Numerical Methods
Lesson 9

Dr. Jose Feliciano Benitez

Universidad de Sonora

Dr. Benitez Homepage: www.jfbenitez.science
Course page: http://jfbenitez.ddns.net:8080/Courses/MetodosNumericos

Goals for this Lesson

- Selected topics from chapters 6, 7, and 9.

Screenshot

Maximum Likelihood Estimators

- Estimation of unknown p.d.f. parameters for a model and corresponding dataset.

6.1 ML estimators

Consider a random variable x distributed according to a p.d.f. $f(x;\theta)$. Suppose the functional form of $f(x;\theta)$ is known, but the value of at least one parameter θ (or parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$) are not known. That is, $f(x;\theta)$ represents a composite hypothesis for the p.d.f. (cf. Section 4.1). The method of **maximum likelihood** is a technique for estimating the values of the parameters given a finite sample of data. Suppose a measurement of the random variable x has been repeated n times, yielding the values $x_1, \ldots, x_n$. Here, x could also represent a multidimensional random vector, i.e. the outcome of each individual observation could be characterized by several quantities.

If the hypothesized p.d.f. and parameter values are correct, one expects a high probability for the data that were actually measured. Conversely, a parameter value far away from the true value should yield a low probability for the measurements obtained. Since the dx_i do not depend on the parameters, the same reasoning also applies to the following function L,

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) \tag{6.2}$$

called the **likelihood function**. Note that this is just the joint p.d.f. for the x_i, although it is treated here as a function of the parameter, θ. The x_i, on the other hand, are treated as fixed (i.e. the experiment is over).

Notes:

- *{x1, x2, ….}* is a set of measurements
- θ is an **unknown** parameter of the p.d.f.
- When the chosen value for θ is the near the correct one then L is maximum. In practice we use the ***log L(θ)*** value to search for the maximum
- If the the log L is differentiable then the parameters can be determined by the condition:
  d log L(θ) /dθ = 0

Ilustration

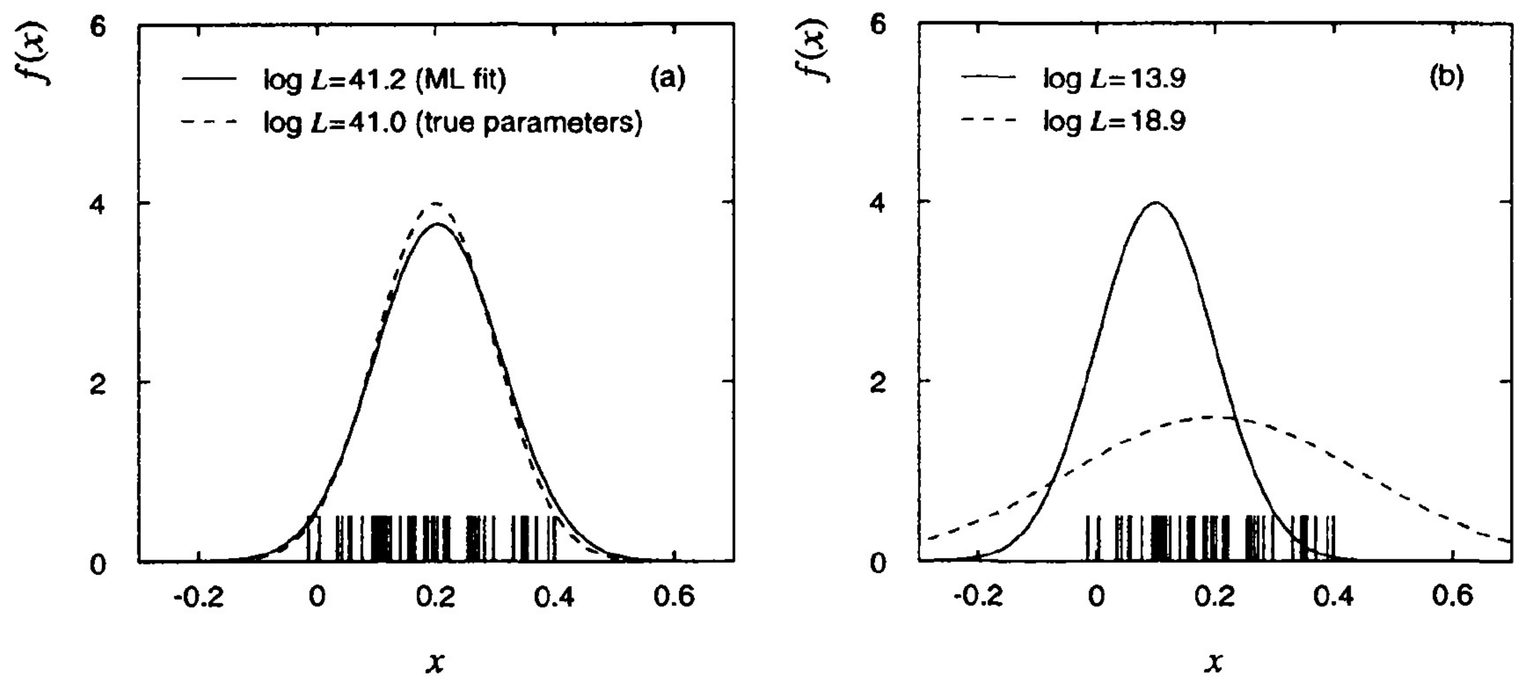


Fig. 6.1 A sample of 50 observations of a Gaussian random variable with mean $\mu = 0.2$ and standard deviation $\sigma = 0.1$. (a) The p.d.f. evaluated with the parameters that maximize the likelihood function and with the true parameters. (b) The p.d.f. evaluated with parameters far from the true values, giving a lower likelihood.

Example

6.2 Example of an ML estimator: an exponential distribution

Suppose the proper decay times for unstable particles of a certain type have been measured for n decays, yielding values $t_i, \ldots, t_n$, and suppose one chooses as a hypothesis for the distribution of t an exponential p.d.f. with mean τ:

$$f(t;\tau) = \frac{1}{\tau} e^{-t/\tau}. \qquad (6.4)$$

The task here is to estimate the value of the parameter τ. Rather than using the likelihood function as defined in equation (6.2) it is usually more convenient to use its logarithm. Since the logarithm is a monotonically increasing function, the parameter value which maximizes L will also maximize $\log L$. The logarithm has the advantage that the product in L is converted into a sum, and exponentials in f are converted into simple factors. The **log-likelihood function** is thus

$$\log L(\tau) = \sum_{i=1}^{n} \log f(t_i;\tau) = \sum_{i=1}^{n} \left(\log \frac{1}{\tau} - \frac{t_i}{\tau} \right). \qquad (6.5)$$

Maximizing $\log L$ with respect to τ gives the ML estimator $\hat{\tau}$,

$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} t_i. \qquad (6.6)$$

Screenshot

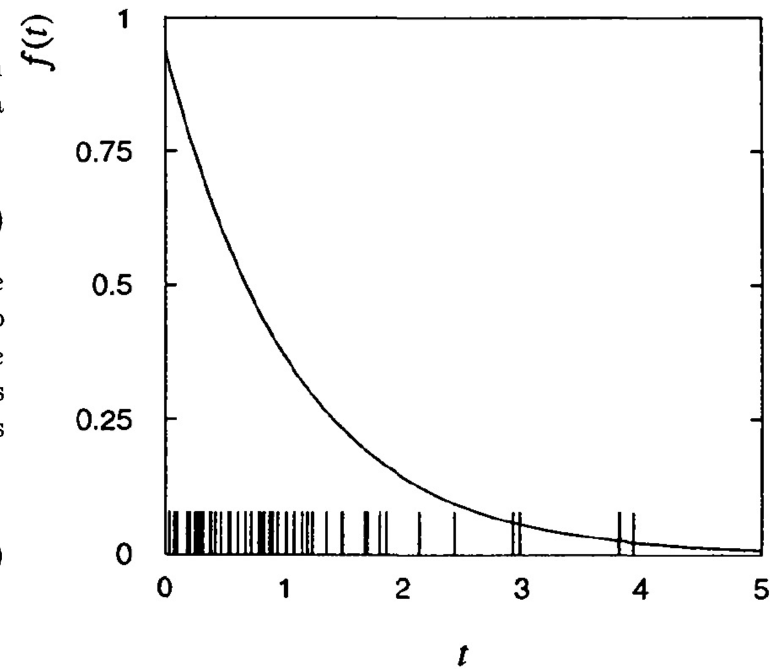


Fig. 6.2 A sample of 50 Monte Carlo generated observations of an exponential random variable t with mean $\tau = 1.0$. The curve is the result of a maximum likelihood fit, giving $\hat{\tau} = 1.062$.

Example

6.3 Example of ML estimators: μ and σ^2 of a Gaussian

Suppose one has n measurements of a random variable x assumed to be distributed according to a Gaussian p.d.f. of unknown μ and σ^2. The log-likelihood function is

$$\log L(\mu, \sigma^2) = \sum_{i=1}^{n} \log f(x_i; \mu, \sigma^2) = \sum_{i=1}^{n} \left(\log \frac{1}{\sqrt{2\pi}} + \frac{1}{2} \log \frac{1}{\sigma^2} - \frac{(x_i - \mu)^2}{2\sigma^2} \right). \tag{6.10}$$

Setting the derivative of $\log L$ with respect to μ equal to zero and solving gives

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{6.11}$$

Computing the expectation value as done in equation (6.7) gives $E[\hat{\mu}] = \mu$, so $\hat{\mu}$ is unbiased. (As in the case of the mean lifetime estimator $\hat{\tau}$, $\hat{\mu}$ here happens to be a sample mean, so one knows already from Sections 2.5 and 5.2 that it is an unbiased estimator for the mean μ.) Repeating the procedure for σ^2 and using the result for $\hat{\mu}$ gives

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})^2. \tag{6.12}$$

Binned Maximum Likelihood

- For data sets where the number of measurements is large.

6.10 Maximum likelihood with binned data

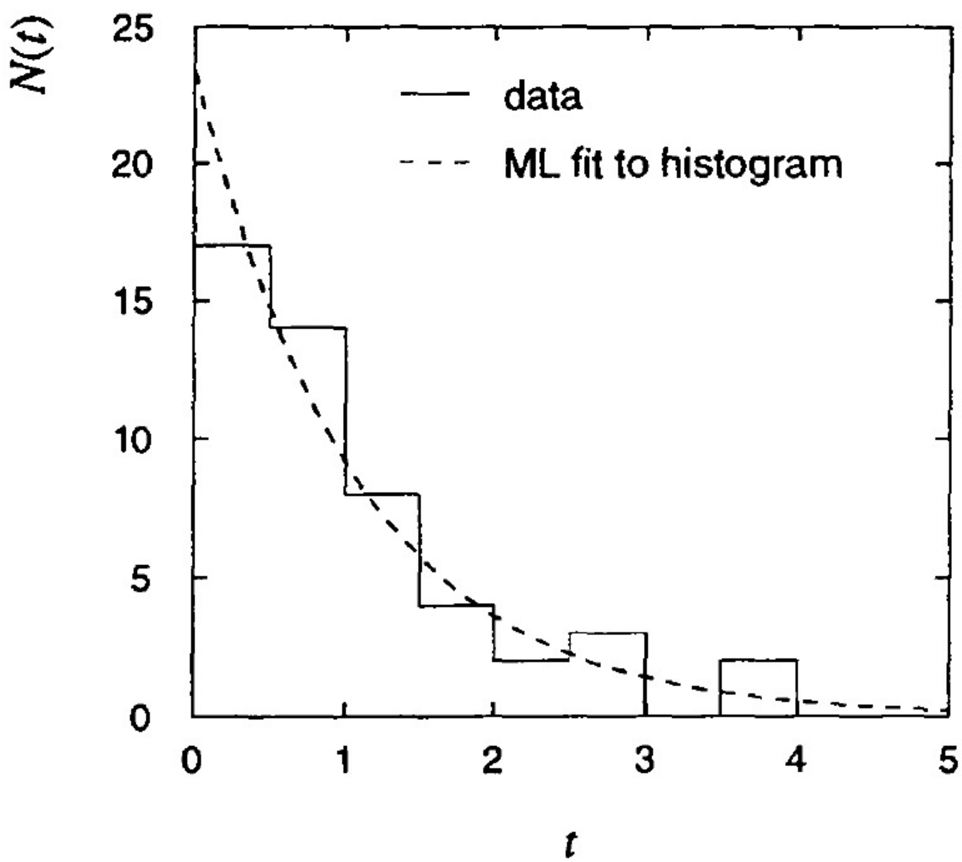
Consider n_{tot} observations of a random variable x distributed according to a p.d.f. $f(x;\boldsymbol{\theta})$ for which we would like to estimate the unknown parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$. For very large data samples, the log-likelihood function becomes difficult to compute since one must sum $\log f(x_i;\boldsymbol{\theta})$ for each value x_i. In such cases, instead of recording the value of each measurement one usually makes a histogram, yielding a certain number of entries $\mathbf{n} = (n_1, \ldots, n_N)$ in N bins. The expectation values $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_N)$ of the numbers of entries are given by

$$\nu_i(\boldsymbol{\theta}) = n_{\mathrm{tot}} \int_{x_i^{\mathrm{min}}}^{x_i^{\mathrm{max}}} f(x;\boldsymbol{\theta})\,dx, \tag{6.40}$$

where x_i^{min} and x_i^{max} are the bin limits. One can regard the histogram as a single measurement of an N-dimensional random vector for which the joint p.d.f. is given by a multinomial distribution, equation (2.6),

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{N} n_i \log \nu_i(\boldsymbol{\theta}), \tag{6.42}$$

Example of binned ML



Range: 0 - 5
Bin size: $\Delta t = 0.5$
Number of bins: n = 10
Original sample size: 50

Unbinned result: 1.062
Binned result : 1.069
True value: 1.000

Fig. 6.10 Histogram of the data sample of 50 particle decay times from Section 6.2 with the ML fit result.

7

The method of least squares

Consider now a set of N independent Gaussian random variables y_i, $i = 1, \ldots, N$, each related to another variable x_i, which is assumed to be known without error. For example, one may have N measurements of a temperature $T(x_i)$ at different positions x_i. Assume that each value y_i has a different unknown mean, λ_i, and a different but known variance, σ_i^2. The N measurements of y_i can be equivalently regarded as a single measurement of an N-dimensional random vector, for which the joint p.d.f. is the product of N Gaussians,

$$g(y_1, \ldots, y_N; \lambda_1, \ldots, \lambda_N, \sigma_1^2, \ldots, \sigma_N^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(\frac{-(y_i - \lambda_i)^2}{2\sigma_i^2} \right). \quad (7.1)$$

Suppose further that the true value is given as a function of x, $\lambda = \lambda(x; \boldsymbol{\theta})$, which depends on unknown parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$. The aim of the method of least squares is to estimate the parameters $\boldsymbol{\theta}$. In addition, the method allows for a simple evaluation of the goodness-of-fit of the hypothesized function $\lambda(x; \boldsymbol{\theta})$. The basic ingredients of the problem are illustrated in Fig. 7.1.

Taking the logarithm of the joint p.d.f. and dropping additive terms that do not depend on the parameters gives the log-likelihood function,

$$\log L(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{i=1}^{N} \frac{(y_i - \lambda(x_i; \boldsymbol{\theta}))^2}{\sigma_i^2}. \quad (7.2)$$

creenshot

- In case of N **independent** random variables y_i where their mean values and uncertainties have been estimated and are related/mapped to another set of values x_i.
- There may exist a relation between the means $\lambda(x,\theta)$ dependent on some parameter of interest q.
- The parameter q may be estimated by minimizing the function:

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^{N} \frac{(y_i - \lambda(x_i; \boldsymbol{\theta}))^2}{\sigma_i^2},$$

- It is possible that $\lambda(x,\theta)$ is itself a p.d.f. and the y_i are counts in bin i. In this case $\sigma_i^2 = \lambda(x_i,\theta)$

Ilustraion of Least Squares

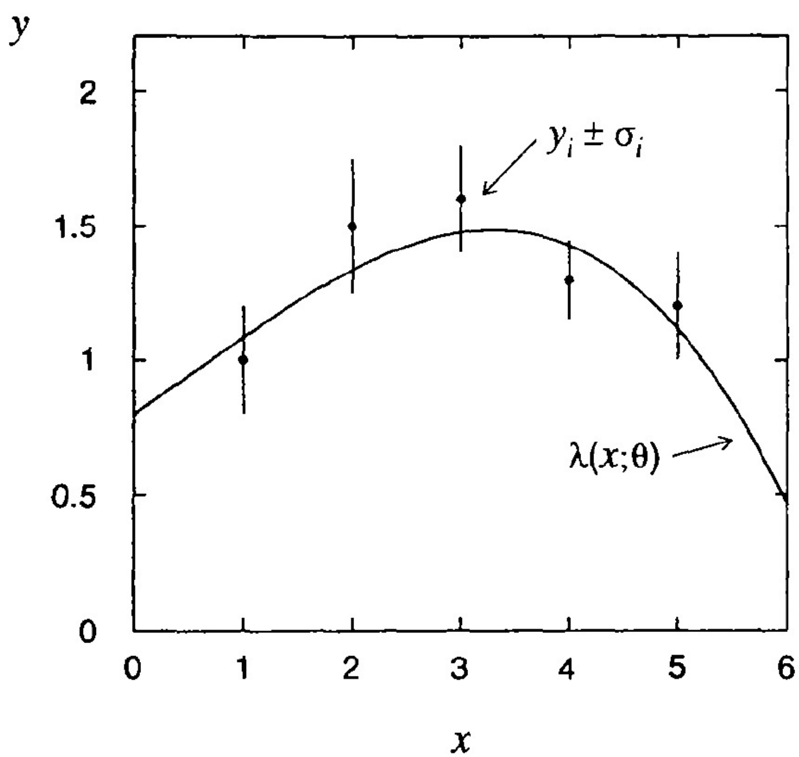
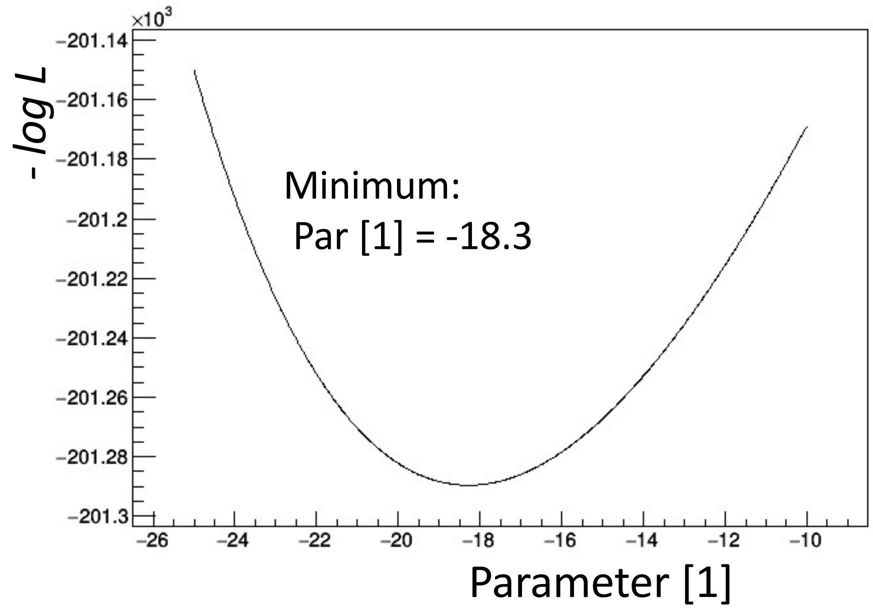
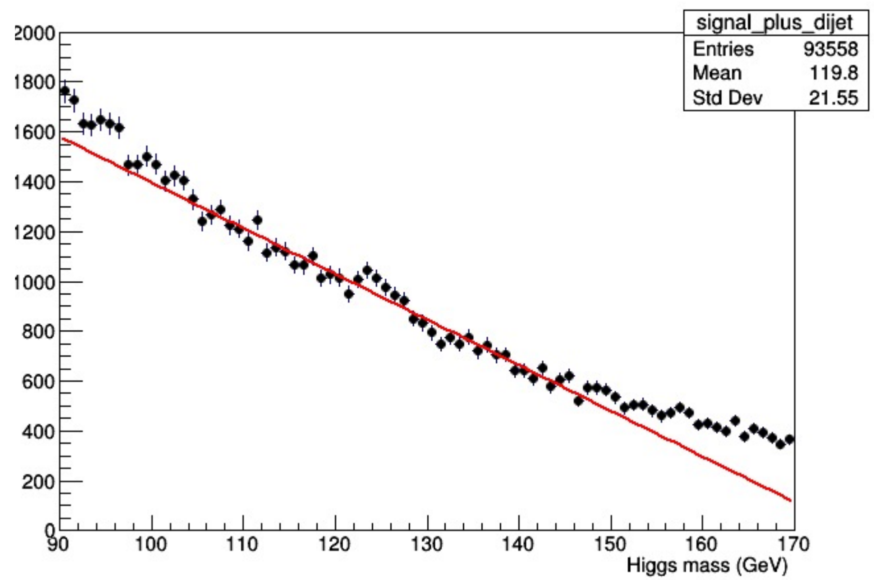


Fig. 7.1 Ingredients of the least squares problem: N values $y_1, \ldots, y_N$ are measured with errors $\sigma_1, \ldots, \sigma_N$ at the values of x given without error by $x_1, \ldots, x_N$. The true value λ_i of y_i is assumed to be given by a function $\lambda_i = \lambda(x_i;\boldsymbol{\theta})$. The value of $\boldsymbol{\theta}$ is adjusted to minimize the value of χ^2 given by equation (7.3).

Binned ML fit to data from Lesson7

- Use code in Lesson9/ fit_signal_dijet.C with **background only** model:
  `TF1 Model("Model","[0]+[1]*(x-[3])",0,200);`
- fit range: 110 – 140
  Parameter [3] = 110
  Parameter [0] = bin content at 110





Binned ML fit to data from Lesson7

- Use code in Lesson9/ fit_signal_dijet.C with **background + signal** model:
  `TF1 Model("Model","[0]+[1]*(x-[3]) + [4]*exp(-0.5*(x-125)*(x-125)/8*8)",0,200)`
- fit range: 110 – 140
  Parameter [3] = 110
  Parameter [0] = bin content at 110
  Parameter [1] = -18.3