3

Parameter estimation

- A simple hypothesis is a proposition without parameters while a composite hypothesis depends on one or more parameters
- Assuming the composite hypothesis to be true, our task is to estimate the parameter values from the data at our disposal
- The Bayesian procedure is quite simple in principle: just use Bayes' theorem to invert the likelihood of the data, given the parameters, to obtain the posterior distribution of the parameters, given the data
- Note that the plausibility of the hypothesis itself is not investigated here; that investigation is called model selection
- For a Frequentist, a parameter has a fixed, but unknown, value and is thus not a random variable. This means that Bayes' theorem cannot be used so that Frequentist parameter estimation has to go via the construction of a statistic

Formal outline (Bayesian)

- Given a parameterized model, the data ***d*** can be described by a likelihood function

$$p(\boldsymbol{d}|\boldsymbol{a}, \boldsymbol{s}, I)$$

- The parameters of interest are denoted by ***a***

- The parameters ***s*** are often needed to describe detector effects like efficiency, acceptance, *etc.* These are called nuisance parameters

- Given the priors for ***a*** and ***s*** the posterior is

$$p(\boldsymbol{a}, \boldsymbol{s}|\boldsymbol{d}, I) \propto p(\boldsymbol{d}|\boldsymbol{a}, \boldsymbol{s}, I)\, p(\boldsymbol{a}, \boldsymbol{s}|I)$$

- Integration over the parameters ***s*** gives the desired result

$$p(\boldsymbol{a}|\boldsymbol{d}, I) = \int p(\boldsymbol{a}, \boldsymbol{s}|\boldsymbol{d}, I)\, \mathrm{d}\boldsymbol{s}$$

- This marginalization over ***s*** is a very elegant way to propagate uncertainties in ***s*** to the parameters ***a*** (systematic error propagation)

Likelihood or sampling distribution?

The conditional probability density

$$p(\boldsymbol{d}|\boldsymbol{a})$$

is called a

- Likelihood when it is regarded as a function of the parameters ***a*** for fixed data ***d***. This is how it is used in Bayes' theorem

$$p(\boldsymbol{a}|\boldsymbol{d}) \propto p(\boldsymbol{d}|\boldsymbol{a})\ p(\boldsymbol{a})$$

Note that the likelihood is <u>not</u> a probability density

- Sampling distribution when it is regarded as a function of the data ***d*** for fixed parameters ***a***. This is how it is used to generate data in a Monte Carlo, for example

The simplest example imaginable

- We draw one measurement $x = 3$ from a Gauss distribution with known width $\sigma = 5$ but with unknown mean μ
- What is the best estimate of μ?
- We will now answer this question using both Bayesian and Frequentist inference

The Bayesian answer

- Assuming a uniform prior the posterior is

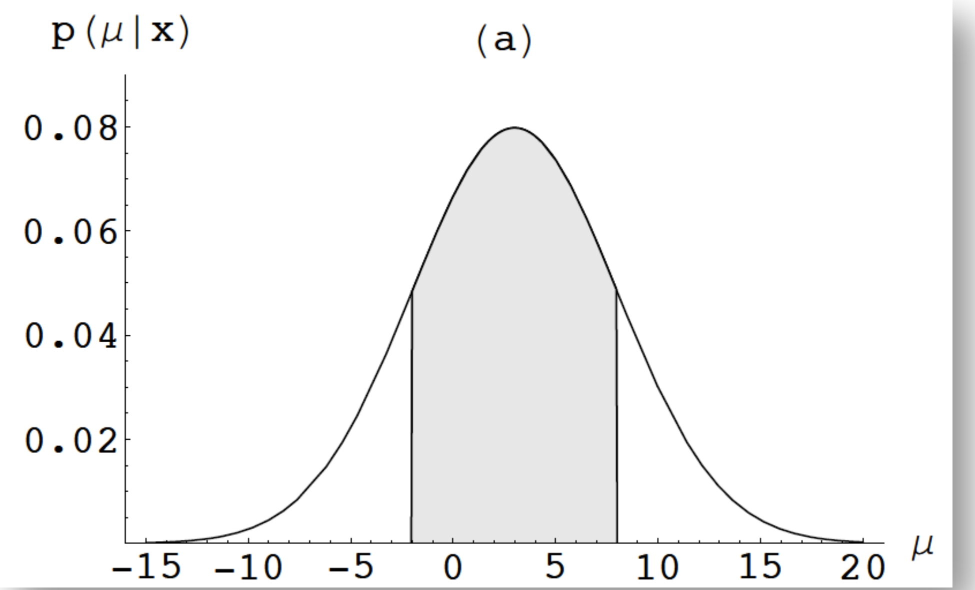
$$p(\mu|x) = C\, p(x|\mu)\, p(\mu) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

- Here we plot the posterior for $x = 3$ and $\sigma = 5$

- The shaded area shows the 68% credible interval

$$\mu = 3 \pm 5 \quad (68.3\%\ \mathrm{CL})$$

$$P(-2 < \mu < 8) = 68.3\%$$



The Frequentist answer

- Cannot use Bayes' Theorem so must go via a statistic
- A good statistic to estimate μ is the sample mean which reduces in our case (one measurement) to the measurement X itself
- Here X is a random variable, Gaussian distributed with unkown μ and known $\sigma = 5$
- The realisations $\{x_1, x_2, \ldots\}$ of X are not random variables
- We can make probability statements about X but not x

- For our Gauss we can make the probabilistic statement

$$P(\mu - \sigma < X < \mu + \sigma) = 68.3\%$$

- Straight forward manipulation of inequalities gives

$$P(X - \sigma < \mu < X + \sigma) = 68.3\%$$

- It is tempting to substitute our observation $x = 3$

$$P(x - \sigma < \mu < x + \sigma) = P(-2 < \mu < 8) = 68.3\%$$

- Forbidden since now the argument of *P*() is not a RV !
- In fact, it would allow for probability inversion without specifying a prior (which is nonsense)

$$P(X - \sigma < \mu < X + \sigma) = 68.3\%$$

This statement tells us, as Frequentists, that

To each observation x is associated a confidence interval $x \pm \sigma$ corresponding to a pre-defined confidence level (here 68%)

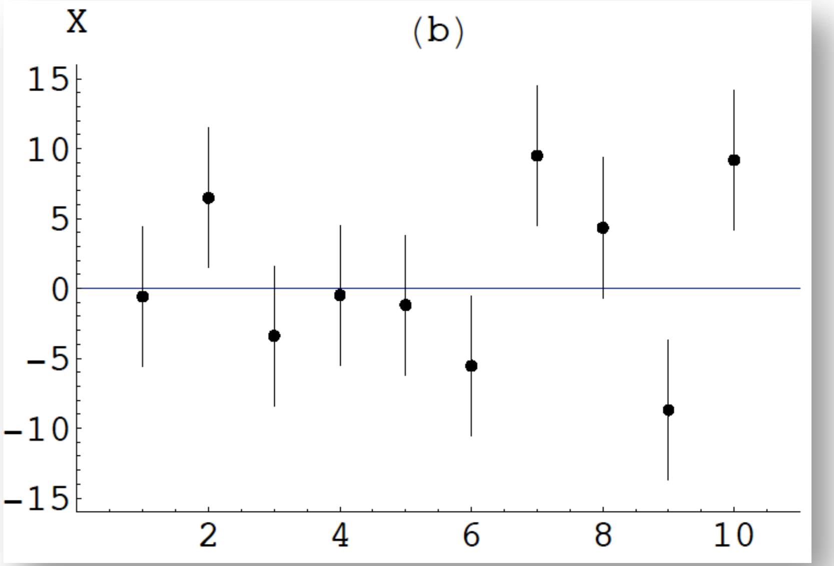
The confidence intervals of repeated observations will contain the unknown μ in 68% of the cases

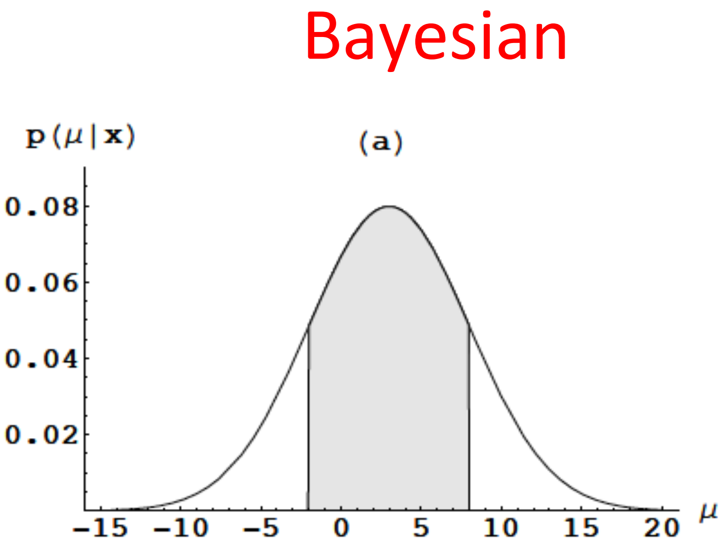
First 10 observations

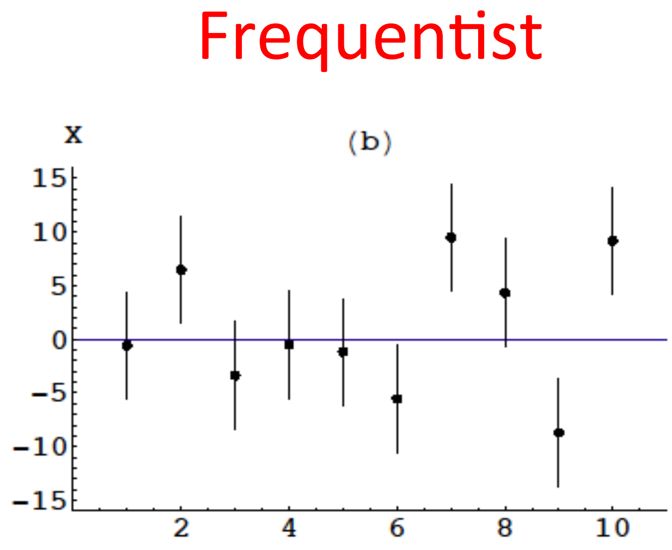
Five out of ten confidence intervals contain the unknown μ

Frequentist coverage

For an infinite number of observations this is 68%







Both schools will report $\mu = 3 \pm 5$

- In the Bayesian case it is clear what it means
- In the Frequentist case it is less clear (I leave it as an exercise for you to formulate it correctly …)
- At least you should now understand that the statement

$$P(-2 < \mu < 8) = 68.3\%$$

does not exist in the Frequentist world!

Get the most probable parameter value

- We are often interested in the mode of the posterior
- Obtained by minimising the log posterior $L(\boldsymbol{a}) = -\ln p(\boldsymbol{a}|\boldsymbol{d})$
- This is done by solving the equation $\mathrm{d}L(\hat{\boldsymbol{a}})/\mathrm{d}\boldsymbol{a} = 0$
- Taylor expansion yields the approximation (write-up p.17)

$$p(\boldsymbol{a}|\boldsymbol{d}) \approx C \exp[-\tfrac{1}{2}(\boldsymbol{a} - \hat{\boldsymbol{a}})\boldsymbol{H}(\boldsymbol{a} - \hat{\boldsymbol{a}})]$$

- ***H*** is the Hessian matrix of second derivatives $H_{ij} \equiv \dfrac{\partial^2 L(\hat{\boldsymbol{a}})}{\partial a_i \partial a_j}$

This is, in fact, what MINUIT is doing numerically for you

- The second order approximation in parameter space is

$$p(\boldsymbol{a}|\boldsymbol{d}) \approx C \exp[-\tfrac{1}{2}(\boldsymbol{a} - \hat{\boldsymbol{a}})\boldsymbol{H}(\boldsymbol{a} - \hat{\boldsymbol{a}})]$$

- The constant *C* can be adjusted to either $C = p(\hat{\boldsymbol{a}}|\boldsymbol{d})$
- Or to a value that normalises the right hand side
- This yields a multivariate Gauss in parameter space

$$p(\boldsymbol{a}|\boldsymbol{d}) \approx \frac{1}{\sqrt{(2\pi)^n\,|\boldsymbol{V}|}} \exp\left[-\tfrac{1}{2}(\boldsymbol{a} - \hat{\boldsymbol{a}})\boldsymbol{V}^{-1}(\boldsymbol{a} - \hat{\boldsymbol{a}})\right]$$

- Covariance matrix is the inverse of the Hessian $\boldsymbol{V} \equiv \boldsymbol{H}^{-1}$
- Note that MINUIT errors are the sqrt of the diagonal elements

$$a_i \pm \sigma_i \quad \text{with} \quad \sigma_i \equiv \sqrt{V_{ii}}$$

- Our innocent looking formula…

$$p(\boldsymbol{a}|d) = \frac{\int p(\boldsymbol{d}|\boldsymbol{a},\boldsymbol{s})\,p(\boldsymbol{a},\boldsymbol{s}|I)\,\mathrm{d}\boldsymbol{s}}{\iint p(\boldsymbol{d}|\boldsymbol{a},\boldsymbol{s})\,p(\boldsymbol{a},\boldsymbol{s}|I)\,\mathrm{d}\boldsymbol{a}\mathrm{d}\boldsymbol{s}}$$

… may in fact hide an enormous amount of numerical computation to evaluate the integrals, in particular when the estimation problem is multi-dimensional

- Considerable simplifications occur when
  - ⇒ Variables are independent (densities then factorise)
  - ⇒ Distributions are Gaussian (easy to marginalise)
  - ⇒ Model is linear in the parameters (easy to minimise)

Simplification I: Independent Measurements

Posterior   Product of likelihoods   Prior

$$p(\boldsymbol{a}|\boldsymbol{d}) = C\left[\prod_{i=1}^{n} p_i(d_i|\boldsymbol{a})\right] p(\boldsymbol{a}|I)$$

$$L(\boldsymbol{a}) = -\ln[p(\boldsymbol{a}|\boldsymbol{d})] = -\ln(C) - \ln[p(\boldsymbol{a}|I)] - \sum_{i=1}^{n} \ln[p_i(d_i|\boldsymbol{a})]$$

Normalization   Log prior   Log likelihood

When the prior is uniform the log prior is constant and only the log likelihood is relevant ⇨ log-likelihood minimisation

Simplification II: Gaussian Sampling Distribution

$$L(\boldsymbol{a}) = \text{constant} + \tfrac{1}{2}\chi^2 = \text{constant} + \tfrac{1}{2}\sum_{i=1}^{n}\left[\frac{d_i - \mu_i(\boldsymbol{a})}{\sigma_i}\right]^2$$

⇨ chi-squared or least squares minimization

Simplification III: Linear function

$$f(x;\boldsymbol{a}) = \sum_{\lambda=1}^{m} a_\lambda f_\lambda(x)$$

⇨ minimization by matrix inversion (write-up p.49-50)

$$\hat{\boldsymbol{a}} = \boldsymbol{W}^{-1}\boldsymbol{b} \qquad W_{\lambda\mu} = \sum_{i=1}^{n} w_i f_\lambda(x_i) f_\mu(x_i) \qquad b_\mu = \sum_{i=1}^{n} w_i d_i f_\mu(x_i)$$

Gaussian sampling

- We want to estimate the mean μ of a Gaussian distribution with <u>known</u> σ (experimental resolution) from which a sample of N measurements is drawn

- The likelihood of one datum is

$$p(x_i|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right]$$

- If the data are independent the total likelihood is

$$p(\boldsymbol{x}|\mu,\sigma) = \prod_{i=1}^{N} p(x_i|\mu,\sigma)$$

- A uniform prior for μ gives for the posterior

$$p(\mu|\boldsymbol{x}, \sigma) = p(\boldsymbol{x}|\mu, \sigma)$$

- The negative logarithm of the posterior is then

$$L(\mu) = \text{Const.} + \tfrac{1}{2} \sum_{i=1}^{N} \left(\frac{x_i - \mu}{\sigma} \right)^2$$

- From this we find

$$\frac{\mathrm{d}L}{\mathrm{d}\mu} = 0 \;\rightarrow\; \mu_0 = \bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

$$\frac{\mathrm{d}^2 L}{\mathrm{d}\mu^2} = \frac{N}{\sigma^2}$$

Hessian

- So we obtain $\mu_0 = \bar{x} \pm \dfrac{\sigma}{\sqrt{N}}$

Estimate of Gaussian mean μ when σ is unknown

- As before, our likelihood is

$$p(\boldsymbol{x}|\mu,\sigma) = \frac{1}{(\sigma\sqrt{2\pi})^N} \exp\left[-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{x_i-\mu}{\sigma}\right)^2\right]$$

- From Bayes' theorem we get the posterior

$$p(\mu,\sigma|\boldsymbol{x}) = p(\boldsymbol{x}|\mu,\sigma)\,p(\mu,\sigma)$$

- Assume a Jeffrey's prior

$$p(\mu,\sigma) = 1/\sigma \text{ for } \sigma > 0 \text{ and } 0 \text{ otherwise}$$

- Integration over σ gives the posterior for μ

$$p(\mu|\boldsymbol{x}) = \int_0^\infty p(\mu,\sigma|\boldsymbol{x})\,\mathrm{d}\sigma$$

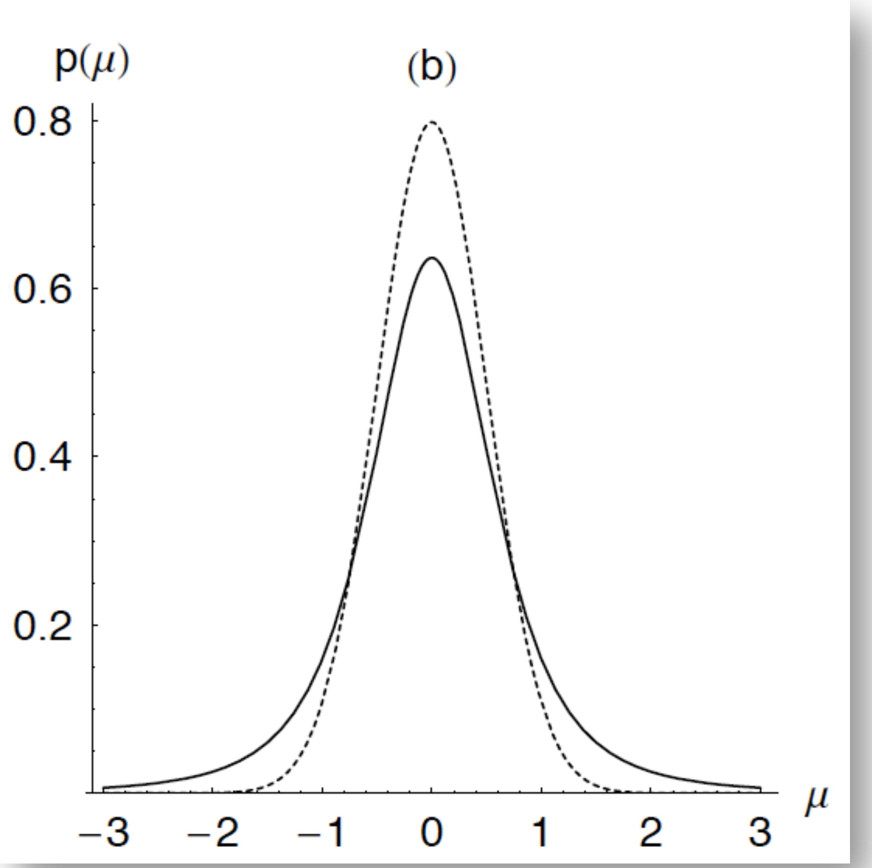
The result is the Student-t distribution

$$p(\mu|\boldsymbol{x}) \propto \left[V + N(\bar{x} - \mu)^2\right]^{-N/2}$$

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$V = \sum_{i=1}^{N}(x_i - \bar{x})^2$$

$$\mu = \bar{x} \pm \sqrt{\frac{V}{N(N-3)}}$$



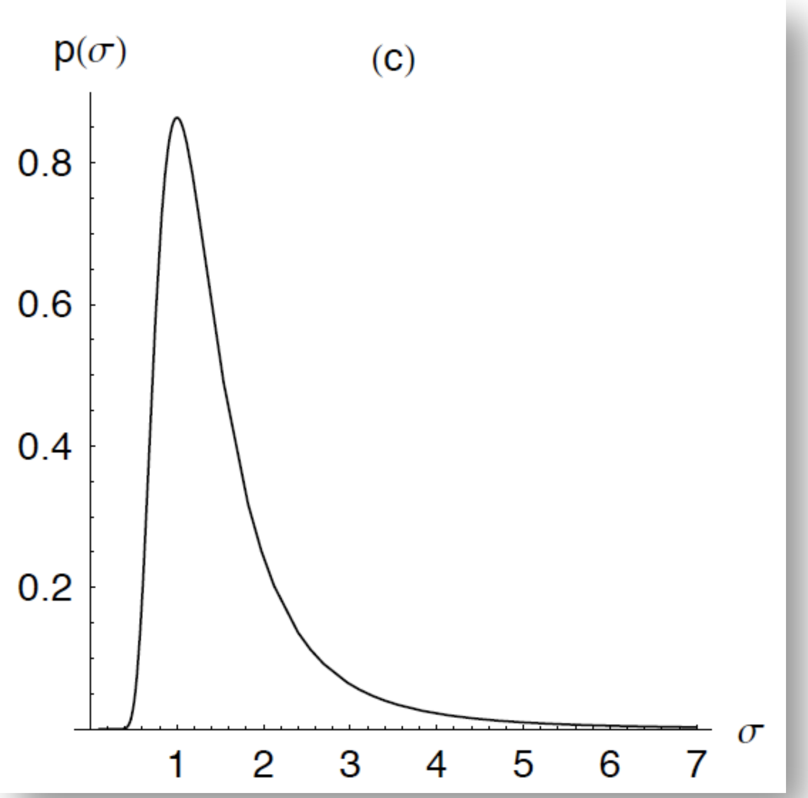
- We can also integrate over μ to get the posterior of σ

$$p(\sigma|\boldsymbol{x}) = \int_{-\infty}^{\infty} p(\mu, \sigma|\boldsymbol{x})\, \mathrm{d}\mu \propto \frac{1}{\sigma^N} \exp\left(-\frac{V}{2\sigma^2}\right)$$

- The result is the χ^2 distribution for $N-1$ degrees of freedom

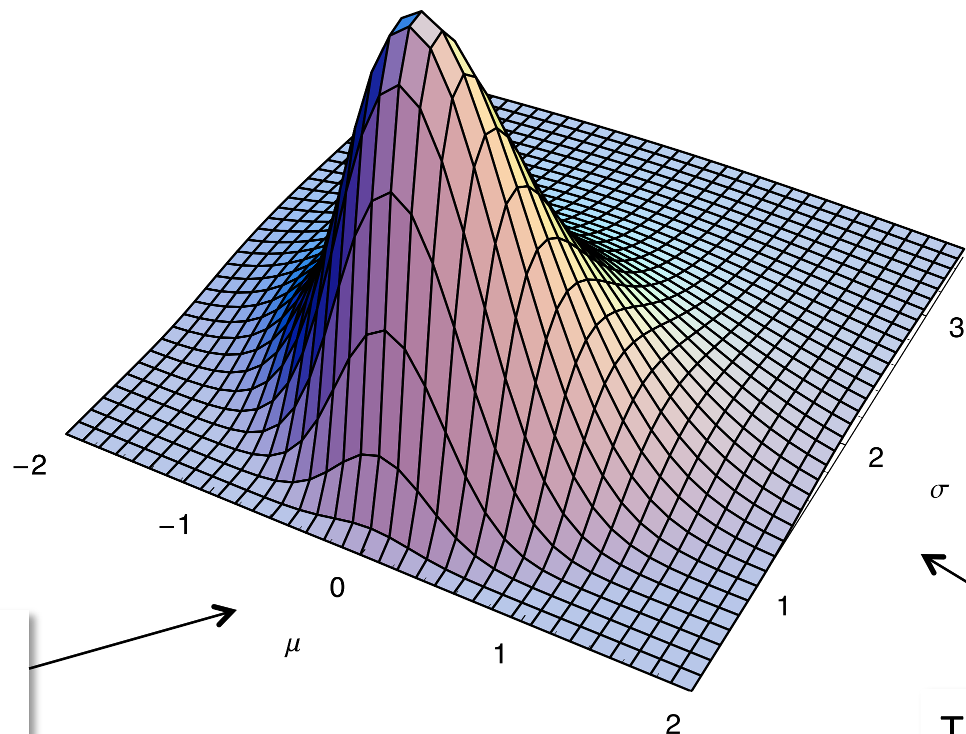
- Mode:

$$\hat{\sigma} = \sqrt{\frac{V}{N}}$$



Here is the joint posterior distribution of μ and σ from four samples drawn from a normal distribution

$$p(\mu, \sigma | \boldsymbol{x}) \propto \frac{1}{\sigma^{N+1}} \exp\left[-\frac{V + N(\bar{x} - \mu)^2}{2\sigma^2}\right]$$



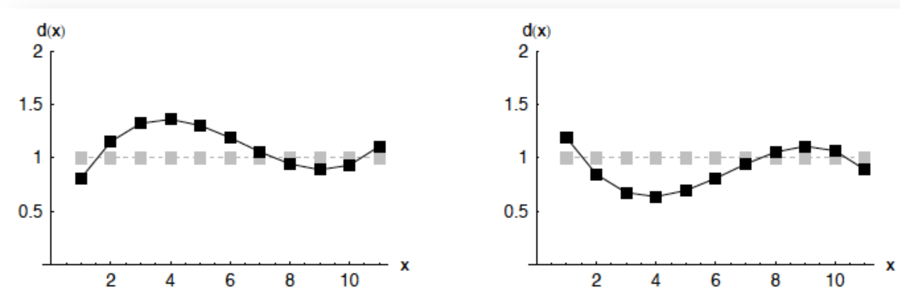
This projection is the Student-t distribution

This projection is the chi2 distribution

Correlated data errors

In the write-up it is described how to properly handle correlated errors (systematic bias) in the data

Lots of mathematical detail so it cannot be presented here …



where C is a normalisation constant and $p(N|I)$ is the prior for the D^0 yield N. The posterior for N is now obtained by integrating (6.28) over the background parameters $\boldsymbol{a}$. As explained in Section 3.1 this yields a one-dimensional Gauss with a variance given by the diagonal element $\sigma^2 = V_{NN}$ of the covariance matrix. Thus we have

$$p(N|\boldsymbol{d}, I) = \frac{C}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{N-\hat{N}}{\sigma}\right)^2\right] p(N|I) = C\,\mathcal{G}(N;\hat{N},\sigma)\,p(N|I) \qquad (6.29)$$

where we have introduced the short-hand notation $\mathcal{G}$ for a one-dimensional Gaussian distribution. In (6.29), $\hat{N} = -0.36$ and $\sigma = 0.74$ as obtained from the fit to the data.

As a last step we encode in the prior $p(N|I)$ our knowledge that N is positive definite

$$P(N|I) \propto \theta(N) \qquad \text{with} \qquad \theta(N) = \begin{cases} 0 & \text{for } N < 0 \\ 1 & \text{for } N \geq 0. \end{cases} \qquad (6.30)$$

Inserting (6.30) in (6.29) and integrating over N to calculate the constant C we find

$$p(N|\boldsymbol{d}, I) = \mathcal{G}(N;\hat{N},\sigma)\,\theta(N)\left[\int_0^\infty \mathcal{G}(N;\hat{N},\sigma)\,\mathrm{d}N\right]^{-1}. \qquad (6.31)$$

The posterior distribution is thus a truncated Gaussian with mean and variance as obtained from the fit. This Gaussian is set to zero for $N < 0$ and re-normalised to unity for $N \geq 0$. The upper limit ($N_{\max}$), corresponding to a given confidence level (CL) is then calculated by (numerically) solving the equation

$$\mathrm{CL} = \int_0^{N_{\max}} p(N|\boldsymbol{d}, I)\,\mathrm{d}N = \int_0^{N_{\max}} \mathcal{G}(N;\hat{N},\sigma)\,\mathrm{d}N \left[\int_0^\infty \mathcal{G}(N;\hat{N},\sigma)\,\mathrm{d}N\right]^{-1}. \qquad (6.32)$$

Using the numbers quoted above this gives $N(D^0 + \bar{D}^0) < 1.5$ per event at 98% CL.

6.5 Correlated data errors

Data are often subject to sources of uncertainty which cause a simultaneous fluctuation of more than one data point. We will call these correlated uncertainties **systematic**, in contrast to point to point un-correlated errors, which we will call **statistical**.

To propagate the systematic uncertainties to the parameters $\boldsymbol{\theta}$ of interest one often offsets the data by each systematic error in turn, redo the analysis, and then add the deviations from the optimal values $\hat{\boldsymbol{\theta}}$ in quadrature. Such an intuitive *ad hoc* procedure (**offset method**) has no sound theoretical foundation and may even spoil your result by assigning errors which are far too large, see [19] for an illustrative example and also Exercise 6.10 below.

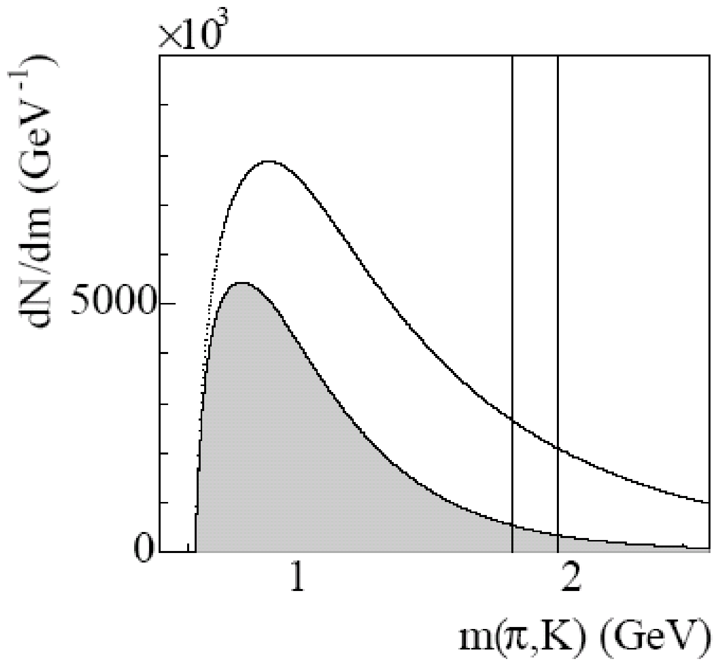
To take systematic errors into account we will include them in the data model. This can of course be done in many ways, depending on the experiment being analysed. Here we restrict ourselves to a linear parametrisation which has the advantage that it is easily incorporated in any least squares minimisation procedure. This model, as it stands,

52

A few examples where the least squares machinery does not really work

1. No signal observed

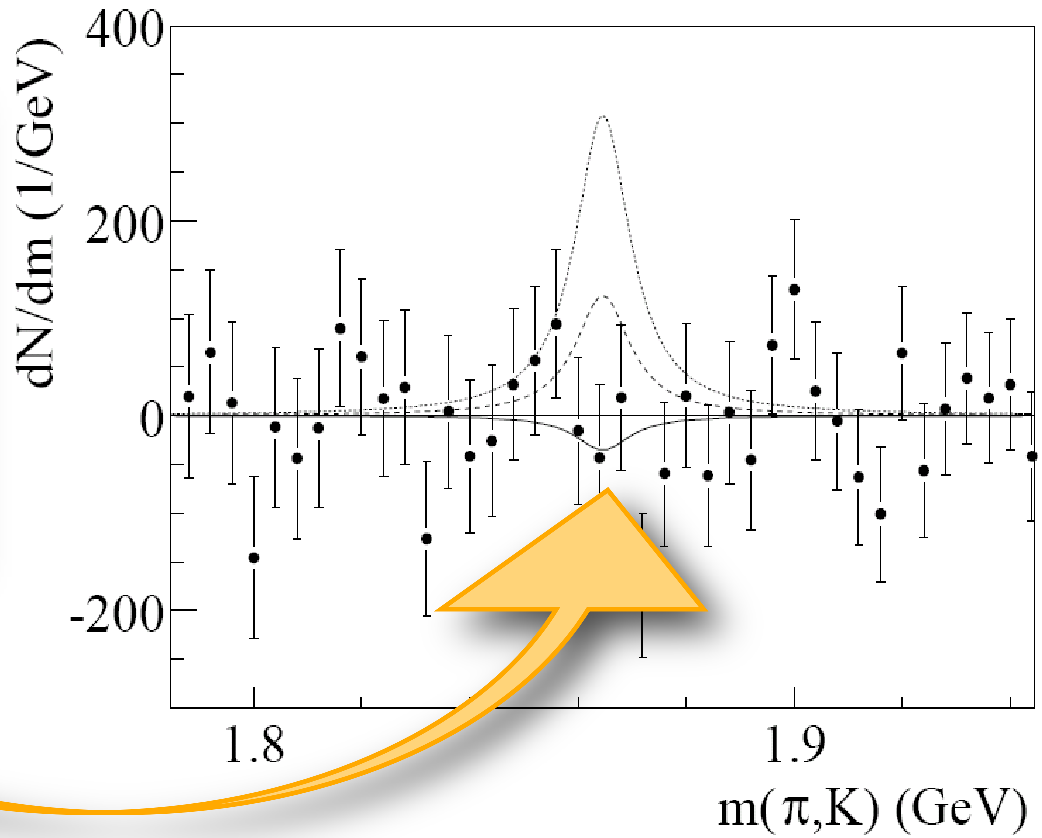
- NA49: search for $D^0 \rightarrow K\pi$ in 4×10^6 Pb-Pb events at the CERN SPS
- Large combinatorial background in the invariant mass spectrum is due to 1400 charged tracks per event



NA49 D^0 yield

$N(D^0) = -0.36 \pm 0.74$

No signal is observed after background subtraction



⇨ Use Bayesian method to get the upper limit

- Parameters $\boldsymbol{\theta} = (N, \boldsymbol{a}) \begin{cases} N & \mathrm{D}^0 \text{ yield} \\ \boldsymbol{a} & \text{background parameters} \end{cases}$
- Write likelihood as Gaussian in parameter space

$$p(\boldsymbol{d}|\boldsymbol{\theta}, I) = \frac{1}{\sqrt{(2\pi)^n\,|\boldsymbol{V}|}} \exp\left[-\tfrac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\boldsymbol{V}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right]$$

- Assume uniform prior in the background parameters

$$p(\boldsymbol{\theta}|I) = p(N, \boldsymbol{a}|I) \propto p(N|I)$$

- The posterior now becomes

$$p(\boldsymbol{\theta}|\boldsymbol{d}, I) = \frac{C}{\sqrt{(2\pi)^n\,|\boldsymbol{V}|}} \exp\left[-\tfrac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\boldsymbol{V}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right] p(N|I)$$

- Integrate over background (nuisance) parameters ***a***

$$p(N|\boldsymbol{d},I) = \frac{C}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{N-\hat{N}}{\sigma}\right)^2\right] p(N|I)$$

- Now encode that *N* is positive definite

$$P(N|I) \propto \theta(N) \qquad \text{with} \qquad \theta(N) = \begin{cases} 0 & \text{for } N < 0 \\ 1 & \text{for } N \geq 0 \end{cases}$$

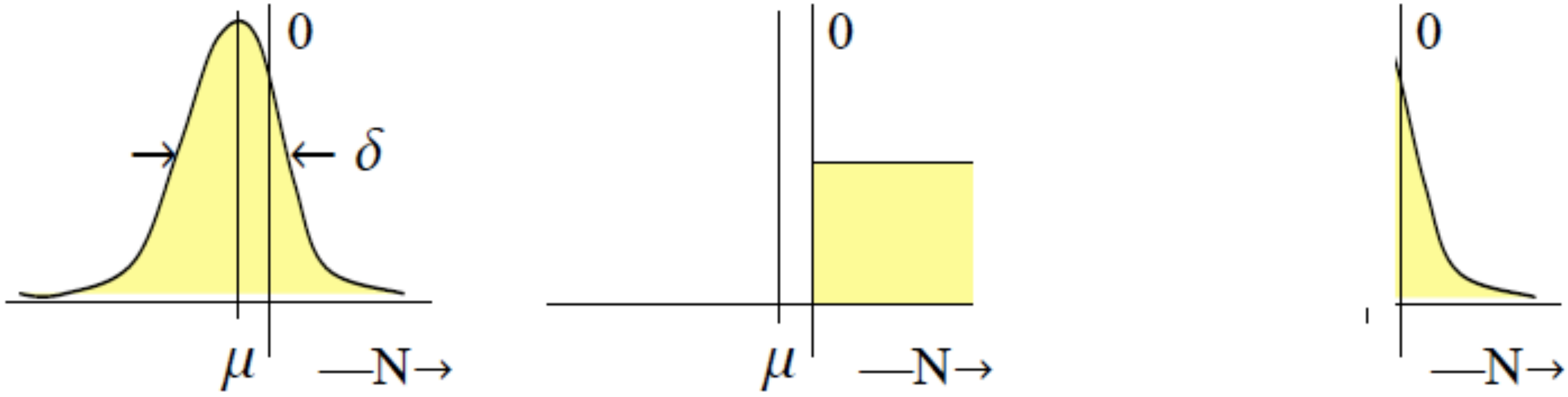
- Calculate upper limit by solving the equation

$$\mathrm{CL} = \int_0^{N_{\max}} p(N|\boldsymbol{d},I)\,\mathrm{d}N$$

- This gave the result

$$N(\mathrm{D}^0) < 1.5 \text{ per event at } 98\% \text{ CL}$$

Posterior for N(D0) per event



Likelihood × Prior = Posterior

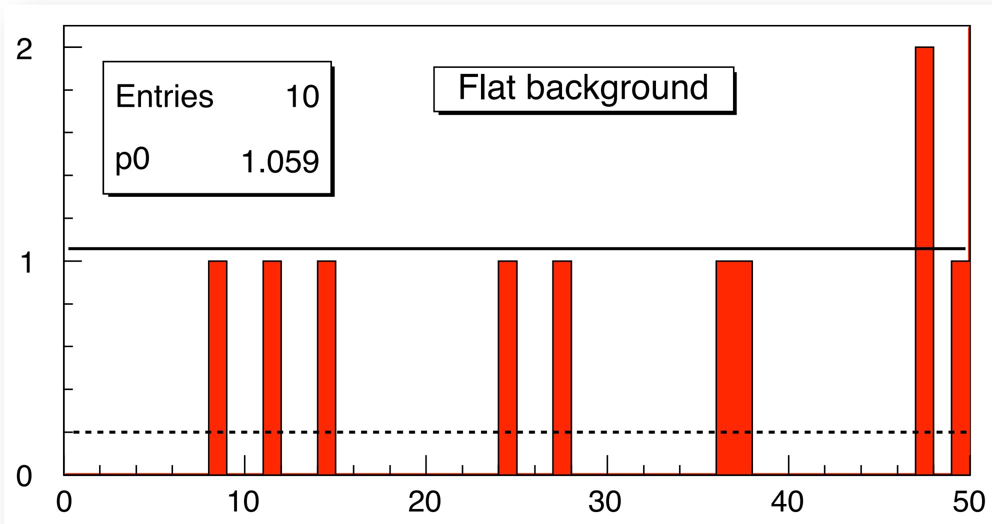
Probability to observe the data if there are N D0 per event. Obtain the likelihood from χ^2 fit

Encodes our prior knowledge that $N > 0$ (and nothing else)

The $N > 0$ tail of the likelihood, normalized to unity

2. Sparsely filled histogram

Average rate per bin is 10/50 = 0.2



Least squares fit to the data gives 1.06 counts per bin

Reason for the discrepancy

- Contents are Poisson and not Gauss distributed
- Empty bins are ignored

$$\chi^2 = \sum \frac{(n_i - R)^2}{n_i}$$

Start from the correct likelihood!

- $$p(\boldsymbol{n}|R) = \prod_i \frac{R^{n_i}}{n_i!}\, e^{-R} \propto p(R|\boldsymbol{n}) \quad \text{(uniform prior)}$$
- This gives for the log posterior

$$L = -\ln(p) = \sum_i \left[R - n_i \ln(R)\right] + \text{constant}$$

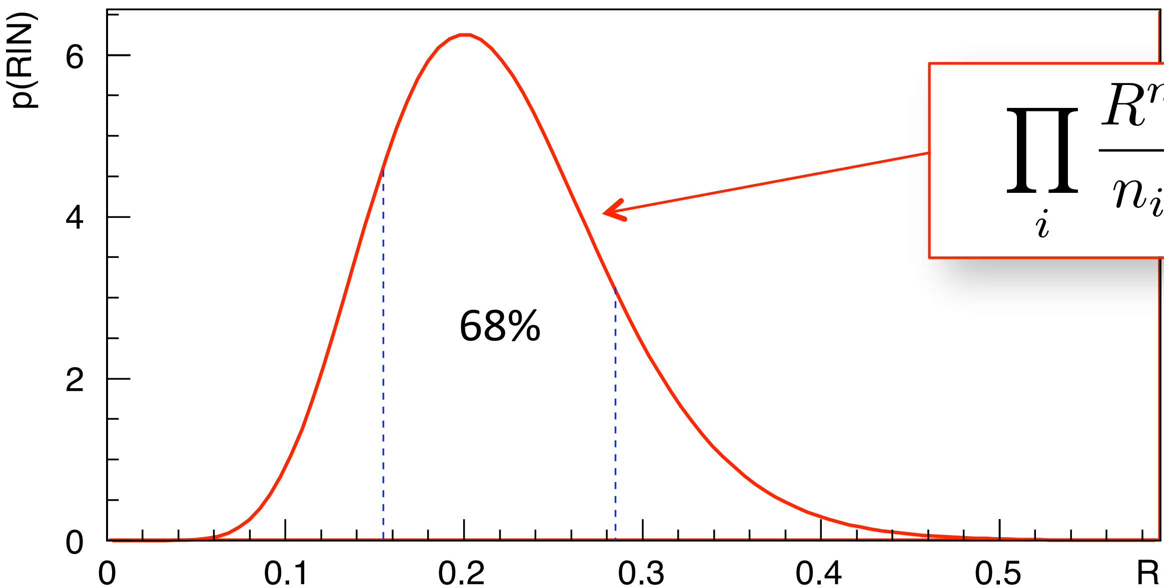
- Set derivative to zero to get the mode

$$\frac{\mathrm{d}L}{\mathrm{d}R} = \sum_i \left(1 - \frac{n_i}{R}\right) = 0$$

- Solve this for R, *et voila* the correct result

$$\hat{R} = \frac{\sum n_i}{n_{\text{bins}}} = \frac{10}{50} = 0.2$$

You can also communicate the result by showing the posterior and giving the 68% credible interval …



$$\prod_i \frac{R^{n_i}}{n_i!}\, e^{-R}$$

$$R = 0.20^{+0.08}_{-0.04} \qquad (68\%\ \mathrm{CL})$$

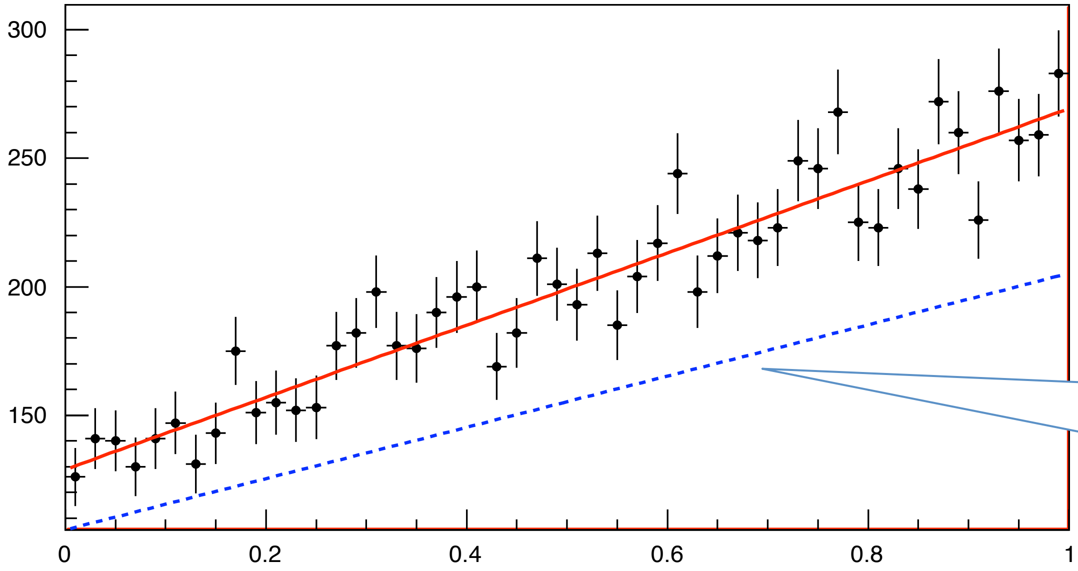
$$P(0.16 < R < 0.28) = 0.68$$

3. Normalisation uncertainty

- Let data float within quoted normalisation error Δ
- Add normalisation parameter and penalty chi-squared

$$\chi^2(N,\boldsymbol{\theta}) = \sum_i \left[\frac{N d_i - f_i(\boldsymbol{\theta})}{\sigma_i}\right]^2 + \left(\frac{N-1}{\Delta}\right)^2$$

Avoid collapse to N = 0



The fit completely misses the data!

Start from the correct likelihood!

- Normalisation factors affect <u>both</u> data and errors so that the likelihood is

$$p(\boldsymbol{d}|N,\boldsymbol{\theta},I) = \prod_i \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\frac{Nd_i - f_i(\boldsymbol{\theta})}{N\sigma_i}\right]^2\right\}$$

- Gaussian prior with width *Δ* for *N*

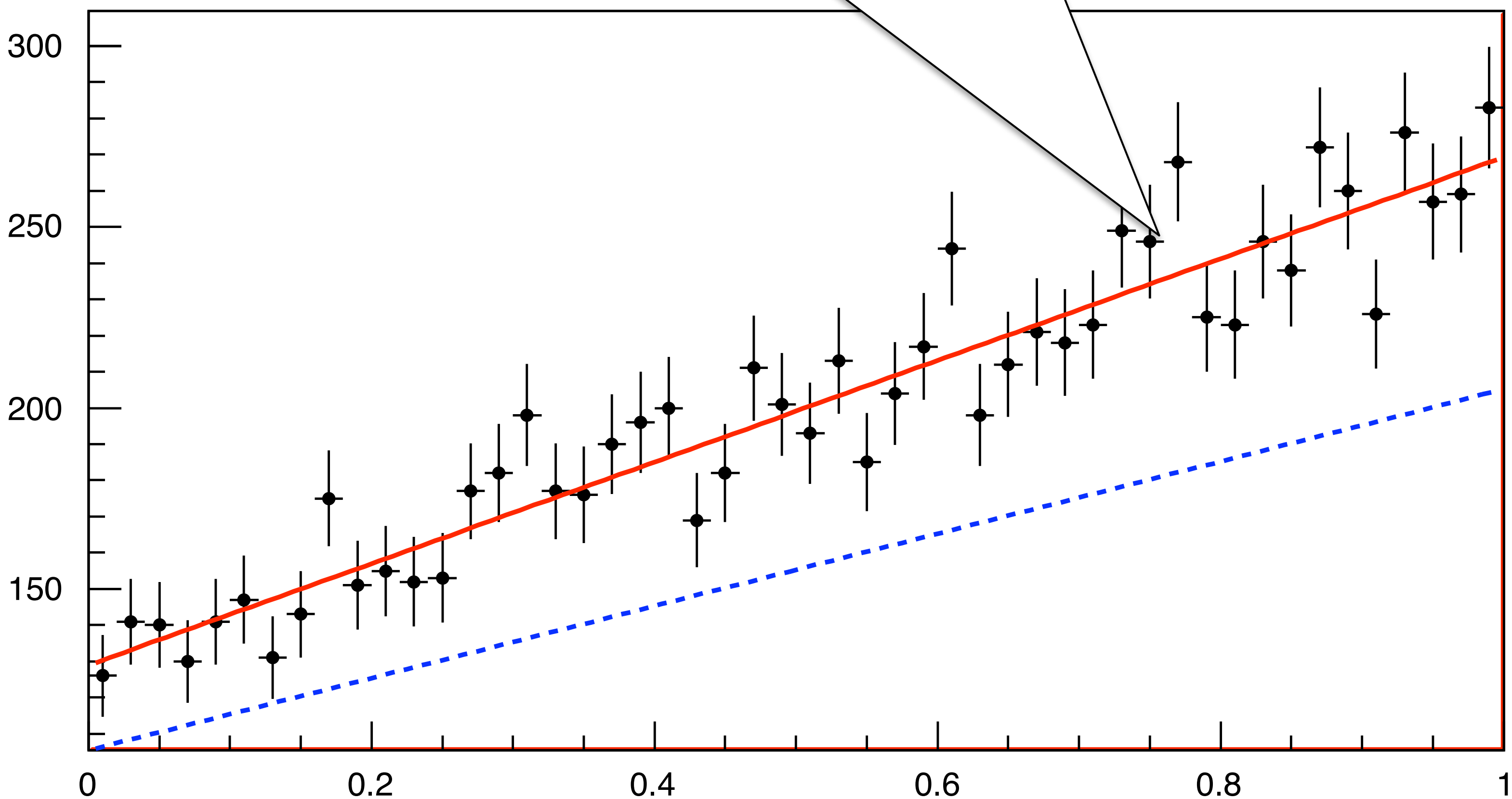
$$p(N|I) \propto \exp\left[-\frac{1}{2}\left(\frac{N-1}{\Delta}\right)^2\right]$$

This is not a magic penalty χ^2 it is a Bayesian prior

- The log posterior is

$$\chi^2(N,\boldsymbol{\theta}) = \sum_i \left[\frac{Nd_i - f_i(\boldsymbol{\theta})}{N\sigma_i}\right]^2 + \left(\frac{N-1}{\Delta}\right)^2$$

Now we get a well behaved fit ...



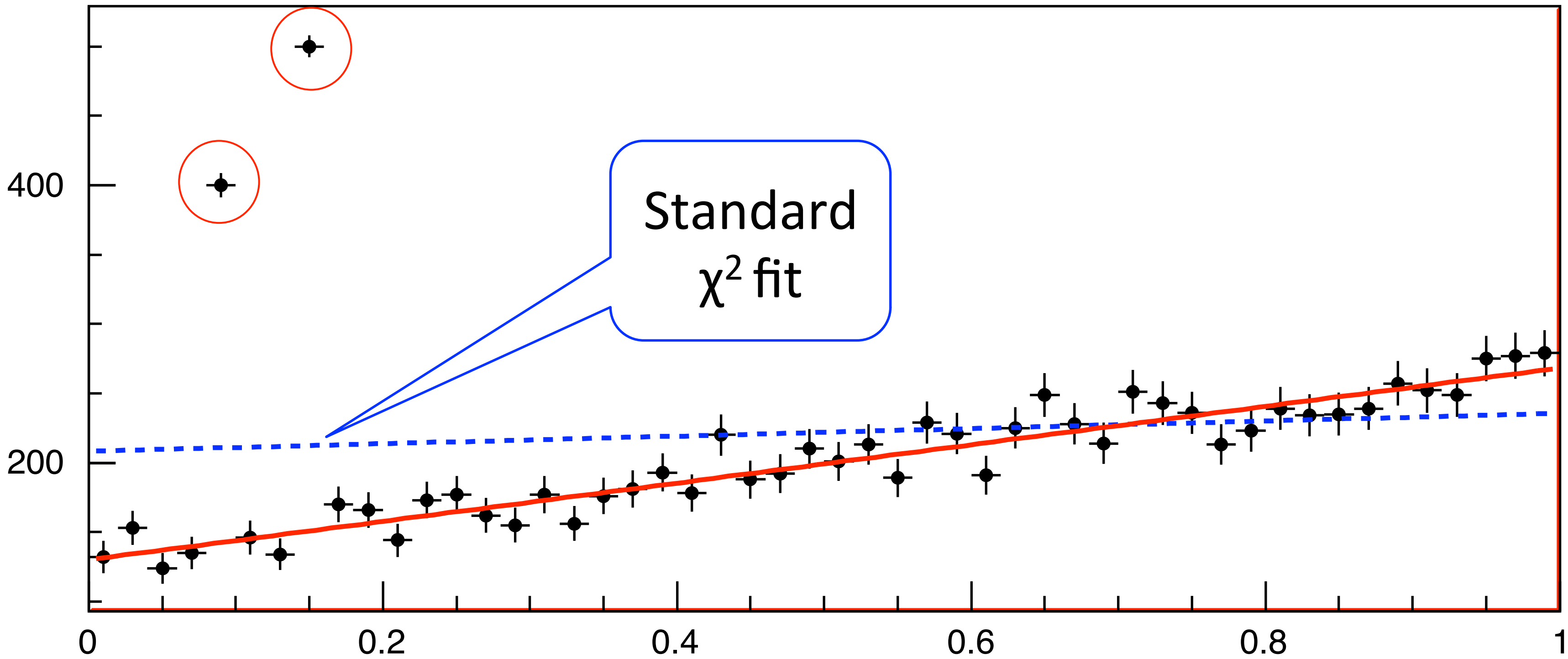
4. Uncertain experimental errors

- Rough estimate of the effect of under-estimated experimental errors, scale them by the factor

$$\alpha = \sqrt{\chi^2/\mathrm{ndf}}$$

- You lose the comparison between the data and the model because, by construction, the model is now compatible with the data
- Scaling the experimental errors by a factor α implies that the fitted parameter covariance matrix should be scaled by α^2

- While underestimated errors are a nuisance, outliers are far more dangerous
- This is because Gaussians carry little probability in the tails so that a few std outlier can exert an enormous pull



- Try to account for uncertain experimental errors by distributing them towards larger values
- Here is the ansatz of a quadratic fall-off

$$p(\sigma|\sigma_0) = \frac{\sigma_0}{\sigma^2} \qquad \text{for} \qquad \sigma \geq \sigma_0$$

Published error

- Expand in σ to get the likelihood

$$p(\boldsymbol{d}|\sigma_0) = \int_0^\infty p(\boldsymbol{d}|\sigma)\, p(\sigma|\sigma_0)\, \mathrm{d}\sigma$$

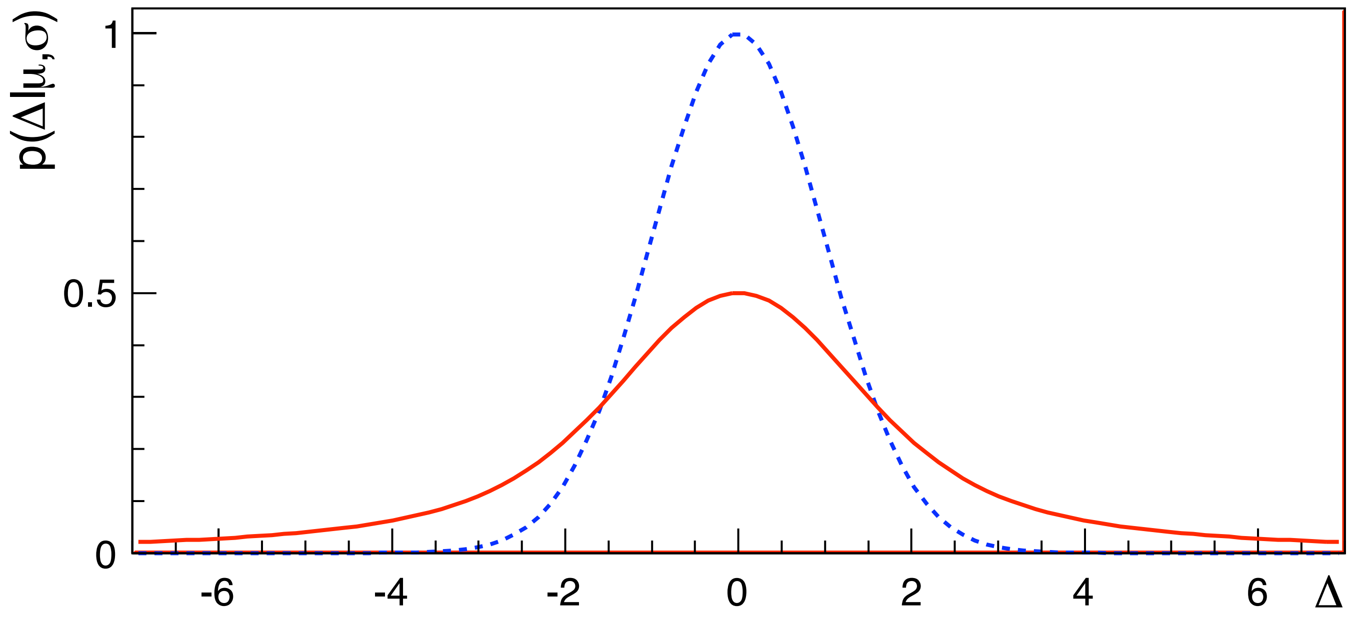
- For Gaussian data this becomes

$$\frac{1}{\sqrt{2\pi}} \int_{\sigma_0}^\infty \frac{\sigma_0}{\sigma^3} \exp\left[-\frac{(d-\mu)^2}{2\sigma^2}\right] \mathrm{d}\sigma$$

- Likelihood of one data point now is

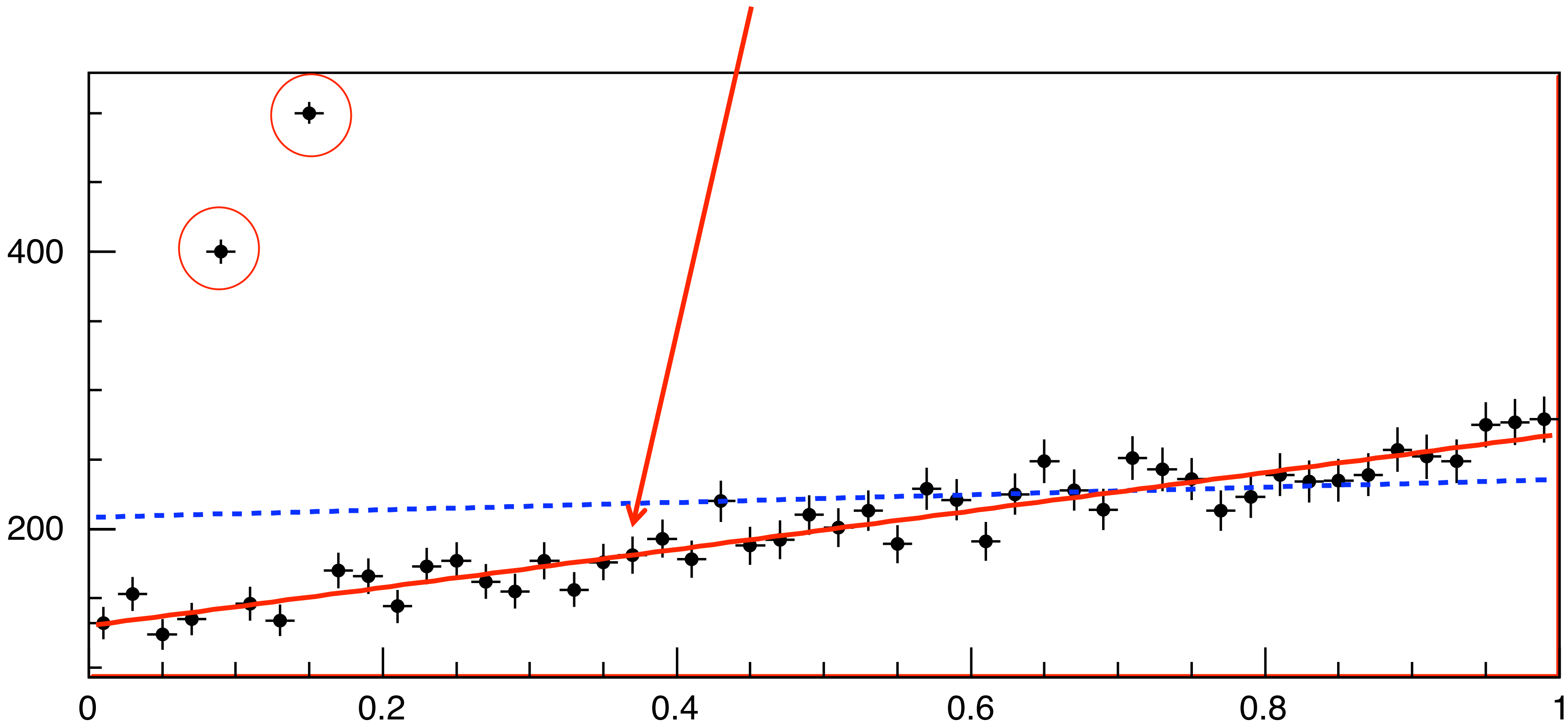
$$p(d,|\mu,\sigma_0) = \frac{1}{\sigma_0\sqrt{2\pi}}\left[\frac{1-\exp(-\Delta^2/2)}{\Delta^2}\right] \qquad \Delta = \frac{d-\mu}{\sigma_0}$$

- Has much longer tails than the original Gauss



Much reduced sensitivity to outliers!

$$L = -\sum_i \log \left[\frac{1 - \exp(-\Delta_i^2/2)}{\Delta_i^2} \right]$$



What we have learned

- Bayesian parameter estimation is just using Bayes' theorem to obtain the posterior distribution of the parameters from the likelihood
- In this sense there is no such thing as a `fit' to the data
- The most probable value of the parameter (mode) can be obtained by minimising the (minus) log posterior using minimisation programs like MINUIT
- Error intervals can simply be read off from the posterior distribution, unlike the complicated construction of such intervals in the Frequentist case

- ✓ <u>Lecture 1</u>

  Basics of logic and Bayesian probability calculus

- ✓ <u>Lecture 2</u>

  Probability assignment

- ✓ <u>Lecture 3</u>

  Parameter estimation

- <u>Lecture 4</u>

  Glimpse at a Bayesian network, and Model selection