Bayesian analysis of inconsistent measurements of neutron cross sections

Kenneth M. Hanson

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

Bayesian Inference and Maximum Entropy Workshop, August 7-12, 2005

Los Alamos NATIONAL LABORATORY

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

LA-UR-05-6367

Overview

- Outliers
- Bayesian treatment
- Physical interpretation of likelihood
  - potential, force
- Outlier data
  - Pu-239 fission cross sections at 14.7 MeV
- Discrepant data sets
  - Am-243 fission cross-section data

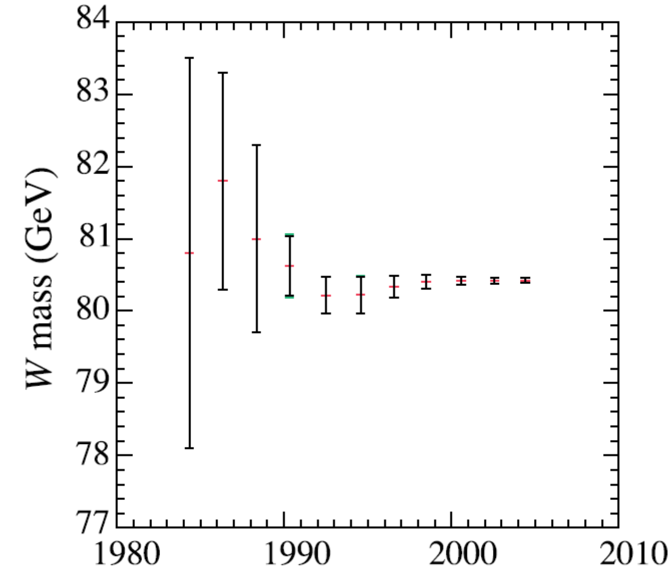
- Acknowledgment and thanks to my colleagues
  Toshihiko Kawano and Patrick Talou

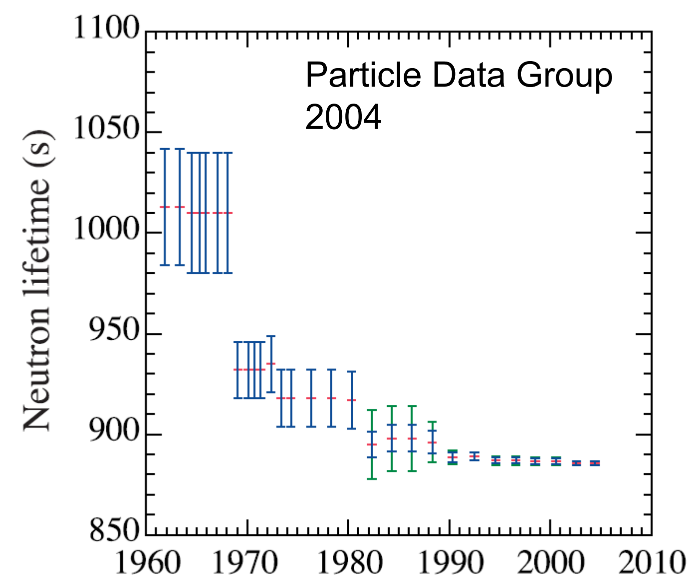
Outliers

- Outliers often caused by mistakes made in taking data or analysis
- Mistakes happen!
  - ask any experimentalist
  - experience and care can reduce number of mistakes, but not eliminate them
- Question is, how do we cope with outliers?
  - be careful, outlier may be real (could mean Nobel prize)
  - traditional approach: identify outliers and drop them from analysis
    - iterative process; may be difficult to decide which data are outliers
    - data are either in or out
  - Bayesian approach: include in likelihood function as long tail
    - iterative process (because pdf is not unimodal), but automatic
    - includes all data
    - weight of each datum is regulated by how well supported by other data

History of particle-properties measurements

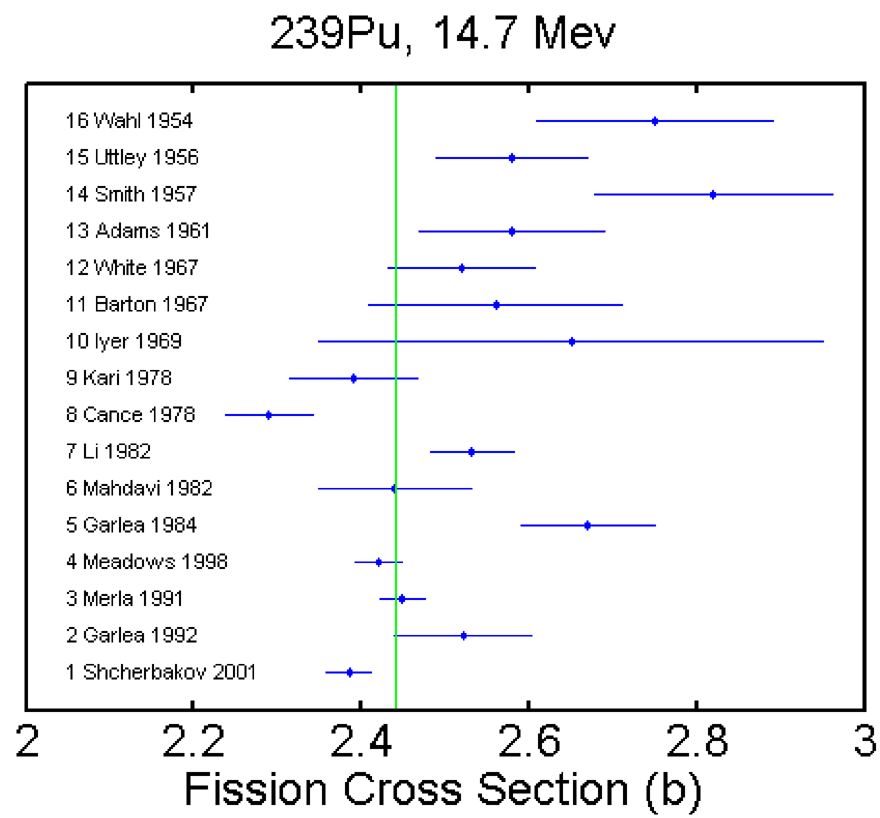
- Plots show histories of two “constants” of fundamental particles
- Mass of W boson
  - logically ordered history
  - all within error bar of last measurement
- Neutron lifetime
  - unsettling history
  - periodic jumps with periods of extreme agreement
  - measurements before 1980 disagree with latest ones
  - systematic errors hard to estimate
  - plot demonstrates human aspects of experimental physics





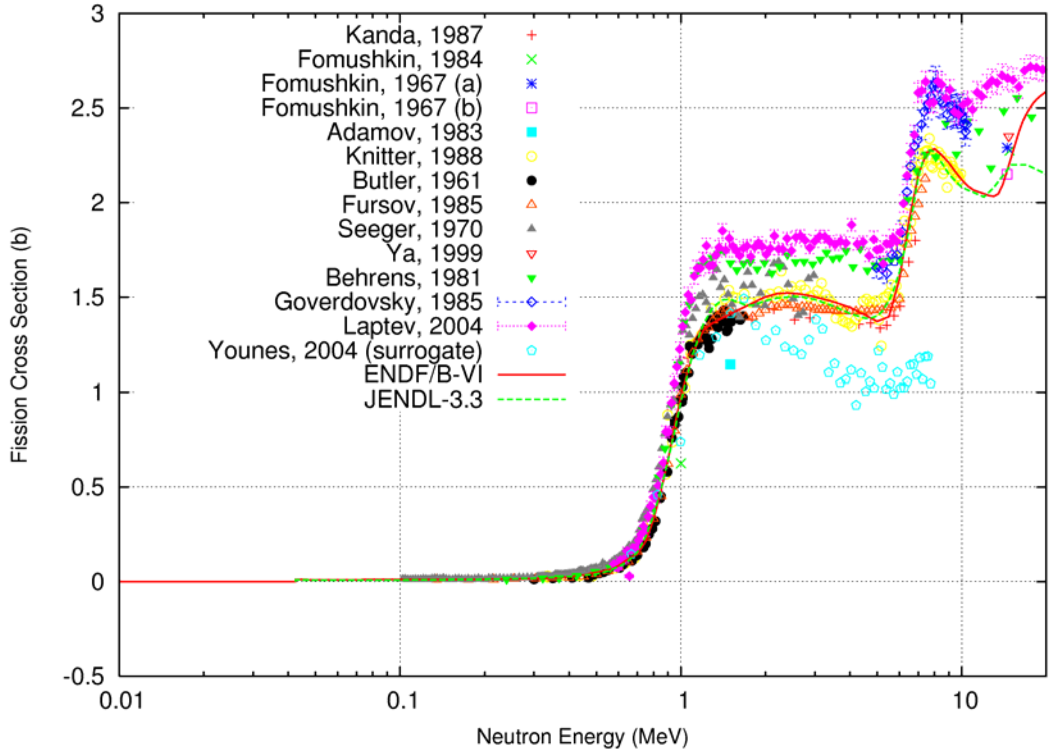
Neutron fission cross section data for ^{239}Pu

- 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



Neutron fission cross-section data

^{243}Am fission cross section



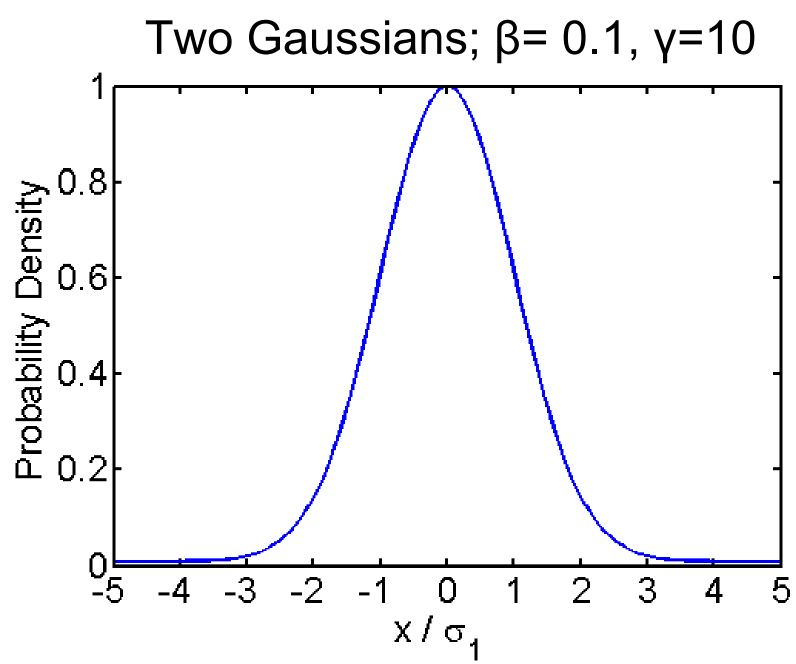
plot from P. Talou

- Neutron cross sections measured by many experimenters
  - sometimes data sets do not agree
  - often little information about uncertainties, esp. systematic errors
  - many data, many experiments – opportunity to learn about data

Outliers

- Long history in Bayesian analysis (outliers and robust estimation)
  - deFinnetti (61), Box and Tiao (68), O'Hagan (79), Berger (91), and many more
  - Hanson and Wolf (92), Sivia (96), Press (97), Dose and von der Linden (99), Fröhner (00)
- Types of likelihood functions, generally have long tail
  - long tail includes possibility of large deviation from true value
  - exact form doesn't seem to matter much
- Simple model: likelihood is mixture of two Gaussians

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

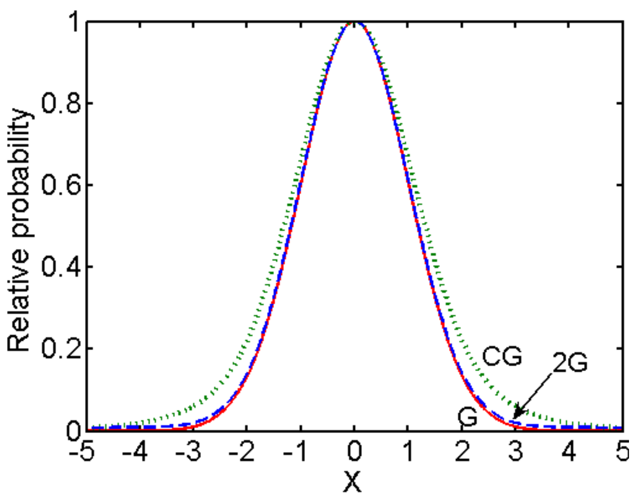


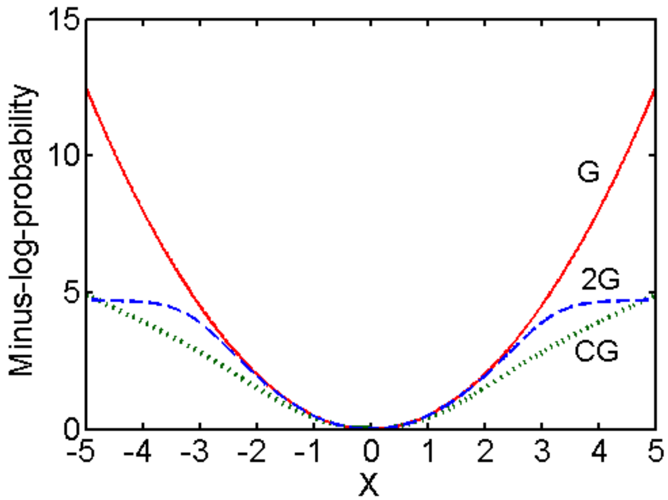
Physical analogy of probability

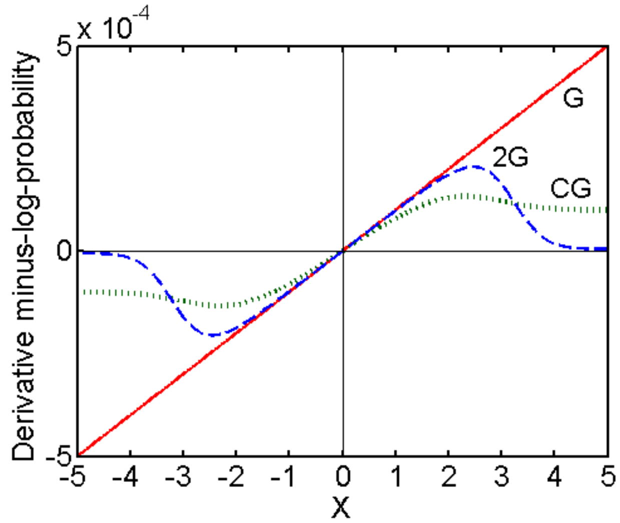
- Think of $\varphi = -\log\{p(\boldsymbol{a} \mid \boldsymbol{y}, \boldsymbol{x})\}$ as a physical potential
  - generally useful notion
  - Gaussians yield quadratic φ
    - linear force law $\nabla\varphi \propto -\boldsymbol{a}$
    - each datum pulls on fit model with force that increases linearly with residual
  - helpful in designing algorithms, e.g., Hamiltonian hybrid MCMC
  - gives meaning to **inferential force of datum**
- Outlier-tolerant likelihoods
  - generally have long tails, restoring force eventually decreases for large residuals

Physical analogy of probability

- Outlier-tolerant likelihoods
  - generally have long tails
  - restoring force eventually decreases for large residuals

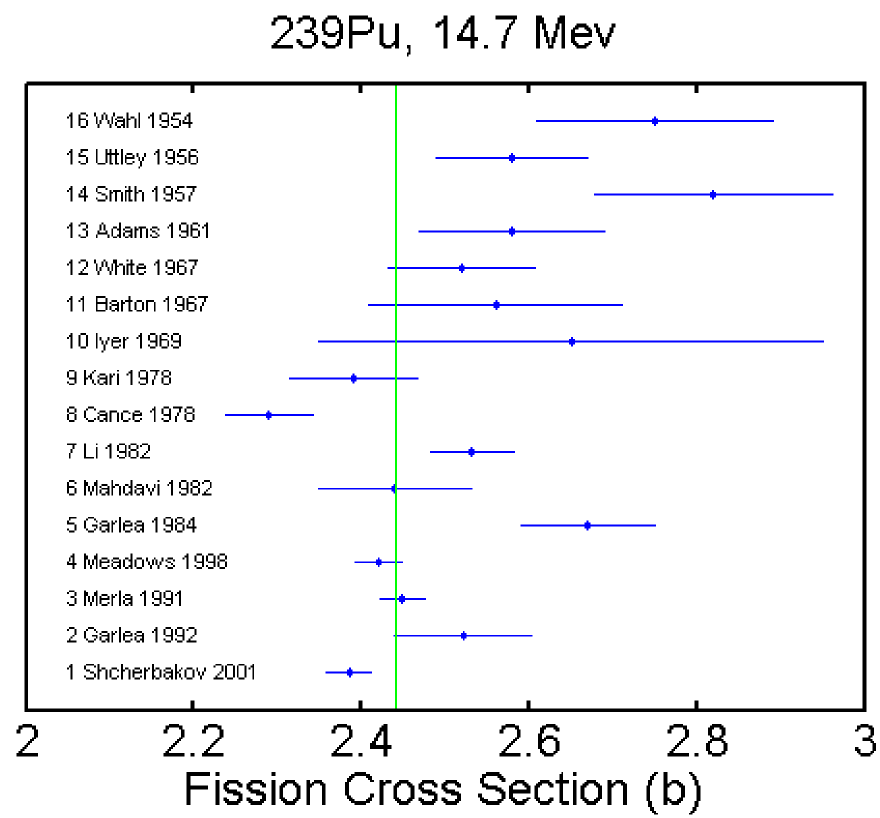






Neutron fission cross section data for ^{239}Pu

- 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
  - outliers?; three data contribute 24 to χ^2, more than half

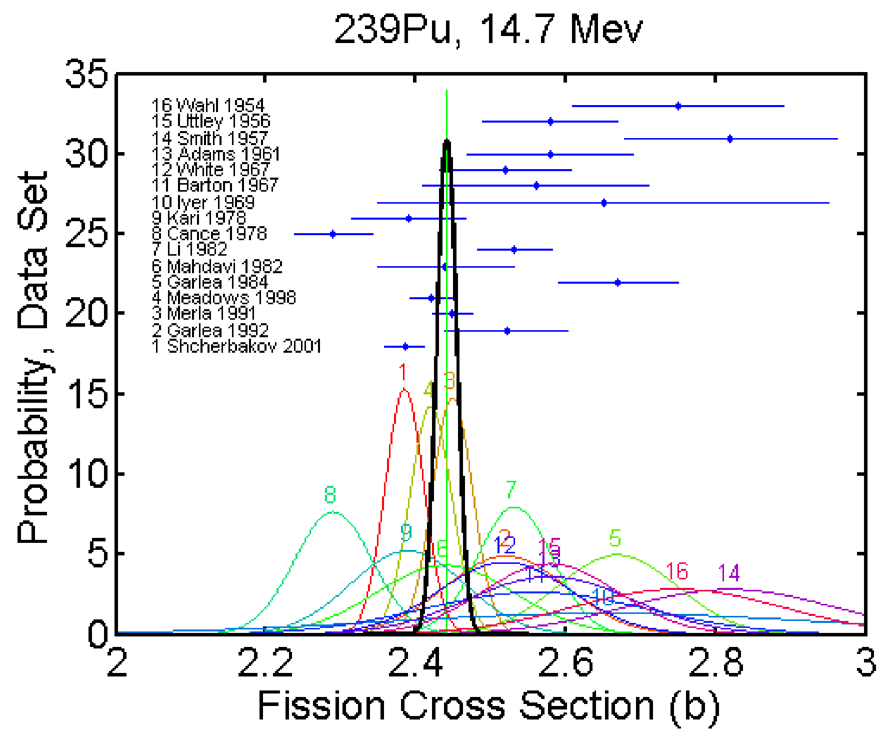


^{239}Pu cross sections – Gaussian likelihood

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

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

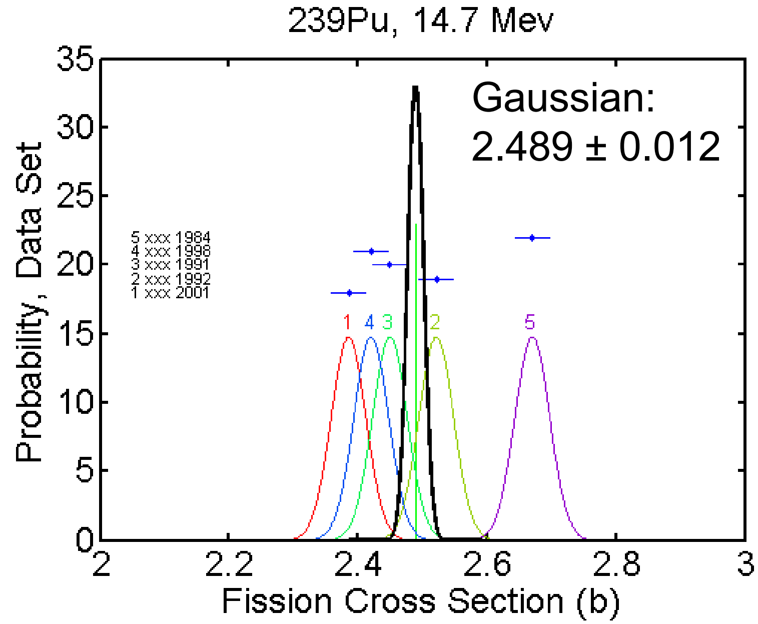
  - independent of where data lie!
    which doesn't make sense

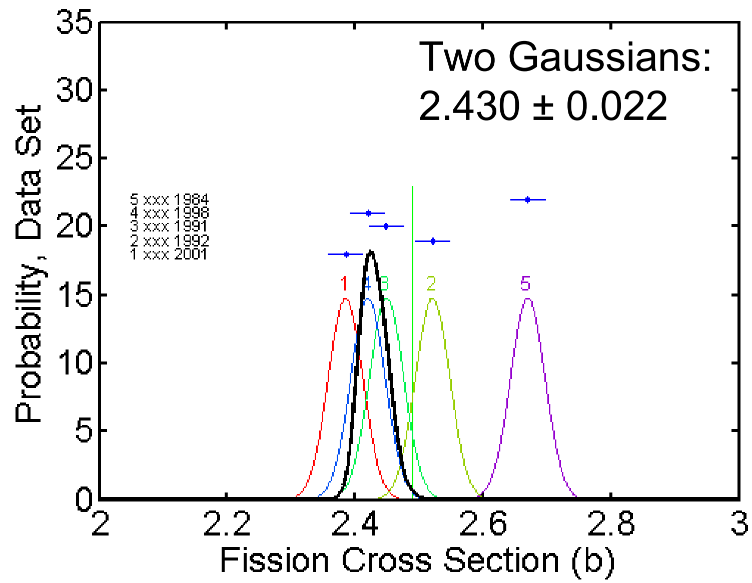


Gaussian: 2.441 ± 0.013

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

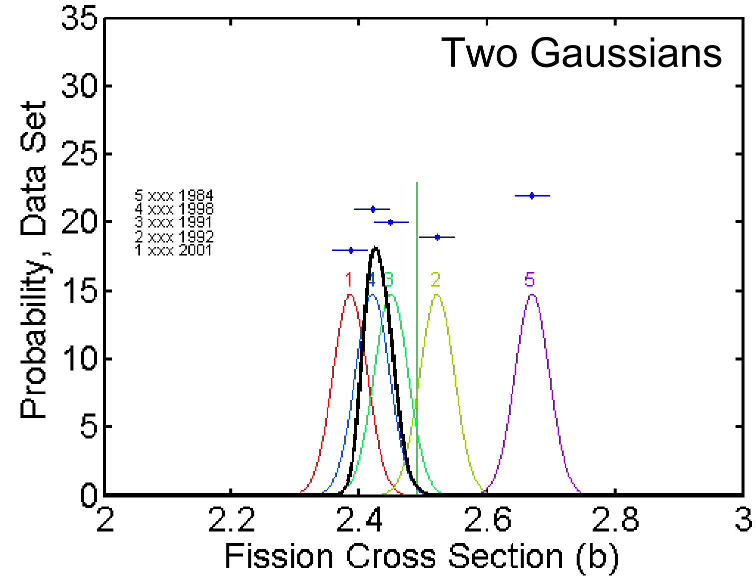
- Use just latest five measurements
- To exaggerate outlier problem, set all standard errors = 0.027
- Compare results from alternative likelihoods:
  - Gaussian: 2.489 ± 0.012
    $\chi^2 = 69.9$, $p = 2\times10^{-14}$ for 4 DOF
  - two Gaussians: 2.430 ± 0.022
- For two-Gaussian likelihood:
  - result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

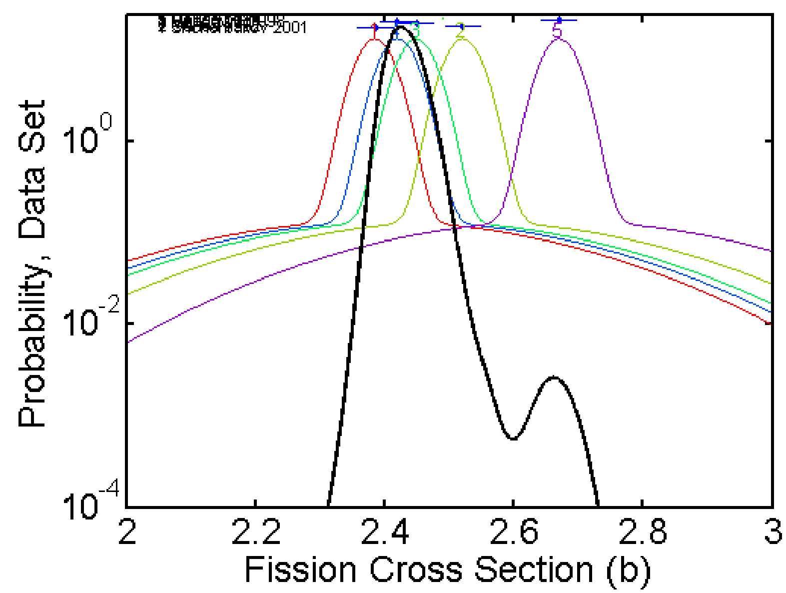




^{239}Pu cross sections – outlier-tolerant likelihood

- 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

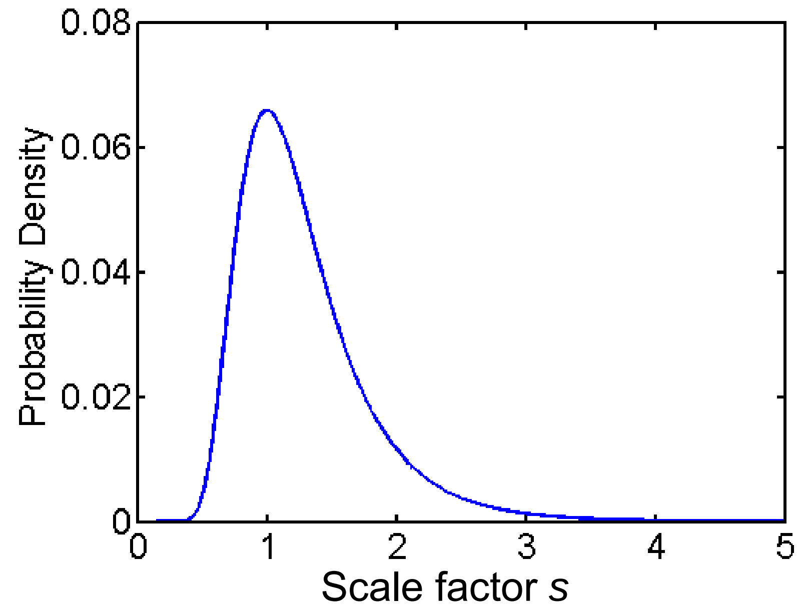
- When data disagree a lot, we may question whether quoted standard errors are correct
- Scale all σ by factor *s*: $\sigma = s\,\sigma_0$
- Then marginalize over *s*

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

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

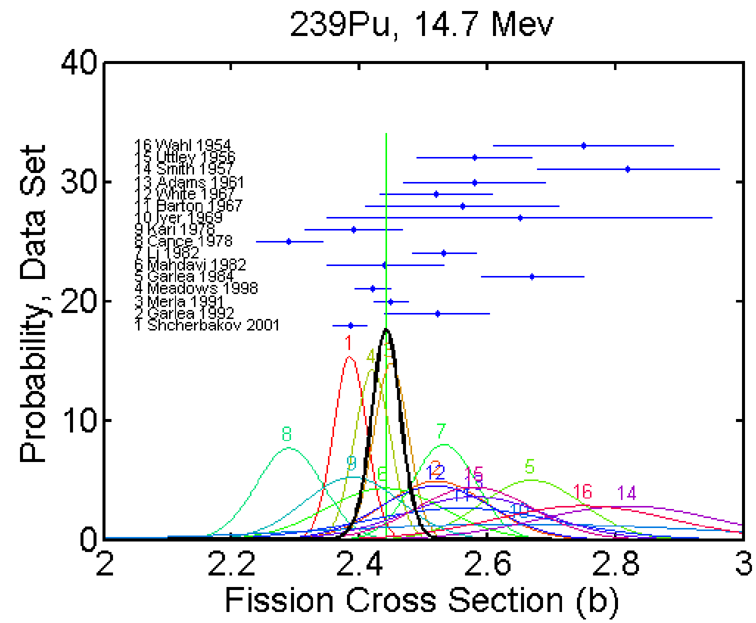
$$p(\boldsymbol{a}\,|\,\boldsymbol{d}) \propto \int p(\boldsymbol{d}\,|\,\boldsymbol{a}, s)\,p(\boldsymbol{a})\,p(s)\,\mathrm{d}s$$

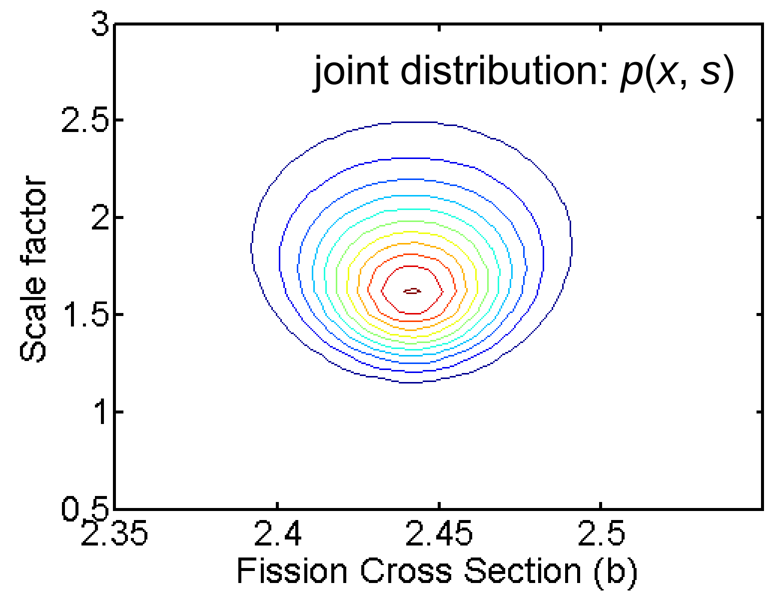
- For prior $p(s)$, either use noninformative (flat in log(*s*)) or one like shown in plot
- Let the data decide!
- This is called **hierarchical model** because one pdf depends on another pdf



^{239}Pu cross sections – scale uncertainties

- Accommodate large dispersion in data by scaling all σ by factor *s*:
  $$\sigma = s\,\sigma_0\ ;\quad \sigma_0 = \text{quoted stand. err.}$$
- For likelihood for *n* points, use Gaussian with scaled σ
  $$p(\mathbf{d}\,|\,x,s) \propto \frac{1}{s^n}\exp\left(-\frac{\chi_0^2}{2s^2}\right)$$
- 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*:
  $$p(x\,|\,\mathbf{d}) \propto \int p(\mathbf{d}\,|\,x,s)\,p(x)\,p(s)\,\mathrm{d}s$$
  to get posterior for *x* (top plot)
- Result: 2.441 ± 0.024; very plausible



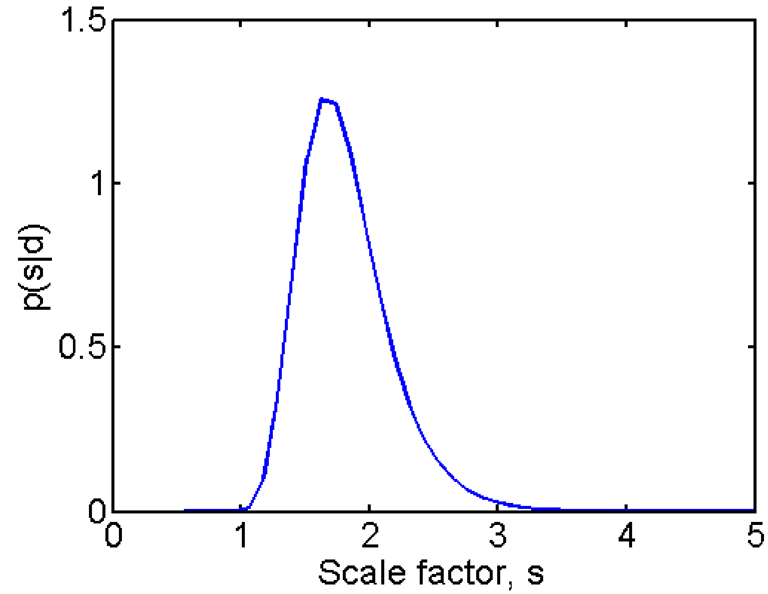


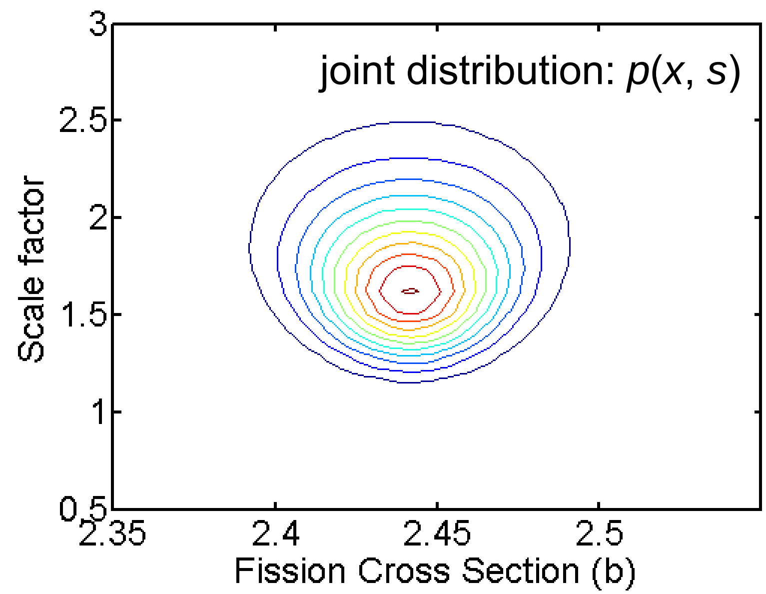
^{239}Pu cross sections – scale uncertainties

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

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

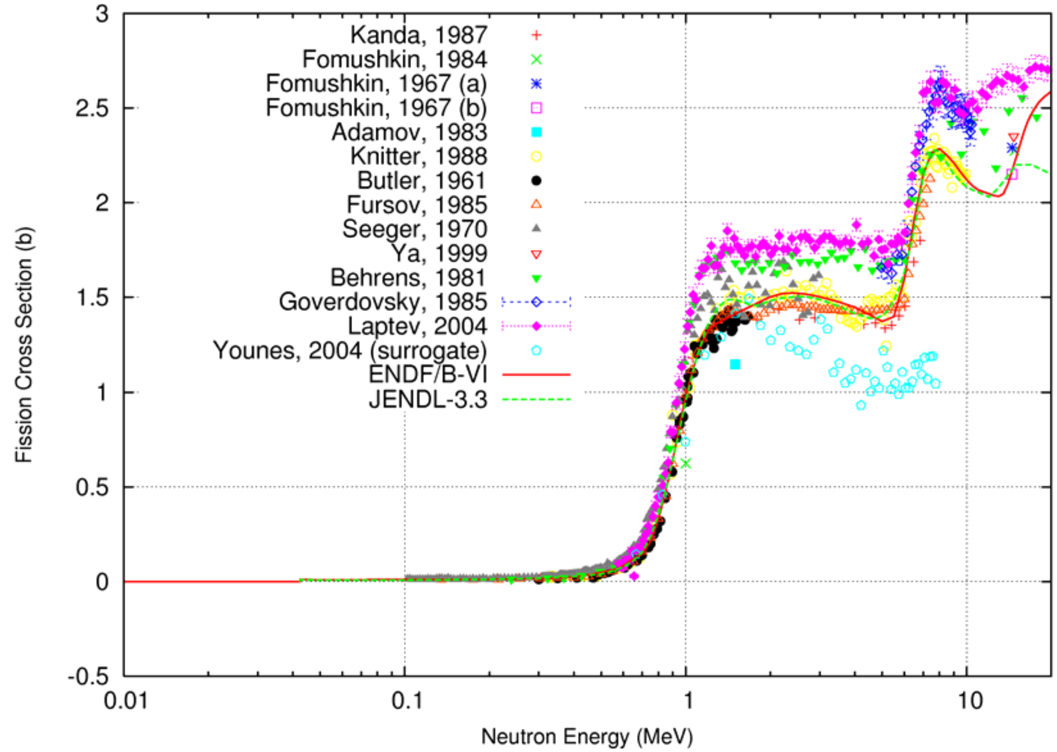
- Plot (top) shows result
  - maximum at about 1.7, $\approx \sqrt{\frac{\chi^2}{\mathrm{DOF}}}$ for original fit
  - however, this result is different than just scaling σ to make χ^2 per DOF unity
  - it allows for a distribution in s, taking into account that s is uncertain
- This model can be extended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





Neutron fission cross-section data

^{243}Am fission cross section



plot from P. Talou

- Observe in this plot
  - principle cause of discrepancies could be in normalization
  - systematic uncertainty in normalization

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

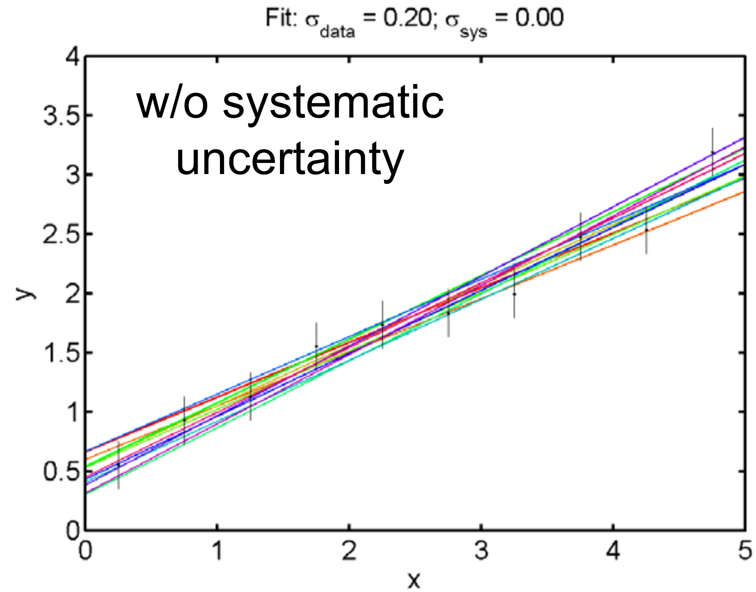
- Writing $p(a,b,\Delta \mid \boldsymbol{y},\boldsymbol{x}) \propto \exp\{-\varphi\}$ and assuming normal distributions

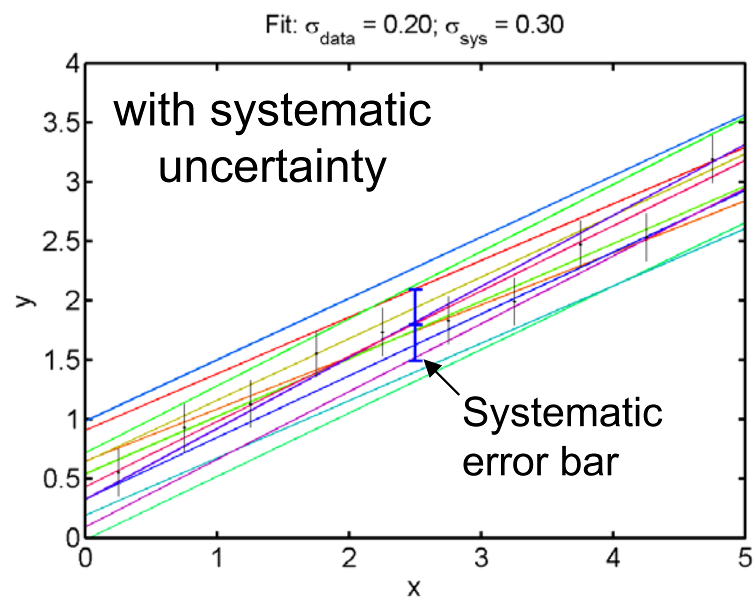
$$2\varphi = \sum \frac{\left(y_i - f(x_i; a,b) - \Delta\right)^2}{\sigma_i^2} + \frac{\Delta^2}{\sigma_\Delta^2}$$

- 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

- Linear model: $y = a + bx + \Delta$
- Simulate 10 data points, $\sigma_y = 0.2$
  exact values: $a = 0.5 \quad b = 0.5$
- Introduce systematic offset Δ with uncertainty $\sigma_\Delta = 0.3$
- Determine parameters, *a*, *b*, and offset Δ; marginalize over Δ
- Colored lines are model realizations drawn from parameter uncertainty pdf
- Systematic uncertainty has effect of introducing additional variation (uncertainty) in vertical direction





Probabilistic model for normalization error

- Represent common uncertainty in measurements by systematic error in normalization factor c: $cx_i = m_i + \varepsilon_i$
  - where the ε_i represent the random fluctuations
- Following same development as before, where prior $p(c)$ assumed normal with expected value of 1 and σ_c = rms uncertainty in normalization
- Writing $p(cx, c \mid \boldsymbol{m}) \propto \exp\{-\varphi\}$

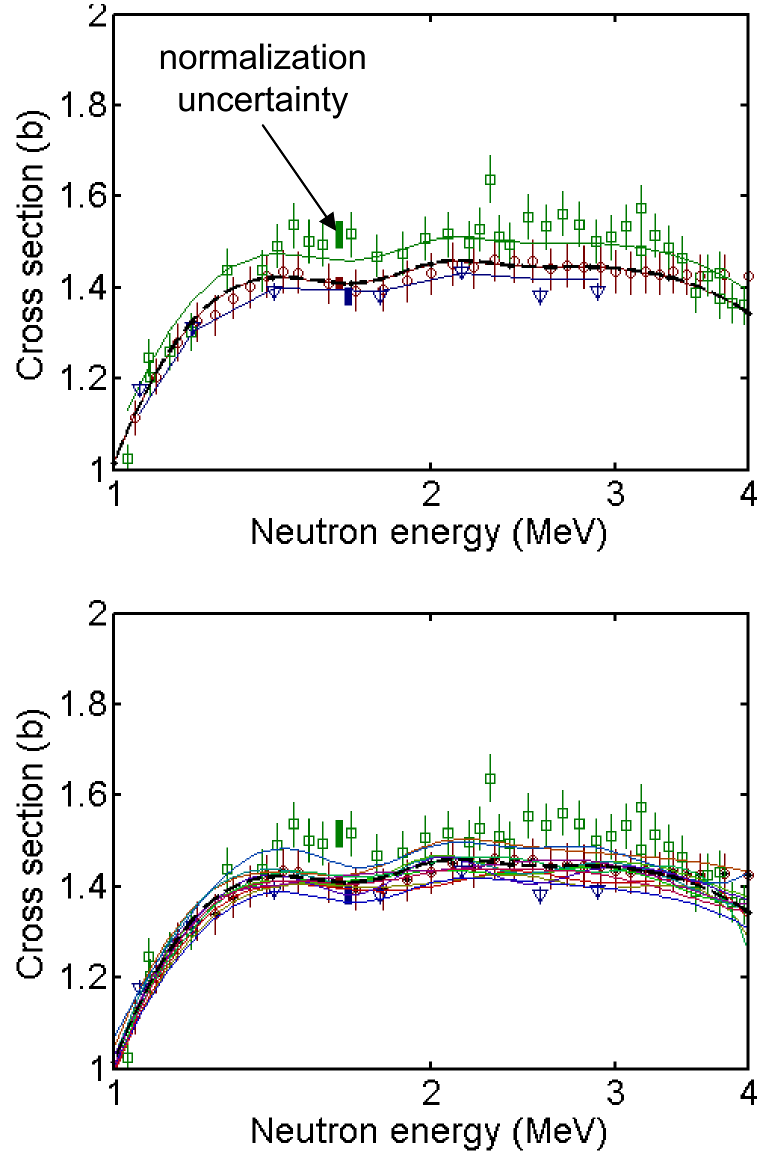
$$2\varphi = \sum_i \frac{(cx - m_i)^2}{\sigma_i^2} + \frac{(c-1)^2}{\sigma_c^2}$$

- Divide $p(cx, c)$ by Jacobian $J = 1/c$ to get $p(x, c)$, which is a log-normal distribution
- $p(x|\boldsymbol{m})$ obtained by numerical integration: $p(x \mid \boldsymbol{m}) = \int p(x, c \mid \boldsymbol{m})\, \mathrm{d}c$

Systematic uncertainties – normalization

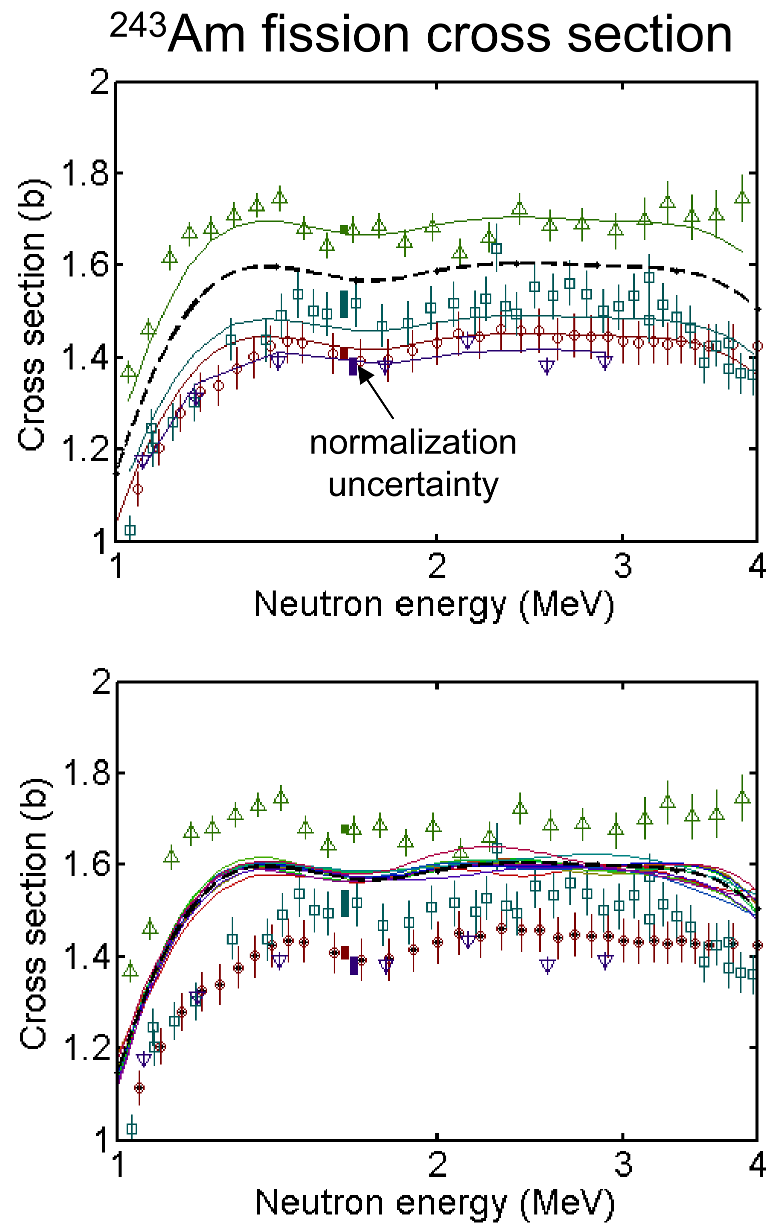
- Consider energy range 1 – 4 MeV
- Three data sets; agree somewhat
- Normalization error of each data set = 1.4%, 2.8%, 1.8% (vertical bars)
- Scale each data set with stated error
- Fit cubic splines, 9 knots; min χ^2
- Top graph – black line is estimate
  - colored lines show normalization of each data set
  - at 2 MeV, nxs = 1.453 ± 0.021
- Bottom graph – posterior samples
  - plausible, but uncertainty smaller than dispersion in data suggests

^{243}Am fission cross section



Discrepant data sets – Gaussian likelihood

- Four data sets; one disagrees in normalization with others by >10 σ
- Normalization error in data sets = 1.4%, 2.8%, 1.8%, (0.9%)
- Treat normalization as systematic effect
- Likelihood: exp($-\chi^2/2$), Gaussian
- Prior on scale factor is Gaussian with stated uncertainties
- nxs(2 MeV) = 1.588 ± 0.016
  - new data set moves result by 7 times their combined error !

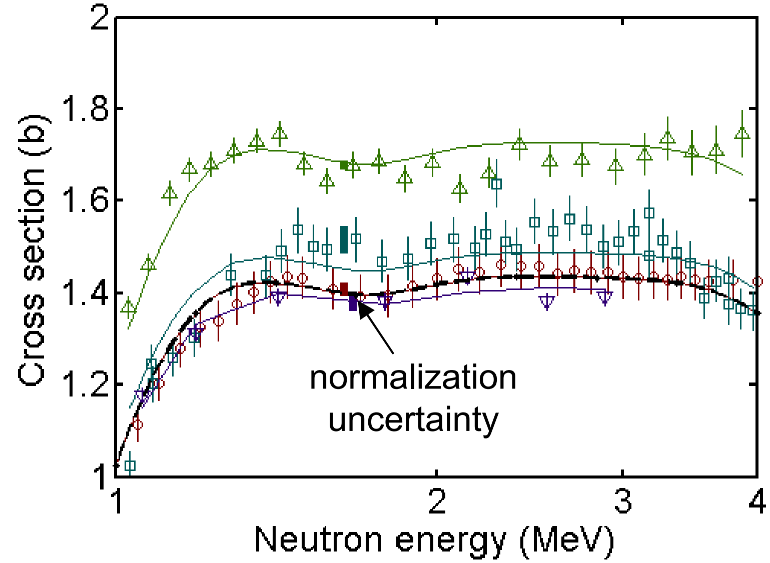


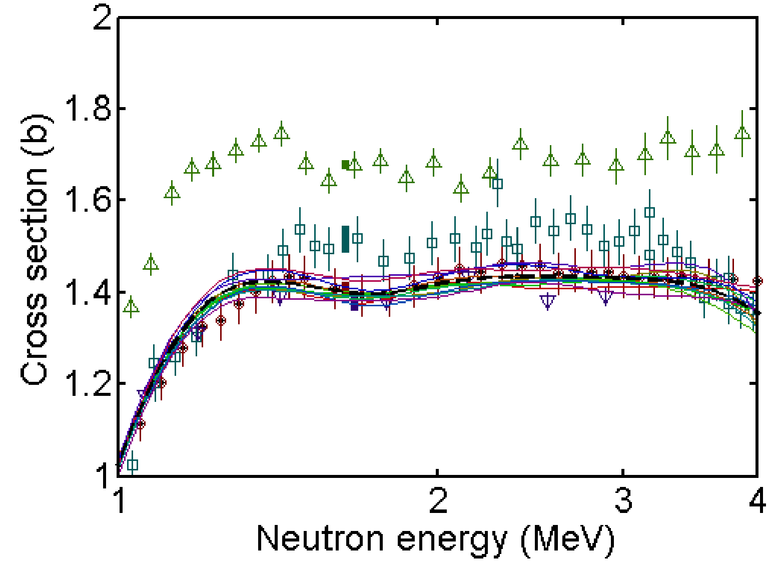
Discrepant data sets – Cauchy-Gaussian mix

- Normalization error in data sets = 1.4%, 2.8%, 1.8%, (0.9%)
- Treat normalization as systematic effect
- Likelihood: Gaussian
- Use outlier-tolerant prior on scale factors to include possibility of gross error in normalization:
  - 0.67*Cauchy + 0.33*Gaussian mixture
- Normalization of outlying data set has no influence on result, but its **shape** is included
- nxs(2 MeV) = 1.418 ± 0.021

  very plausible result

^{243}Am fission cross section





Future work

- Systematic uncertainties
  - include possibility of scaling the quoted uncertainties
  - use informative priors based on knowledge of experiments:
    how done, techniques used, who did them
  - do thorough analysis of what kinds of uncertainties are typical and include them
- Treatment of outliers
  - systematically investigate various choices for form of long-tailed likelihood function
  - balance ability to ameliorate effects of outliers with undue increase in posterior variance
- Do global analysis on original data (often in the form of ratios to standard cross sections with smaller error bars)

Bibliography

- A. O'Hagan, "On outlier rejection phenomena in Bayes inference," *J. Roy. Statist. Soc.* **B 41**, 358-367 (1979)
- F H Fröhner, "Bayesian evaluation of discrepant experimental data," *Maximum Entropy and Bayesian Methods - 1988*, J Skilling, ed., 467–474 (Dordrecht: Kluwer Academic, 1989)
- K. M. Hanson and D. R. Wolf, "Estimators for the Cauchy distribution," *Maximum Entropy and Bayesian Methods - 1993,* G. R. Heidbreder, ed., pp. 255-263 (Kluwer Academic, Dordrecht, 1996)
- D. S. Sivia, "Dealing with duff data," *MAXENT 96: Proceedings of the Maximum Entropy* Conference, M. Sears et al., eds., 131–137 (Port Elizabeth: N.M.B. Printers, 1996)
- W.H. Press, "Understanding data better with Bayesian and global statistical methods," *Unsolved problems in astrophysics*, J. N. Bahcall and J. P. Ostriker, eds., 49–60 (Princeton, Princeton University, 1997)
- V. Dose and W. von der Linden, "Outlier tolerant parameter estimation," *Maximum Entropy and Bayesian Methods - 1998,* von der Linden W et al., eds., 47–56 (Dordrecht: Kluwer Academic, 1999)

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