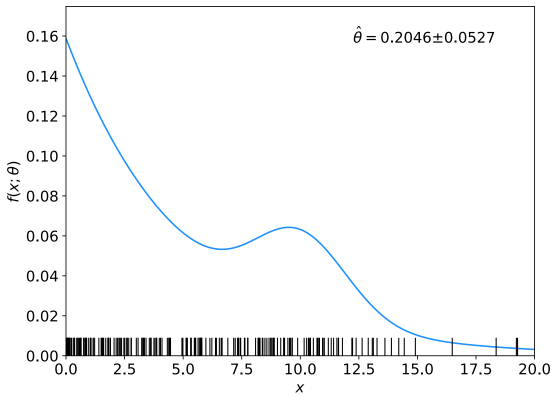
Discussion session notes 14 Dec 2020

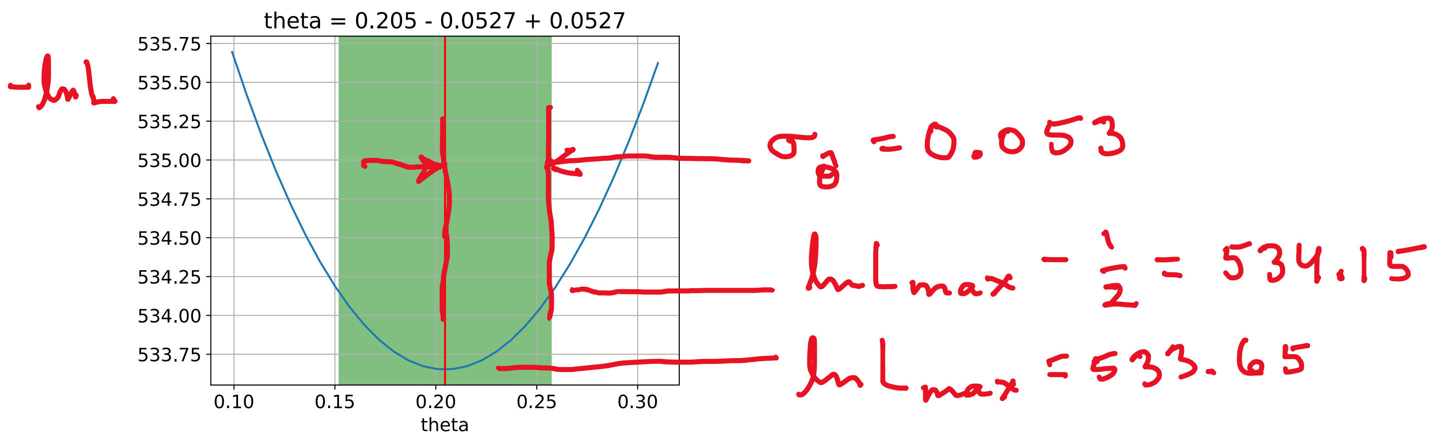
G. Cowan / RHUL Physics
PH4515 Problem Sheet 8

1a) [6 marks]   Running the program mlFit.py produces the following plots:

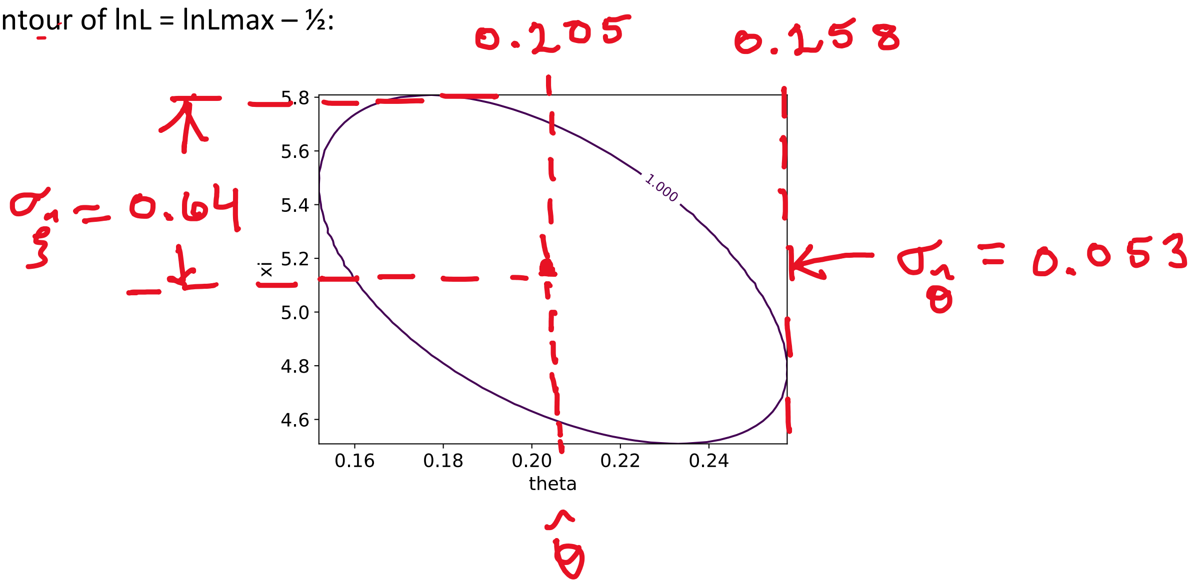
A fit of the pdf:



A scan of -lnL versus theta:



A contour of lnL = lnLmax – ½:



1b) [6 marks]

$$1\,b)\quad V^{-1}_{ij} = -\int \frac{\partial^2 \ln P(\vec{x}|\vec{\theta})}{\partial\theta_i\,\partial\theta_j}\, P(\vec{x}|\vec{\theta})\, d\vec{x}$$

Use $P(\vec{x}|\vec{\theta}) = \prod_{i=1}^{n} f(x_i;\vec{\theta})$ i.i.d. sample

$$\Rightarrow V^{-1}_{ij} = -\int \frac{\partial^2 \sum_{k=1}^{n} \ln f(x_k;\vec{\theta})}{\partial\theta_i\,\partial\theta_j} \times \prod_{l=1}^{n} f(x_l;\vec{\theta})\, dx_l$$

$$= -\sum_{k=1}^{n} \underbrace{\int \frac{\partial^2 \ln f(x_k;\vec{\theta})}{\partial\theta_i\,\partial\theta_j}\, f(x_k;\vec{\theta})\, dx_k}_{\text{all } n \text{ terms equal}} \cdot \prod_{l\neq k} \underbrace{\int f(x_l;\vec{\theta})\, dx_l}_{=1}$$

$$= -n \int \frac{\partial^2 \ln f(x;\vec{\theta})}{\partial\theta_i\,\partial\theta_j}\, f(x;\vec{\theta})\, dx$$

$\propto n$

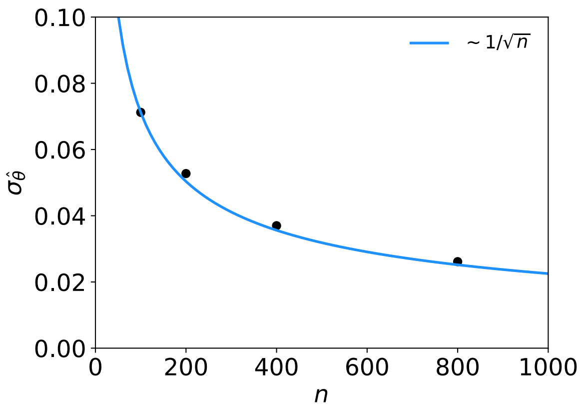
$V V^{-1} = I \Rightarrow$ if $V^{-1} \propto n$, $V \propto \frac{1}{n}$,

$\sigma_{\hat{\theta}_i} = \sqrt{V_{ii}} \propto \frac{1}{\sqrt{n}}$ for all i

1(c) [6 marks] Running mlFit.py with different numbers of events gave:

numVal	thetaHat	sigma_thetaHat
100	0.197218	0.071219
200	0.204551	0.052736
400	0.160808	0.036985
800	0.198224	0.026129

A plot of sigma_thetaHat versus numVal is shown below. The standard deviation of the estimator is seen to decrease as $1/\sqrt{n}$, as expected.



1(d) [6 marks] The results of the fit with different combinations of parameters adjustable are:

Free	Fixed	sigma_thetaHat
theta	mu, sigma, xi	0.044535
theta, xi	mu, sigma	0.052736
theta, xi, sigma	mu	0.064456
theta, xi, sigma, mu	--	0.085786

As can be seen, the standard deviation of the estimator of theta increases when it is fitted simultaneously with an increasing number of other adjustable parameters.

Discussion Session Problem 1: The binomial distribution is given by

$$P(n; N, \theta) = \frac{N!}{n!(N-n)!} \theta^n (1-\theta)^{N-n} ,$$

where n is the number of 'successes' in N independent trials, with a success probability of θ for each trial. Recall that the expectation value and variance of n are $E[n] = N\theta$ and $V[n] = N\theta(1-\theta)$, respectively. Suppose we have a single observation of n and using this we want to estimate the parameter θ.

1(a) Find the maximum likelihood estimator $\hat{\theta}$.

1(b) Show that $\hat{\theta}$ has zero bias and find its variance.

1(c) Suppose we observe $n = 0$ for $N = 10$ trials. Find the upper limit for θ at a confidence level of CL = 95% and evaluate numerically.

1(d) Suppose we treat the problem with the Bayesian approach using the Jeffreys prior, $\pi(\theta) \propto \sqrt{I(\theta)}$, where

$$I(\theta) = -E\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right]$$

is the expected Fisher information. Find the Jeffreys prior $\pi(\theta)$ and the posterior pdf $p(\theta|n)$ as proportionalities.

1(e) Explain how in the Bayesian approach how one would determine an upper limit on θ using the result from (d). (You do not actually have to calculate the upper limit.)

Explain briefly the differences in the interpretation between frequentist and Bayesian upper limits.

Solution:

1(a) The likelihood function is given by the binomial distribution evaluated with the single observed value n and regarded as a function of the unknown parameter θ:

$$L(\theta) = \frac{N!}{n!(N-n)!} \theta^n (1-\theta)^{N-n} .$$

The log-likelihood function is therefore

$$\ln L(\theta) = n \ln \theta + (N-n) \ln(1-\theta) + C ,$$

where C represents terms not depending on θ. Setting the derivative of $\ln L$ equal to zero,

$$\frac{\partial \ln L}{\partial \theta} = \frac{n}{\theta} - \frac{N-n}{1-\theta} = 0 ,$$

we find the ML estimator to be

$$\hat{\theta} = \frac{n}{N} .$$

1(b) We are given the expectation and variance of a binomial distributed variable as $E[n] = N\theta$ and $V[n] = N\theta(1-\theta)$. Using these results we find the expectation value of $\hat{\theta}$ to be

$$E[\hat{\theta}] = E\left[\frac{n}{N}\right] = \frac{E[n]}{N} = \frac{N\theta}{N} = \theta \; ,$$

and therefore the bias is $b = E[\hat{\theta}] - \theta = 0$. Similarly we find the variance to be

$$V[\hat{\theta}] = V\left[\frac{n}{N}\right] = \frac{1}{N^2}V[n] = \frac{N\theta(1-\theta)}{N^2} = \frac{\theta(1-\theta)}{N} \; .$$

1(c) Suppose we observe $n = 0$ for $N = 10$ trials. The upper limit on θ at a confidence level of $\mathrm{CL} = 1 - \alpha$ is the value of θ for which there is a probability α to find as few events as we found or fewer, i.e.,

$$\alpha = P(n \leq 0; N, \theta) = \frac{N!}{0!(N-0)!}\theta^0(1-\theta)^{N-0} \; .$$

Solving for θ gives the 95% CL upper limit

$$\theta_{\mathrm{up}} = 1 - \alpha^{1/N} = 1 - 0.05^{1/10} = 0.26 \; .$$

1(d) To find the Jeffreys prior we need the second derivative of $\ln L$,

$$\frac{\partial^2 \ln L}{\partial \theta^2} = -\frac{n}{\theta^2} - \frac{N-n}{(1-\theta)^2} \; .$$

The expected Fisher information is therefore

$$I(\theta) = -E\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right] = \frac{N\theta}{\theta^2} + \frac{N(1-\theta)}{(1-\theta)^2} = \frac{N}{\theta} + \frac{N}{1-\theta} = \frac{N}{\theta(1-\theta)} \; .$$

The Jeffreys prior is therefore

$$\pi(\theta) \propto \frac{1}{\sqrt{\theta(1-\theta)}} \; .$$

Using this in Bayes theorem to find the posterior pdf gives

$$p(\theta|n) \propto L(n|\theta)\pi(\theta) \propto \frac{\theta^n(1-\theta)^{N-n}}{\sqrt{\theta(1-\theta)}} = \theta^{n-1/2}(1-\theta)^{N-n-1/2} \; .$$

1(e) To find a Bayesian upper limit on θ one simply integrates the posterior pdf so that a specified probability $1-\alpha$ is contained below θ_{up}, i.e.,

$$1 - \alpha = \int_0^{\theta_{\mathrm{up}}} p(\theta|n)\, d\theta \; ,$$

solving for θ_{up} gives the upper limit.

A frequentist upper limit as found in (c) is a function of the data designed to be greater than the true value of the parameter with a fixed probability (the confidence level) regardless of the parameter's actual value. A Bayesian interval can be regarded as reflecting a range for the parameter where it is believed to lie with a fixed probability (the credibility level). Note that with the Jeffreys prior, one may not necessary use the degree of belief interpretation of the interval, but rather take it to have a certain probability to cover the true θ (which in general will depend on θ).

Simplified "Errors on Errors" Model

The model in Lectures 11-3, 11-4

Details in: G. Cowan, *Statistical Models with Uncertain Error Parameters*, Eur. Phys. J. C (2019) 79:133, arXiv:1809.05778

makes a distinction between the $\sigma_{y,i}$ (~statistical errors), which are known, and the $\sigma_{u,i}$ ~systematic errors), which are treated as adjustable parameters.

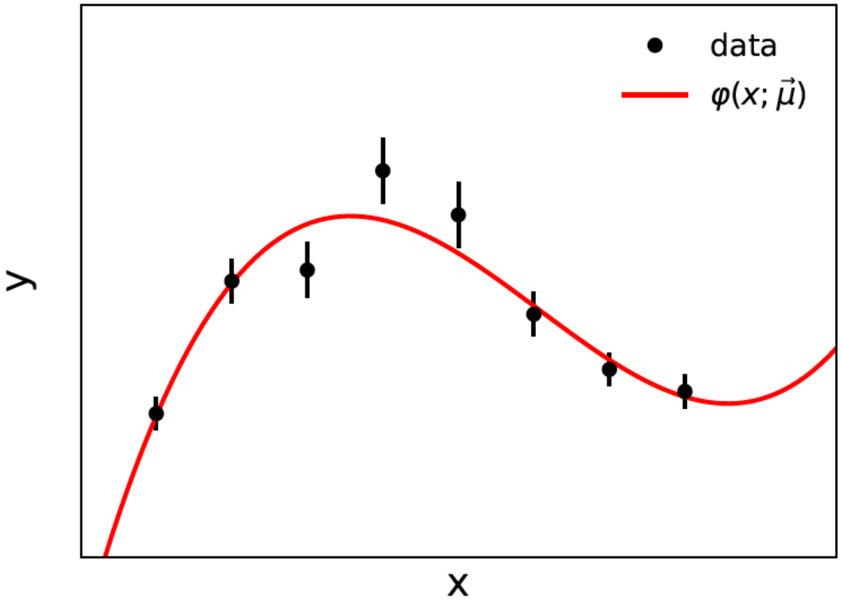
Here we show a simplified model that does not distinguish between statistical and systematic errors.

Curve fitting, averages

Suppose independent
$y_i \sim$ Gauss, $i = 1,\ldots,N$, with

$$E[y_i] = \varphi(x_i; \boldsymbol{\mu})$$

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



$\boldsymbol{\mu}$ are the parameters in the fit function $\varphi(x;\boldsymbol{\mu})$.

If we take the σ_i as known, we have the usual log-likelihood

$$\ln L(\boldsymbol{\mu}) = -\frac{1}{2} \sum_{i=1}^{N} \frac{(y_i - \varphi(x_i; \boldsymbol{\mu}))^2}{\sigma_i^2}$$

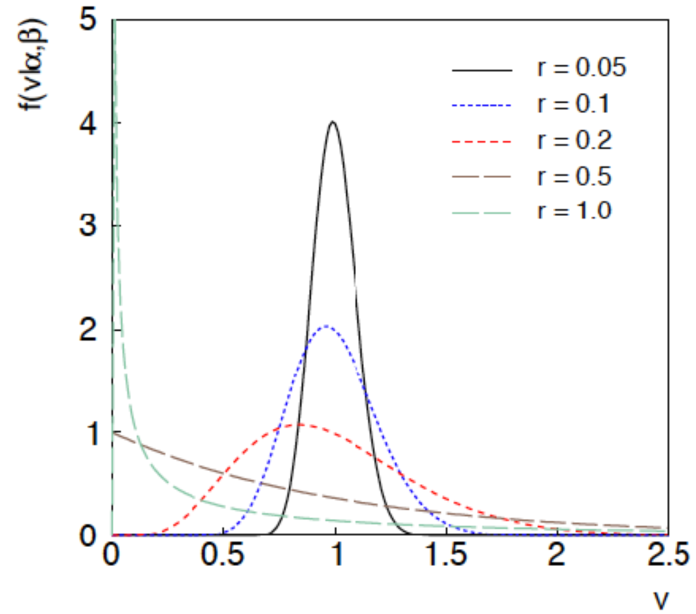
which leads to the Least Squares estimators for $\boldsymbol{\mu}$.

Model with uncertain σ_i^2

If the σ_i^2 are uncertain, we can take them as adjustable parameters.

The estimated variances $v_i = s_i^2$ are modeled as gamma distributed.

The likelihood becomes



$$L(\boldsymbol{\mu}, \boldsymbol{\sigma^2}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-(y_i - \varphi(x_i;\boldsymbol{\mu}))^2/2\sigma_i^2} \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} v_i^{\alpha_i - 1} e^{-\beta_i v_i}$$

Want $E[v_i] = \sigma_i^2$ $\quad \frac{\sigma_{s_i}}{E[s_i]} \approx r_i$ $\quad (s_i = \sqrt{v_i})$

→ $\alpha_i = \frac{1}{4r_i^2}$ $\quad \beta_i = \frac{\alpha_i}{\sigma_i^2}$

Profile log-likelihood

One can profile over the σ_i^2 in close form.

The log-profile-likelihood is

$$\ln L'(\boldsymbol{\mu}) = \ln L(\boldsymbol{\mu}, \widehat{\widehat{\boldsymbol{\sigma}^2}}) = -\frac{1}{2}\sum_{i=1}^{N}\left(1+\frac{1}{2r_i^2}\right)\ln\left[1+2r_i^2\frac{(y_i-\varphi(x_i;\boldsymbol{\mu}))^2}{v_i}\right]$$

Quadratic terms replace by sum of logs.

Equivalent to replacing Gauss pdf for y_i by Student's t, $\nu_{\text{dof}} = 1/2r_i^2$

Confidence interval for $\boldsymbol{\mu}$ becomes sensitive to goodness-of-fit (increases if data internally inconsistent).

Fitted curve less sensitive to outliers.

Simple program for Student's t average: stave.py