Computing and Statistical Data Analysis
Stat 8: More Parameter Estimation

London Postgraduate Lectures on Particle Physics;

University of London MSci course PH4515

Glen Cowan
Physics Department
Royal Holloway, University of London
g.cowan@rhul.ac.uk
www.pp.rhul.ac.uk/~cowan

Course web page:

www.pp.rhul.ac.uk/~cowan/stat_course.html

ML with binned data

Often put data into a histogram: $\vec{n} = (n_1, \ldots, n_N),\ n_{\text{tot}} = \sum_{i=1}^{N} n_i$

Hypothesis is $\vec{\nu} = (\nu_1, \ldots, \nu_N),\ \nu_{\text{tot}} = \sum_{i=1}^{N} \nu_i$ where

$$\nu_i(\vec{\theta}) = \nu_{\text{tot}} \int_{\text{bin } i} f(x; \vec{\theta})\, dx$$

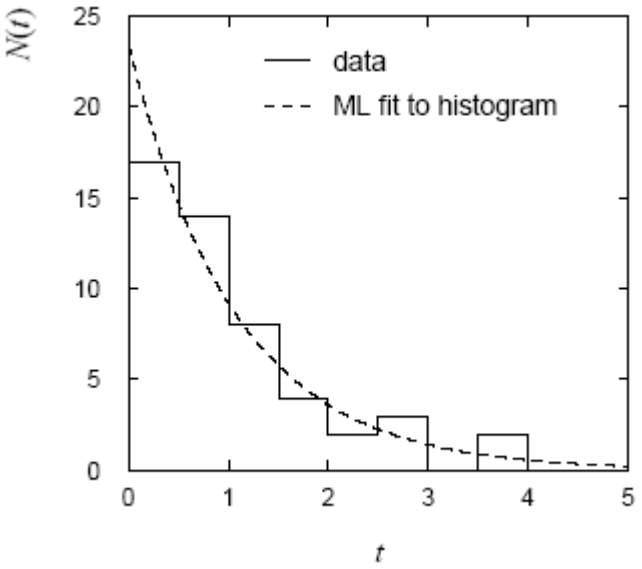
If we model the data as multinomial (n_{tot} constant),

$$f(\vec{n}; \vec{\nu}) = \frac{n_{\text{tot}}!}{n_1! \ldots n_N!} \left(\frac{\nu_1}{n_{\text{tot}}}\right)^{n_1} \cdots \left(\frac{\nu_N}{n_{\text{tot}}}\right)^{n_N}$$

then the log-likelihood function is: $\ln L(\vec{\theta}) = \sum_{i=1}^{N} n_i \ln \nu_i(\vec{\theta}) + C$

ML example with binned data

Previous example with exponential, now put data into histogram:



$\hat{\tau} = 1.07 \pm 0.17$

(1.06 ± 0.15 for unbinned ML with same sample)

Limit of zero bin width → usual unbinned ML.

If n_i treated as Poisson, we get extended log-likelihood:

$$\ln L(\nu_{\text{tot}}, \vec{\theta}) = -\nu_{\text{tot}} + \sum_{i=1}^{N} n_i \ln \nu_i(\nu_{\text{tot}}, \vec{\theta}) + C$$

Relationship between ML and Bayesian estimators

In Bayesian statistics, both θ and x are random variables:

$$L(\theta) = L(\vec{x}|\theta) = f_{\text{joint}}(\vec{x}|\theta)$$

Recall the Bayesian method:

Use subjective probability for hypotheses (θ);

before experiment, knowledge summarized by prior pdf $\pi(\theta)$;

use Bayes' theorem to update prior in light of data:

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

Posterior pdf (conditional pdf for θ given x)

ML and Bayesian estimators (2)

Purist Bayesian: $p(\theta \mid x)$ contains all knowledge about θ.

Pragmatist Bayesian: $p(\theta \mid x)$ could be a complicated function,

→ summarize using an estimator $\hat{\theta}_{\text{Bayes}}$

Take mode of $p(\theta \mid x)$, (could also use e.g. expectation value)

What do we use for $\pi(\theta)$? No golden rule (subjective!), often represent 'prior ignorance' by $\pi(\theta)$ = constant, in which case

$$\hat{\theta}_{\text{Bayes}} = \hat{\theta}_{\text{ML}}$$

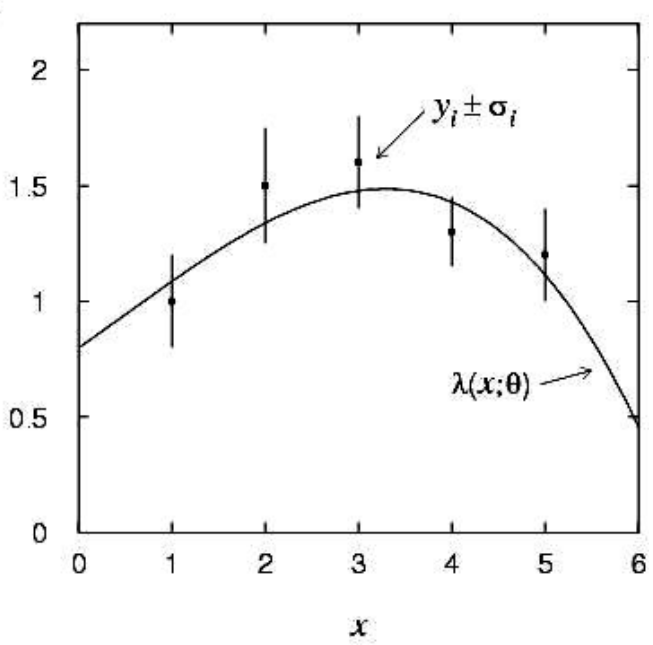
But... we could have used a different parameter, e.g., $\lambda = 1/\theta$, and if prior $\pi_\theta(\theta)$ is constant, then $\pi_\lambda(\lambda)$ is not!

'Complete prior ignorance' is not well defined.

The method of least squares

Suppose we measure N values, y_1, ..., y_N, assumed to be independent Gaussian r.v.s with

$$E[y_i] = \lambda(x_i; \theta) \ .$$



Assume known values of the control variable x_1, ..., x_N and known variances

$$V[y_i] = \sigma_i^2 \ .$$

We want to estimate θ, i.e., fit the curve to the data points.

The likelihood function is

$$L(\theta) = \prod_{i=1}^{N} f(y_i; \theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(y_i - \lambda(x_i; \theta))^2}{2\sigma_i^2}\right]$$

The method of least squares (2)

The log-likelihood function is therefore

$$\ln L(\theta) = -\frac{1}{2}\sum_{i=1}^{N}\frac{(y_i - \lambda(x_i;\theta))^2}{\sigma_i^2} + \text{terms not depending on } \theta$$

So maximizing the likelihood is equivalent to minimizing

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

Minimum defines the least squares (LS) estimator $\hat{\theta}$.

Very often measurement errors are ~Gaussian and so ML and LS are essentially the same.

Often minimize χ^2 numerically (e.g. program **MINUIT**).

LS with correlated measurements

If the y_i follow a multivariate Gaussian, covariance matrix V,

$$g(\vec{y}, \vec{\lambda}, V) = \frac{1}{(2\pi)^{N/2}|V|^{1/2}} \exp\left[-\frac{1}{2}(\vec{y} - \vec{\lambda})^T V^{-1} (\vec{y} - \vec{\lambda})\right]$$

Then maximizing the likelihood is equivalent to minimizing

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