Statistical Methods in Particle Physics

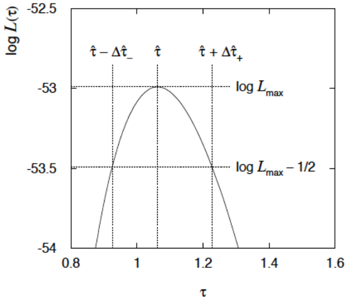
Lecture 6

November 19, 2012

Silvia Masciocchi, GSI Darmstadt
s.masciocchi@gsi.de

Winter Semester 2012 / 13

Outline

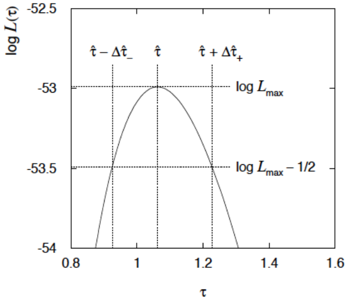


- Estimators
- Estimators for mean, and variance

- The likelihood function
- Maximum likelihood estimators
- Examples: parameters of exponential and Gaussian pdfs

- Variance of ML estimators
- Difference methods:
  - Analytic
  - Monte Carlo
  - The RCF bound
  - Graphical method

The usual start ...



Consider n independent observations of a random variable x:

$\rightarrow$ sample of size n

Equivalently, take a single observation of an n-dimensional vector:

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

The x_i are independent $\rightarrow$ the joint pdf for the sample is:

$$f_{\text{sample}}(\vec{x}) = f(x_1)\, f(x_2) \ldots f(x_n)$$

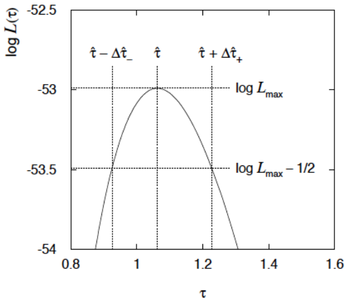
TASK: given a data sample, infer properties of f(x)

$\rightarrow$ construct functions of the data to estimate various properties of f(x) (like mean, variance)

Often, the form of f(x) is hypothesized: value of the parameter(s) is unknown!

$\rightarrow$ given form of f(x;θ) and data sample, estimate θ

Example of parameter(s)



The parameters of a pdf are constants that characterize its shape. For example:

$$f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}$$

Random variable **parameter**

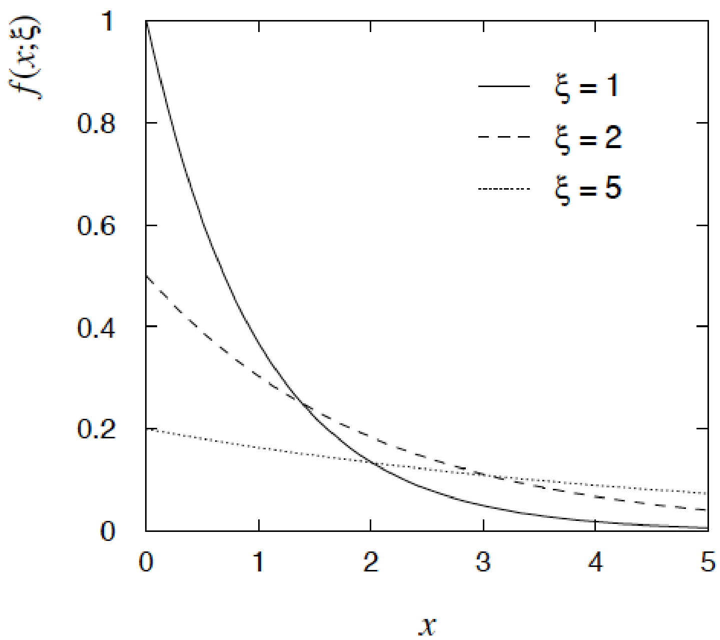
Example: the exponential distribution describes the decay time of an unstable particle measured in its rest frame:

θ = lifetime

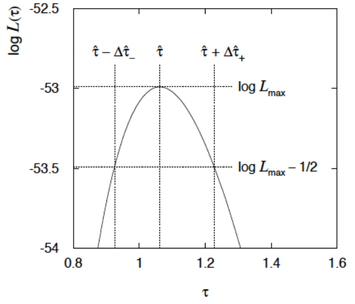
e.g.: neutron (udd) 881.5±1.5 s

Λ (uds) 2.63±0.02 x 10^{-10} s

Λ_c (udc) 2.00±0.02 x 10^{-13} s



Parameter estimation



Suppose we have a sample of observed values: $\vec{x} = (x_1, \ldots, x_n)$

We want to find some function of the data to estimate the parameter(s):

$$\boxed{\hat{\theta}(\vec{x})}$$

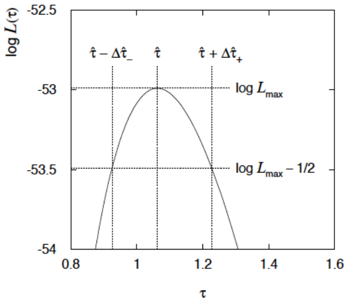
← Estimator written with a hat

We say:

'Estimator' for the function of $(x_1, \ldots, x_n)$. Statistic is used to estimate some property of a pdf. Notation: the hat $\hat{\theta}(\vec{x})$ is a function of a (vector) random variable → it is itself a random variable, characterized by a pdf g($\hat{\theta}$), mean variance …

'*Estimate*' for the value of the estimator with a particular data set.

Estimators



How do we construct an estimator $\hat{\theta}(\vec{x})$?

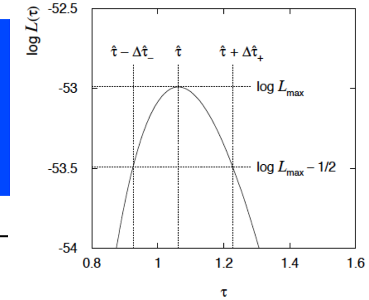
There is no golden rule on how to construct an estimator !!

Construct estimators to satisfy (in general conflicting) criteria

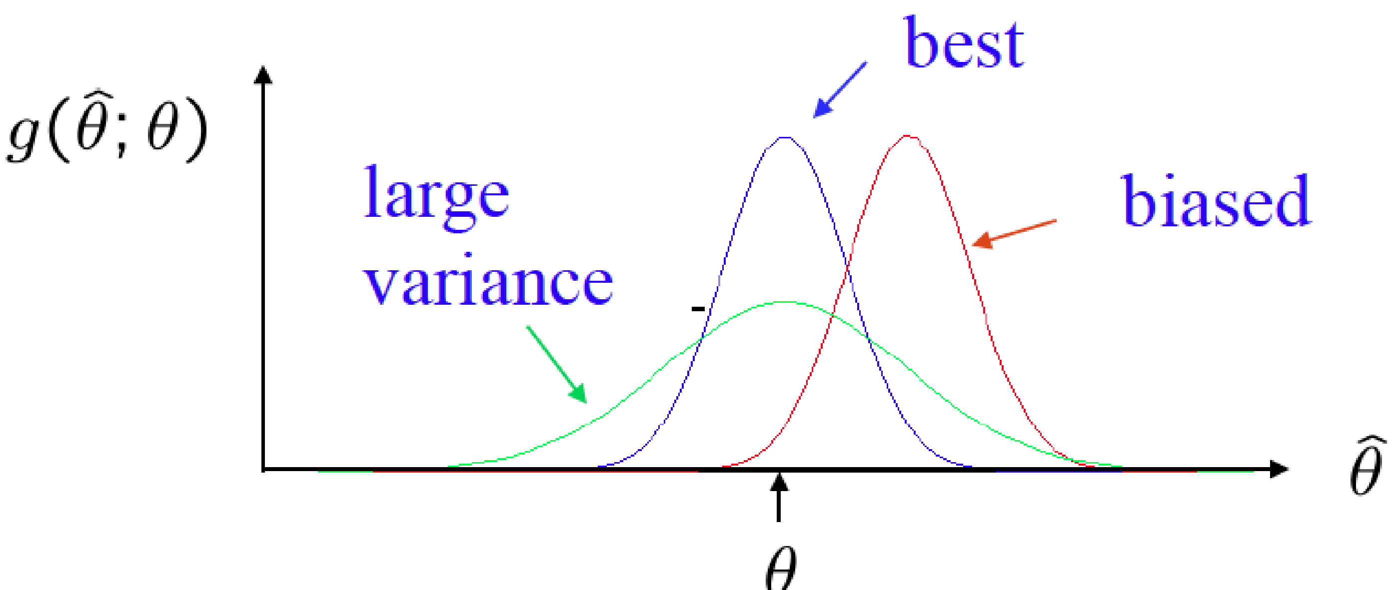
First: require consistency: $\lim_{n \to \infty} \hat{\theta} = \theta$

i.e. as size of sample increases, estimate converges to true value

Properties of estimators



If we were to repeat the entire measurement, the estimates from each measurement would follow a pdf $g(\hat{\theta};\theta)$:



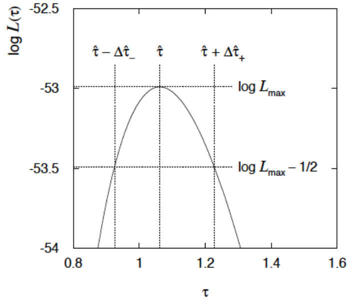
We want small (or zero) **bias** (systematic error): $b = E[\hat{\theta}] - \theta$

→ average of repeated measurements should tend to true value

And we want a small **variance** (statistical error): $V[\hat{\theta}] = \sigma_{\hat{\theta}}^2$

→ small bias and variance are in general conflicting criteria

Properties of estimators - 2



For many estimators we will have:

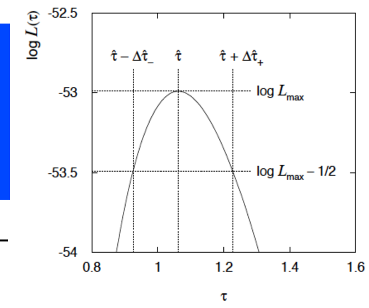
$$\sigma_{\hat{\theta}} \propto \frac{1}{\sqrt{n}} \qquad \mathrm{b} \propto \frac{1}{n}$$

Sometimes consider the mean squared error:

$$\mathrm{MSE} = V[\hat{\theta}] + \mathrm{b}^2$$

In general there is a trade-off between bias and variance.
Often require minimum variance among estimators with 0 bias.

Estimator for the mean (expectation value)



Parameter: $\mu = E[x]$

Sample: n measurements of x: $x_1, \ldots, x_n$

Estimator: $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \equiv \bar{x}$ "sample mean"

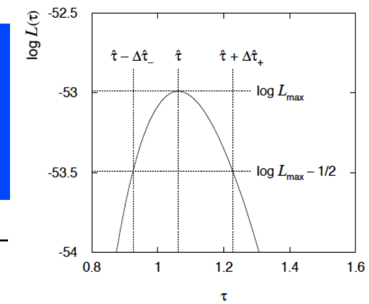
Compute expectation value and variance of the estimator $\hat{\mu}$

DO !

We find: $b = E[\hat{\mu}] - \mu = 0$ $\rightarrow$ $\hat{\mu}$ is an unbiased estimator for μ

if $\sigma = V[x] \rightarrow V[\hat{\mu}]$

Estimator for the mean (expectation value)



Parameter: $\mu = E[x]$

Sample: n measurements of x: $x_{1,} \ldots, x_n$

Estimator: $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \equiv \bar{x}$ "sample mean"

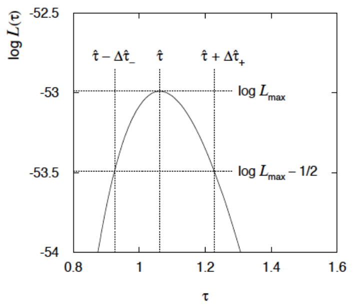
Compute expectation value and variance of the estimator $\hat{\mu}$

DO !

We find: $b = E[\hat{\mu}] - \mu = 0$ $\rightarrow$ $\hat{\mu}$ is an unbiased estimator for μ

$$\text{if } \sigma = V[x] \rightarrow V[\hat{\mu}] = \frac{\sigma^2}{n} \qquad \left(\sigma_{\hat{\mu}} = \frac{\sigma}{\sqrt{n}} \right)$$

Estimator for the variance

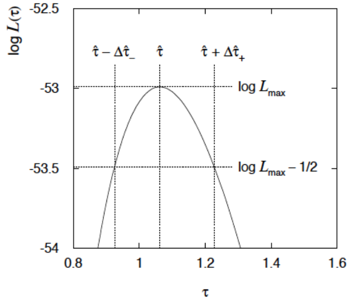


Parameter: $\sigma^2 = V[x]$

Estimator: $\widehat{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \equiv s^2$ “sample variance”

We find: $\boxed{\text{DO !}}$ $b = E[\widehat{\sigma^2}] - \sigma^2 =$

Estimator for the variance



Parameter: $\sigma^2 = V[x]$

Estimator: $\widehat{\sigma^2} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \equiv s^2$ "sample variance"

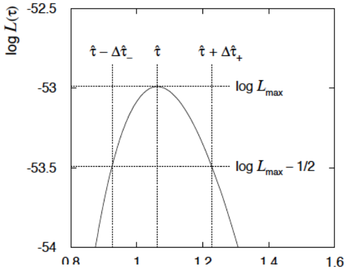
We find: **DO !** $b = E[\widehat{\sigma^2}] - \sigma^2 = 0$ factor n-1 makes this so

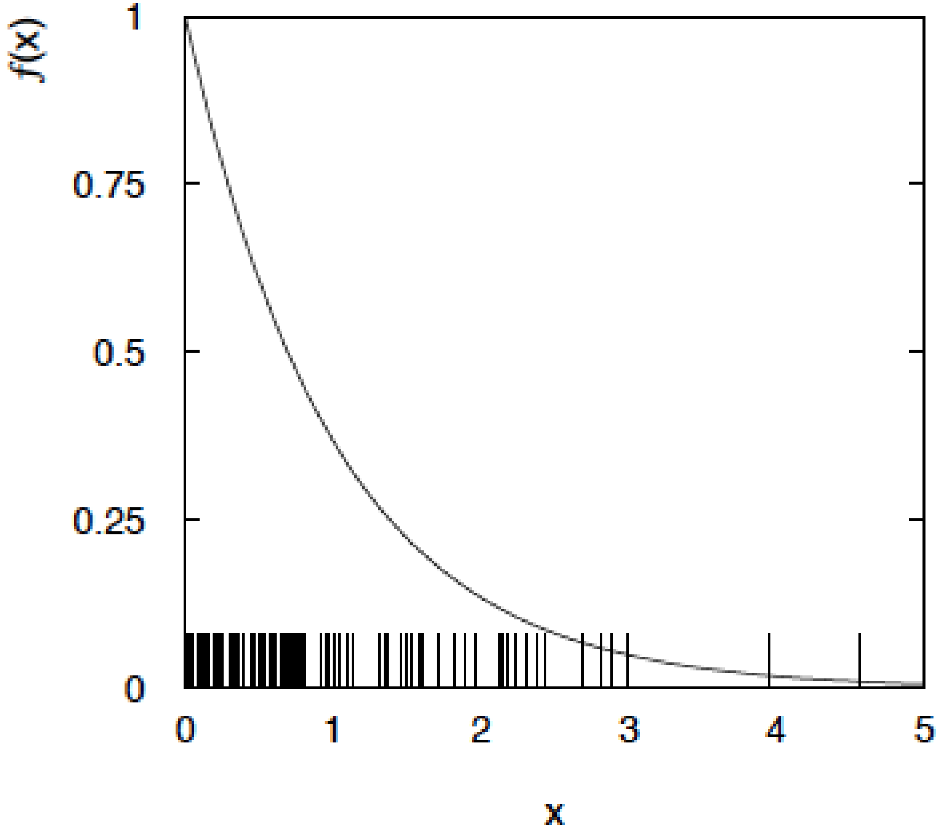
No bias !

$$V[\widehat{\sigma^2}] = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\mu_2^2\right)$$, where

$$\mu_k = \int (x-\mu)^k f(x)\,dx$$ k-th central moment

Example of estimator for mean

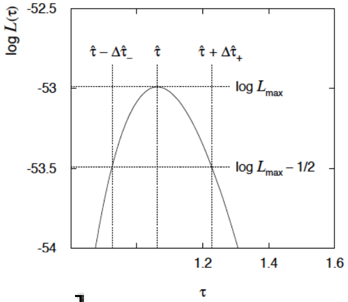




Data sample of $n = 100$ values from MC with $\mu = 1, \;\; \sigma^2 = 1.$

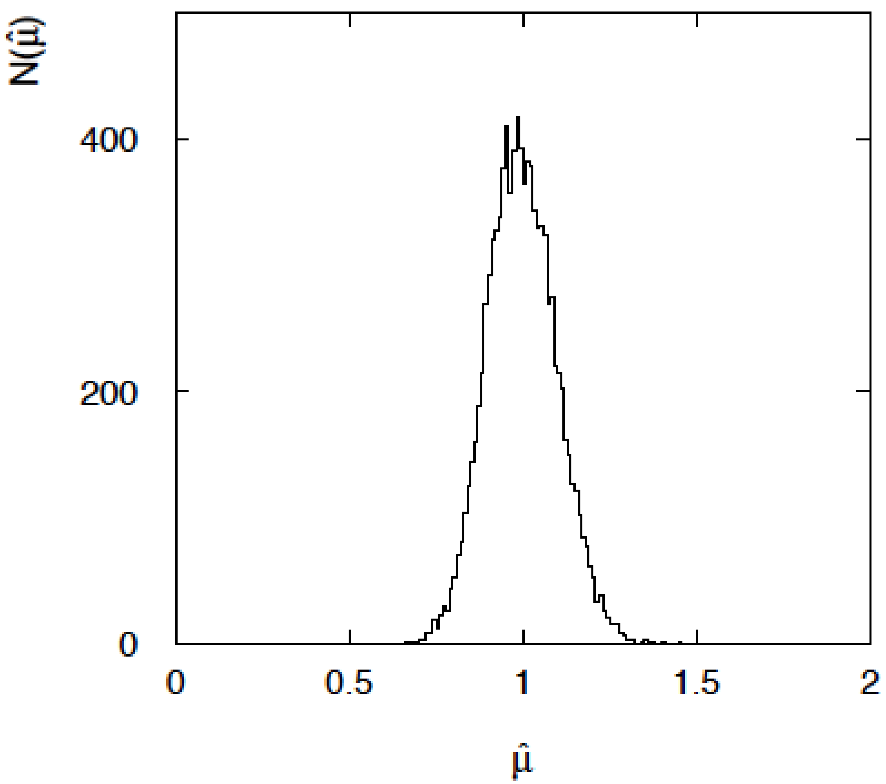
$$\hat{\mu} = \overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = 1.073$$

Example of estimator for mean - 2



Now repeat the experiment 10^4 times with $n = 100$ values each,

enter the sample mean for each experiment into histogram:



$\overline{\hat{\mu}} = 0.9981$ ($\hat{\mu}$ unbiased)

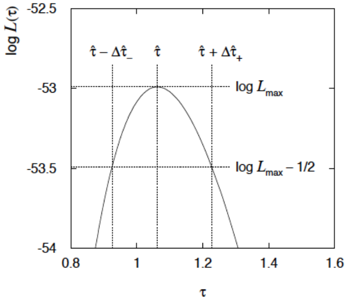
Sample standard deviation

of $\hat{\mu}$ values $= 0.0995$

$$\approx \frac{\sigma}{\sqrt{n}}$$

N.B. pdf of $\hat{\mu}$ approximately Gaussian (Central Limit Theorem).

The likelihood function



Suppose the entire result of an experiment (set of measurements) is a collection of numbers x, and suppose the joint pdf for the data x is a function that depends on a set of parameters θ:

$$f(\vec{x}\,;\vec{\theta}\,)$$

Now evaluate this function with the data obtained and regard it as a function of the parameter(s). This is the

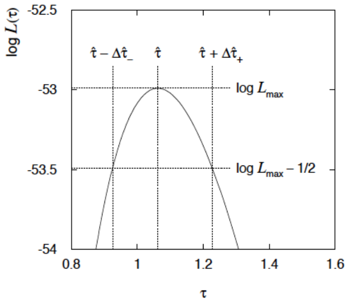
likelihood function:

$$\boxed{L(\vec{\theta}\,) = f(\vec{x}\,;\vec{\theta}\,)}$$

x constant

For θ close to true value, expect high probability of the data we got.
For θ far away from the true value, low probability to have observed what we did !

Independent and identically distributed data



Consider n independent observations of x:
where x follows f(x;θ). The joint pdf for the whole data sample is:

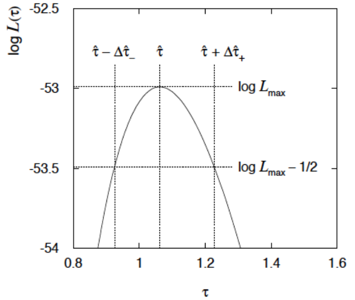
$$f(x_1,\ldots,x_n) = \prod_{i=1}^{n} f(x_i;\theta)$$

In this case the likelihood function is:

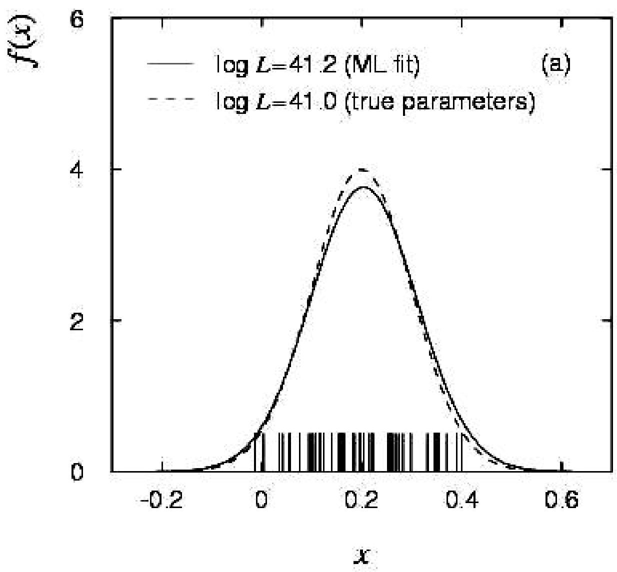
$$L(\vec{\theta}) = \prod_{i=1}^{n} f(x_i;\theta)$$

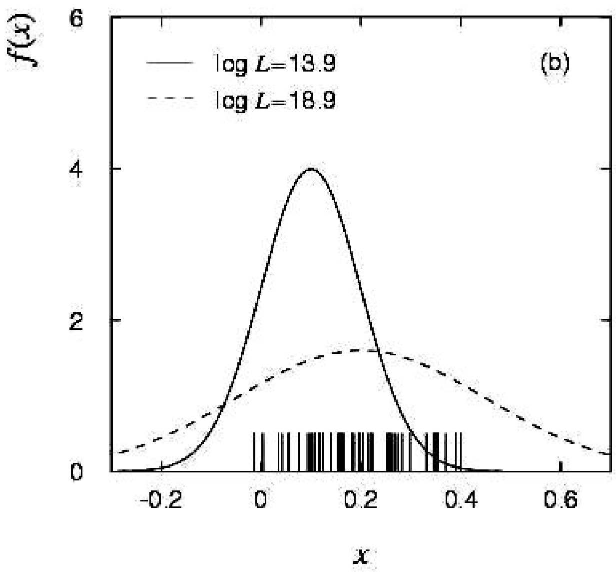
x_i constant

Maximum likelihood estimators



If the hypothesized θ is close to the true value, then we expect a high probability to get data like that which we actually found.

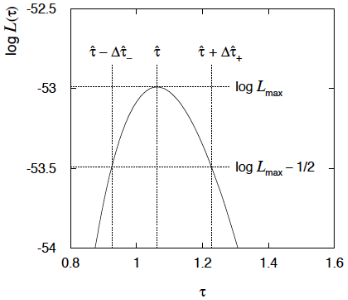




So we define the maximum likelihood (ML) estimator(s) to be the parameter value(s) for which the likelihood is maximum

ML estimators not guaranteed to have any 'optimal' properties, but in practice they are very good

Maximum likelihood estimators



Define ML estimator $\hat{\theta}$ as the value of θ that maximizes L(θ).

We write the estimator as $\hat{\theta}$ with the hat, to distinguish from the true value θ, which may forever remain unknown.

For m parameters, usually find solution $\hat{\theta}_1, \ldots, \hat{\theta}_m$ by solving

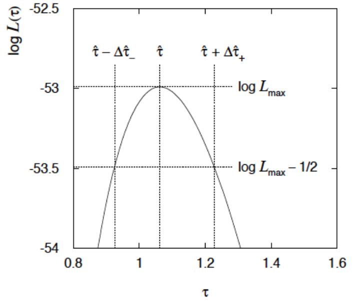
$$\frac{\partial L}{\partial \theta_i} = 0 \qquad i = 1, \ldots, m$$

Sometimes L(θ) has more than one local maximum:

→ take the highest one

* no binning of data ('all information used')

ML example: parameter of expenential pdf



Consider the exponential pdf: $$f(t;\tau) = \frac{1}{\tau} e^{-t/\tau}$$

And suppose we have i.i.d. data: $t_1, \ldots, t_n$

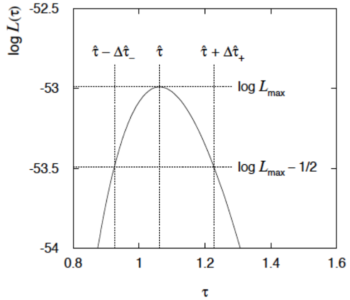
The likelihood function is

$$L(\tau) = \prod_{i=1}^{n} \frac{1}{\tau} e^{-t_i/\tau}$$

The value of τ for which L(τ) is maximum also gives the maximum value of its logarithm (the log-likelihood function):

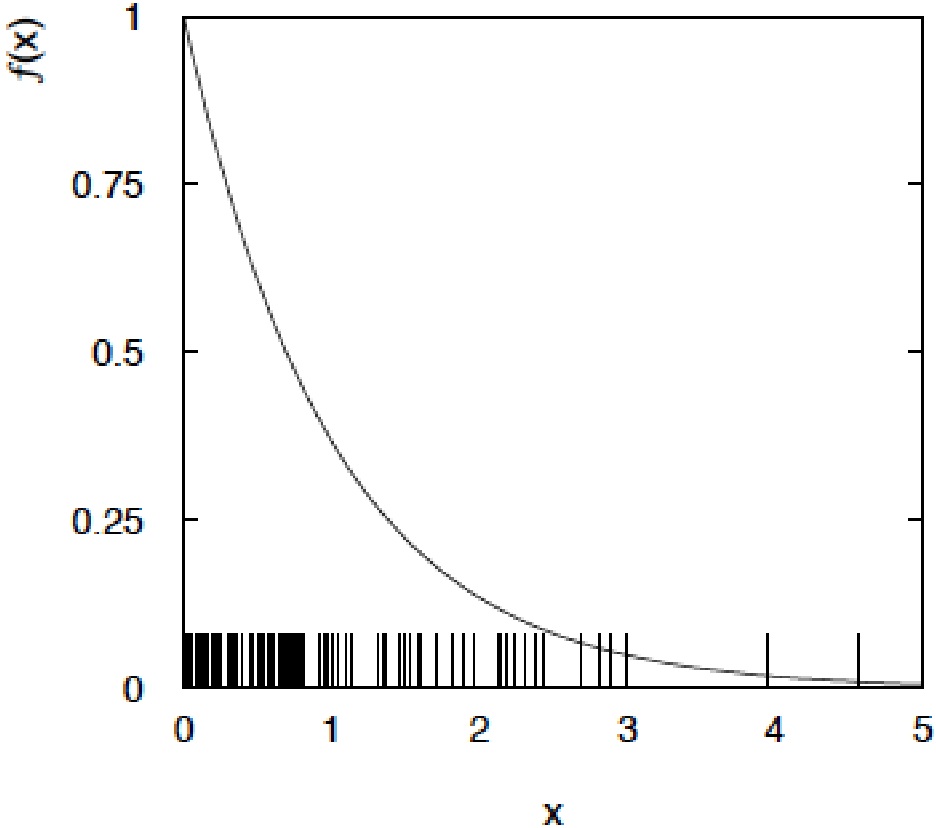
$$\ln L(\tau) = \sum_{i=1}^{n} \ln f(t_i;\tau) = \sum_{i=1}^{n} \left(\ln \frac{1}{\tau} - \frac{t_i}{\tau} \right)$$

ML example: parameter of exponential pdf



Find its maximum by setting $\frac{\partial \ln L(\tau)}{\partial \tau} = 0$

$$\rightarrow \quad \hat{\tau} = \frac{1}{n}\sum_{i=1}^{n} t_i$$



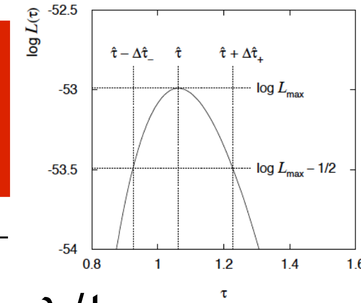
Monte Carlo test:

Generate 50 values using τ = 1:

We find the ML estimate:

$$\hat{\tau} = 1.062$$

Functions of ML estimators



Suppose we had written the exponential pdf as $f(t;\lambda) = \lambda e^{-\lambda/t}$
i.e. we use $\lambda = 1/\tau$ (decay constant). What is the ML estimator for λ?

For a function α(θ) of a parameter θ, it does not matter whether we express L as a function of α or θ.

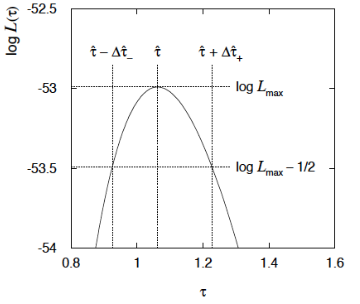
The ML estimator of a function α(θ) is simply $\hat{\alpha} = \alpha(\hat{\theta})$

So for the decay constant we have: $\hat{\lambda} = \frac{1}{\hat{\tau}} = \left(\frac{1}{n} \sum_{i=1}^{n} t_i \right)^{-1}$

Caveat: $\hat{\lambda}$ is biased, even though $\boxed{\hat{\tau} \text{ is unbiased}}$
Can show: $E[\hat{\lambda}] = \lambda \frac{n}{n-1}$ (bias → 0, for n → ∞)

SHOW

ML example: parameters of Gaussian pdf



Consider independent $x_1, \ldots, x_n$, with $x_i \sim$ Gaussian $(\mu, \sigma^2$ unknown$)$

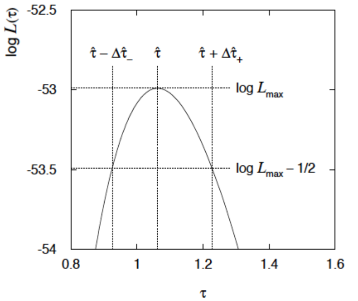
$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

The log likelihood function is:

$$\ln L(\mu, \sigma^2) = \sum_{i=1}^{n} \ln f(x_i; \mu, \sigma^2)$$

$$= \sum_{i=1}^{n} \left(\ln \frac{1}{\sqrt{2\pi}} + \frac{1}{2} \ln \frac{1}{\sigma^2} - \frac{(x_i - \mu)^2}{2\sigma^2} \right)$$

ML example: parameters of Gaussian pdf - 2



Set derivatives with respect to μ, σ^2 to zero and solve:

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

We already know that the estimator for μ is unbiased (see slide 8).

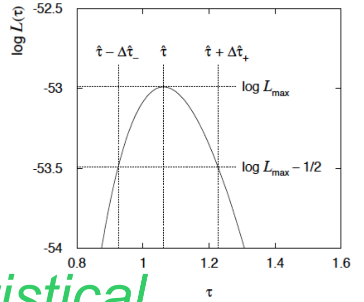
But we find, however: $E[\widehat{\sigma^2}] = \frac{n-1}{n}\sigma^2$

so, the ML estimator for σ^2 has a bias, but $b \to 0$ for $n \to \infty$.
Recall, however, that

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$$

is an unbiased estimator for the variance of ANY pdf.

Variance of estimator: analytic method



Having estimated our parameter we now need to report its "statistical error", i.e. how widely distributed would estimates be if we were to repeat the entire measurement many times.

Recall the estimator for the mean of exponential: $\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n} t_i$

How wide is the pdf $g(\hat{\tau};\tau,n)$?

$$V[\hat{\tau}] = E[\hat{\tau}^2] - (E[\hat{\tau}])^2$$

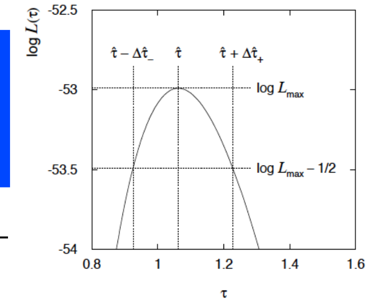
$$= \int \ldots \int \left(\frac{1}{n}\sum_{i=1}^{n} t_i\right)^2 \frac{1}{\tau}e^{-t_1/\tau} \ldots \frac{1}{\tau}e^{-t_n/\tau} dt_1 \ldots dt_n$$

$$- \left(\int \ldots \int \left(\frac{1}{n}\sum_{i=1}^{n} t_i\right) \frac{1}{\tau}e^{-t_1/\tau} \ldots \frac{1}{\tau}e^{-t_n/\tau} dt_1 \ldots dt_n\right)^2$$

$$= \frac{\tau^2}{n}.$$

The variance of $\hat{\tau}$ s n times smaller than the variance of t

Variance of estimator: analytic method



IMPORTANT :

$V[\hat{\tau}]$, $\sigma_{\hat{\tau}}$ are functions of the true (unknown) τ

Estimate using:

$$\hat{\sigma}_{\hat{\tau}} = \frac{\hat{\tau}}{\sqrt{n}}$$

Often given as STATISTICAL ERROR, e.g.

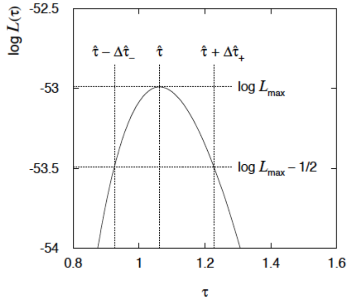
$$\hat{\tau} \pm \hat{\sigma}_{\hat{\tau}} = 1.062 \pm 0.150$$

Meaning: ML estimate for τ is 1.062

ML estimate for the σ of $g(\hat{\tau};\tau,n)$ is 0.150

If $g(\hat{\tau};\tau,n)$ is Gaussian, $[\hat{\tau} - \hat{\sigma}_{\hat{\tau}}, \hat{\tau} + \hat{\sigma}_{\hat{\tau}}]$ same as "68% confidence interval" (more on this soon)

Variance of estimators: Monte Carlo method



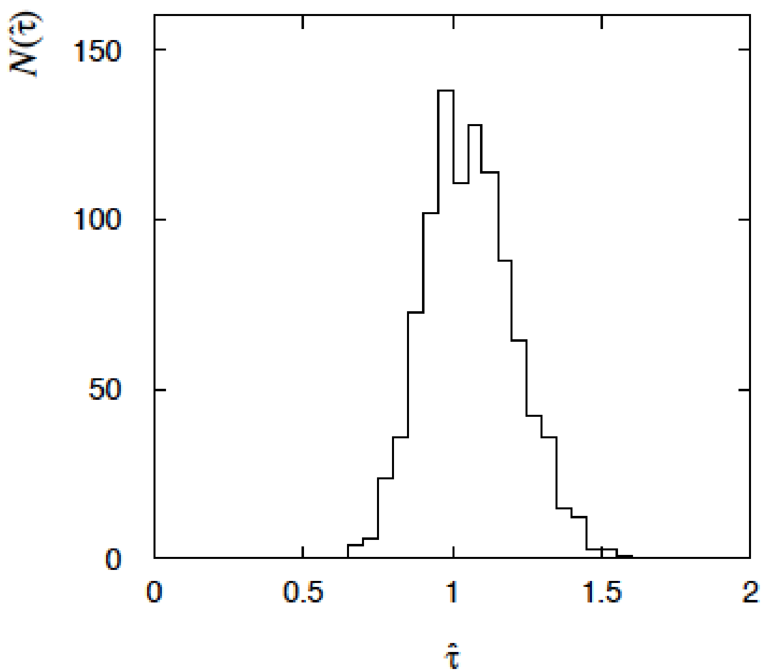
Often the form of $\hat{\theta}, g(\hat{\theta}; \theta, n)$ not known explicitly.

$\rightarrow$ simulate the entire experiment many times with a **Monte Carlo** program.

For the exponential example (slide 17), we had $\hat{\tau} = 1.062$. Take it as "true". Generate 1000 samples (experiments) of n=50 values. Compute $\hat{\tau}$ for each experiment and histogram:

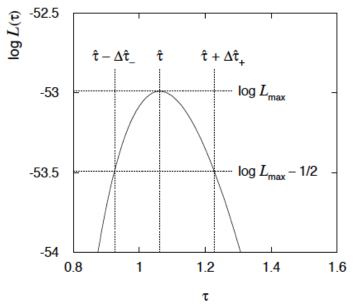
Sample variance of estimates gives:

$$\hat{\sigma}_{\hat{\tau}} = 0.151$$



Note distribution of estimates is roughly Gaussian (central limit theorem) – (almost) always true for ML in large sample limit

Variance of estimators from information inequality



A lower bound on the variance of ANY estimator (not just ML) is:

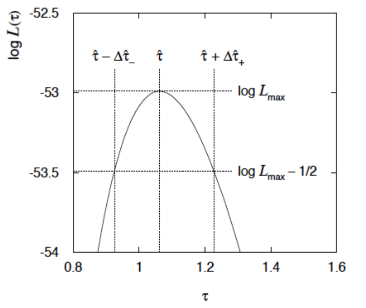
$$V[\hat{\theta}] \geq \left(1 + \frac{\partial b}{\partial \theta}\right)^2 \Big/ E\left[-\frac{\partial^2 \ln L}{\partial \theta^2}\right]$$

Minimum Variance Bound (MVB)

This is the Rao-Cramer-Frechet inequality (information inequality).
If equality is met, $\hat{\theta}$ is said to be efficient.

→ ML estimators are (almost always) efficient for large n,
Often assume this to be true and use RCF bound to estimate

Variance of estimators from information inequality



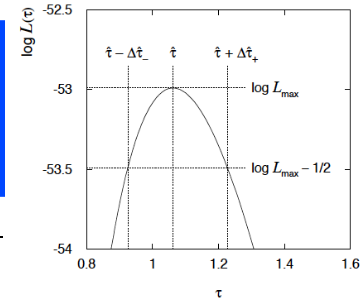
Often the bias b is small, and equality either holds exactly or is a good approximation (e.g. large data sample limit). Then,

$$V[\hat{\theta}] \approx -1/E\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right]$$

Estimate this using the 2nd derivative of lnL at its maximum (function of the true parameters):

$$\hat{V}[\hat{\theta}] = -\left(\frac{\partial^2 \ln L}{\partial \theta^2}\right)^{-1}\Bigg|_{\theta=\hat{\theta}}$$

Variance of estimators: graphical method



Expand $\ln L(\theta)$ about its maximum $\hat{\theta}$:

$$\ln L(\theta) = \ln L(\hat{\theta}) + \left[\frac{\partial \ln L}{\partial \theta}\right]_{\theta=\hat{\theta}} (\theta - \hat{\theta}) + \frac{1}{2!}\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right]_{\theta=\hat{\theta}} (\theta - \hat{\theta})^2 + \ldots$$

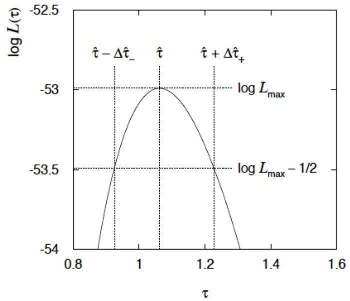
First term is $\ln L_{max}$, second term is zero, third term use information inequality (assume equality):

$$\ln L(\theta) \approx \ln L_{max} - \frac{(\theta-\hat{\theta})^2}{2\widehat{\sigma^2_{\hat{\theta}}}}$$

i.e. $$\ln L(\hat{\theta} \pm \hat{\sigma}_{\hat{\theta}}) \approx \ln L_{max} - \frac{1}{2}$$

→ to get $\hat{\sigma}_{\hat{\theta}}$, change θ away from $\hat{\theta}$ until $\ln L$ decreases by 1/2

Example of variance by graphical method



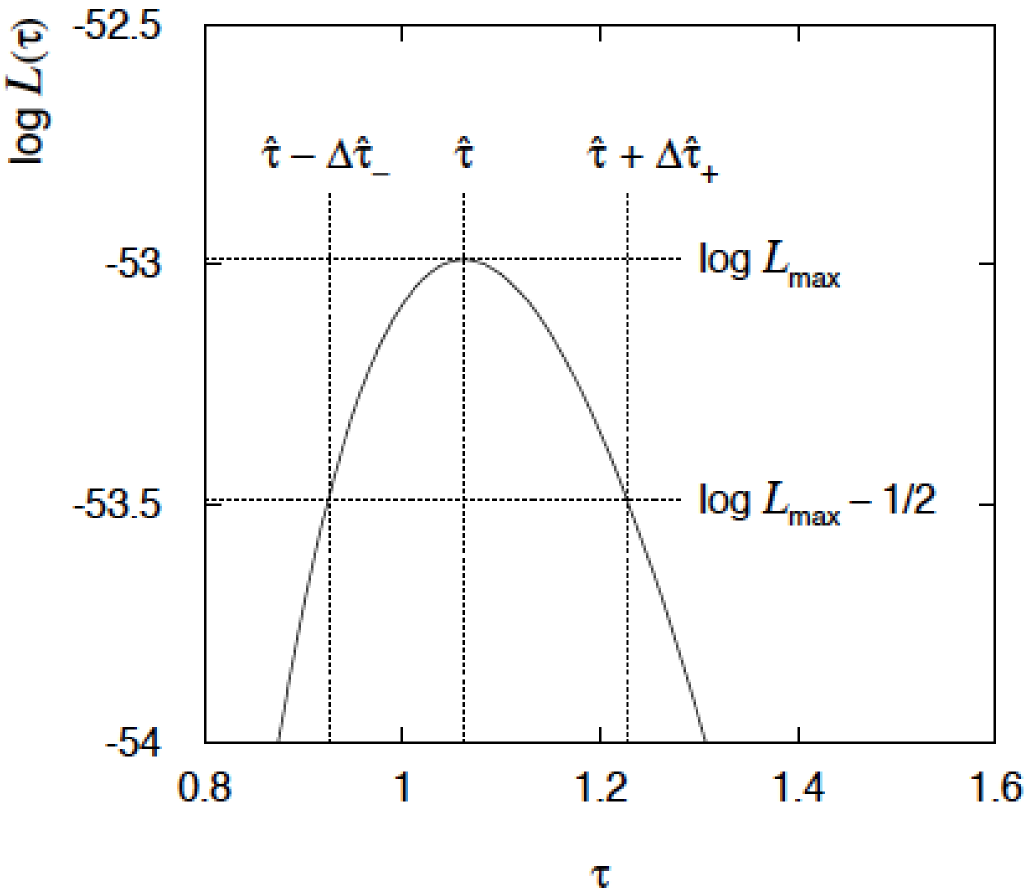
ML example with exponential:

$$\hat{\tau} = 1.062$$

$$\Delta\hat{\tau}_{\text{minus}} = 0.137$$

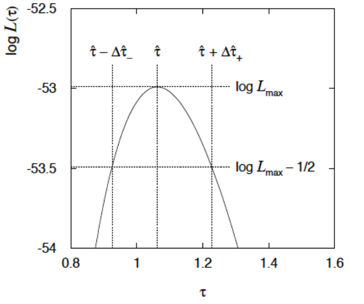
$$\Delta\hat{\tau}_{\text{plus}} = 0.165$$

$$\sigma_{\hat{\tau}} \approx \Delta\hat{\tau}_{\text{minus}} \approx \Delta\hat{\tau}_{\text{plus}} \approx 0.15$$



Not quite parabolic ln L since finite sample size (n=50)

Wrapping up



- Estimators
- Estimators for mean, and variance

- The likelihood function
- Maximum likelihood estimators
- Examples: parameters of exponential and Gaussian pdfs

- Variance of ML estimators
- Difference methods:
  - Analytic
  - Monte Carlo
  - The RCF bound
  - Graphical method