Lecture 2

Estimators:

- use estimator $\hat{\vec{\theta}}$ to estimate the true parameters $\vec{\theta}$ from a sample $\vec{x}$ following some distribution $p(x|\vec{\theta})$
- $\hat{\vec{\theta}}$ should be
  i) unbiased $\vec{b} \equiv \langle \hat{\vec{\theta}} \rangle - \vec{\theta} \overset{!}{=} 0$
  ii) consistent $\lim_{n \to \infty} \hat{\vec{\theta}} = \vec{\theta}$
  iii) efficient $\mathrm{var}(\hat{\vec{\theta}})$ small (often conflicts with (i))

e.g. estimators $\hat{\mu}$ for mean of a sample

	unbiased	consistent	efficient
$\hat{\mu} = \frac{1}{n} \sum x_i$	✓	✓	✓
$\hat{\mu} = \frac{1}{n-1} \sum x_i$	×	✓	×
$\hat{\mu} = x_1$	✓	×	×
$\hat{\mu} = 0$	×	×	✓

e.g. best estimator for variance of sample:

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

Maximum likelihood estimator:

MLE: $\hat{\vec{\theta}} = \arg\max_{\vec{\theta}} \mathcal{L}(\vec{\theta}|\vec{x})$

- not necessarily best estimator
- Can be biased
- consistent
- most efficient for $n \to \infty$
- invariant under coord. transf.
- e.g. for a sample of Gaussian distributed values x_i:

$$\mathcal{L}(\mu, \sigma | \vec{x}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \prod_i \exp\left[\frac{(x_i - \mu)^2}{2\sigma^2}\right]$$

$$\frac{\partial \mathcal{L}}{\partial \mu} \stackrel{!}{=} 0 \Leftrightarrow \frac{\partial \ln \mathcal{L}}{\partial \mu} = 0 \Rightarrow \hat{\mu} = \frac{1}{n}\sum_i x_i \qquad \equiv \text{best estimator}$$

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} \stackrel{!}{=} 0 \Leftrightarrow \frac{\partial \ln \mathcal{L}}{\partial \sigma^2} = 0 \Rightarrow \hat{\sigma^2} = \frac{1}{n}\sum_i (x_i - \bar{x})^2 \quad \text{biased! (still consistent)}$$

- relation to Bayesian parameter estimation:

$$p(\vec{\theta}|\vec{x}) = \frac{\mathcal{L}(\vec{x}|\vec{\theta})\,p(\vec{\theta})}{\int \mathcal{L}(\vec{x}|\vec{\theta})\,p(\vec{\theta})\,d\vec{\theta}}$$

for flat prior $p(\vec{\theta})$: most probable posterior $\equiv$ MLE

<u>Uncertainty of MLE:</u>

i) MC method (Bootstrapping)

- repeat toy experiments and determine MLE $\hat{\vec{\theta}}$ for each
- use sample standard deviation as uncertainty $\hat{\sigma}_{\hat{\theta}}$

ii) Cramér-Rao bound

scalar: $$\mathrm{var}(\hat{\theta}) \geq \underbrace{\frac{1}{I(\theta)}}_{\text{Fisher inform.}} \cdot \underbrace{\left(\frac{\partial \langle\hat{\theta}\rangle}{\partial \theta}\right)^2}_{=1 \text{ if unbiased}}$$

multivariate: $$\mathrm{cov}(\hat{\vec{\theta}}) \geq \underbrace{\frac{\partial \langle\hat{\vec{\theta}}\rangle}{\partial \vec{\theta}} \left[I(\vec{\theta})\right]^{-1} \left(\frac{\partial \langle\hat{\vec{\theta}}\rangle}{\partial \vec{\theta}}\right)^T}_{=1 \text{ if unbiased}}$$

Fisher information: $$I(\vec{\theta})_{ij} := \left\langle \frac{\partial \ln \mathcal{L}}{\partial \theta_i} \frac{\partial \ln \mathcal{L}}{\partial \theta_j} \right\rangle \stackrel{\text{regularity}}{=} \left\langle -\frac{\partial^2 \ln \mathcal{L}}{\partial \theta_i \partial \theta_j} \right\rangle$$

- Cramér-Rao bound only gives lower threshold for any estimator's variance
- if = instead of ≥ : estimator (fully) efficient

  true for MLE for $n \to \infty$

- for large n: $\mathcal{L} \approx$ Gaussian, unbiased

$$\Rightarrow \mathrm{var}(\hat{\theta}) \simeq \frac{1}{\left\langle -\frac{\partial^2 \ln \mathcal{L}}{\partial^2 \theta} \right\rangle} \simeq -\frac{1}{\left.\frac{\partial^2 \ln \mathcal{L}}{\partial^2 \theta}\right|_{\theta=\hat{\theta}}}$$

iii) "Graphical" method

- for large n: $\ln \mathcal{L} \approx$ parabola

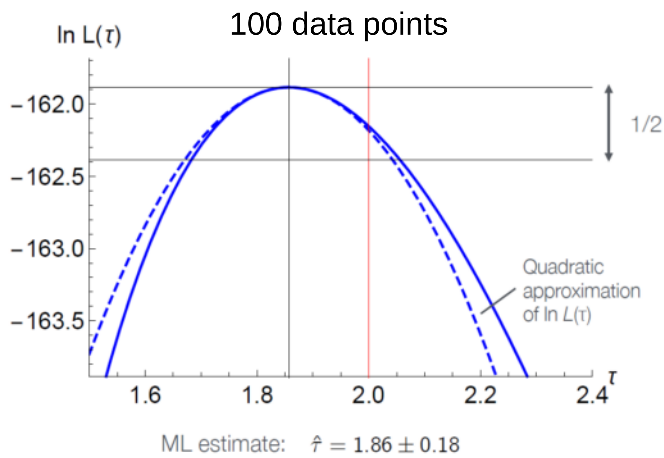
$$\Rightarrow \ln \mathcal{L}(\theta) \simeq \ln \mathcal{L}(\hat{\theta}) + \frac{1}{2\hat{\sigma}^2_{\hat{\theta}}} (\theta - \hat{\theta})^2$$

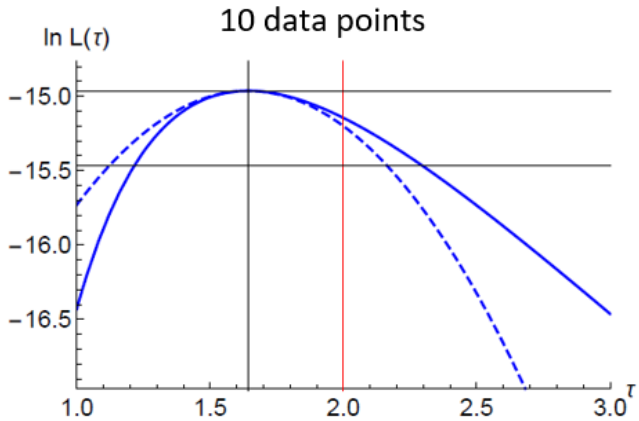
- estimate $\hat{\sigma}_{\hat{\theta}}$ via $\boxed{\ln \mathcal{L}(\hat{\theta} \pm z\hat{\sigma}_{\hat{\theta}}) \simeq \ln \mathcal{L}_{max} - \frac{z^2}{2}}$

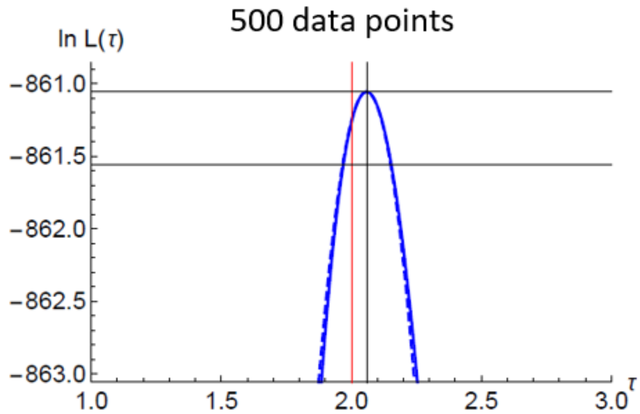
e.g. n data points sampled from

$$p(x|\tau) = \frac{1}{\tau} \exp(-\tau x) \quad , \tau = 2$$

$$\Rightarrow \mathcal{L}(\tau|\vec{x}) = \frac{1}{\tau} \prod_i \exp(-\tau x_i)$$



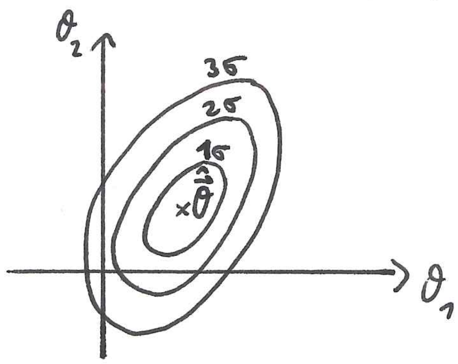




K. Reygers: Statistical Methods in Particle Physics WS 2017/18

- multivariate:

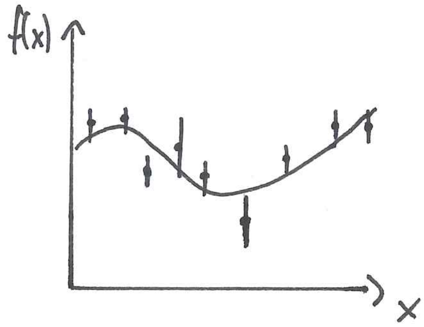
$z \cdot \sigma$ error contours at: $\ln \mathcal{L}(\vec{\theta}) = \ln \mathcal{L}_{max} - \frac{z^2}{2}$



- Graphical method also used when $\mathcal{L}$ not Gaussian
- if $\mathcal{L}$ is Gaussian, this becomes identical to the Cramér-Rao bound

ML Fit:

- consider data points $\{x_i, y_i\}$ measuring some fct. $f(x|\vec{\theta})$ and Gaussian uncertainties



$$\mathcal{L}(\vec{\theta}|\{x_i, y_i\}) \propto \prod_i \exp\left[-\frac{1}{2}\left(\frac{y_i - \langle y_i \rangle}{\sigma_i}\right)^2\right]$$

$$\Rightarrow \ln \mathcal{L} = -\frac{1}{2} \underbrace{\sum_i \left(\frac{y_i - f(x_i|\vec{\theta})}{\sigma_i}\right)^2}_{\chi^2} + C$$

- $\hat{\vec{\theta}}$ maximizes $\mathcal{L} \Leftrightarrow \hat{\vec{\theta}}$ minimizes χ^2 ! (for Gaussian uncertainties)
- error ellipses at: $\chi^2(\vec{\theta}) = \chi^2_{min} + z^2$

Goodness of fit:

- if f correctly describes data, then $\chi^2_{min} = \sum_i \left(\frac{y_i - f(x_i|\hat{\vec{\theta}})}{\sigma_i}\right)^2$ follows a χ^2-distribution:

$p(\chi^2_{min}|ndf) \propto (\chi^2_{min})^{\frac{ndf}{2}-1} e^{-\frac{\chi^2_{min}}{2}}$ where ndf = # data points − # fit params

- $\langle \chi^2_{min} \rangle = ndf$
- assign χ^2 p-value: $p = \int_{\chi^2_{min}}^{\infty} p(t|ndf)\,dt$

where p is the probability to observe data with at least χ^2_{min} (if f is true)

p is not the probability that f is true