma^2}_{\widehat{\theta}} = \left(-1 \Big/ \frac{\partial^2 \log L}{\partial \theta^2}\right)\Bigg|_{\theta=\widehat{\theta}}$

Variance of ML estimator

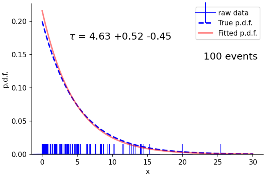
Taylor expansion

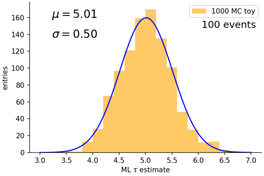
Let's take a single parameter θ and expand log-likekihood in Taylor series about the ML estimate:

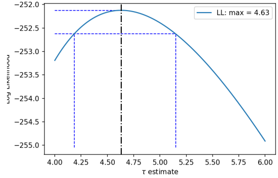
$$\log L(\theta) = \log L(\widehat{\theta}) + \left[-\frac{\partial \log L}{\partial \theta} \right] \Bigg|_{\theta=\widehat{\theta}} (\theta - \widehat{\theta}) + \frac{1}{2!} \left[\frac{\partial^2 \log L}{\partial \theta^2} \right] \Bigg|_{\theta=\widehat{\theta}} (\theta - \widehat{\theta})^2 + \ldots \qquad (6)$$

By definition, $\log L(\widehat{\theta}) = \log L_{\max}$ and $\frac{\partial \log L}{\partial \theta}\Big|_{\theta=\widehat{\theta}} = 0$, and so:

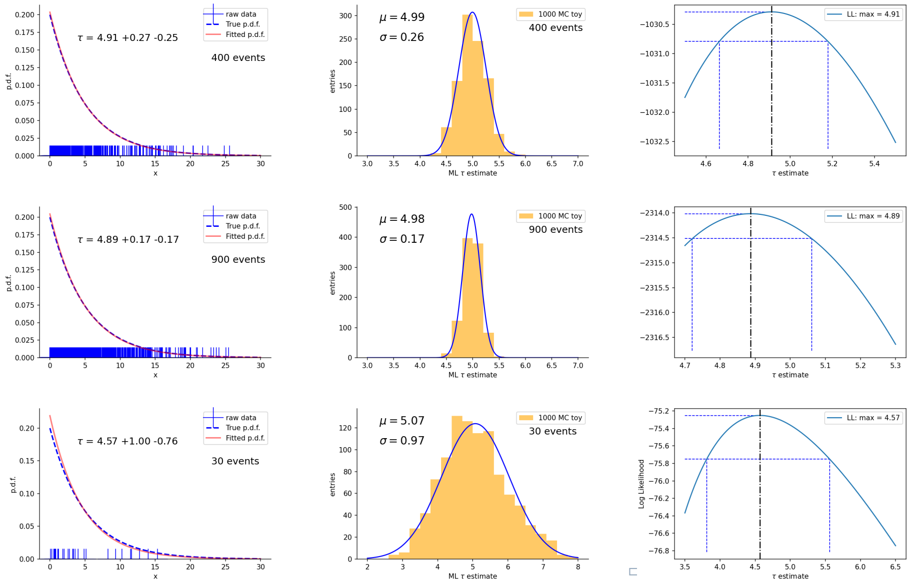
$$\log L(\theta) = \log L_{\max} - \frac{(\theta - \widehat{\theta})^2}{2\widehat{\sigma^2}_{\widehat{\theta}}} \quad \text{or} \quad \log L(\theta \pm \widehat{\sigma}_{\widehat{\theta}}) = \log L_{\max} - \frac{1}{2}. \qquad (7)$$







Variance of ML estimator



Extended Maximum Likelihood

Let $f(x;\theta)$ is a p.d.f. of a known form but unknown parameter θ (more generally $\boldsymbol{\theta} = (\theta_1, ...\theta_m)$). Let $x_1, x_2, ..., x_n$ be a sample of n events drawn from the above p.d.f. Now, let us treat n as a Poisson random variable with mean ν.

- The likelihood takes the form:

$$L(\boldsymbol{\theta}) = \frac{\nu^n}{n!} e^{-\nu} \prod_{i=1}^{n} f(x_i;\boldsymbol{\theta}) = \frac{e^{-\nu}}{n!} \prod_{i=1}^{n} \nu f(x_i;\boldsymbol{\theta}) \tag{8}$$

  called the **extended likelihood function**.

- If ν is a function of $\boldsymbol{\theta}$ then log-likelihood is:

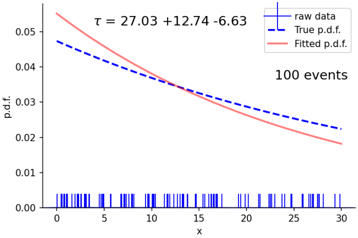
$$\log L(\boldsymbol{\theta}) = n \ln \nu(\boldsymbol{\theta}) - \nu(\boldsymbol{\theta}) + \sum_{i=1}^{n} \ln f(x_i;\theta) = -\nu(\boldsymbol{\theta}) + \sum_{i=1}^{n} \ln\left(\nu(\boldsymbol{\theta}) f(x_i;\theta)\right), \tag{9}$$

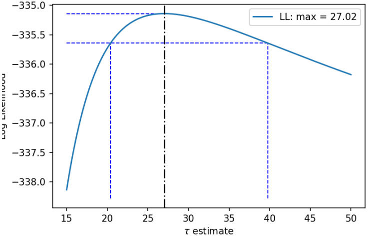
  where terms not depending on $\boldsymbol{\theta}$ have been dropped.

Extended Maximum Likelihood

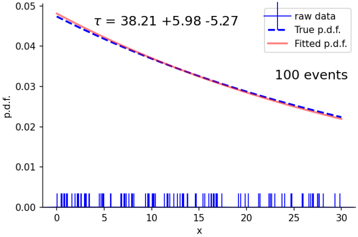
mean lifetime: $\tau = 40$, $\nu = \nu_0(1 - e^{-T/\tau})$

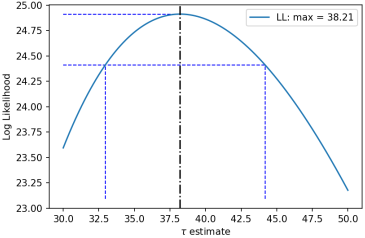
$\log L = \sum \ln f(x_i)$





$\log L = -\nu + \sum \ln(\nu f(x_i))$

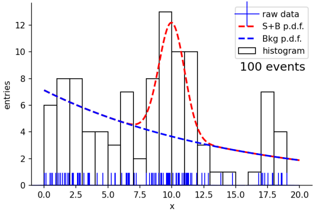




Extended Maximum Likelihood

sample composition

In a trivial case when ν is independent from $\boldsymbol{\theta}$, the derivative of $\log L$ w.r.t. ν gives the estimator $\widehat{\nu} = n$. (show!)
However, it is often the case that different classes of events with known p.d.f.'s contribute to the observed distribution and our task is to estimate yields of the individual components.



In a general case of having m contributions, a $\log L$ can be defined as:

$$\log L(\boldsymbol{\mu}) = -\sum_{j=1}^{m} \mu_j + \sum_{i=1}^{n} \ln\left(\sum_{j=1}^{m} \mu_j f_j(x_i)\right), \tag{10}$$

where the vector $\boldsymbol{\mu} = (\mu_1, ..., \mu_m)$ represents directly the yields of individual contributions. Of course, the fit is capable of estimating all μ_j parameters only if the corresponding p.d.f.'s are different. In general, such a fit results in correlated estimates.

Binned Maximum Likelihood fit

ML fit as we've discussed it, uses mesured quantities directly and provides *efficient* estimators. But this approach is not always practical. When a dataset is large it may be very computationally intense. The commonly used alternative is the **binned maximum likelihood** fit.

- The data sample of n_{tot} events has to be histogrammed, yielding a certain number of entries $\boldsymbol{n} = (n_1, ..., n_N)$ in N bins.
- The expectation values $\boldsymbol{\nu} = (\nu_1, ..., \nu_N)$ for the bin are given by:

$$\nu_i(\boldsymbol{\theta}) = n_{\mathrm{tot}} \int_{x_i^{\mathrm{min}}}^{x_i^{\mathrm{max}}} f(x; \boldsymbol{\theta}) dx, \tag{11}$$

  where $x_i^{\mathrm{min/max}}$ are the bin limits.
- If n_{tot} is fixed and one is interested in the *shape* of the distribution, the joint p.d.f. is given by the multinomial distribution:

$$f_{\mathrm{joint}}(\boldsymbol{n}, \boldsymbol{\nu}) = \frac{n_{\mathrm{tot}}!}{n_1! ... n_N!} \left(\frac{\nu_1}{n_{\mathrm{tot}}} \right)^{n_1} ... \left(\frac{\nu_N}{n_{\mathrm{tot}}} \right)^{n_N}, \tag{12}$$

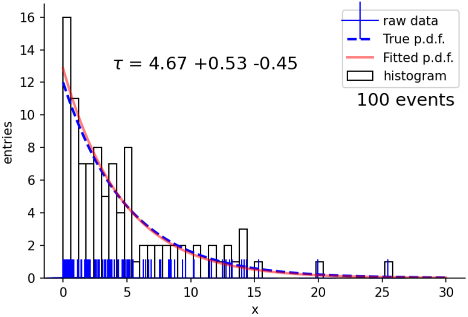
Binned Maximum Likelihood fit

- The resulting log-likelihood (ignoring spurious terms) reads:

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{N} n_i \ln \nu_i(\boldsymbol{\theta}). \tag{13}$$

- In the limit of large number of bins, the binned ML approaches the standard one. However, for coarser binning may result in suboptimal results.

- $\tau = 5.0$
- $n_{\mathrm{tot}} = 100$
- $\sigma_{\widehat{\tau}} = 0.5$
- $N = 50$



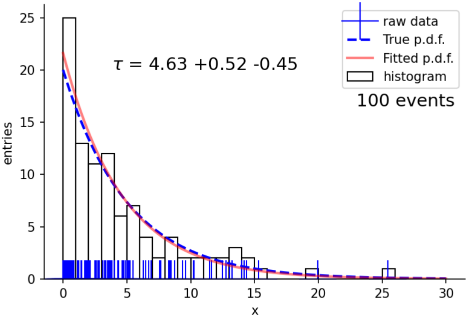
Binned Maximum Likelihood fit

- The resulting log-likelihood (ignoring spurious terms) reads:

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{N} n_i \ln \nu_i(\boldsymbol{\theta}). \tag{14}$$

- In the limit of large number of bins, the binned ML approaches the standard one. However, for coarser binning may result in suboptimal results.

- $\tau = 5.0$
- $n_{\mathrm{tot}} = 100$
- $\sigma_{\widehat{\tau}} = 0.5$
- $N = 30$



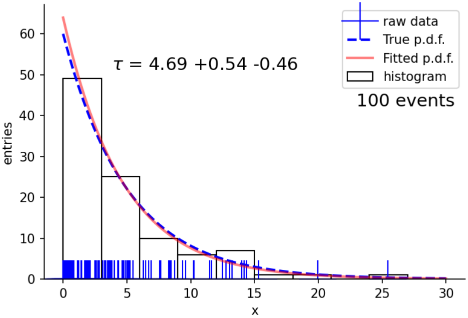
Binned Maximum Likelihood fit

- The resulting log-likelihood (ignoring spurious terms) reads:

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{N} n_i \ln \nu_i(\boldsymbol{\theta}). \tag{15}$$

- In the limit of large number of bins, the binned ML approaches the standard one. However, for coarser binning may result in suboptimal results.

- $\tau = 5.0$
- $n_{\text{tot}} = 100$
- $\sigma_{\widehat{\tau}} = 0.5$
- $N = 10$



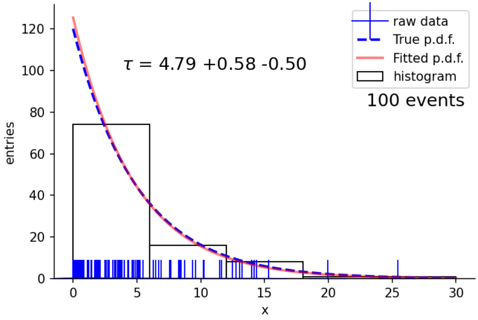
Binned Maximum Likelihood fit

- The resulting log-likelihood (ignoring spurious terms) reads:

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{N} n_i \ln \nu_i(\boldsymbol{\theta}). \tag{16}$$

- In the limit of large number of bins, the binned ML approaches the standard one. However, for coarser binning may result in suboptimal results.

- $\tau = 5.0$
- $n_{\mathrm{tot}} = 100$
- $\sigma_{\widehat{\tau}} = 0.5$
- $N = 5$



Binned Maximum Likelihood fit

extended binned log-likelihood

We may take an alternative approach whereby the total number of events is itself considered a Poisson-distributed random variable with the mean ν_{tot}.

- The expectation values $\boldsymbol{\nu} = (\nu_1, ..., \nu_N)$ now depend on ν_{tot} and $\boldsymbol{\theta}$:

$$\nu_i(\nu_{\text{tot}}, \boldsymbol{\theta}) = \nu_{\text{tot}} \int_{x_i^{\min}}^{x_i^{\max}} f(x; \boldsymbol{\theta}) dx. \tag{17}$$

- Content of each bin becomes a Poisson random variable. We are now concerned both with the *shape* and the *normalisation* of the sample distribution. The joint p.d.f. is given by:

$$f_{\text{joint}}(\boldsymbol{n}, \boldsymbol{\nu}) = \prod_{i=1}^{N} \frac{\nu_i^{n_i}}{n_i!} e^{-\nu_i} \tag{18}$$

Binned Maximum Likelihood fit

extended binned log-likelihood

- Taking the logarithm of the joint p.d.f. and dropping terms that do not depend on the parameters gives:

$$\log L(\nu_{\mathrm{tot}}, \boldsymbol{\theta}) = -\nu_{\mathrm{tot}} + \sum_{i=1}^{N} n_i \ln \nu_i(\nu_{\mathrm{tot}}, \boldsymbol{\theta}), \qquad (19)$$

- which is the binned version of the extended log-likelihood function.
- Here, similar conclusions apply as the ones drawn for the unbinned ML fit.

Example of a 2D ML fit

Data events have been collected in an experiment yielding a scalar random variable x:

- The sample consists of a mixture of *signal* and *background* events with known p.d.f.'s.
- Signal p.d.f.: $f_S(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$,
- Background p.d.f.: $f_B(x;\lambda) = \frac{x}{\lambda^2} e^{-x/\lambda}$, $(\lambda = 5)$,
- The varaible x has been recorded in the range $(0, 30)$.
- No assumption is made about the yields N_S & N_B.

We want to:

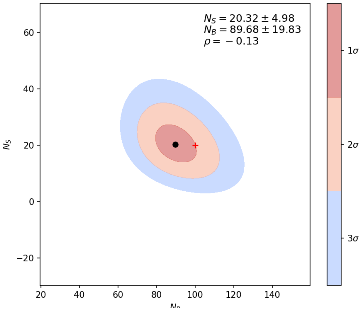
1. Estimate number of signal N_S and background N_B events in the observed sample,
2. assess the error on the fitted N_S, N_B.

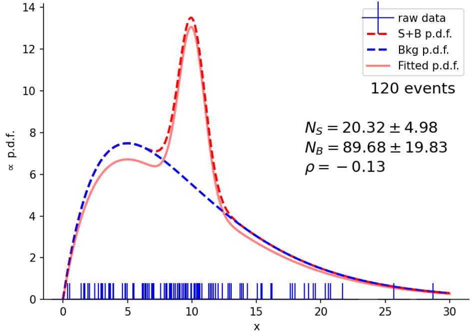
We use the extended ML fit:

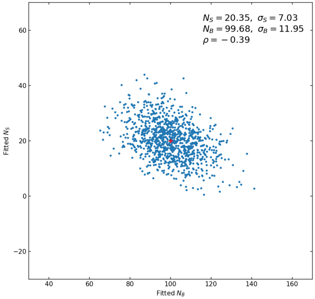
$$\log L(N_S, N_B) = -N_S - N_B + \sum_{i=1}^{n} \ln\left(N_S f_S(x_i;\mu,\sigma) + N_B f_B(x_i;\lambda)\right). \quad (20)$$

Example of a 2D ML fit

- $\lambda = 5.0$, $\mu = 10.0$, $\sigma = 1.0$.
- Uncertainties from the $\log L$ profile $\nearrow$
- Scatter plot of the fit for 1000 MC $\searrow$ events
- Example fit: $\widehat{V} = \left[-\frac{\partial^2 \log L}{\partial\theta_i \partial\theta_j}\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \right]^{-1}$

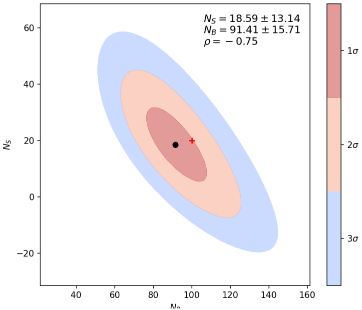


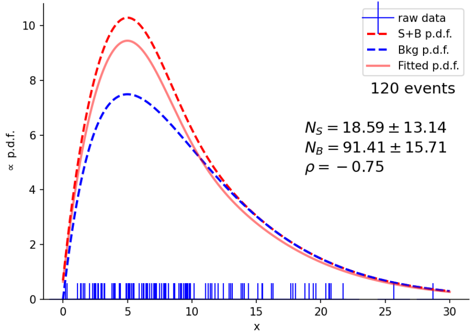


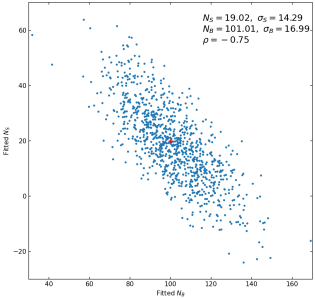


Example of a 2D ML fit

- $\lambda = 5.0,\ \mu = 5.0,\ \sigma = 3.0.$
- Uncertainties from the $\log L$ profile $\nearrow$
- Scatter plot of the fit for 1000 MC $\searrow$ events
- Example fit: $\widehat{V} = \left[-\frac{\partial^2 \log L}{\partial\theta_i \partial\theta_j}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \right]^{-1}$







The method of least-squares (LS)

Consider fitting a function to a set of n measuremnets (x_i, y_i), where y_i is assumed to be Gaussian random centered about the true value $\lambda_i(x_i, \boldsymbol{\theta})$ with standard deviation σ_i. $\boldsymbol{\theta} = (\theta_1, ..., \theta_m)$ are the function parameters we seek to estimate. The meaasurements are assumed to be mutually independent.
The joint p.d.f. is given by:

$$g(\boldsymbol{y}; \boldsymbol{\lambda}, \boldsymbol{\sigma}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y_i - \lambda_i)^2}{2\sigma_i^2}\right) \tag{21}$$

Taking the logarithm and ignoring terms not dependent on $\boldsymbol{\lambda}$ gives the log-likelihood function:

$$\log L(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{i=1}^{n} \frac{(y_i - \lambda(x_i; \boldsymbol{\theta}))^2}{\sigma_i^2} \tag{22}$$

$$\text{or:} \quad \chi^2(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{(y_i - \lambda(x_i; \boldsymbol{\theta}))^2}{\sigma_i^2}, \tag{23}$$

where the latter χ^2 needs to be minimized in order to find LS estimators $\widehat{\boldsymbol{\theta}}$.

The method of least-squares (LS)

General case of correlated Gaussian

The $\chi^2(\boldsymbol{\theta})$ function can be defined also for more general case of correlated measurements, as long as they can be described by a n-dimensional Gaussian:

$$\chi^2(\boldsymbol{\theta}) = \sum_{i,j=1}^{n} (y_i - \lambda(x_i;\boldsymbol{\theta}))\,(V^{-1})_{ij}\,(y_j - \lambda(x_j;\boldsymbol{\theta})) \tag{24}$$

which reduces to E.q. 23 when V is diagonal.

Linear LS fit

Let us consider a special case when predictions λ depend *linearly* on $\boldsymbol{\theta}$.

$$\lambda(x_i;\boldsymbol{\theta}) = \sum_{j=1}^{m} a_j(x_i)\theta_j = \sum_{j=1}^{m} A_{ij}\theta_j = (A\boldsymbol{\theta})_i \tag{25}$$

The χ^2 can be written as:

$$\chi^2(\boldsymbol{\theta}) = (\boldsymbol{y}-\boldsymbol{\lambda})^T V^{-1}(\boldsymbol{y}-\boldsymbol{\lambda}) = (\boldsymbol{y}-A\boldsymbol{\theta})^T V^{-1}(\boldsymbol{y}-A\boldsymbol{\theta}), \tag{26}$$

and the χ^2 minimum w.r.t. $\boldsymbol{\theta}$ can be found analytically!

$$\frac{\partial\chi^2}{\partial\boldsymbol{\theta}} = -2(A^T V^{-1}\boldsymbol{y} - A^T V^{-1}A\boldsymbol{\theta}) = 0 \tag{27}$$

$$\widehat{\boldsymbol{\theta}} = (A^T V^{-1}A)^{-1}A^T V^{-1}\boldsymbol{y}, \tag{28}$$

i.e. the solution is a linear combination of the measurements y_i, provided matrix $A^T V^{-1}A$ can be inverted (is not singular).
Gauss-Markov theorem: $\widehat{\boldsymbol{\theta}}$ is unbiased and has minimum variance independently of n and individual measurements p.d.f.'s.

Linear LS fit

The $\text{cov}[\widehat{\boldsymbol{\theta}}]$ is obtained using error propagation:

$$\text{cov}[\widehat{\boldsymbol{\theta}}] \equiv U = \frac{\partial \widehat{\boldsymbol{\theta}}}{\partial \boldsymbol{y}} V \frac{\partial \widehat{\boldsymbol{\theta}}}{\partial \boldsymbol{y}}^T = (A^T V^{-1} A)^{-1} \tag{29}$$

Note that we reproduce the RCF bound:

$$(U^{-1})_{ij} = (A^T V^{-1} A)_{ij} = \frac{1}{2} \left[\frac{\partial^2 \chi^2}{\partial \theta_i \partial \theta_j} \right]_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = - \left[\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j} \right]_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \tag{30}$$

$$\begin{aligned} \chi^2(\boldsymbol{\theta}) &= \chi^2(\widehat{\boldsymbol{\theta}}) + \frac{1}{2} \sum_{i,j=1}^{m} \left[\frac{\partial^2 \chi^2}{\partial \theta_i \partial \theta_j} \right]_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} (\theta_i - \widehat{\theta}_i)(\theta_j - \widehat{\theta}_j) \\ &= \chi^2(\widehat{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})^T U^{-1} (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}), \end{aligned} \tag{31}$$

gives us the contour of one standard deviation from LS estimates:

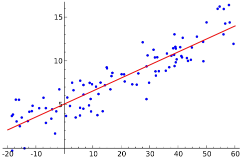
$$\chi^2(\widehat{\boldsymbol{\theta}} \pm \boldsymbol{\sigma}) = \chi^2(\widehat{\boldsymbol{\theta}}) + 1 = \chi^2_{\text{min}} + 1 \tag{32}$$

The contour holds even for non-linear parameters in analogy to the $\log L$ Taylor expansion.

Linear LS fit

A super-simple example - linear regression

Assume a straight line fit through n measurements (x_i, y_i), all with the same error σ:



$$\chi^2(p0, p1) = \sum_{i=1}^{n} \frac{(y_i - (p0 + p1 x_i))^2}{\sigma^2} \tag{33}$$

$$A = \begin{pmatrix} 1 & x_1 \\ . & . \\ . & . \\ 1 & x_n \end{pmatrix}, \quad V = \begin{pmatrix} \sigma^2 & \dots & 0 \\ \dots & \sigma^2 & \dots \\ 0 & \dots & \sigma^2 \end{pmatrix}, \quad U = \frac{\sigma^2}{n \sum x_i^2 - (\sum x_i)^2} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix} \tag{34}$$

When unknown, σ can be estimated from the fit itself: $\widehat{\sigma^2} = \sum_{i=1}^{n} \frac{(y_i - (\widehat{p0} + \widehat{p1} x_i))^2}{n-2}$

Linear LS approximation (linear expansion) has a lot of applications, allowing for solutions to problems with very large number of parameters using linear algebra. For convergence, it often involves multiple iterations of the fit!

Least Squares with binned data

LS is also used to fit p.d.f. to data binned in a histogram. Let $f(x;\boldsymbol{\theta})$ be the hypothetical p.d.f. and y_i contents of bin i. The number of entries predicted in the bin $\lambda_i = E[y_i]$ is:

$$\lambda_i(\boldsymbol{\theta}) = n \int_{x_i^{\min}}^{x_i^{\max}} f(x;\boldsymbol{\theta}) dx = n p_i(\boldsymbol{\theta}). \tag{35}$$

Making an approximation of Gaussian distribution of y_i the χ^2 is given as:

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^{N} \frac{(y_i - n p_i(\boldsymbol{\theta}))^2}{n p_i(\boldsymbol{\theta})}, \tag{36}$$

where $\sqrt{n p_i(\boldsymbol{\theta})}$ is the standard deviation for the Poisson distribution. Sometimes the **modified** LS (MLS) is used instead:

$$\chi^2_{\mathrm{M}}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \frac{(y_i - n p_i(\boldsymbol{\theta}))^2}{y_i}. \tag{37}$$

Careful! This is fine for large statistics. Think of poorly populated or empty bins!

Least Squares with binned data

It might be tempting to use LS method in order to fit the total yield of the distribution by introducing additional free parameter ν:

$$\lambda_i(\boldsymbol{\theta}, \nu) = \nu \int_{x_i^{\min}}^{x_i^{\max}} f(x; \boldsymbol{\theta}) dx = \nu p_i(\boldsymbol{\theta}). \tag{38}$$

However, minimising the χ^2 by setting $\frac{\partial \chi^2}{\partial \nu} = 0$ results in biased estimates of n. We get[1]:

- $\widehat{\nu}_{\mathrm{LS}} = n + \frac{\chi^2}{2}$,
- $\widehat{\nu}_{\mathrm{MLS}} = n - \chi^2$.

Note: This can be corrected for, but should be considered at all times.

[1] Exercise: Show it by explicit calculation

Putting it to work...

1D fit of the signal yield

N data events have been collected in an experiment yielding a scalar random variable x:

- The sample consists of a mixture of *signal* and *background* events with known p.d.f.'s.
- Background p.d.f.: $f_B(x) = \frac{1}{\tau} e^{-x/\tau}$, $\tau = 5$,
- Signal p.d.f.: $f_S(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$, $\mu = 10$, $\sigma = 3$,
- The varaible x has been recorded in the range $(0, 30)$.
- No assumption about background yield can be made: $N = N_S + N_B$.

Our task is to:

1. Estimate number of signal events N_S in the observed sample,
2. assess the error of the N_S estimate from the $\log L$ or χ^2 profile.

MIND: This is NOT an extended fit.

Putting it to work...

1D fit of the signal yield

We shall use three strategies:

1. the *Unbinned Maximum Likelihood* fit:
   `Unbinned ML Python notebook template in Colab`
2. the *Binned Maximum Likelihood* fit; 10 bins over (0, 30):
   `Binned ML Python notebook template in Colab`
3. the *Binned Least Squares* (NOT modified) fit; 10 bins over (0, 30):
   `Binned LS Python notebook template in Colab`

and four data samples:

1. `pickled data sample 1 from GitHub`
2. `pickled data sample 2 from GitHub`
3. `pickled data sample 3 from GitHub`
4. `pickled data sample 4 from GitHub`

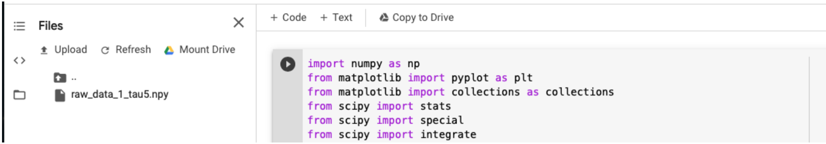
All shall be executed on the *Google Colaboratory* platform.

Detailed instructions

⟶ Click on one of the Python notebook links in order to open it in *Google Colaboratory*.

⟶ Click to download the assigned dataset file from GitHub.

⟶ Click on one of the *Files* icon on the left bar of your *Colab* interface. If you cannot see any datafiles, click on the *Upload* button and select previously downloaded file. As a result you should see:



Hints part 1: fit the N_s, $\log L$ or χ^2 function

$\longrightarrow$ You need p.d.f. normalization factors in the range (0,30). For this purpose calculate `scales` and `scaleb` just as it is done in the main part of the Python script. You will need `xmin`, `xmax`, `tau0`, `mu0` and `sigma0`.

$\longrightarrow$ You need the total number of collected events. For the unbinned ML this is just the length of the `data` vector (`len(data)`). For the binned methods loop over the `hist` array and sum all entries. `nbins=len(hist)` gives you the number of bins.

$\longrightarrow$ For the unbinned ML you need to loop over the `data` array, for each entry calculate the normalized p.d.f.'s (`gauss` & `decay`) and acumulate $\log L$ according to Eq. (25) of lecture 4 and using the combined S+B p.d.f.

$\longrightarrow$ For the binned methods you need to loop over the bins, the `hist` array (e.g. `for k in range(nbins)`). You need to get the prediction for the bin by integrating the normalized p.d.f., see Eq. (13) of lecture 5. Bin k is delimitted by `binsy[k]` and `binsy[k+1]`.

$\longrightarrow$ For the binned ML increment the $\log L$ using Eq. (13) of lecture 5 and the combined S+B p.d.f.

$\longrightarrow$ For the binned LS increment the χ^2 using Eq. (36) of lecture 5 and the combined S+B p.d.f.

$\longrightarrow$ NOTE: We are fitting just one free parameter, N_S (coded as `mus`). Make sure you properly define the combined S+B p.d.f. using N_S and the total number of collected events.

Hints part 2: estimate the error on N_s using $\log L$ or χ^2 profile

$\longrightarrow$ The code provides you with the `pl` & `ll_array` arrays which contain the estimated N_S and the corresponding value of either $\log L$ or χ^2, respectively.
$\longrightarrow$ Your task is to find values of N_S corresponding to $+1\sigma$ and -1σ about the fitted value (coded as `sigma_neg` & `sigma_pos`).
$\longrightarrow$ For the purpose, recall Eq. (7) and Eq. (32) of lecture 5.

Thank you

Back-up

Least Squares with binned data

χ^2 estimators of the yield

The total yield of the distribution parameterised by an additional free parameter ν: $\lambda_i = \nu p_i$. We have n events distributed over N bins.
Minimising the χ^2 by setting $\frac{\partial \chi^2}{\partial \nu} = 0$ results in biased estimates of n:

LS:

$$
\begin{aligned}
&\chi^2 = \sum^N \frac{(y_i - \nu p_i)^2}{\nu p_i}, \quad \frac{\partial \chi^2}{\partial \nu} = -2 \sum^N \frac{(y_i - \nu p_i) p_i}{\nu p_i} - \sum^N \frac{(y_i - \nu p_i)^2 p_i}{(\nu p_i)^2} = 0 \\
&-2 \sum^N (y_i - \nu p_i) - \sum^N \frac{(y_i - \nu p_i)^2}{\nu p_i} = -2n + 2\nu - \chi^2 = 0 \\
&\Longrightarrow \widehat{\nu}_{\mathrm{LS}} = n + \frac{\chi^2}{2}
\end{aligned} \tag{39}
$$

MLS:

$$
\begin{aligned}
&\chi^2 = \sum^N \frac{(y_i - \nu p_i)^2}{y_i}, \quad \frac{\partial \chi^2}{\partial \nu} = -2 \sum^N \frac{(y_i - \nu p_i) p_i}{y_i} = 0 \Longrightarrow \nu \sum^N \frac{p_i^2}{y_i} = 1 \\
&\chi^2 = \sum^N y_i - 2\nu \sum^N p_i + \nu^2 \sum^N \frac{p_i^2}{y_i} = n - 2\nu + \nu = n - \nu \\
&\Longrightarrow \widehat{\nu}_{\mathrm{MLS}} = n - \chi^2
\end{aligned} \tag{40}
$$