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

Los Alamos
NATIONAL LABORATORY

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

LA-UR-05-5680

Tutorial 1
Bayesian approach

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

Uncertainty quantification

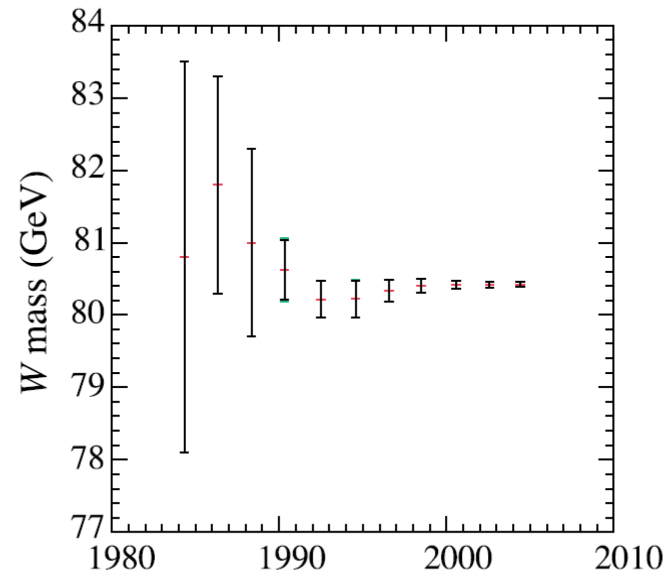
We need to know uncertainty in data:

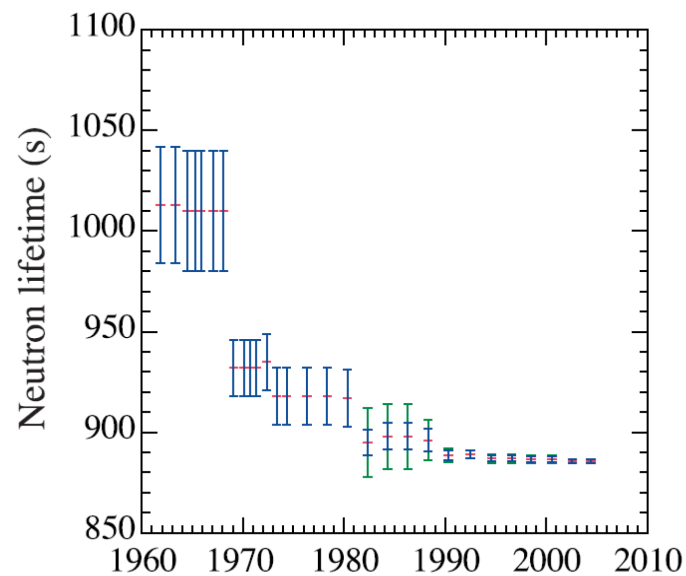
- To determine agreement among data, or between data and theory
- Inference about validity of models requires knowing degree of uncertainty
- We typically assume uncertainty described by a Gaussian pdf
  - often a good approximation
  - width of Gaussian characterized by its standard deviation σ
  - σ provides the metric for uncertainty about data
  - when combining measurements, weight by inverse variance σ^{-2}

- Nomenclature – uncertainty or error?
  - error – state of believing what is incorrect; wrong belief; mistake
  - uncertainty – lack of certainty, sureness; vagueness
  - **uncertainty analysis** seems to convey appropriate meaning

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 wrt last (best?) measurement
- Neutron lifetime
  - disturbing history
  - periodic jumps with periods of extreme agreement
  - most earlier measurements disagree with latest ones
  - plot demonstrates possible sociological and psychological aspects of experimental physics

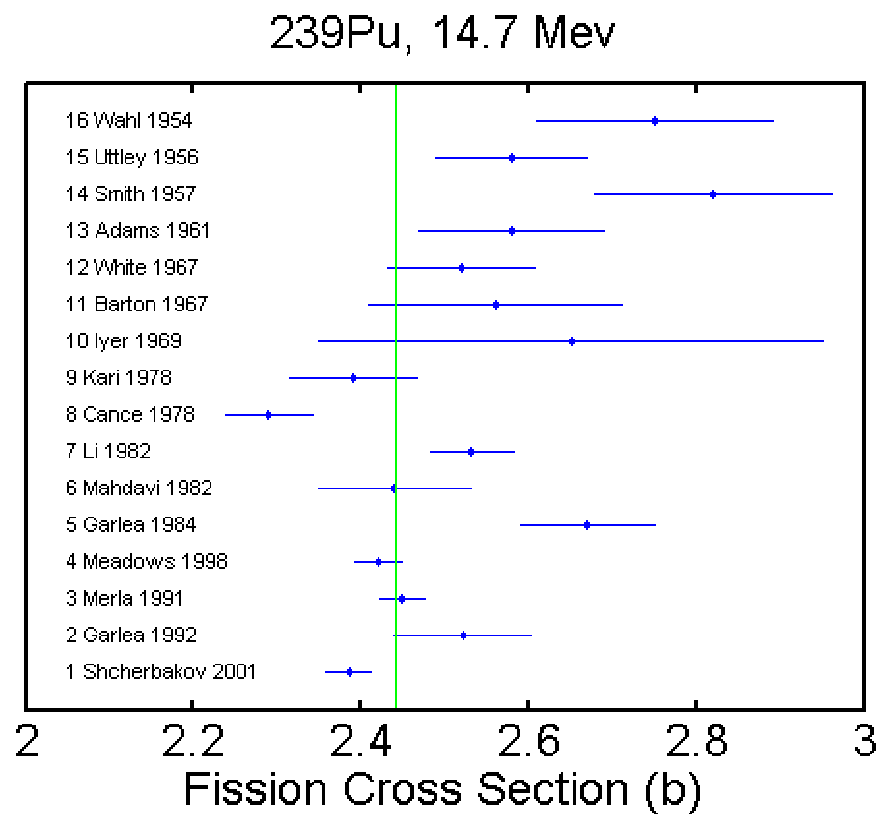




plots from Particle Data Group 2004

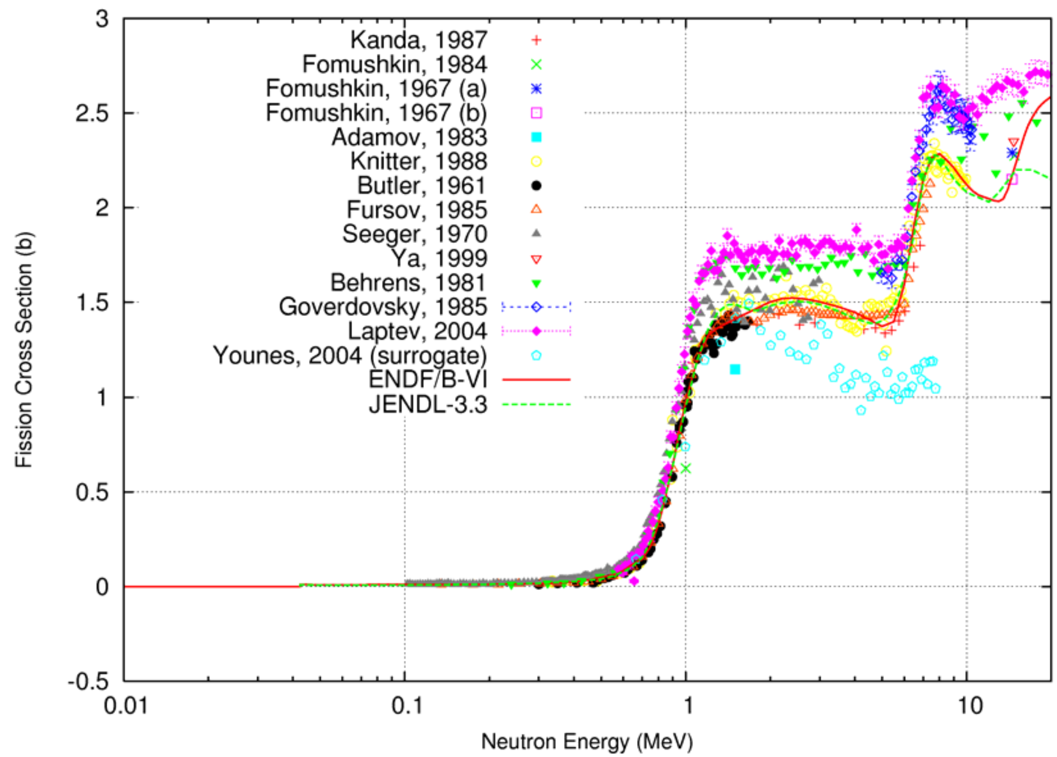
Neutron fission cross section data for ^{239}Pu

- 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



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
  - a thorough analysis must go back to original data and consider all discrepancies

Bayesian analysis of experimental data

- Bayesian approach
  - focus is as much on uncertainties in parameters as on their best (estimated) value
  - provides means for coping with Uncertainty Quantification (UQ)
  - quantitative support of scientific method
  - use of prior knowledge, e.g., previous experiments, modeling expertise, subjective
  - experiments should provide decisive information
  - model-based analysis
  - model checking –
    does model agree with experimental evidence?
- Goal is to estimate **model parameters and their uncertainties**

Bayesian approach to model-based analysis

- Models
  - used to describe and analyze physical world
  - parameters inferred from data
- Bayesian analysis
  - uncertainties in parameters described by probability density functions (pdf)
  - prior knowledge about situation may be incorporated
  - quantitatively and logically consistent methodology for making inferences about models
  - open-ended approach
    - can incorporate new data
    - can extend models and choose between alternatives

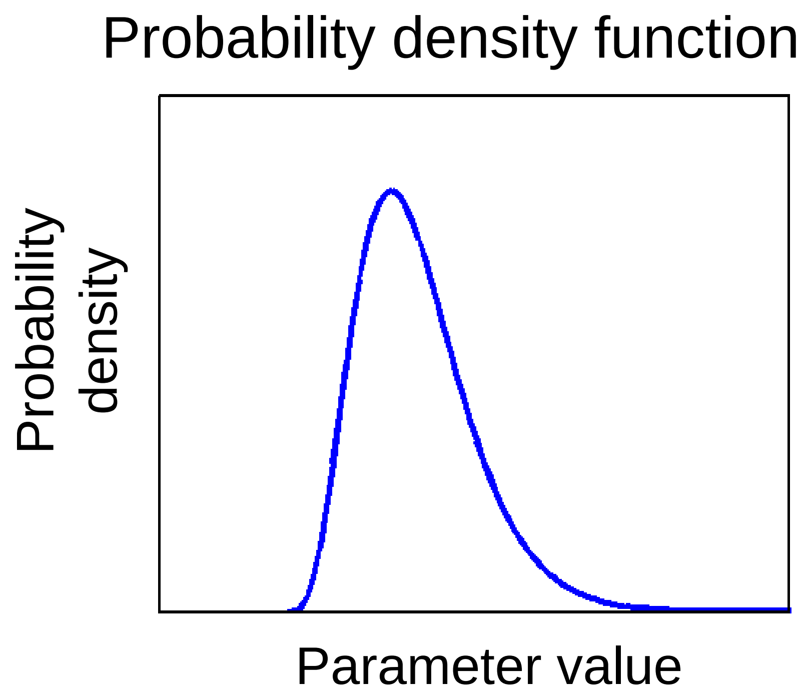
Bayesian approach to model-based analysis

- Bayesian formalism provides framework for modeling
  - choice of model is up to analyst (as in any analysis)
  - many ways to do it
  - calling an analysis Bayesian does not distinguish it

- Because it is a Bayesian analysis does not necessarily mean it is a good analysis; it can also be bad or inappropriate

Uncertainties and probabilities

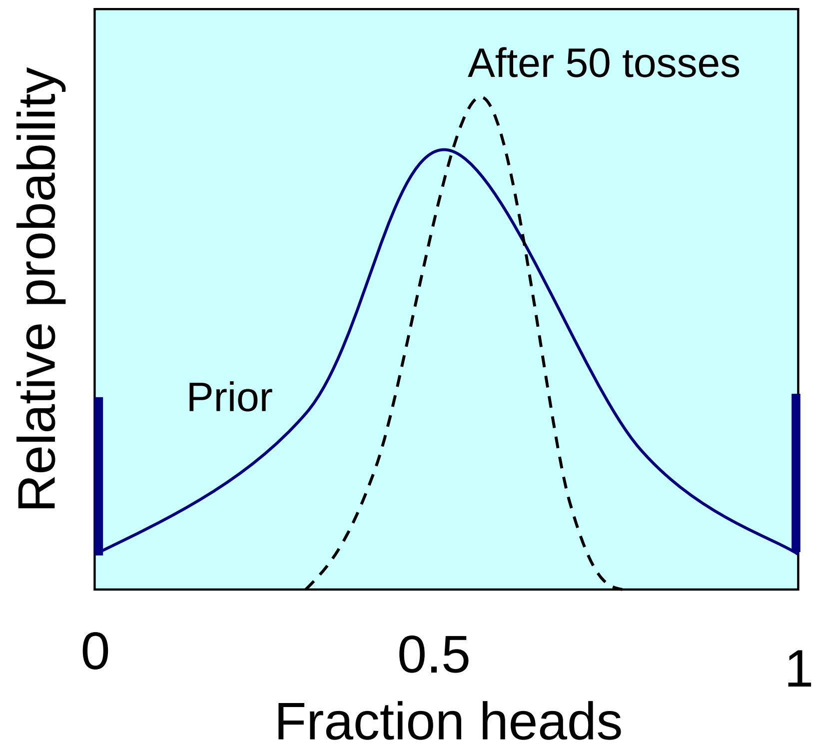
- Uncertainties in parameters are characterized by probability density functions (pdf)
- Probability interpreted as quantitative measure of "**degree of belief**"
- This interpretation is referred to as "subjective probability"
  - different for different people with different knowledge
  - changes with time
  - in science, we seek consensus, avoid bias
- Rules of classical probability theory apply
  - provides firm foundation with mathematical rigor and consistency



Subjective probability can be quantitative

Example – coin toss

- Hypothesis: for a specific coin, fraction of tosses that come up heads = 50%
- Hypothesis seems so reasonable that you might believe it is true
- On basis of this subjective probability, you might be willing to bet with 1:1 odds
- Before any tosses, you might have a prior as shown
- After 50 tosses, you would know better whether coin is fair



Coherent bet quantifies subjective probability

- A property of the Gaussian distribution is that random draws from it will fall inside the interval from -σ to +σ 68% of time
- Suppose, on basis of what you know, you specify the standard error σ of your measurement of a quantity, assuming Gaussian
- If you truly believe in the value of σ you have assigned, you should be willing to accept a bet, randomly chosen between two options:
  - 2:1 bet that a much more accurate measurement would differ from your measured value by **less** than one σ
  - OR 1:2 bet that a much more accurate measurement would differ from your measured value by **more** than one σ
- Your willingness to take bet either way makes this a **coherent bet**
- As physicists, we should make honest effort to assign uncertainties in this spirit, and communicate what we have done

Rules of probability

- Continuous variable x; $p(x)$ is a probability density function (**pdf**)
- **Normalization**: $\int p(x)dx = 1$
- Decomposition of **joint distribution** into conditional distribution:

$$p(x, y) = p(x \mid y)\, p(y)$$

where $p(x \mid y)$ is **conditional** pdf (probability of x given y)
  - if $p(x \mid y) = p(x)$, x is independent of y
- **Bayes law** follows:

$$p(y \mid x) = \frac{p(x \mid y)\, p(y)}{p(x)}$$

- **Marginalization**:

$$p(x) = \int p(x, y)\, dy = \int p(x \mid y)\, p(y)\, dy$$

is probability of x, without regard for y (nuisance parameter)

Rules of probability

- Change of variables: if **x** transformed into **z**, $\mathbf{z} = f(\mathbf{x})$, the pdf in terms of **z** is

$$p(\mathbf{z}) = \left|\mathbf{J}\right|^{-1} p(\mathbf{x})$$

where **J** is the Jacobian matrix for the transformation:

$$\mathbf{J} = \begin{pmatrix} \frac{\partial z_1}{\partial x_1} & \cdots & \frac{\partial z_3}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial z_1}{\partial x_3} & \cdots & \frac{\partial z_3}{\partial x_3} \end{pmatrix}$$

Bayesian analysis of experimental data

- Bayes rule

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

  - where
    d is the vector of measured data values
    a is the vector of parameters for model that predicts the data
  - $p(\boldsymbol{d}\,|\,\boldsymbol{a},I)$ is called the **likelihood** (of the data given the true model and its parameters)
  - $p(\boldsymbol{a}\,|\,I)$ is called the **prior** (on the parameters ***a***)
  - $p(\boldsymbol{a}\,|\,\boldsymbol{d},I)$ is called the **posterior** – fully describes final uncertainty in the parameters
  - *I* stands for whatever **background information** we have about the situation, results from previous experience, our expertise, and the model used
  - denominator provides normalization: $\boldsymbol{p}(\boldsymbol{d})=\int \boldsymbol{p}(\boldsymbol{d}\,|\,\boldsymbol{a})\,\boldsymbol{p}(\boldsymbol{a})\,d\boldsymbol{a}$
    i.e., is integral of numerator

Auxiliary information – *I*

All relevant information about the situation may be brought to bear:

- Details of experiment
  - laboratory set up, experiment techniques, equipment used
  - potential for experimental technique to lead to mistakes
  - expertise of experimenters

  } more subjective
- Relationship between measurements and theoretical model
- History of kind of experiment
- Appropriate statistical models for likelihood and prior
- Experience and expertise

- We usually leave *I* out of our formulas, but keep it in mind

Likelihood

- Form of the likelihood $p(\boldsymbol{d} \mid \boldsymbol{a}, I)$ depends 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; then marginalize it out

Priors

- Noncommittal prior
  - uniform pdf; $p(\theta)$ = const. when θ is offset parameter
  - uniform in log(θ); $p(\log\theta)$ = const. when θ is scale parameter
  - choose pdf with maximum entropy, subject to known constraints
- Physical principles
  - cross sections are nonnegative $\Rightarrow p(\theta) = 0$ when $\theta < 0$
  - invariance arguments, symmetries
- Previous experiments
  - use posterior from previous measurements for prior
  - Bayesian updating
- Expertise
  - elicit pdfs from experts in the field, avoiding common info sources
  - elicitation, an established discipline, may be useful in physical sciences

Priors

- Conjugate priors
  - for many forms of likelihood, there exist companion priors that make it easy to integrate over the variables
  - these priors facilitate analytic solutions for posterior
  - example: for the Poisson likelihood in n and λ, the conjugate prior is a Gamma distribution in λ with parameters α and β, which determine the position and width of the prior
  - conjugate priors can be useful and their parameters can often be chosen to create a prior close to what the analyst has in mind
  - however, in the context of numerical solution of complicated overall models, they loose their appeal

Posterior

- Posterior $p(\boldsymbol{a} \mid \mathbf{d}, I)$
  - net result of a Bayesian analysis
  - summarizes our state of knowledge
  - it provides fully quantitative description of uncertainties
  - usual practice is to characterize posterior in terms of an estimated value of the variables and their variance
- Visualization
  - difficult to visualize directly because it is a density distribution of many variables (dimensions)
  - Monte Carlo allows us to visualize the posterior through it effect on the model that has been used in the analysis

Visualization of uncertainties

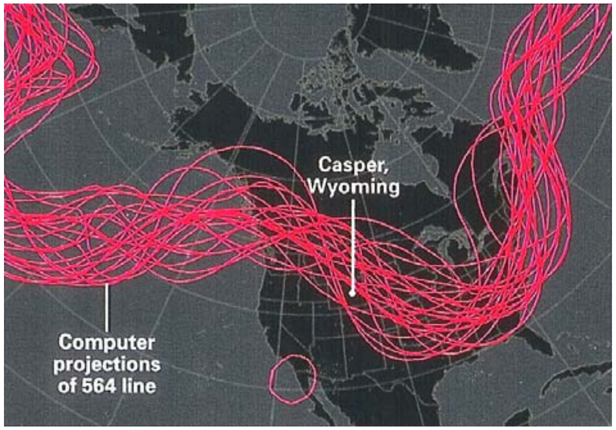
- Visualization plays an important role in developing an understanding of a model and communicating its consequences
- Monte Carlo is often a good choice – choose sets of parameters from their uncertainty distribution and visualize corresponding outputs from the model
- Random sampling from posterior is typically done
- Quasi-random sampling is noteworthy alternative; it provides more uniform sets of samples

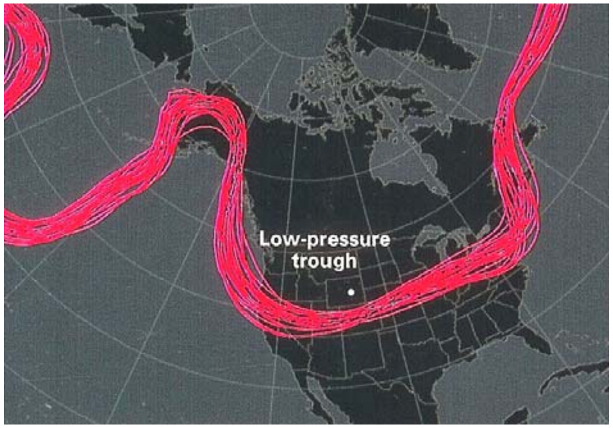
Probability in weather forecasting

- Metrological forecast for Oct. 30, 2003 for Casper, Wyoming
- 22 predictions of 564 line (500 mb) obtained by varying input conditions; indicate plausible outcomes
- Density of lines conveys certainty/probability of winter storms

7 days ahead

564 line; predictive of winter storms





1 day ahead

4 days ahead

what happened?
20-inches of snow!

National Geographic, June 2005

Posterior – quantitative results

- Quantitative results are obtained by characterizing the posterior:
  - mean (first moment): $\hat{x} = \langle x \rangle = \int x\, p(x)\, dx$
    - mean minimizes quadratic cost function
  - maximum (peak position); same as mean if pdf symmetric
  - standard deviation (second moment): $\sigma_x = \sqrt{\int \left(x - \langle x \rangle\right)^2 p(x)\, dx}$
    - **standard error**
  - covariance matrix: $\mathrm{cov}(x, y) = \mathbf{C}_{xy} = \int \left(x - \langle x \rangle\right)\left(y - \langle y \rangle\right) p(x, y)\, dx dy$
    - correlation matrix: $\mathrm{corr}(x, y) = \mathbf{R}_{xy} = \sigma_{xy}^2 / \sigma_x \sigma_y$
  - credible (confidence) interval, e.g., 95% credible interval
- Means for estimating these include:
  - can use calculus if posterior is in convenient analytic form
  - second-order approximation around peak (numerical)
  - Monte Carlo (numerical)

Higher-order inference

- One can make inferences about models, not just parameters
- The posterior for a model is

$$p(M \mid \boldsymbol{d}) = \int p(\boldsymbol{a}, M \mid \boldsymbol{d})\, d\boldsymbol{a} = \int p(\boldsymbol{a}, M \mid \boldsymbol{d})\, d\boldsymbol{a}$$

$$\propto \int p(\boldsymbol{d} \mid \boldsymbol{a}, M)\, p(\boldsymbol{a}, M)\, d\boldsymbol{a}$$

$$= p(M) \int p(\boldsymbol{d} \mid \boldsymbol{a}, M)\, p(\boldsymbol{a} \mid M)\, d\boldsymbol{a}$$

  - the final integral is the normalizing denominator in original Bayes law for $p(\boldsymbol{a}|\boldsymbol{d})$; it is called the **evidence**
  - while the evidence is not needed for parameter inference, it is required for model inference
- May be used for **model selection**, e.g., deciding between two or more models
  - e.g., how many terms to include in a functional analysis

Summary

In this tutorial:

- Need for uncertainty quantification
- Bayesian fundamentals
  - subjective probability, nevertheless quantifiable
  - Bayesian use of probability theory
  - posterior sampling
  - visualization of uncertainties – Monte Carlo
  - higher-order inference