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
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 – three 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
Tutorial 2
Bayesian modeling
Peelle’s Pertinent Puzzle (1987)

Overview:

• Paradoxical result produced by strong correlations in uncertainties
• Probabilistic view of PPP
• Specific probabilistic model for PPP elucidates how correlations in uncertainties arise
• Plausible experimental situation consistent with PPP result
• Bayesian approach to coping with uncertainty in model
• With probabilistic modeling, you can go beyond simple linear, additive models
• PPP underlines the need to specify how uncertainties contribute to reported data
Peelle’s pertinent puzzle

• Robert Peelle (ORNL) posed the PPP in 1987:
  Given two measurements of same quantity $x$:
  \[ m_1 = 1.5; \quad m_2 = 1.0, \]
  each with independent standard error of 10% ,
  and fully correlated standard error of 20% .
  Weighted average using least-squares is $x = 0.88 \pm 0.22$

• Peelle asks “under what conditions is this result reasonable?”
• By extension, if this not reasonable, what answer is appropriate?
• PPP is pertinent – its effect has been observed in nuclear data
  evaluation for decades
• Comment – PPP description of errors is ambiguous, which leads
  to numerous plausible interpretations
PPP in cross-section evaluation

• Although the PPP problem may seem academic, it has significant real-world consequences in cross-section evaluation
  ▶ historically, fits to several data sets fall below lowest measurements

from Pronyaev, INDC(NDS)-438, p. 163 (2003)

note large data discrepancies
Standard solution to PPP

• The solution given in PPP is based on standard matrix equations for least-squares result:
  
  estimated value \( x = (G^T C^{-1} G)^{-1} G^T C^{-1} m \)
  
  covariance in estimate \( V = (G^T C^{-1} G)^{-1} \)

  where the sensitivity matrix is \( G = [1.0 \ 1.0] \)
  
  and the measurements are the vector \( m = [1.5 \ 1.0]^T \)

  with covariance matrix \( C = \begin{pmatrix} 1.5^2 \star (0.1^2 + 0.2^2) & 1.5 \star 1.0 \star 0.2^2 \\ 1.5 \star 1.0 \star 0.2^2 & 1.0^2 \star (0.1^2 + 0.2^2) \end{pmatrix} \)

• Result is \( x = 0.88 \pm 0.22 \)

• This result is smaller than both measurements, which seems implausible
Probabilistic view of standard PPP solution

- Consider the probability density function (pdf) for the variables
  \[ x = [x_1 \ x_2]^T \]
  \[ p(x | m) \propto \exp \left\{ -\frac{1}{2} (x - m)^T C^{-1} (x - m) \right\} \]
  where measurements are \( m = [1.5 \ 1.0]^T \) and their covariance matrix is
  \[ C = \begin{pmatrix}
                  1.5^2 \ast (0.1^2 + 0.2^2) & 1.5 \ast 1.0 \ast 0.2^2 \\
                  1.5 \ast 1.0 \ast 0.2^2 & 1.0^2 \ast (0.1^2 + 0.2^2)
               \end{pmatrix} \]

- For \( x = x_1 = x_2 \) (diagonal of 2D pdf), \( p(x|m) \) is normal distribution centered at 0.88
Probabilistic model for additive error

- Represent common uncertainty in measurements by systematic additive offset \( \Delta: \ x_1 = m_1 + \epsilon_1 + \Delta; \ x_2 = m_2 + \epsilon_2 + \Delta \)
  - where the \( \epsilon_i \) represent the random fluctuations
- Bayes law gives joint pdf for \( x \) and \( \Delta \)
  \[
p(x, \Delta | m) = p(m | x, \Delta) p(x) p(\Delta)
\]
  where priors \( p(x) \) is uniform and \( p(\Delta) \) assumed normal (\( \sigma_\Delta = 0.2 \))
- Writing \( p(x, \Delta | m) \propto \exp\{ -\phi \} \) and assuming normal distributions
  \[
  2\phi = \frac{(x_1 - m_1 - \Delta)^2}{\sigma_1^2} + \frac{(x_2 - m_2 - \Delta)^2}{\sigma_2^2} + \frac{\Delta^2}{\sigma_\Delta^2}
  \]
  where \( \sigma_1 = 0.1 \times m_1; \ \sigma_2 = 0.1 \times m_2; \ \sigma_\Delta = 0.2 \)
- Pdf for \( x \) obtained by integration: \( p(x | m) = \int p(x, \Delta | m) d\Delta \)
- This model equivalent to
  \[
p(x | m) \propto \exp\left\{ -\frac{1}{2} (x - m)^T C^{-1} (x - m) \right\}
  \]
Plausible experimental scenario

- Under what conditions is PPP result reasonable?
- Suppose that
  - measurements made in intervals shown
  - from experience with apparatus, we know background increases linearly in time
    - background subtraction for \( m_1 \) is 1.5 times larger than for \( m_2 \); leads to stated covariance matrix
- For this scenario, the additive model is appropriate, and the PPP solution, 0.88, is the correct answer
Probabilistic model for normalization error

- Represent common uncertainty in measurements by systematic error in normalization factor $c$: $cx = m_1 + \varepsilon_1$; $cx = m_2 + \varepsilon_2$
  - where the $\varepsilon_i$ represent the random fluctuations
- Following same development as before, where prior $p(c)$ assumed normal with expected value of 1 and $\sigma_c = 0.2$
- Writing $p(cx, c \mid m) \propto \exp \{-\varphi\}$
  $$2\varphi = \frac{(cx - m_1)^2}{\sigma_1^2} + \frac{(cx - m_2)^2}{\sigma_2^2} + \frac{(c - 1)^2}{\sigma_c^2}$$
  - where $\sigma_1 = 0.1 \cdot m_1$; $\sigma_2 = 0.1 \cdot m_2$; $\sigma_c = 0.2$
- Divide $p(cx, c)$ by Jacobian $J = 1/c$ to get $p(x, c)$, which is a log-normal distribution
- $p(x)$ obtained by numerical integration: $p(x \mid m) = \int p(x, c \mid m) dc$
- This approach promoted by D. Smith (1991)
Probabilistic view of normalization error

- Consider the probability density function (pdf) for variables \( x = [x_1 \ x_2]^T \)

\[
\chi^2 = \left( \frac{cx_1 - m_1}{m_1 \rho_1} \right)^2 + \left( \frac{cx_2 - m_2}{m_2 \rho_2} \right)^2 + \left( \frac{c-1}{\sigma_c} \right)^2;
\]

\( \sigma_c = \rho_c; \)

where measurements are \( m = [1.5 \ 1.0]^T \)

- also, divide \( p(cx, c) \) by Jacobian \( J = 1/c \) to get \( p(x, c) \),

- for \( x = x_1 = x_2 \) (diagonal of 2D pdf), \( p(x|m) \) is not a simple normal distribution

- max at: \( x_{\text{max}} = 1.074 \)

- posterior mean and rmsd:

\[
x = 1.200 \pm 0.276
\]
Probabilistic model for normalization error

- Compare pdfs for two models for correlated effect:
  - A – additive offset
  - B – normalization factor
- Observe significant difference in two results
  - emphasizes need to know which kind of effect leads to correlation
- Probabilistic modeling is capable of handling a variety of known effects
But which model should we use?

- Ambiguity in specifying source of correlation leads to uncertainty about which model to use.
- Bayesian approach can handle model uncertainty
  \[ p(x \mid m) = \int p(x, M \mid m) \, dM \]
  \[ = \int p(x \mid m, M) \, p(M) \, dM \]
  \[ = \frac{1}{2} p(x \mid m, M_1) + \frac{1}{2} p(x \mid m, M_2) \]
  for two equally likely models \( M_1 \) and \( M_2 \)
- Answer is average **both** pdfs!!
  \[ x = 1.04 \pm 0.30 \]

solid black line is average of A and B
An alternative approach

- Devinder Sivia offers an variation on this approach
- Use data to help decide which model to use

\[ p(x \mid m) = \sum_i p(x, M_i \mid m) \]

\[ = \sum_i p(x \mid m, M_i) p(M_i \mid m) \]

\[ = w_1 p(x \mid m, M_1) + w_2 p(x \mid m, M_2) \]

where \( w_i \) is proportional to the evidence integral for \( p(M_i \mid m) \)

- Answer is: \( x = 0.96 \pm 0.27 \)
- Comment: relative weights depend heavily on resp. priors; perhaps not a good situation

Conclusions

• PPP result is consistent with plausible experimental scenario
  ➤ in which correlated (systematic) error contributes additively to result
• Ambiguous statement of the PPP leads to other interpretations
  ➤ some of which yield more plausible answers
• Analysts need better information to analyze data without guessing

• Probabilistic modeling can cope with various known uncertainty effects
Conclusions

• **Experimenters – please provide measurement details**

• Some of the details needed:
  ► specify standard errors as precisely as possible, indicating where uncertainties in their assessment lie
  ► specify components in uncertainties and whether they are
    • independent, or correlated, e.g., systematic errors
    • given relative to measured quantities or inferred values
    • additive (background subtraction) or multiplicative (normalization)

• **Correlation matrix by itself is not enough**

• Another issue in PPP is inconsistency between two measurements: one can cope with this discrepancy by introducing notion that the true errors may differ from quoted errors, i.e., treatment of outliers
Monte Carlo techniques

Monte Carlo – represent pdf by a set of point samples
- Typically use MC to draw samples from posterior for parameters, which are fed into model to get prediction; **predictive distribution**
- Visualization of pdf, uncertainty
- Numerical calculations
  - estimation of mean, standard deviation, correlations
  - integration, marginalization
- Quasi-Monte Carlo – select points with more uniform distribution
  - provide more accurate estimates for fixed number of samples
  - often deterministic point sets
- Markov chain Monte Carlo
  - draw random samples for numerically-defined pdf
  - facilitates inference through numerical calculations
Voronoi analysis

- Voronoi diagram
  - partitions domain into polygons
  - points in $i$th Voronoi region are closest to $i$th generating point, $x_i$
  - boundaries often obtained by geometrical construction
- Monte Carlo technique for Voronoi analysis
  - randomly throw large number of points $z_k$ into region
  - compute distance of each $z_k$ to all generating points $\{x_i\}$
  - $z_k$ belongs to Voronoi region of closest $x_j$
  - can compute volume, first moment, radial moments, identify neighbors, ...
- Readily extensible to high dimensions
Centroidal Voronoi Tessellation

- Plot shows 13 random points (·) and the centroids of their Voronoi regions (×)
- A point set is called a Centroidal Voronoi Tessellation (CVT) when the generating points \( z^j \) coincide with the centroids their Voronoi regions; a CVT minimizes
\[
\sum_j \int_{V_j} |z^j - x|^2 \, dx
\]
- Algorithm (McQueen)
  - start with arbitrary set of generating points
  - perform Voronoi analysis using Monte Carlo
  - move each generating point to its Voronoi centroid
  - iterate lasts two steps until convergence
- Final CVT points are uniformly distributed
CVT for multi-variate normal distribution

- CVT algorithm works for an arbitrary density function, e.g., a normal distribution
- In above MC algorithm for Voronoi analysis, simply draw random numbers from desired distribution
- Plots show starting random point set and final CVT set
- Radii of points are rescaled to achieve desired average variance along axes
- CVT points appear uniformly distributed within constraint of adhering to unit-variance normal distribution
- This kind of distribution may have benefits for MC calculations and visualizations
Sampling from correlated normal distribution

• Want to draw samples from multi-variate normal distribution with known covariance $C_x$

• Important to include correlations among uncertainties, i.e., off-diagonal elements

• Algorithm:
  ▶ perform eigenanalysis of covariance matrix of $d$ dimensions
    \[ C_x = U\Lambda U^T \]
    where $U$ is orthogonal matrix of eigenvectors and $\Lambda$ is the diagonal matrix of eigenvalues
  ▶ draw $d$ samples from uncorrelated unit-variance normal distr., $\xi_i$
  ▶ scale this vector by $\lambda_i^{1/2}$
  ▶ transform vector into parameter space using the eigenvector matrix
  ▶ to summarize, fluctuations are given by: $\Delta x = U\Lambda^{1/2}\xi$
Sampling from correlated normal distribution

Proof of algorithm:

- Want to draw samples from multi-variate normal distribution with specified covariance $C_x$

- Algorithm:
  
  fluctuations given by: $\Delta x = U \Lambda^{1/2} \xi$

  where $\xi_i$ randomly drawn from uncorrelated normal pdf and $U$ and $\Lambda$ come from an eigenanalysis of $C_x$: $C_x = U \Lambda U^T$

  where $U$ is orthogonal matrix of eigenvectors and $\Lambda$ is the diagonal matrix of eigenvalues

- Proof:
  
  Covariance of an ensemble of $x$ vectors is

  $$C = \left\langle \Delta x \Delta x^T \right\rangle = \left\langle U \Lambda^{1/2} \xi \xi^T \Lambda^{1/2} U^T \right\rangle$$

  $$= U \Lambda^{1/2} \left\langle \xi \xi^T \right\rangle \Lambda^{1/2} U^T = U \Lambda U^T = C_x$$

  thus, the fluctuations $\Delta x$ have the desired covariance
Neutron cross sections

- Plot shows
  - measured fission cross sections for neutrons on $^{239}$Pu; red data points
  - inferred cross sections; blue line
  - weighted average in 30 energy bins (groups); green histogram

- PARITSN code simulates neutron transport based on multigroup, discrete-ordinates method
  - uses 30 energy bins (groups)
  - calculates criticality for specified configuration of fissile-material
  - establish dependence of criticality experiment to cross sections
Neutron cross sections - uncertainties

- Analysis of measured cross sections yields a set of evaluated cross sections
- Uncertainties in evaluated cross sections are $\sim 1.4-2.4\%$
- Covariance matrix important
- Strong positive correlations caused by normalization uncertainties in each experiment

![Standard error in cross sections](image1)

![Correlation matrix](image2)
JEZEBEL – criticality experiment

- **JEZEBEL experiment (1950-60)**
  - fissile material $^{239}$Pu
  - measure neutron multiplication as function of separation of two hemispheres of material
  - summarize criticality with neutron multiplication factor, $k_{\text{eff}} = 0.9980 \pm 0.0019$
  - very accurate measurement

- Our goal – use highly accurate JEZEBEL measurement to improve our knowledge of $^{239}$Pu cross sections
JEZEBEL – sensitivity analysis

- PARITSN code calculates $k_{\text{eff}}$ on basis of neutron cross sections
- Sensitivity of $k_{\text{eff}}$ to cross sections found by perturbing cross section in each energy bin by 1% and observing increase in $k_{\text{eff}}$
- Observe that 1% increase in all cross sections results in 1% increase in $k_{\text{eff}}$, as expected
Bayesian update

- For data linearly related to the parameters, the Bayesian (aka Kalman) update for Gaussian distributions is

\[
C_1^{-1} x_1 = C_0^{-1} x_0 + S_y^T C_y^{-1} S_y (y - y_0)
\]

\[
C_1^{-1} = C_0^{-1} + S_y^T C_y^{-1} S_y
\]

- \(x_0\) and \(x_1\) are parameter vectors before and after update
- \(C_0\) and \(C_1\) are their covariance matrices
- \(y\) and \(C_y\) are the measured data vector and its covariance
- \(y_0\) is the value of \(y\) for \(x_0\)
- \(S_y\) is the matrix of the sensitivity of \(y\) to \(x\); \(\partial y / \partial x\)

- For the JEZEBEL case, \(y\) is a scalar (\(k_{\text{eff}}\)), \(C_y\) is a scalar (variance), and \(S_y\) is a vector
Updated cross sections

- Plot shows uncertainties in cross sections before and after using JEZEBEL measurement
- Modest reduction in uncertainties; follows energy dependence of sensitivity
- Correlation matrix is significantly altered
- Strong negative correlations introduced by integral constraint of matching JEZEBEL’s $k_{\text{eff}}$
  - reduction in uncertainties in future prediction depends on how closely its sensitivity matches JEZEBEL’s
Linear-response model – output uncertainty

• Assume outputs of a model are linearly related to perturbations in the inputs,
  \[ \delta y = S_y^T \delta x \]
  where \( S_y \) is sensitivity matrix \( \frac{\partial y}{\partial x} \)

• The covariance in the output \( y \) is
  \[ C_y = S_y^T C_x S_y \]
  when output \( y \) is a scalar,
  the covariance \( C_y \) is a scalar (variance),
  and \( S_y \) is a vector

• If linear model is sufficient and one knows \( S_y \), then predictive distribution is easily characterized

• For complex simulations, \( S_y \) is not usually known
Uncertainty in subsequent simulations

- Our goal is to use updated cross sections in new calculations
  - expect that integral constraint will reduce uncertainties
- Demonstrate usefulness of quasi-MC in form of CVT point sets by “predicting” $k_{\text{eff}}$ measured in JEZEBEL
  - for this demo, assume linear model with known sensitivity vector
    - under this assumption, we can calculate exact answer and compare to MC-style sampling to obtain predictive distribution
- For a new physical scenario, we would not have sensitivity vector and would have to do full simulation calculation
  - thus, only a modest number of function evaluations can be done
Accuracy of predicted $k_{eff}$ and its uncertainty

- Prediction based on liner model with known sensitivities
  - only 30 sample sets allowed for neutronics calc. because of time
  - check accuracy of predicted mean and standard deviation
- Conclude – CVT is more accurate than random sampling

**Performance summary from 1000 runs, each with set of 30 sample vectors; ‘rot’ indicates single sample set randomly rotated to achieve each new one**

<table>
<thead>
<tr>
<th></th>
<th>est. mean $k_{eff}$ (avg.)</th>
<th>est. std. dev. $k_{eff}$ (avg.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rms dev.</td>
<td>rms dev.</td>
</tr>
<tr>
<td>random</td>
<td>0.99788</td>
<td>0.00037</td>
</tr>
<tr>
<td>random-rot</td>
<td>0.99824</td>
<td>0.00010</td>
</tr>
<tr>
<td>CVT-rot</td>
<td>0.99796</td>
<td>0.00001</td>
</tr>
<tr>
<td>exact-linear</td>
<td>0.99796</td>
<td>-</td>
</tr>
</tbody>
</table>
Summary

In this tutorial:

• Peelles’ pertinent puzzle
  ► impact on cross-section evaluation
  ► probabilistic modeling; additive and multiplicative systematic effects
  ► experimenters need to provide more than correlation matrices

• Monte Carlo
  ► generation of samples with specified covariance matrix
  ► quasi-Monte Carlo – more uniformly spaced points than random
  ► Centroidal Voronoi Tessellation (CVT) algorithm

• Bayesian updating of cross sections to include integral data
  ► JEZEBEL criticality experiment
  ► integral constraint results in negative correlations
  ► CVT point set improves prediction accuracy