# Uncertainty Quantification and Calibration

Laura Swiler

9211, Optimization and Uncertainty Estimation, Manager Scott Mitchell

Department Review, May 04, 2004



Sandia National Laboratories

# Highlights this performance year

- Formulation of Bayesian analysis for CS&E problems
- Developing a research area in Calibration under Uncertainty
- Added UQ capabilities to DAKOTA
- Significant collaboration with 15000, including SEM sensitivity analysis, Agent-based logistics LDRD – working with David Schoenwald, Jean-Paul Watson, and Bill Hart
- PI for both the project on Prognostics for the F-16 Accessory Drive Gearbox, and an LDRD on Prognostics

Sandia National Laboratories

# UQ methods within DAKOTA

- **Global sensitivity indices**
  - Within Latin Hypercube sampling, added the capability to perform simple, partial, and rank correlations
    - Updated some very old Fortran code and used Epetra types to perform Cholesky decomposition, for example
    - These are important metrics for “first cut” analysis of variables that most influence a simulation
  - Variance-based decomposition: decompose total output variance as sum of input variances
    - Requires replicated samples for each input variable
    - Will implement as part of replicated LHS and/or DDACE

# Bayesian Analysis

- **Construct a prior distribution on a parameter (which might be a parameter of a distribution)**
- **The prior distribution should be based on previous experience, engineering judgment**
- **The distribution on the prior is updated with actual data. The resulting updated distribution is called the posterior.**

| Frequentist | Bayesian |
|---|---|
| Assumes there is an unknown but fixed parameter $\theta$ | Assumes a distribution on unknown parameter $\theta$ |
| Estimates $\theta$ with some confidence interval | Uses probability theory, treats $\theta$ as a random variable |

# Bayesian Analysis

- **Why would we use it for Engineering Science problems?**
- **Nice feature of incorporating additional data as it becomes available**
- **We often don't have good estimates: Bayes provides a framework for starting with what we do know, and refining our estimates in a statistically consistent manner**
- **Example: Penetrator reliability (PRIDE LDRD)**
  - Update probability of failure, update parameters in a surrogate model for a trust region
- **Example: Calibration under Uncertainty (CUU): Update our parameter estimates based on experimental data AND uncertainty in a model**

# Bayesian Methods

## Discrete Case

$$p(\theta \mid \mathbf{x}) = \frac{p(\mathbf{x}, \theta)}{p(\mathbf{x})} = \frac{p(\mathbf{x} \mid \theta)\, p(\theta)}{p(\mathbf{x})} = \frac{p(\mathbf{x} \mid \theta)\, p(\theta)}{\sum_{\theta} p(\mathbf{x} \mid \theta)\, p(\theta)}$$

where θ is a parameter(s), x is a data vector, and p is a probability mass function.

$$p(\theta \mid \mathbf{x}) = posterior \propto p(\mathbf{x} \mid \theta)\, p(\theta) = likelihood * prior$$

# Examples

- Use Binomial distribution to model the number of failures, x, in n trials.

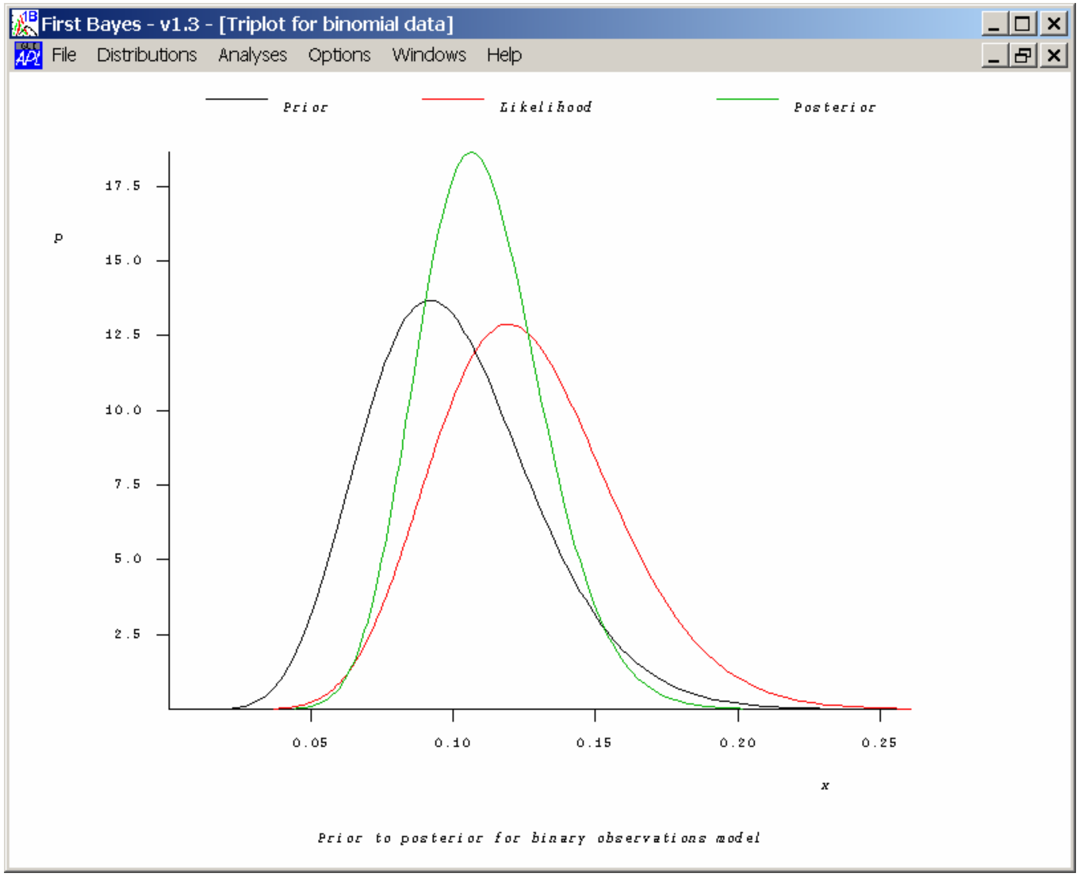
$$f(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

- We obtain data that shows 2 failures in 5 trials

| Prior Probability | Posterior Probability |
|---|---|
| P{θ=0.3}=0.1 | P{θ=0.3}=0.13 |
| P{θ=0.6}=0.9 | P{θ=0.6}=0.87 |

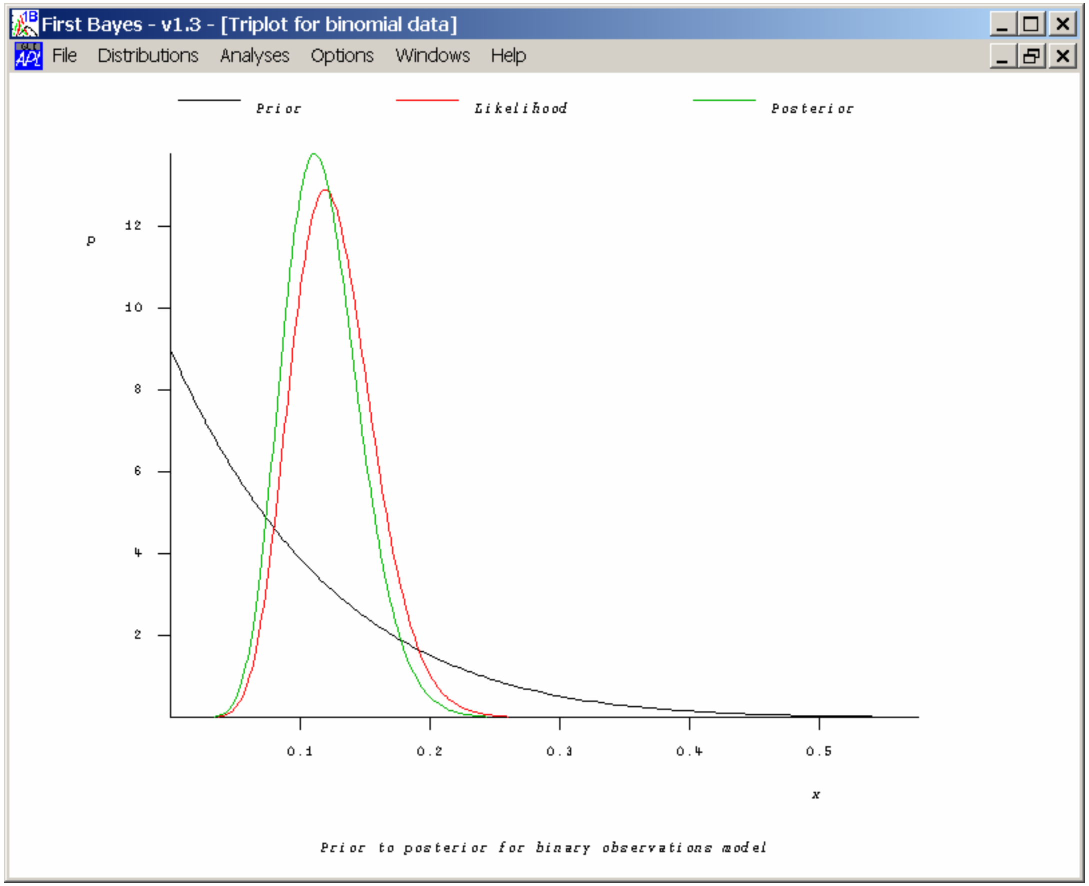
- The posterior distribution reflects the fact that in this set of data, θ = 0.4 which is closer to 0.3 than 0.6 and so the probability of θ=0.3 has risen slightly.

# FirstBayes Software




The dataset is a string of ones and zeros, representing the failure or success of the Rosenbrock function, where failure is defined as a function value > 1000 over the input range $-2 \leq x_1, x_2 \leq 2$. Approximately 10% of the points "fail" according to this threshold.

# FirstBayes Software




If instead we take a prior that is "non-informative" (but still has a mean of 0.1), the prior has a much larger variance and so doesn't influence the posterior as much. Notice that the posterior is much closer to the likelihood function.

# How are posterior distributions calculated?

- In the case of conjugate pairs, one can analytically calculate the posterior distribution
- Most cases are too difficult to calculate analytically, thus we need to go to a sampling method
- Most popular approach is called Markov Chain Monte Carlo (MCMC)
- In MCMC, the idea is to generate a sampling density that is approximately equal to the posterior.  We want the sampling density to be the stationary distribution of a Markov chain.

# Markov Chain Monte Carlo

- How do we generate the Markov chain with the stationary probability that we want?
- Construct a transition probability that will get you there
- Metropolis-Hastings and Gibbs sampling are the most commonly used algorithms
- Both have the idea of a "proposal density" which is used for generating $X_{i+1}$ in the sequence, conditional on $X_i$. The proposal density is often denoted as $Q_Y(Y|X_i)$

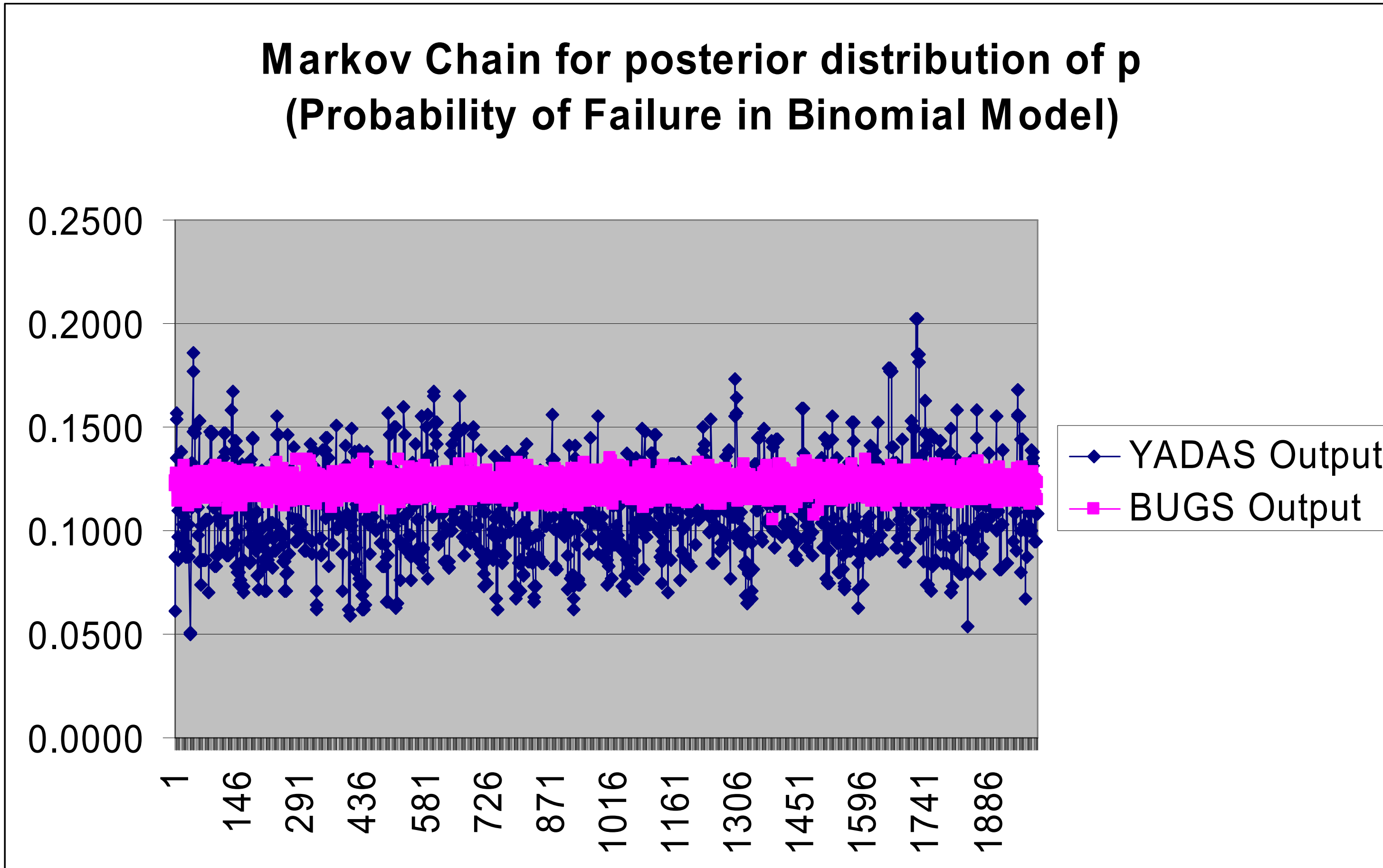
# Metropolis-Hasting

- Basic method: generate a proposed sample from Q, calculate acceptance rate, calculate random number to see if candidate is accepted

$$\alpha(X,Y) = \min(1, \frac{f_X(Y) q_Y(Y \mid X_i)}{f_X(X) q_X(X_i \mid Y)})$$

- **Issues:**
  - Does Q, the proposal density, need a special form?
    - Symmetric *Q*(X|Y)=*Q*(Y|X).
    - Independent *Q*(Y|X)=*Q*(Y|)
  - How long do you run the chain, how do you know when it is converged, how long is the burn-in period, etc.?
  - ACCEPTANCE RATE is CRITICAL. Need to tune Q to get an “optimal” acceptance rate, 45-50% for 1-D problems, 23-26% for high dimensional problems

# BUGS and YADAS: Posterior distribution



# Observations about MCMC

- It works best if your prior is well-defined and close to the posterior.
- It is very difficult to tell if the chain has converged to the “true” underlying posterior
- It requires substantial statistical knowledge to formulate the posterior “proposal” distribution correctly
- Each problem requires tuning of the parameters that govern the Markov chain generation – step sizes, “leaping” parameters, etc.

# Some concerns about Bayes

- The Bayesian framework allows one to integrate observed data and prior knowledge:  conceptually very nice.
- It won't work well in cases where there is very little data or lots of data: optimum is where we have some data that is likely to be added to over time.
- In the context of many of the science and engineering problems encountered at Sandia, we need to seriously question the usefulness of the Bayesian approach.

# PRIDE LDRD

## Three possibilities for Bayesian application

- **Estimation of probability of failure**
- **Estimation of hyperparameters that govern a surrogate model in a trust region or over the entire surface**
  - **Experience with a linear regression model:**

$$E[y_i \mid \boldsymbol{\beta}, \mathbf{X}] = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}$$

  - **Bayesian estimates for mean of $\beta$ and for $\sigma^2$ are the same as those obtained by classical regression or by Maximum likelihood estimates**
  - **What does the Bayesian framework buy us? Are we really going to sample from values of the posterior of $\beta$ to use in a simulation?**
- **Multi-level surrogates**
  - **Can we construct a surrogate based on a few high-resolution function evaluations, then update it with many low-resolution function evaluations or vice-versa? This is a promising area.**

Sandia National Laboratories

# Calibration under Uncertainty

- **Idea: Want to account for both experimental uncertainty AND model uncertainty in the determination of model parameter values**
- Building on the work of Kennedy and O'Hagan.
- Formulate a relationship between observations, "true" process, and model output as: $z_i = \zeta(x_i) + e_i = \rho\, \eta(x_i,t_i) + \delta(x_i) + e_i$

  where z is the observed data, t is the observed value of parameters $\theta$, $e_i$ is the observation error for the $i^{th}$ observation, $\rho$ is an unknown regression parameter, and $\delta(x)$ is a model discrepancy or model inadequacy function that is independent of the code output $\eta(x,t)$.
- $\eta(x_i,t_i)$ and $\delta(x_i)$ are Gaussian process models. They are distributed with a mean and variance which are functions: e.g., $\eta(x_i,t_i) \sim N\,(h(x)^T\beta_, \; c(x, x'))$ where the covariance is often given as:

$$\mathbf{c(x - x')} = \sigma^2 \exp\{\sum_{j=1}^{q} \omega_j (x_j - x'_j)^2\}$$

# Calibration Under Uncertainty

- Calibration involves calculating a very complicated joint pdf on all of these parameters: $\rho$, $\sigma$, $\omega$, h terms, $\beta$, $\lambda$, and $\theta$.
- Approach is to fix some of these terms, and estimate others.  Even KOH admit at this point, this is not readily tractable.
- The "updating" does not have to be Bayesian – one could use Maximum likelihood as Dennis Cox at Rice does.  This removes problems with generation of the posterior distribution.
- Whole approach is HIGHLY parameterized.

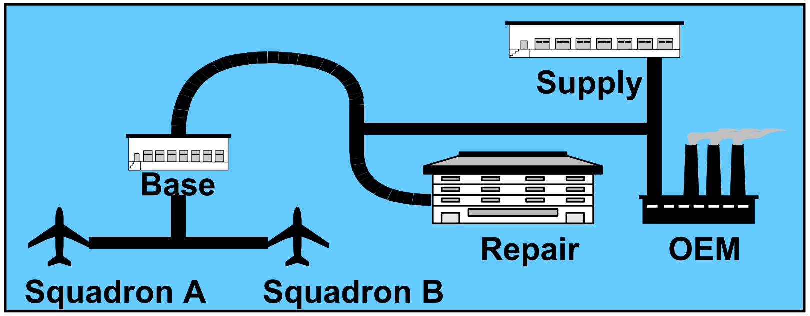
# Current status on CUU

- Serious investigation of Gaussian process models
- Tried a couple of software packages, developing some code in MATLAB to generate predictions
- Talked with Tony Giunta and Tom Paez about the problem of ill-conditioning, how much can Singular Value Decomposition solve?
- Formulating an example with the Rosenbrock function: difficulty separating the GP model error term from the observation error:
- $z_i = \zeta(x_i) + e_i = \rho\, \eta(x_i,t_i) + \delta(x_i) + e_i = GP1+GP2+ e_i$

- 3-4 papers on this work coming out over the next 6 months

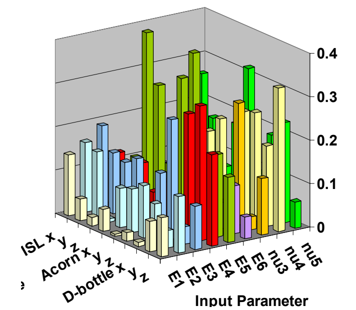
# Sensitivity Analysis for SEM Model

- Approach: Latin Hypercube Sampling (2-loop sampling), examination of correlation coefficients
- Findings thus far: for very large model with thousands of discrete variables, correlation coefficients are not very meaningful. You need to have the model operating near a "optimal" regime (can't be flush with resources or parts, for example) to determine the impact of adding more personnel or inventory.
- At this point, we are tuning the model and experimenting. We would like to have a hierarchical formulation of VBD: vary all MTBF values by 20% vs. all manpower levels by 20% to see which has greater impact.
- Optimization BEFORE sensitivity analysis
- Applicability of results/understanding to David Schoenwald's Agent-Based Logistics LDRD

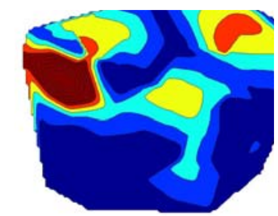


# Sensitivity of Material Parameters on Component Response Environments

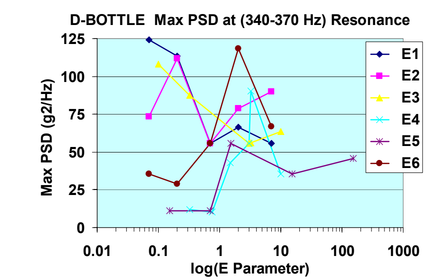
- Life Extension Program




- Partial Correlations: Inputs to Outputs
  - **E3, E4 and E5 are mildly correlated with the outputs.**
  - **E1, E2, and E6 are not significantly correlated with the outputs.**
  - **Parameter correlation varies with location and orientation.**



- Response Surfaces / Contours
  - **Because of the non-linear response, this study did not reveal (flat) regions where changes to the input parameter values have little effect on the calculations. No plateaus.**




- Partial Correlations: Output to Output (low frequency response)
  - **High correlation of the output variables suggests rigid-body mode shape in the structural behavior.**

Sandia National Laboratories

# Service/Publications

- Mentored John Eddy, excellent student from SUNY Buffalo.  He developed a Multi-objective GA which I incorporated into DAKOTA, adding Pareto optimization, warm start, and mixed GAs.
- Developed a modernized standalone version of LHS for UNIX systems, in process of getting that version of LHS copyrighted for inclusion in DAKOTA
- Taking on leadership roles in PRIDE LDRD, CUU work, and CSRF OUU project
- Reviewed 6 Papers
- Interviewed 5 candidates
- Active in Society of Women Engineers
- INFORMS:  Two talks in Oct. 2003 annual conference
- SAMO, Stanford Monte Carlo course

# Conclusions/Future Work

- More UQ methods in DAKOTA, specifically sensitivity measures (variance based decomposition) and sampling methods (quasi Monte-Carlo methods, bootstrapping, importance sampling)
- Continue the Calibration under Uncertainty/ Gaussian Process model/Bayesian work
- Become the resident team expert on Pareto optimization
- Continue work on sensitivity analysis for large discrete models
- LDRD on “Robustness” – combination of evolutionary representations and stochastic dynamic programming