Bayesian analysis in nuclear physics

Ken Hanson

T-16, Nuclear Physics; Theoretical Division
Los Alamos National Laboratory

Tutorials presented at LANSCE
Los Alamos Neutron Scattering Center
July 25 – August 1, 2005

This presentation available at
http://www.lanl.gov/home/kmh/

LA-UR-05-5680

Tutorial 3
Bayesian data analysis

Goals of tutorials

My aim is to

- present overview of Bayesian and probabilistic modeling
- cover basic Bayesian methodology relevant to nuclear physics, especially cross section evaluation
- point way to how to do it

- convince you that
  - Bayesian analysis is a reasonable approach to coping with measurement uncertainty

- Many thanks to my T-16 colleagues
  - Gerry Hale, Toshihiko Kawano, Patrick Talou

Outline – four tutorials

1. **Bayesian approach**
   probability – quantifies our degree of uncertainty
   Bayes law and prior probabilities
2. **Bayesian modeling**
   Peelle's pertinent puzzle
   Monte Carlo techniques; quasi-Monte Carlo
   Bayesian update of cross sections using Jezebel criticality expt.
3. **Bayesian data analysis**
   linear fits to data with Bayesian interpretation
   uncertainty in experimental measurements; systematic errors
   treatment of outliers, discrepant data
4. **Bayesian calculations**
   Markov chain Monte Carlo technique
   analysis of Rossi traces; alpha curve
   background estimation in spectral data

Slides and bibliography

- These slides can be obtained by going to my public web page:
  http://public.lanl.gov/kmh/talks/
  - link to **tutorial slides**
  - short **bibliography** relevant to topics covered in tutorial
  - other presentations, which contain more detail about material presented here
- Noteworthy books:
  - D. Sivia, *Data Analysis: A Bayesian Tutorial* (1996); lucid pedagogical development of the Bayesian approach with an experimental physics slant
  - D. L. Smith, *Probability, Statistics, and Data Uncertainties in Nuclear Science and Technology* (1991); lots of good advice relevant to cross-section evaluation
  - G. D'Agostini, *Bayesian Reasoning in Data Analysis: A Critical Review*, (World Scientific, New Jersey, 2003); Bayesian philosophy
  - A. Gelman et al., *Bayesian Data Analysis* (1995); statisticians' view
  - W. R. Gilks et al., *Markov Chain Monte Carlo in Practice* (1996); basic MCMC text

Types of measurement uncertainties

- Generally two major types of uncertainties
  - ► random uncertainty – different for each measurement of same quantity
    - in repeated measurements, get a different answer each time
    - often assumed to be statistically independent, but aren't always
  - ► systematic uncertainty – same for each measurement within a group
    - component of measurements that remains unchanged
    - for example, caused by error in calibration or zeroing
    - this kind of uncertainty needs more attention
- Nomenclature varies
  - ► physics – random uncertainty and systematic uncertainty
  - ► statistics – random and bias
    - metrology standards (NIST, ASME, ISO) –
      random and systematic uncertainties (now)
  - ► trend toward quoting **standard error**

Measurement uncertainties in cross sections

In cross-section experiments, sources of uncertainties include:

- Random uncertainties
  - counting statistics for primary process and monitoring process
  - background
- Systematic uncertainties
  - integrated beam intensity
  - target thickness, target impurities
  - detector efficiency
  - count rate corrections
  - geometry
  - corrections for contamination from other processes
- Try to reduce systematic uncertainties through calibration, design
- Random uncertainties usually easy to assess;
  systematic uncertainties require judgment

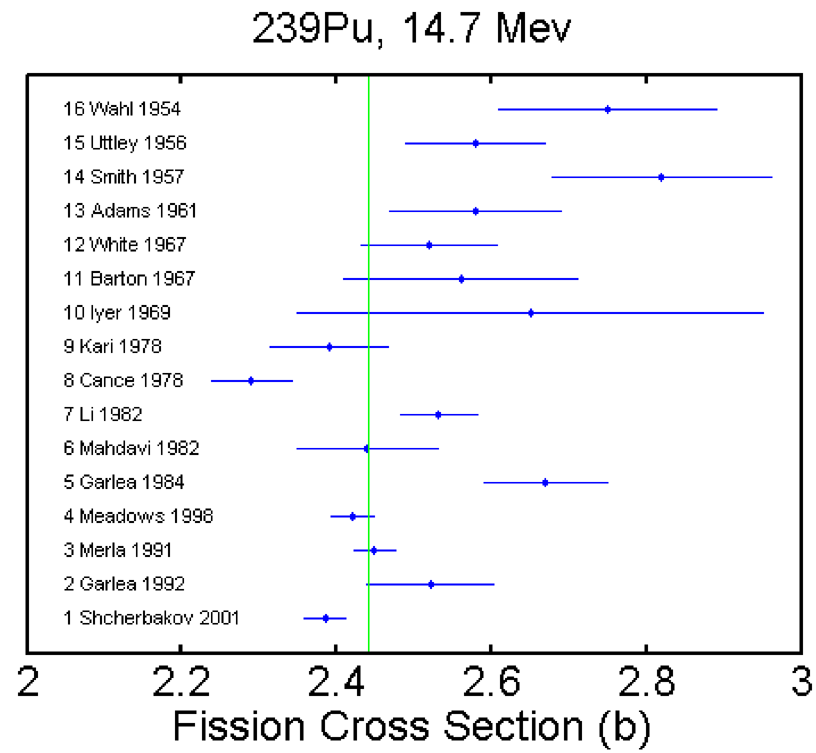
Characterization of measurement uncertainties

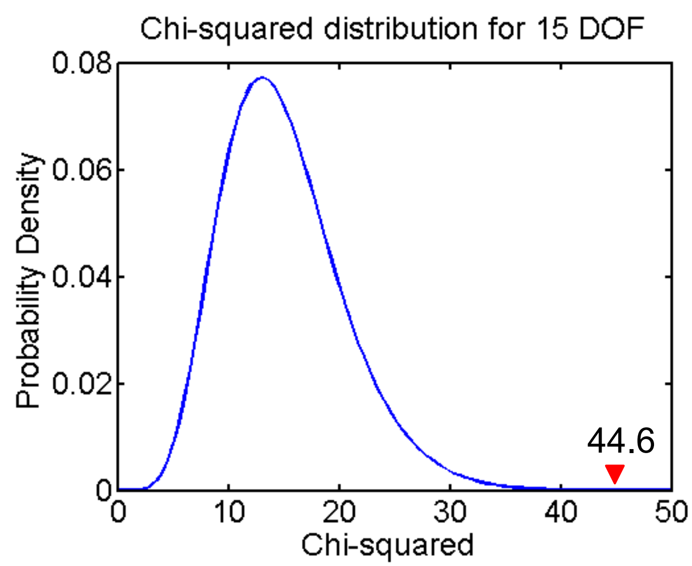
- The best analysis is based on a thorough understanding of probabilistic nature of the fluctuations in the data
- In nuclear physics we are fortunate to have control over measurements; we can calibrate and study apparatus
- Look closely at measurements to characterize random fluctuations
  - shape of pdf
  - standard deviation (variance) of fluctuations,
  - presence of outliers
  - covariance, correlation: $\mathrm{cov}(\boldsymbol{d}) \equiv \mathbf{C}_{\boldsymbol{d}} = \left\langle (\boldsymbol{d} - \hat{\boldsymbol{d}})(\boldsymbol{d} - \hat{\boldsymbol{d}})^{\mathrm{T}} \right\rangle$
  - usually need to assume stationarity, same characteristics everywhere
  - autocorrelation function useful for estimating correlations

$$\rho(l) = \frac{1}{N} \sum_{i=1}^{N} y(i)\, y(i-l)$$

Neutron fission cross section data for ^{239}Pu

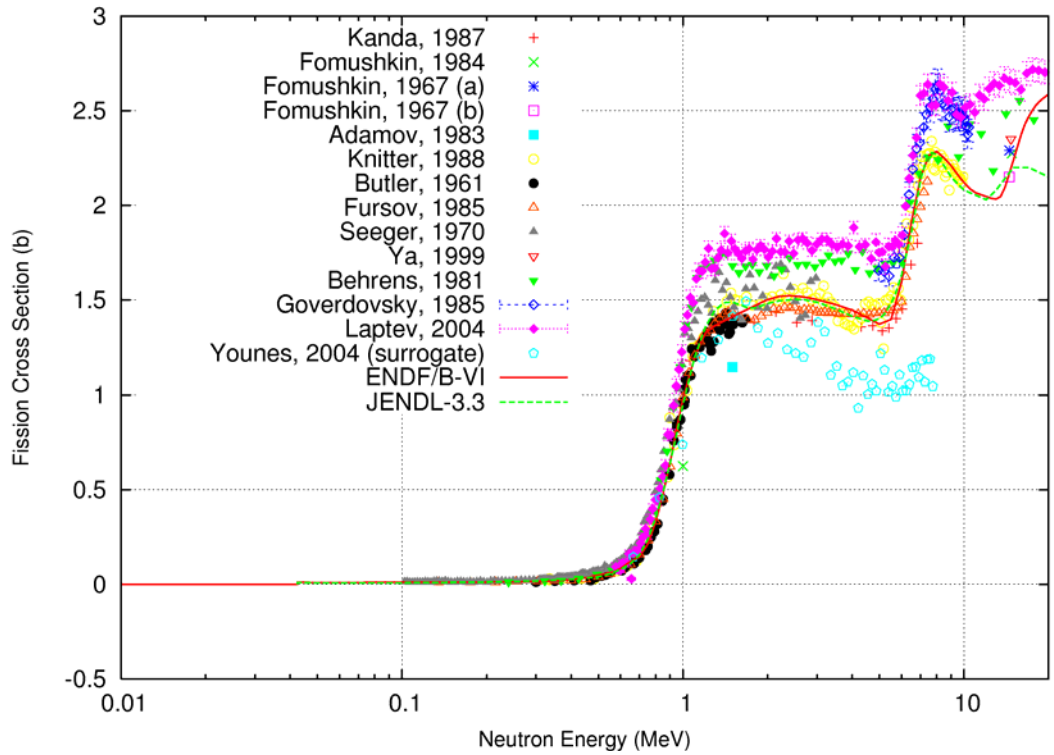
- Graph shows 16 measurements of fission cross-section for ^{239}Pu at 14.7 MeV
- Data exhibit fair amount of scatter
- Quoted error bars get smaller with time
- Minimum $\chi^2 = 44.6$ ($p = 10^{-4}$) indicates a problem
  - dispersion of data larger than quoted error bars by factor $\sqrt{3}$
  - outliers?; three data contribute 24 to χ^2, more than half





Neutron fission cross-section data

^{243}Am fission cross section



plot from P. Talou

- Neutron cross sections measured by many experimenters
  - sometimes data sets differ significantly
  - often little information about uncertainties, esp. systematic errors
  - many directly measure ratios of cross sections, e.g., ^{243}Am/ ^{235}U
  - thorough analysis must take into account all discrepancies

Inference using Bayes rule

- We wish to infer the parameters $\boldsymbol{a}$ of a model M, based on data $\boldsymbol{d}$
- Use Bayes rule, which gives the *posterior*:

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

  - where I represents general information we have about the situation
  - $p(\boldsymbol{d}\,|\,\boldsymbol{a}, M, I)$ is the *likelihood*, the probability of the observed data, given the parameters, model, and general info
  - $p(\boldsymbol{a}\,|\,M, I)$ is the *prior*, which represents what we know about the parameters exclusive of the data
- Note that inference requires specification of the prior

Likelihood

- Form of the likelihood $p(\boldsymbol{d} \,|\, \boldsymbol{a}, I)$ based on how we model the uncertainties in the measurements $\boldsymbol{d}$
- Choose pdf that appropriately describes uncertainties in data
  - Gaussian – good generic choice
  - Poisson – counting experiments
  - Binomial – binary measurements (coin toss …)
- Outliers exist
  - likelihood should have a long tail, i.e., there is some probability of large fluctuation
- Systematic errors
  - caused by effects common to many (all) measurements
  - model by introducing variable that affects many (all) measurements; marginalize out

The model and parameter inference

- We write the model as

$$\boldsymbol{y} = \boldsymbol{y}(\boldsymbol{x}, \boldsymbol{a})$$

  - where $\boldsymbol{y}$ is a vector of physical quantities, which is modeled as a function of the independent variables vector $\boldsymbol{x}$ and $\boldsymbol{a}$ represents the parameter vector for the model
- In inference, the aim is to determine:
  - the parameters $\boldsymbol{a}$ from a set of n measurements d_i of $\boldsymbol{y}$ under specified conditions x_i
  - **and** the uncertainties in the parameter values
- This process is called parameter inference, model fitting (or regression); however, uncertainty analysis is often not done, only parameters estimated

The likelihood and chi-squared

- The form of the likelihood $p(\boldsymbol{d} \,|\, \boldsymbol{a}, I)$ depends on how we model the uncertainties in the measurements $\boldsymbol{d}$
- Assuming the error in each measurement d_i is normally (Gaussian) distributed with zero mean and variance σ_i^2, and that the errors are statistically independent,

$$p(\boldsymbol{d} \,|\, \boldsymbol{a}) \propto \prod_i \exp\left[-\frac{[d_i - y_i(a)]^2}{2\sigma_i^2} \right]$$

- where y_i is the value predicted for parameter set $\boldsymbol{a}$
- The above exponent is one-half chi squared

$$\chi^2 = -2\log\left[p(\boldsymbol{d} \,|\, \boldsymbol{a}) \right] = \sum_i \left[\frac{[d_i - y_i(a)]^2}{\sigma_i^2} \right]$$

- For this error model, likelihood is $p(\boldsymbol{d} \,|\, \boldsymbol{a}) \propto \exp(-\frac{1}{2}\chi^2)$

Likelihood analysis

- For a non-informative **uniform prior**, the posterior is proportional to the likelihood
- Given the relationship between chi-squared and the likelihood, the posterior is

$$p(\boldsymbol{a} \mid \boldsymbol{d}) \propto p(\boldsymbol{d} \mid \boldsymbol{a}) \propto \exp(-\tfrac{1}{2}\chi^2)$$

- Parameter estimation based on **maximum likelihood** is equivalent to that based on **minimum chi squared** (or **least squares**)

Likelihood analysis – chi squared

- When the errors in each measurement are Gaussian distributed and independent, likelihood is related to chi squared:

$$p(\boldsymbol{d} \mid \boldsymbol{a}) \propto \exp(-\tfrac{1}{2}\chi^2) = \exp\left\{-\tfrac{1}{2}\sum_i \left[\frac{[d_i - y_i(\boldsymbol{a})]^2}{\sigma_i^2}\right]\right\}$$

- near minimum, χ^2 is approximately quadratic in the parameters $\boldsymbol{a}$

$$\chi^2(\boldsymbol{a}) = \tfrac{1}{2}\left(\boldsymbol{a} - \hat{\boldsymbol{a}}\right)^{\mathrm{T}} \boldsymbol{K} \left(\boldsymbol{a} - \hat{\boldsymbol{a}}\right) + \chi^2(\hat{\boldsymbol{a}})$$

  - where $\hat{\boldsymbol{a}}$ is the parameter vector at minimum χ^2 and $\boldsymbol{K}$ is the χ^2 curvature matrix (aka the *Hessian*)

- The covariance matrix for the uncertainties in the estimated parameters is

$$\mathrm{cov}(\boldsymbol{a}) \equiv \left\langle (\boldsymbol{a} - \hat{\boldsymbol{a}})(\boldsymbol{a} - \hat{\boldsymbol{a}})^{\mathrm{T}} \right\rangle \equiv \boldsymbol{C} = 2\boldsymbol{K}^{-1}$$

Characterization of chi squared

- Expand vector $\boldsymbol{y}$ around $\boldsymbol{y}^0$, and approximate:

$$y_i = y_i(x_i, \boldsymbol{a}) = y_i^0 + \sum_j \left.\frac{\partial y_i}{\partial a_j}\right|_{a^0} (a_j - a_j^0) + \cdots$$

- The derivative matrix is called the *Jacobian*, $\boldsymbol{J}$
- Estimated parameters $\hat{\boldsymbol{a}}$ minimize χ^2 (MAP estimate)
- As a function of $\boldsymbol{a}$, χ^2 is approximately quadratic in $\boldsymbol{a} - \hat{\boldsymbol{a}}$

$$\chi^2(\boldsymbol{a}) = \tfrac{1}{2}\left(\boldsymbol{a} - \hat{\boldsymbol{a}}\right)^{\mathrm{T}} \boldsymbol{K} \left(\boldsymbol{a} - \hat{\boldsymbol{a}}\right) + \chi^2(\hat{\boldsymbol{a}})$$

  - where $\boldsymbol{K}$ is the χ^2 curvature matrix (aka the *Hessian*);

$$\left[\boldsymbol{K}\right]_{jk} = \left.\frac{\partial^2 \chi^2}{\partial a_j \partial a_k}\right|_{\hat{a}} \quad ; \quad \boldsymbol{K} = 2\boldsymbol{J}\boldsymbol{\Lambda}\boldsymbol{J}^{\mathrm{T}} \; ; \quad \boldsymbol{\Lambda} = \mathrm{diag}(\sigma_1^{-2}, \sigma_2^{-2}, \sigma_3^{-2}, \ldots)$$

- Jacobian useful for finding min. χ^2 , i.e., optimization

Multiple data sets and Gaussian prior

- Analysis of multiple data sets
  - to combine the data from multiple, independent data sets into a single analysis, the combined chi squared is

$$\chi^2_{all} = \sum_k \chi^2_k$$

  - where $p(\boldsymbol{d}_k \mid \boldsymbol{a}, I)$ is the likelihood from kth data set
- Include Gaussian priors through Bayes theorem

$$p(\boldsymbol{a} \mid \boldsymbol{d}, I) \propto p(\boldsymbol{d} \mid \boldsymbol{a}, I)\; p(\boldsymbol{a} \mid I)$$

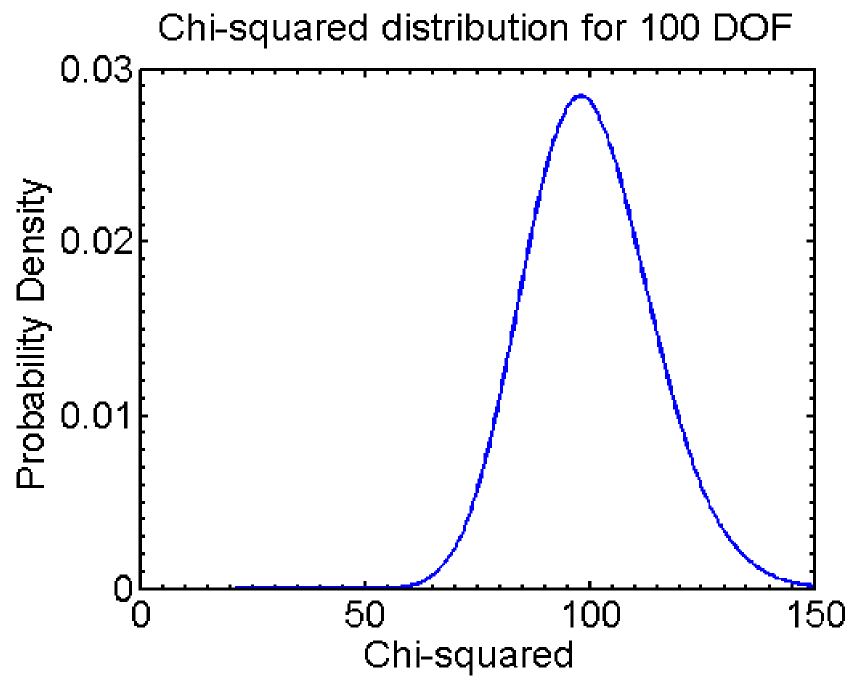
  - for a Gaussian prior on a parameter a_j

$$-\log p(\boldsymbol{a} \mid \boldsymbol{d}, I) = \varphi(\boldsymbol{a}) = \tfrac{1}{2}\chi^2 + \frac{\left(a_j - \tilde{a}_j\right)^2}{2\sigma_j^2}$$

  - where $\tilde{a}_j$ is the default value for a_j and σ_j^2 is assumed variance

Chi-squared distribution

- Plot shows χ^2 distribution for number of degrees of freedom, $\nu = 100$
- Generally,
  - mean = ν
  - rms dev = $\sqrt{2/\nu}$
- Cumulative distribution gives p value, probability of $\chi^2 \geq$ observed value
- p often used a measure of goodness of fit
- Checks self-consistency of models used to explain data (weakly)



Goodness of fit

- Check of minimum chi-squared value only weakly confirms validity of models used
- Chi-squared value depends on numerous factors:
  - assumption that errors follow Gaussian distribution and are statistically independent
  - proper assignment of standard deviation of errors
  - correctness of model used to calculate measured quantity
  - measurements correspond to calculated quantity (proper measurement model)
- Thus, a reasonable chi-squared *p* value does not necessarily mean everything is OK, because there may be compensating effects

Fit linear function to data – minimum χ^2

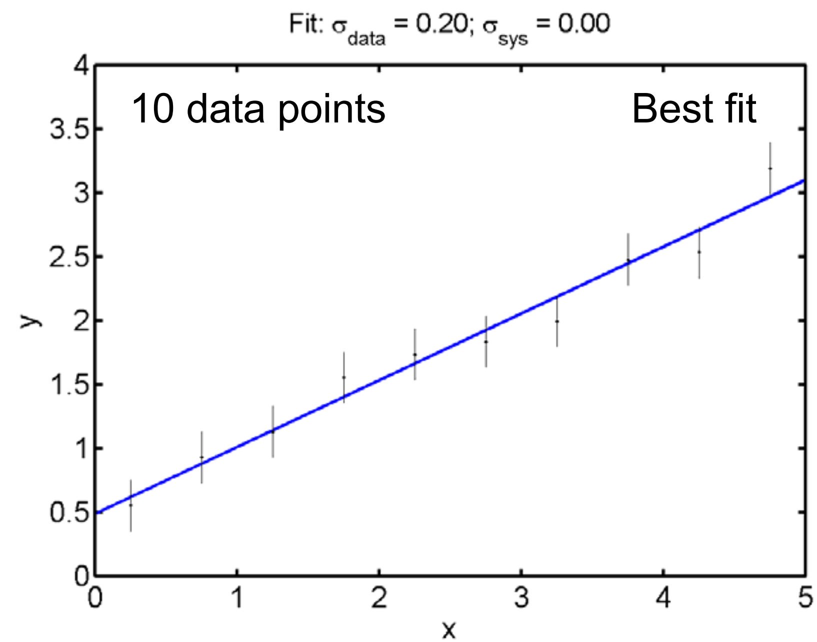
- Linear model: $y = a + bx$
- Simulate 10 data points, $\sigma_y = 0.2$
  exact values: $a = 0.5 \quad b = 0.5$
- Determine parameters, intercept a and slope b, by minimizing chi-squared (standard least-squares analysis)
- Result: $\chi^2_{\min} = 4.04 \quad p = 0.775$

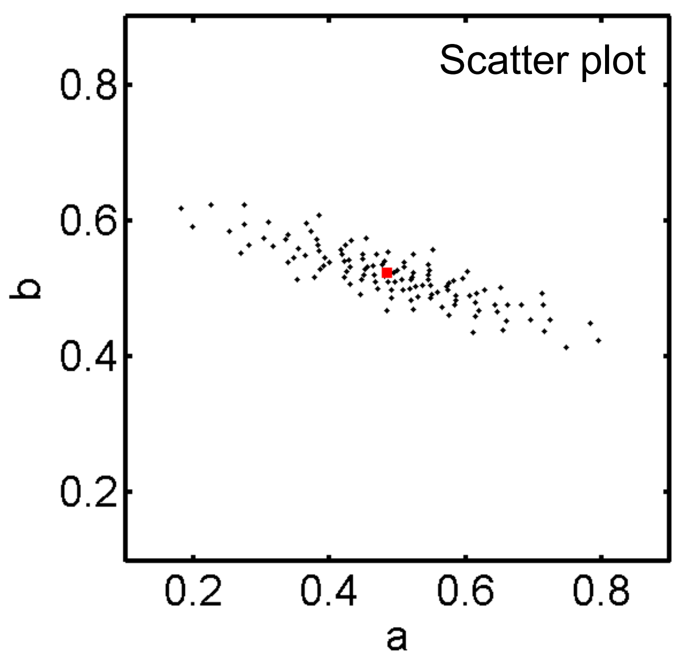
$$\hat{a} = 0.484 \qquad \sigma_a = 0.127$$

$$\hat{b} = 0.523 \qquad \sigma_b = 0.044$$

$$\mathbf{R} = \begin{bmatrix} 1 & -0.867 \\ -0.867 & 1 \end{bmatrix}$$

- Strong correlations between parameters a and b





Sampling from correlated normal distribution

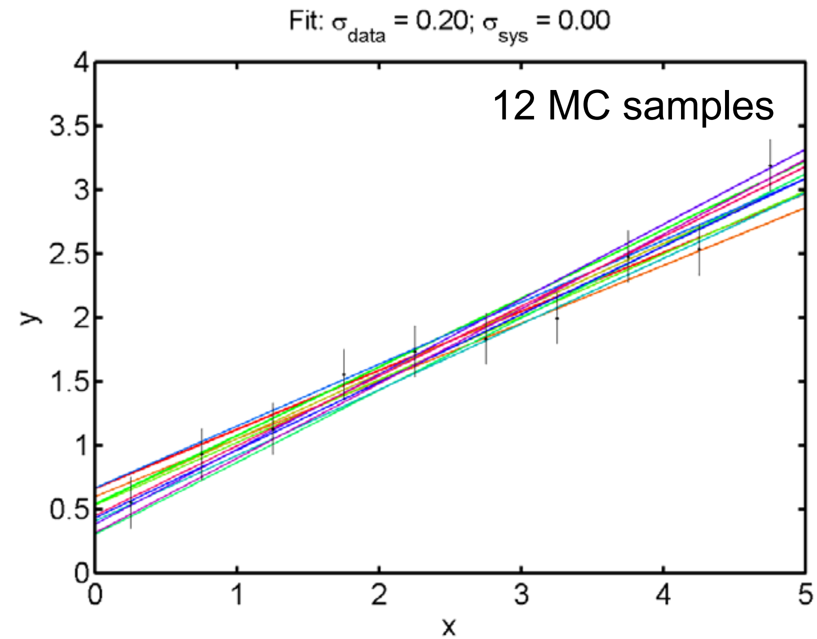
- Want to draw samples $\mathbf{x}$ from multi-variate normal distribution with known covariance $\mathbf{C_x}$
- Important to include correlations among uncertainties, i.e., off-diagonal elements
- Algorithm:
  - perform eigenanalysis of covariance matrix of *d* dimensions

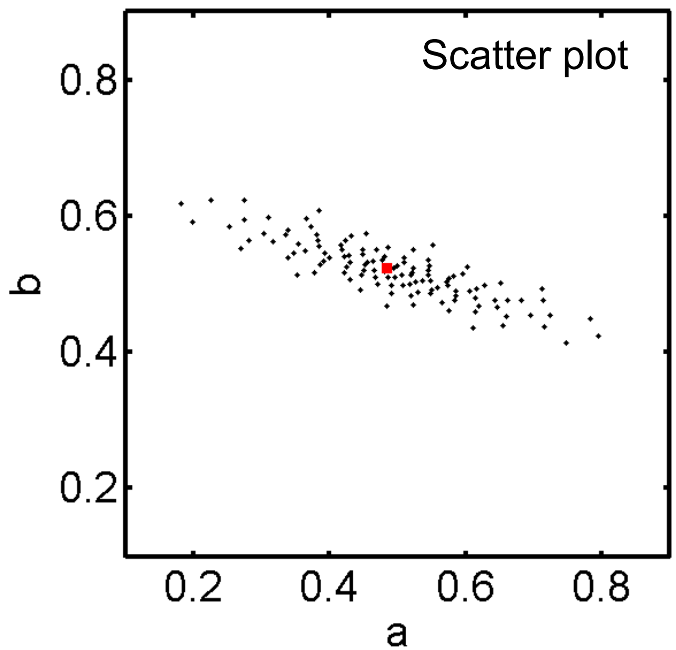
$$\mathbf{C_x} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{T}}$$

    where $\mathbf{U}$ is orthogonal matrix of eigenvectors and
    $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues
  - draw *d* samples from unit variance normal distribution, ξ_i
  - scale this vector by $\lambda_i^{1/2}$
  - transform vector into parameter space using the eigenvector matrix
  - to summarize: $\mathbf{x} = \mathbf{U}\mathbf{\Lambda}^{1/2}\boldsymbol{\xi}$

Linear fit – uncertainty visualization

- Uncertainties in parameters are represented by Gaussian pdf in 2-D parameter space
  - correlations evidenced by tilt in scatter plot
  - points are samples from pdf
- Should focus on implied uncertainties in physical domain
  - model realizations drawn from parameter uncertainty pdf
  - these appear plausible – called **model checking**
  - this comparison to the original data confirms model adequacy
  - called **predictive distribution**



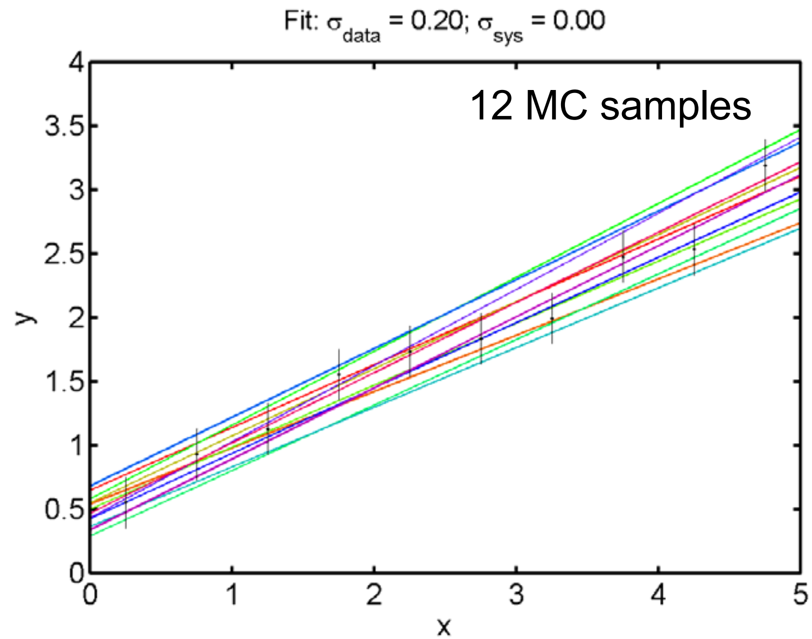


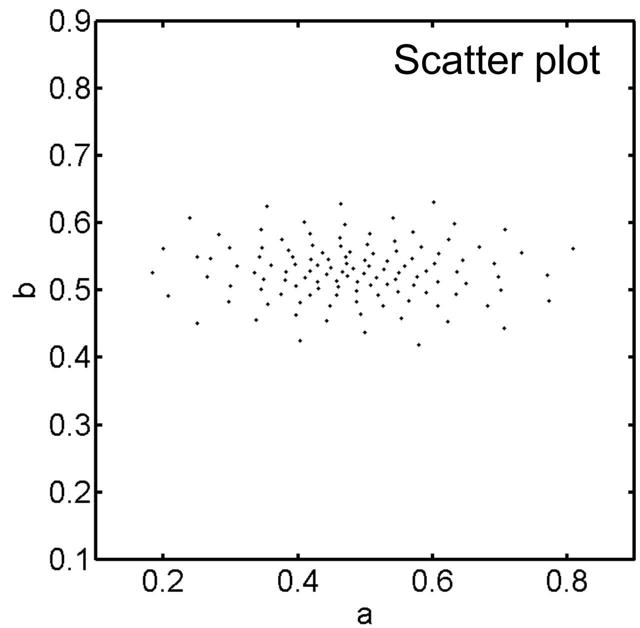
Linear fit – correlations are important

- Plots show what happens if off-diagonal terms of covariance matrix are ignored
- Correlation matrix is

$$\mathbf{R} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- Model realizations show much wider dispersion than consistent with uncertainties in data
- No tilt in scatter plot – uncorrelated
- Correlations are important !





Probabilistic model for additive error

- Represent systematic additive uncertainty in measurements by common additive offset Δ: $y_i = a + bx_i + \varepsilon_i + \Delta = f(x_i; a,b) + \varepsilon_i + \Delta$
  - where the ε_i represent the random fluctuations
- Bayes law gives joint pdf for all the parameters

$$p(a,b,\Delta \mid \boldsymbol{y},\boldsymbol{x}) = p(\boldsymbol{y} \mid a,b,\Delta,\boldsymbol{x})\, p(a)\, p(b)\, p(\Delta)$$

  where priors $p(a)$, $p(b)$ are uniform and $p(\Delta)$ assumed normal

- Writing $p(a,b,\Delta \mid \boldsymbol{y},\boldsymbol{x}) \propto \exp\{-\varphi\}$ and assuming normal distributions

$$2\varphi = \sum \frac{\left(y_i - f(x_i; a,b) - \Delta\right)^2}{\sigma_i^2} + \frac{\Delta^2}{\sigma_\Delta^2}$$

- Pdf for x obtained by integration: $p(a,b \mid \boldsymbol{y},\boldsymbol{x}) = \int p(a,b,\Delta \mid \boldsymbol{y},\boldsymbol{x})\, \mathrm{d}\Delta$

- This model equivalent to standard least-squares approach by including Δ in fit, and using just results for a and b

Linear fit – systematic uncertainty

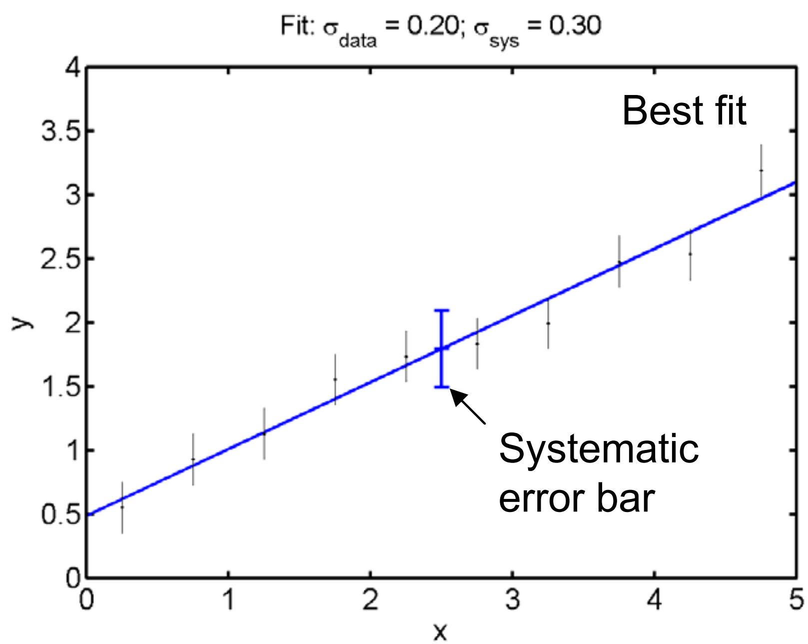
- Introduce systematic offset Δ with uncertainty $\sigma_\Delta = 0.3$
- Linear model: $y = a + bx + \Delta$
- Determine parameters, *a*, *b*, and offset Δ by minimizing chi-squared (standard least-squares analysis)
- Result: $\hat{\Delta} = 0$

$$\hat{a} = 0.484 \qquad \sigma_a = 0.326$$

$$\hat{b} = 0.523 \qquad \sigma_b = 0.044$$

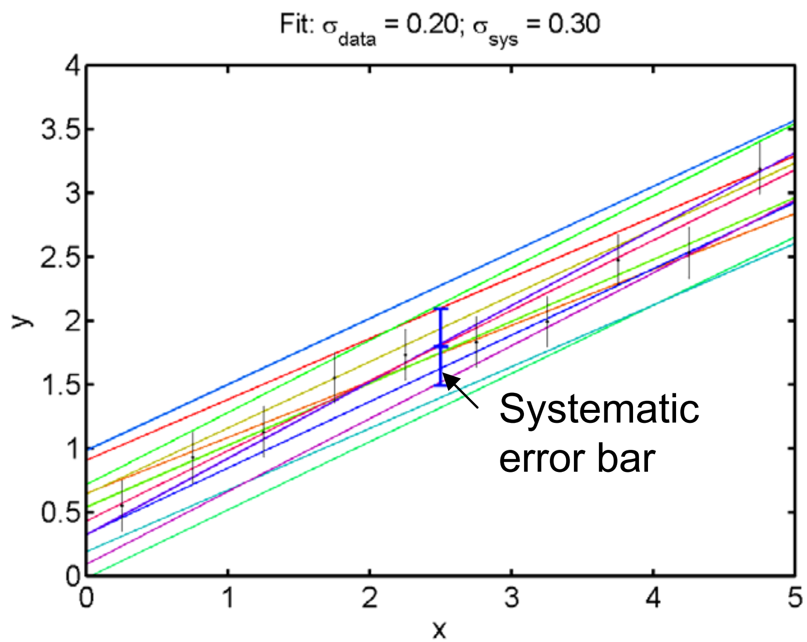
$$\mathbf{R} = \begin{bmatrix} 1 & -0.338 \\ -0.338 & 1 \end{bmatrix}$$

- Same parameters, but σ_a much larger



Linear fit – systematic uncertainty

- Show uncertainties in inferred models
  - colored lines are model realizations drawn from parameter uncertainty pdf
  - these appear plausible, considering additional systematic uncertainty, $\sigma_\Delta = 0.3$



Role of simulated data

- Simulated data are crucially important for testing algorithms
  - treat simulated data as is actual measurements
  - can compare algorithmic results with known true values
  - can test how well algorithm copes with specific data deficiencies
  - aid in debugging computer code, underlying ideas
- Important to mimic real data
  - characteristics of measurement fluctuations (noise)
  - limited resolution (blur) of signal
  - systematic effects

Linear fit to many data

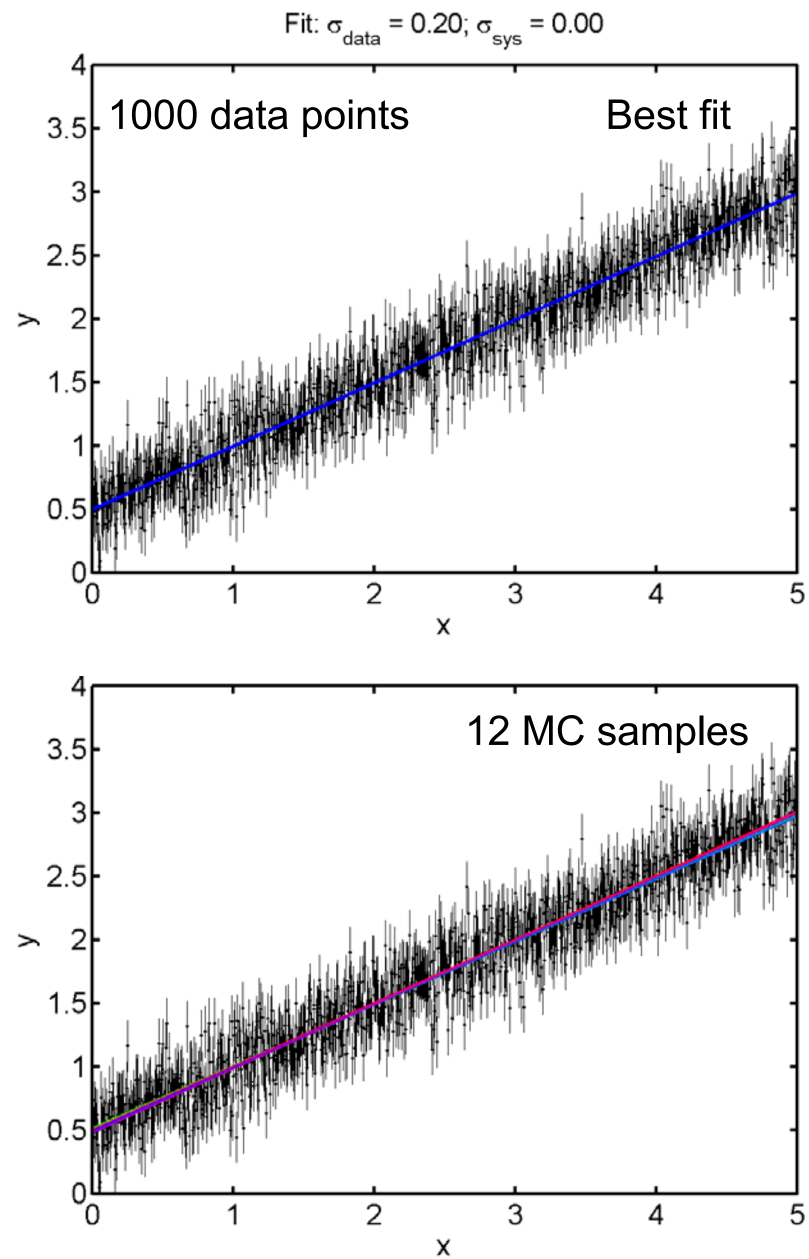
- Linear model: $y = a + bx$
- Simulate 1000 data points, $\sigma_y = 0.2$
  exact values: $a = 0.5 \quad b = 0.5$
- Determine parameters by minimizing chi-squared
- Result: $\chi^2_{\min} = 972.0 \quad p = 0.717$

$$\hat{a} = 0.496 \quad \sigma_a = 0.0126$$

$$\hat{b} = 0.499 \quad \sigma_b = 0.0044$$

$$\mathbf{R} = \begin{bmatrix} 1 & -0.866 \\ -0.866 & 1 \end{bmatrix}$$

- Standard errors are reduced by factor of 10 through data averaging
- Is this reasonable?



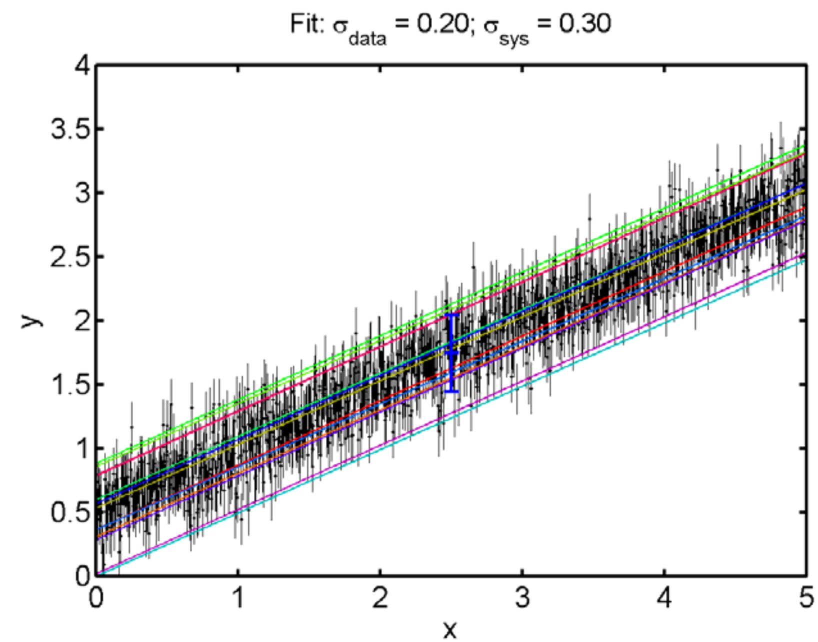
Linear fit to many data - systematic uncertainty

- Introduce systematic offset Δ with uncertainty $\sigma_\Delta = 0.3$
- Linear model: $y = a + bx + \Delta$
- Determine parameters, *a*, *b*, and offset Δ by minimizing chi-squared (standard least-squares analysis)
- Result: $\hat{\Delta} = 0$

$$\hat{a} = 0.496 \qquad \sigma_a = 0.300$$

$$\hat{b} = 0.499 \qquad \sigma_b = 0.0044$$

$$\mathbf{R} = \begin{bmatrix} 1 & -0.036 \\ -0.036 & 1 \end{bmatrix}$$



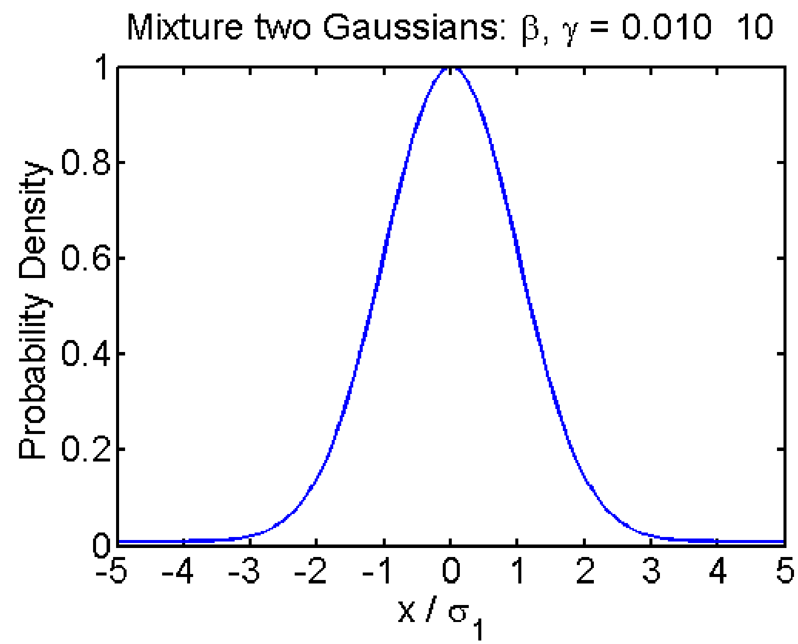
- Same fit, but σ_a dominated by σ_Δ
- Uncertainty in slope still small

Outliers

- Measurements that differ from true value by more than expected
- Often caused by mistakes
  - every experimenter knows mistakes happen!
- Can accommodate in likelihood function by including long tail
- Simple model: likelihood is mixture of two Gaussians

$$(1-\beta)\exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}+\beta\exp\left\{-\frac{(x-m)^2}{2\gamma\sigma^2}\right\}$$

- Long tail includes possibility of large deviation from true value
- Outlier-tolerant analysis generally called "robust estimation"



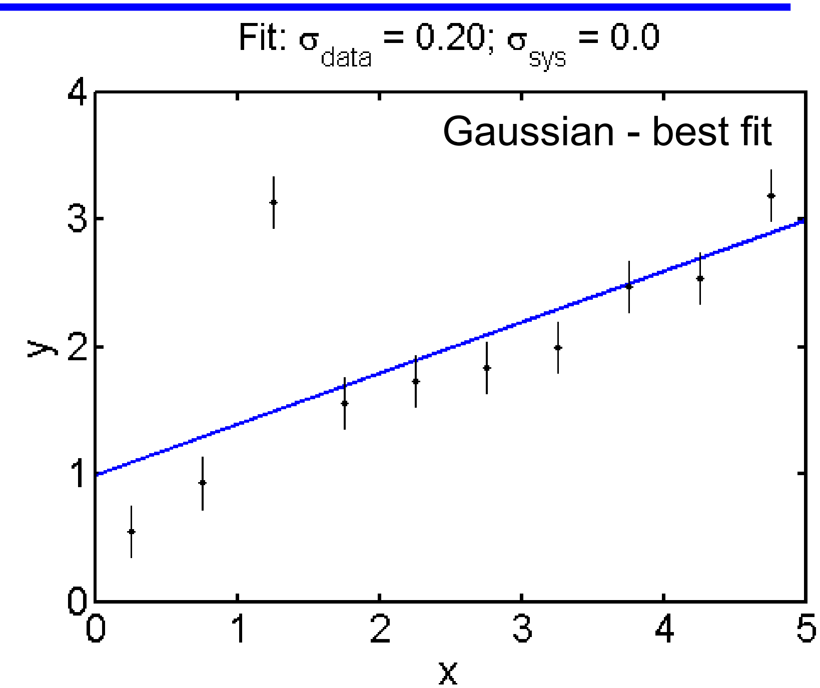
Linear fit – outliers

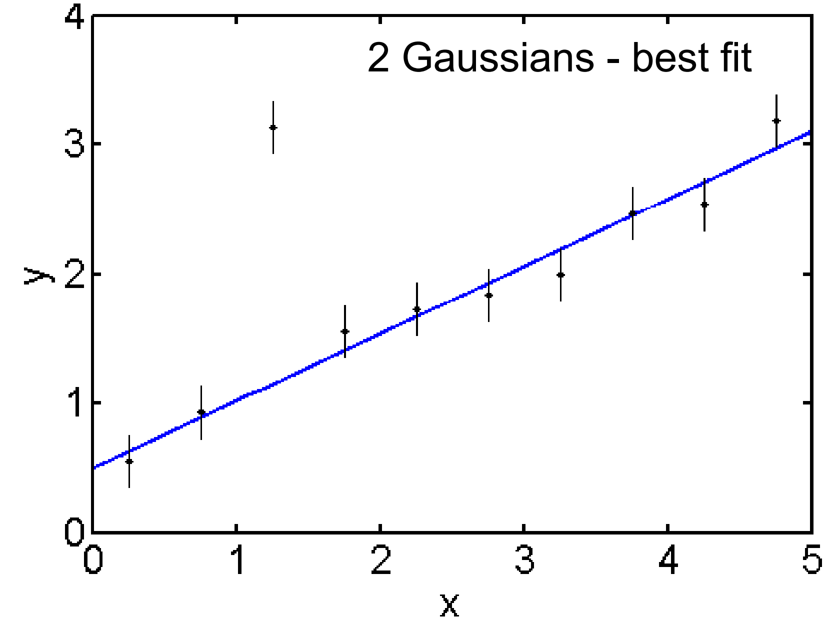
- Outliers pose significant problem for min χ^2 algorithm
- Create outlier by artificially perturbing third point
- Min-χ^2 results in large shift of fitted line: $\chi^2_{\min} = 85.6 \quad p = 10^{-15}$

$$\hat{a} = 0.987 \quad \sigma_a = 0.180$$
$$\hat{b} = 0.402 \quad \sigma_b = 0.062$$

- Two-Gaussian likelihood handles outlier very well
  - fit is nearly the same as before

$$\hat{a} = 0.494 \quad \sigma_a = 0.140$$
$$\hat{b} = 0.520 \quad \sigma_b = 0.043$$



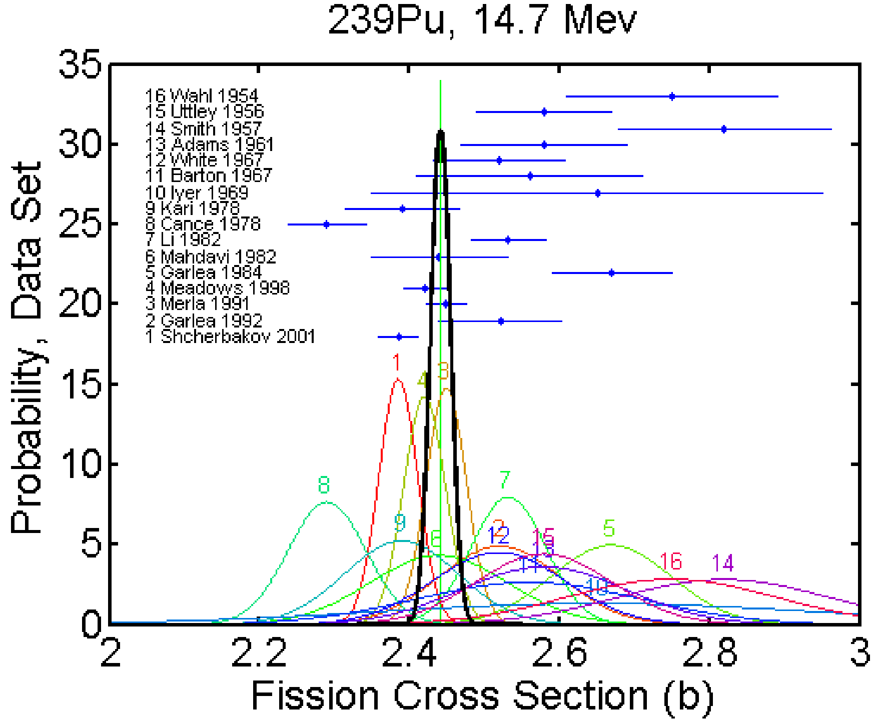


^{239}Pu cross sections – Gaussian likelihood

- With Gaussian likelihood ($\min \chi^2$) yields
  - $\chi^2 = 44.7$, $p = 0.009\%$ for 15 DOF
    2.441 ± 0.013
  - implausibly small uncertainty given three smallest uncerts. ≈ 0.027
- Each datum reduces the standard error of result, even if it does not agree with it!
  - consequence of Gaussian likelihood

$$\sigma^{-2} = \sum_{i=1}^{n} \sigma_i^{-2}$$

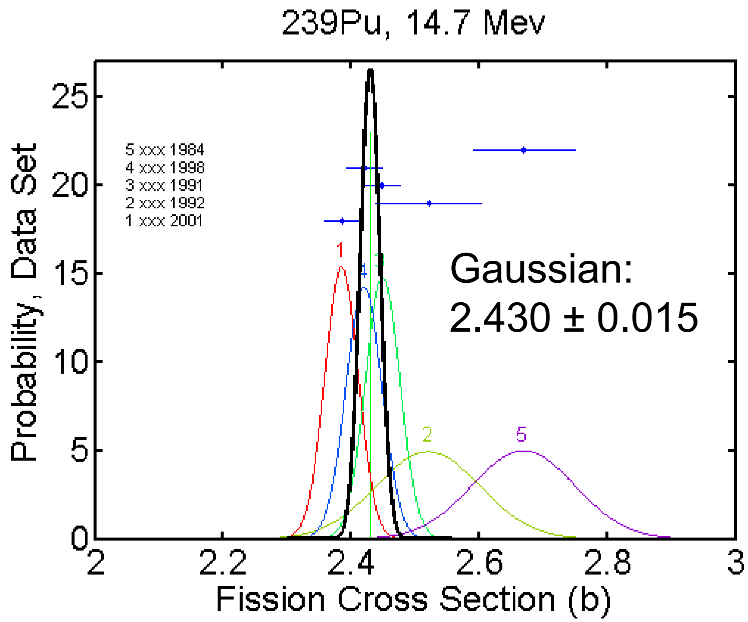
  - independent of where data lie!
    which doesn't make sense

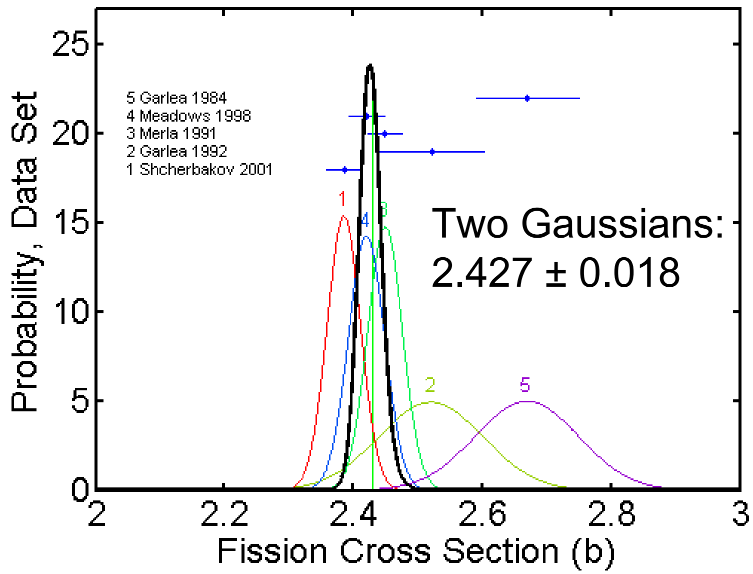


Gaussian: 2.441 ± 0.013

^{239}Pu cross sections – outlier-tolerant likelihood

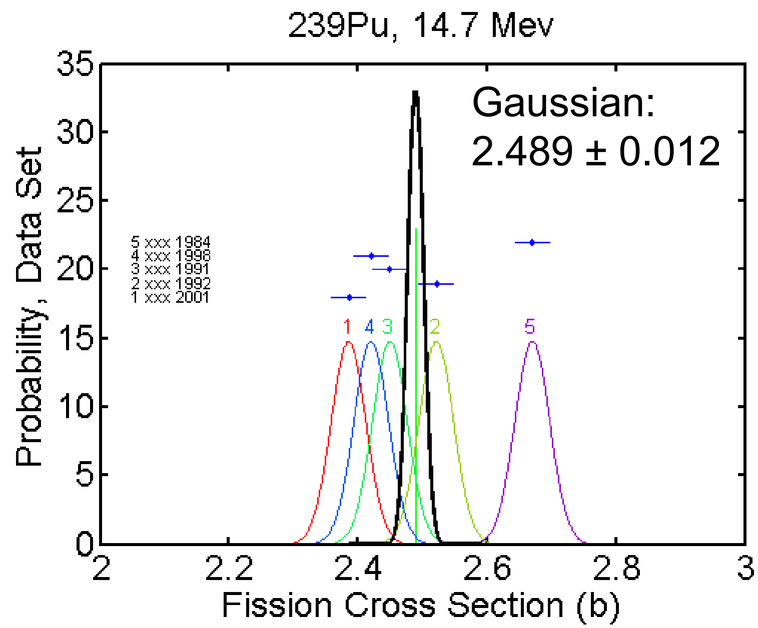
- Use just latest five measurements
- Compare results from alternative likelihoods:
  - Gaussian: 2.430 ± 0.015
    $\chi^2 = 13.88$, $p = 0.8\%$ for 4 DOF
  - two Gaussians: 2.427 ± 0.018
- For two-Gaussian likelihood:
  - result not pulled as hard by outlier
  - σ is not as small, seemingly taking into account discrepant nature of data

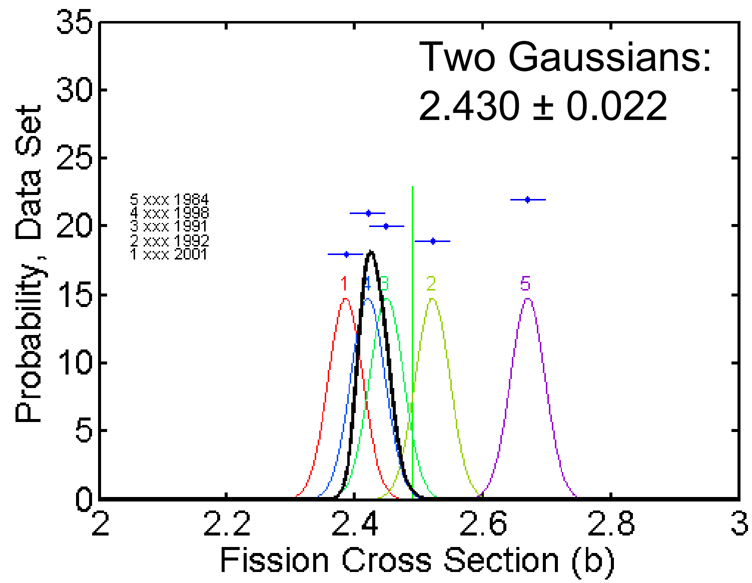




^{239}Pu cross sections – outlier-tolerant likelihood

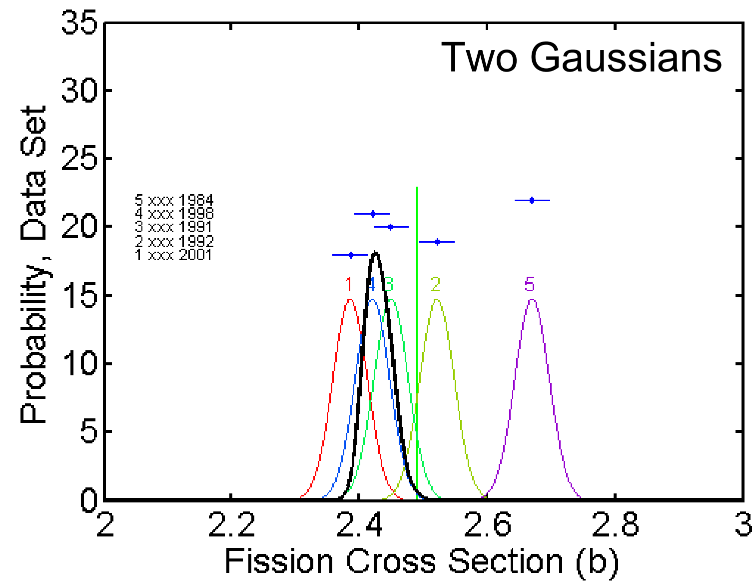
- Use just latest five measurements
- To exaggerate outlier problem, set all standard errors = 0.027
- Compare results from alternative likelihoods:
  - Gaussian: 2.489 ± 0.012
    $\chi^2 = 69.9$, $p = 2\times10^{-14}$ for 4 DOF
  - two Gaussians: 2.430 ± 0.022
- For two-Gaussian likelihood:
  - result is close to cluster of three points; outliers have little effect
  - uncertainty is plausible

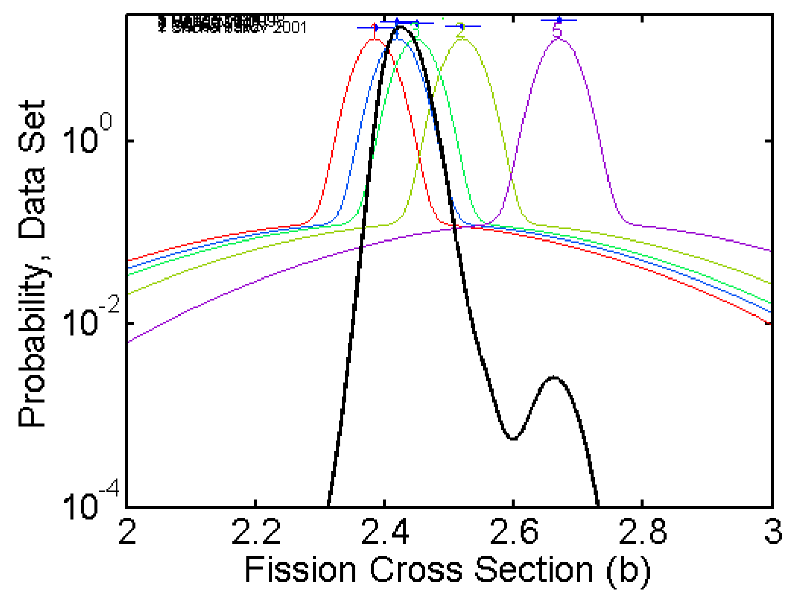




^{239}Pu cross sections – outlier-tolerant likelihood

- To exaggerate outlier problem, set all standard errors = 0.027, using just latest five measurements
- Plot shows pdfs on log scale, which shows what is going on with two-Gaussian likelihood
  - long tail of likelihood function for outlier does not influence peak shape near cluster of three measurements; for single Gaussian, it would make it narrower
  - long tails of likelihood functions from cluster allows outlier to produce a small secondary peak; has little effect on posterior mean





Hierarchical model – scale uncertainties

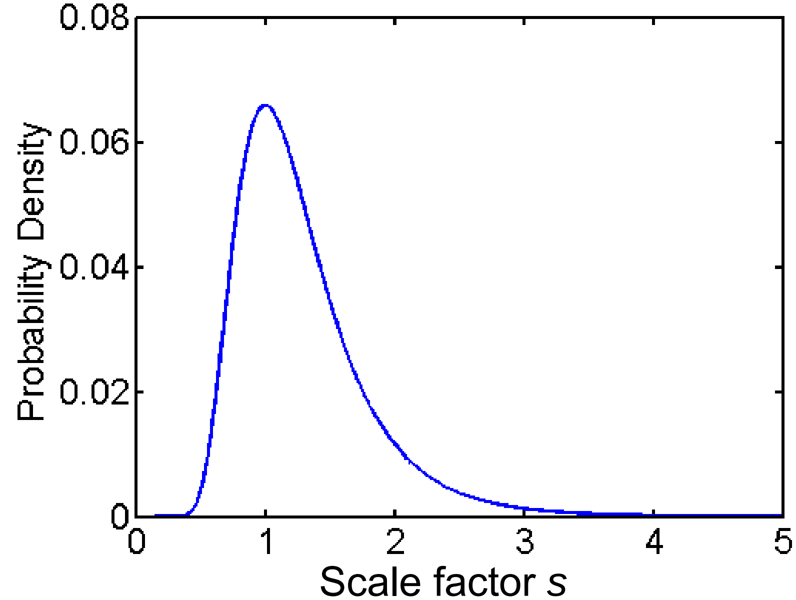
- When data disagree a lot, we may question whether quoted standard errors are correct
- Scale all σ by factor *s*: $\sigma = s\,\sigma_0$
- Then marginalize over *s*

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

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

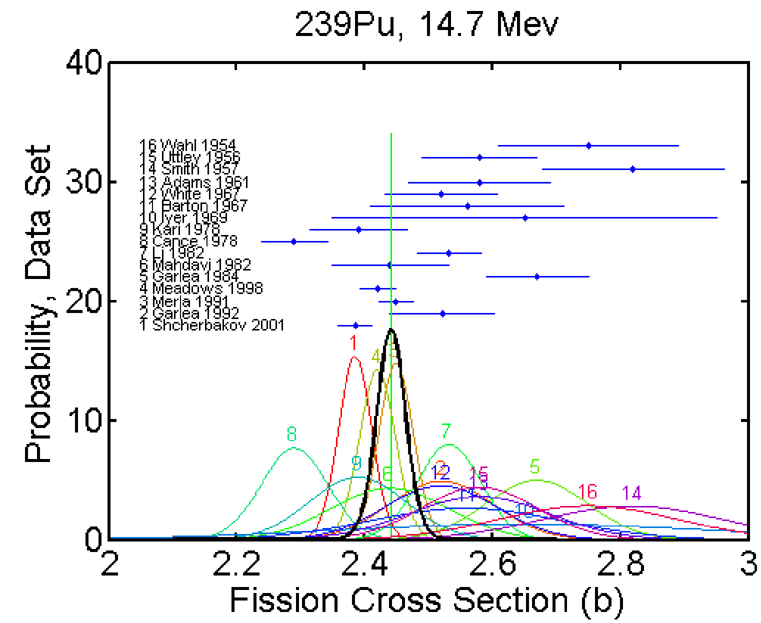
$$p(\boldsymbol{a}\,|\,\boldsymbol{d}) \propto \int p(\boldsymbol{d}\,|\,\boldsymbol{a}, s)\,p(\boldsymbol{a})\,p(s)\,\mathrm{d}s$$

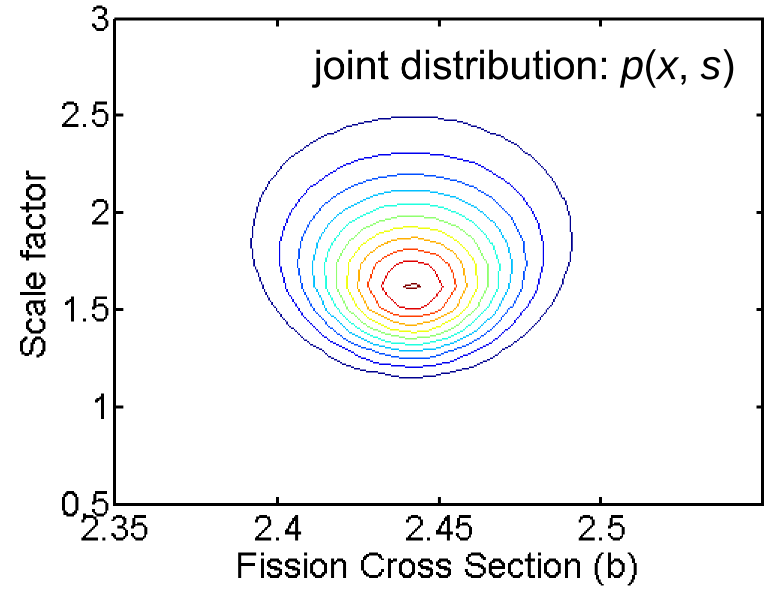
- For prior $p(s)$, either use noninformative (flat in log(s)) or one like shown in plot
- Let the data decide!
- This is called **hierarchical model** because properties of one pdf, the likelihood, are specified by another pdf



^{239}Pu cross sections – scale uncertainties

- Accommodate large dispersion in data by scaling all σ by factor *s*:
  $\sigma = s\,\sigma_0\ ;\ \ \sigma_0$ = quoted stand. err.
- For likelihood, use Gaussian with scaled σ
  $$p(\boldsymbol{d}\,|\,x,s) \propto \frac{1}{s^n}\exp\left(-\frac{\chi_0^2}{2s^2}\right)$$
- For prior $p(s)$, use non-informative prior for scaling parameter $p(s) \propto 1/s$
- Bottom plot shows joint posterior pdf
- Marginalize over *s*:
  $$p(x\,|\,\boldsymbol{d}) \propto \int p(\boldsymbol{d}\,|\,x,s)\,p(x)\,p(s)\,\mathrm{d}s$$
  to get posterior for *x* (top plot)
- Result is: 2.441 ± 0.024;
  very plausible uncertainty



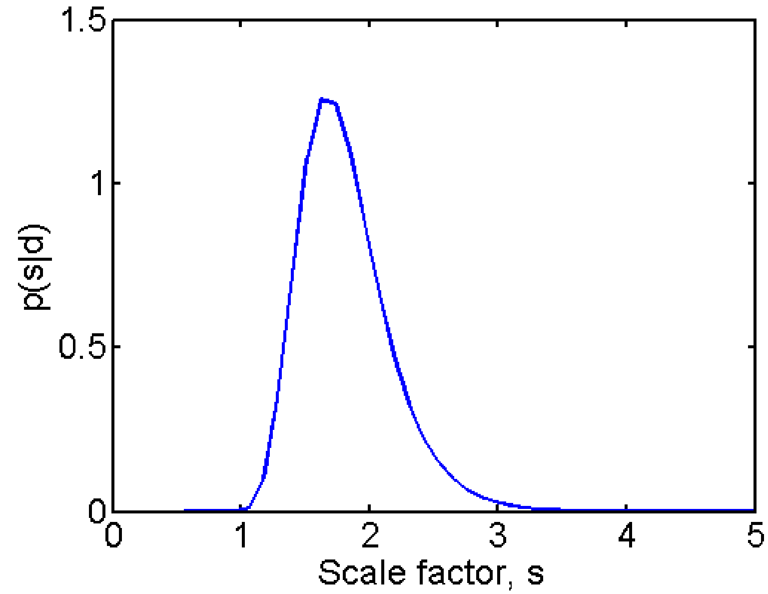


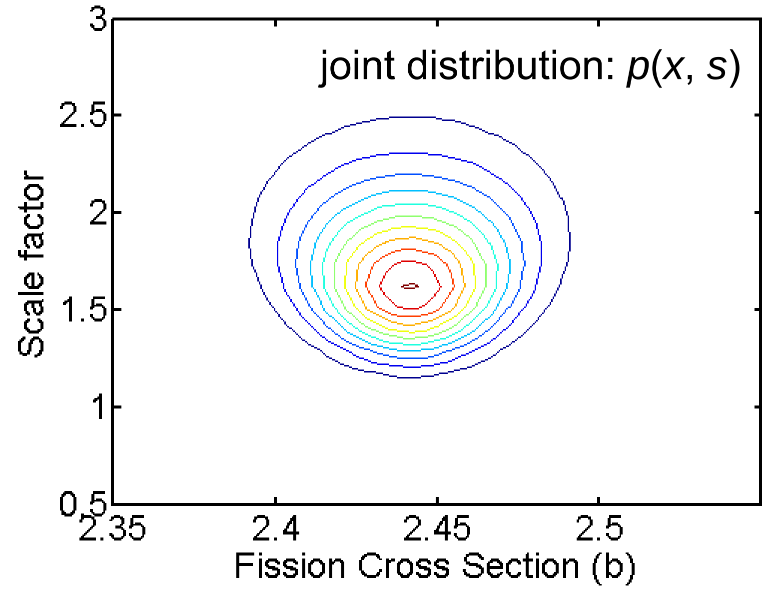
^{239}Pu cross sections – scale uncertainties

- To obtain the posterior for the scaling parameter s, marginalize joint posterior over x:

$$p(s\,|\,\boldsymbol{d}) \propto \int p(\boldsymbol{d}\,|\,x,s)\; p(x)\; p(s)\,\mathrm{d}x$$

- Plot (top) shows result
  - maximum at about 1.7, $\approx \sqrt{\dfrac{\chi^2}{\mathrm{DOF}}}$ for original fit
  - however, this result is different from just scaling σ to make χ^2 per DOF unity
  - it allows for a distribution in s, taking into account that s is uncertain
- This model can be extended to allow each σ_i to be scaled separately
  - prior on s_i could reflect our confidence in quoted σ_i for each experiment





Summary

In this tutorial:

- Types of uncertainties in measurements – random and systematic
- Uniform prior ⇒ likelihood analysis ⇒ χ^2 analysis
- Used straight line fit to illustrate various Bayesian concepts and models
  - posterior sampling; predictive distribution and model checking
  - systematic uncertainties
  - averaging over many measurements
  - outliers
- Studied Pu cross-section data at 14.7 MeV
  - outlier-tolerant likelihood
  - scaling of quoted standard errors using a distribution of scales, which is determined by input data