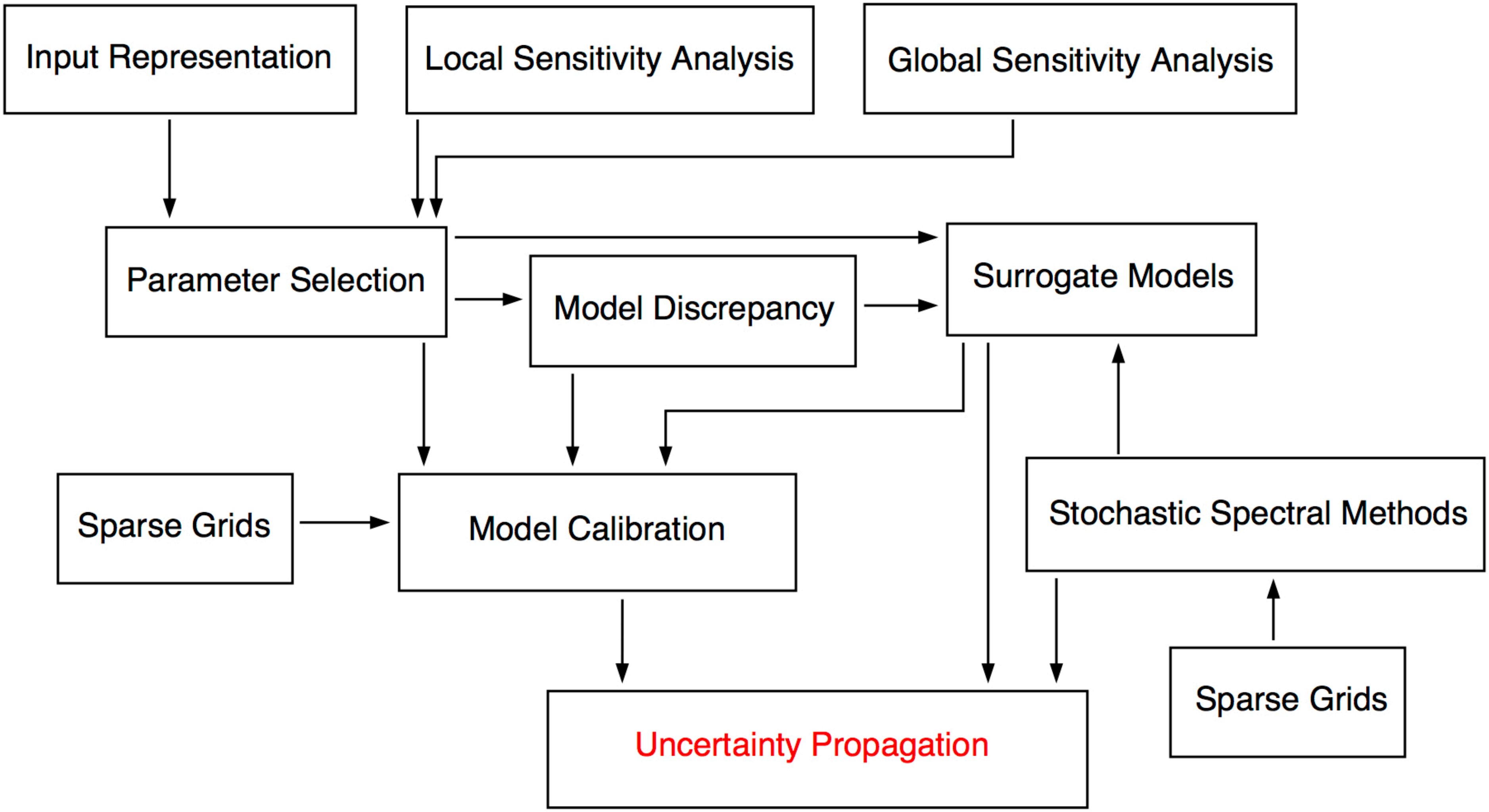
Steps in Uncertainty Quantification

Note: Uncertainty quantification requires synergy between statistics, mathematics and application area.



Uncertainty Propagation

Setting:

- We assume that we have determined distributions for parameters
    - e.g., Bayesian inference, prior experiments, expert opinion

$$\dot{T}_1 = \lambda_1 - d_1 T_1 - (1-\varepsilon) k_1 V T_1$$

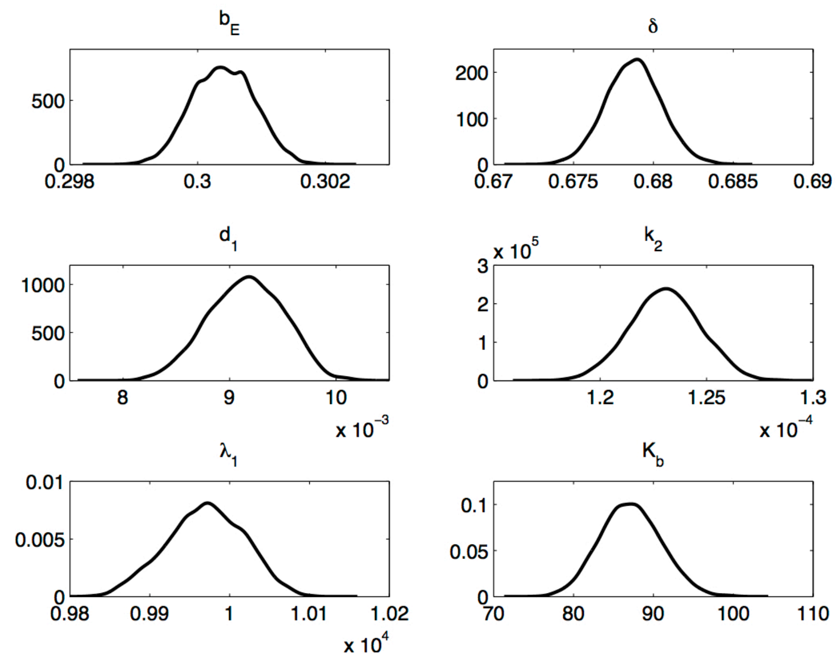
$$\dot{T}_2 = \lambda_2 - d_2 T_2 - (1-f\varepsilon) k_2 V T_2$$

$$\dot{T}_1^* = (1-\varepsilon) k_1 V T_1 - \delta T_1^* - m_1 E T_1^*$$

$$\dot{T}_2^* = (1-f\varepsilon) k_2 V T_2 - \delta T_2^* - m_2 E T_2^*$$

$$\dot{V} = N_T \delta (T_1^* + T_2^*) - cV - [(1-\varepsilon)\rho_1 k_1 T_1 + (1-f\varepsilon)\rho_2 k_2 T_2] V$$

$$\dot{E} = \lambda_E + \frac{b_E (T_1^* + T_2^*)}{T_1^* + T_2^* + K_b} E - \frac{d_E (T_1^* + T_2^*)}{T_1^* + T_2^* + K_d} E - \delta_E E$$



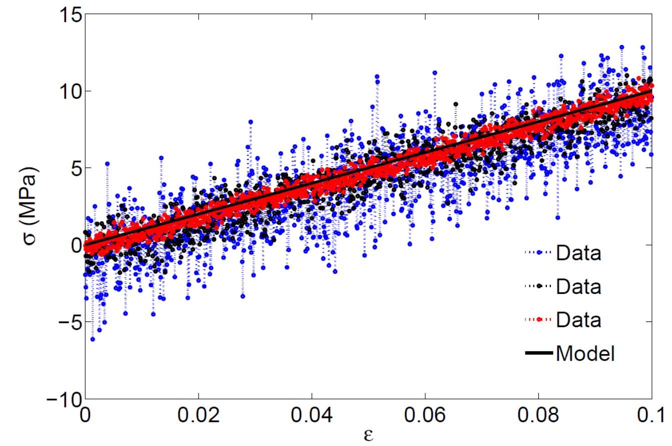
Goal: Construct statistics for quantities of interest

- e.g., Expected viral load in HIV patient with appropriate uncertainty intervals
- Note: Often involves moderate to high-dimensional integration

$$\mathbb{E}[V(t)] = \int_{\mathbb{R}^6} V(t,q)\rho(q)\,dq$$

Forward Uncertainty Propagation: Linear Models

Linear Models: Analytic mean and variance relations



Example: Linear stress-strain relation

$$\Upsilon_i = Ee_i + E_2 e_i^3 + \varepsilon_i \ , \ i = 1, \ldots, n$$

Model Statistics:

Let $\overline{E}, \overline{E}_2$ and $\mathrm{var}(E), \mathrm{var}(E_2)$ denote parameter means and variance. Then

$$\mathbb{E}[Ee_i + E_2 e_i^3] = \overline{E}e_i + \overline{E}_2 e_i^3$$

$$\mathrm{var}[Ee_i + E_2 e_i^3] = e_i^2 \mathrm{var}(E) + e_i^6 \mathrm{var}(E_2) + 2e_i^4 \mathrm{cov}(E, E_2)$$

Response Statistics: Assume measurement errors uncorrelated from model response.

$$\mathbb{E}[\Upsilon_i] = \overline{E}e_i + \overline{E}_2 e_i^3$$

$$\mathrm{var}[\Upsilon_i] = e_i^2 \mathrm{var}(E) + e_i^6 \mathrm{var}(E_2) + 2e_i^4 \mathrm{cov}(E, E_2) + \mathrm{var}(\varepsilon_i)$$

Problem: Models are almost always nonlinearly parameterized

Forward Uncertainty Propagation: Sampling Methods

Strategy 1: Randomly sample from parameter and measurement error distributions and propagate through model to quantify response uncertainty.

Advantages:

- Applicable to nonlinear models.
- Parameters can be correlated and non-Gaussian.
- Straight-forward to apply and convergence rate is independent of number of parameters.
- Can directly incorporate both parameter and measurement uncertainties.

Disadvantages:

- Very slow convergence rate: $\mathcal{O}(1/\sqrt{M})$ where M is the number of samples.
- 100-fold more evaluations required to gain additional place of accuracy.
- This motivates numerical analysis techniques.

Numerical Quadrature

Motivation: Computation of expected values requires approximation of integrals

$$\mathbb{E}[u(t,x)] = \int_{\mathbb{R}^p} u(t,x,q)\rho(q)dq$$

Example: HIV model

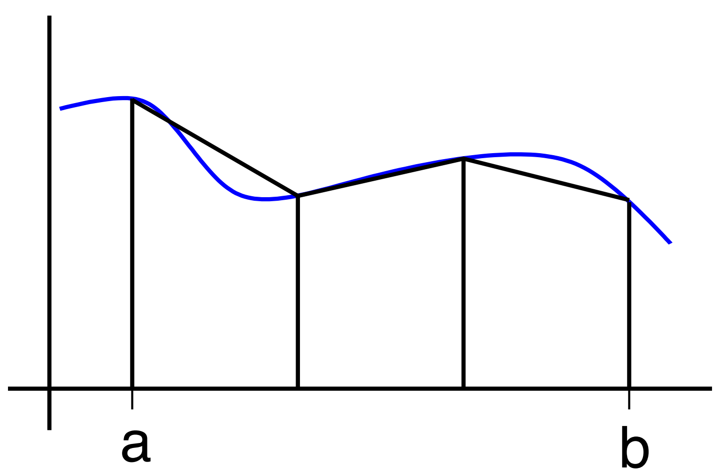
$$\mathbb{E}[V(t)] = \int_{\mathbb{R}^6} V(t,q)\rho(q)dq$$

Numerical Quadrature:

$$\int_{\mathbb{R}^p} f(q)\rho(q)dq \approx \sum_{r=1}^{R} f(q^r)w^r$$

Questions:

- How do we choose the quadrature points and weights?
  - E.g., Newton-Cotes; e.g., trapezoid rule



$$\int_a^b f(q)dq \approx \frac{h}{2}\left[f(a) + f(b) + 2\sum_{r=1}^{R-2} f(q^r)\right]$$

$$q^r = a + hr \ , \ h = \frac{b-a}{R-1}$$

Numerical Quadrature

Motivation: Computation of expected values requires approximation of integrals

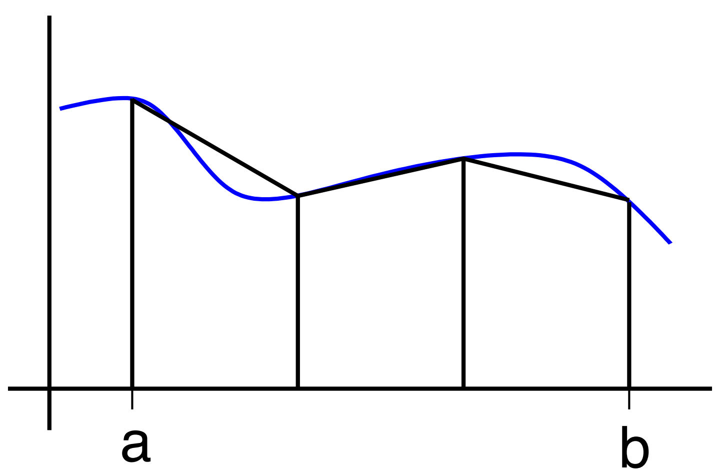
$$\mathbb{E}[u(t,x)] = \int_{\mathbb{R}^p} u(t,x,q)\rho(q)dq$$

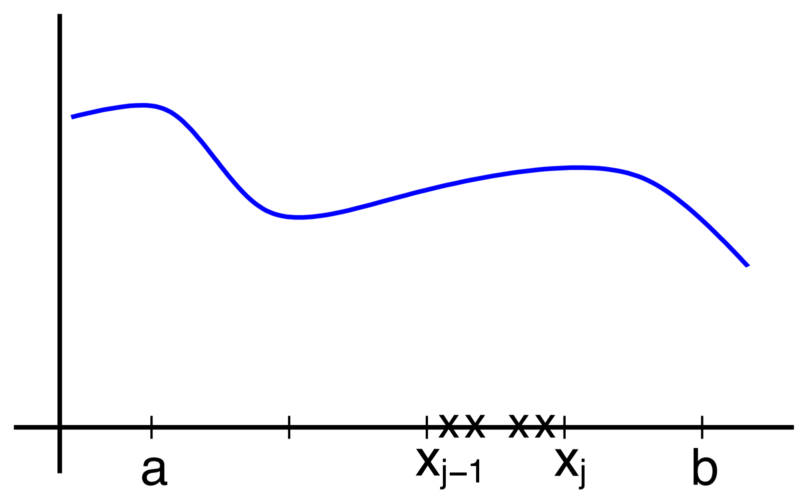
Numerical Quadrature:

$$\int_{\mathbb{R}^p} f(q)\rho(q)dq \approx \sum_{r=1}^{R} f(q^r)w^r$$

Questions:

- How do we choose the quadrature points and weights?
  - E.g., Newton-Cotes, Gaussian algorithms





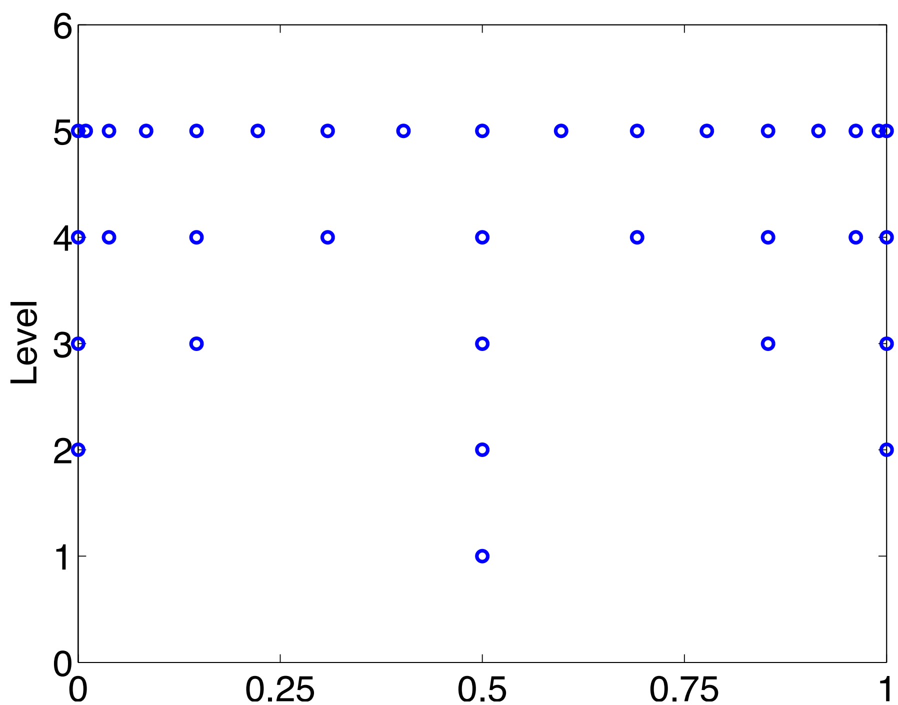
Numerical Quadrature

Numerical Quadrature:

$$\int_{\mathbb{R}^p} f(q)\rho(q)dq \approx \sum_{r=1}^{R} f(q^r)w^r$$

Questions:

- Can we construct nested algorithms to improve efficiency?
    - E.g., employ Clenshaw-Curtis points

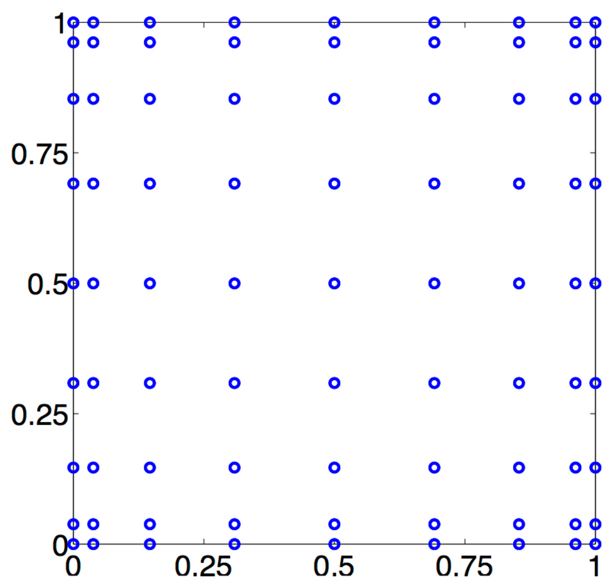


Numerical Quadrature

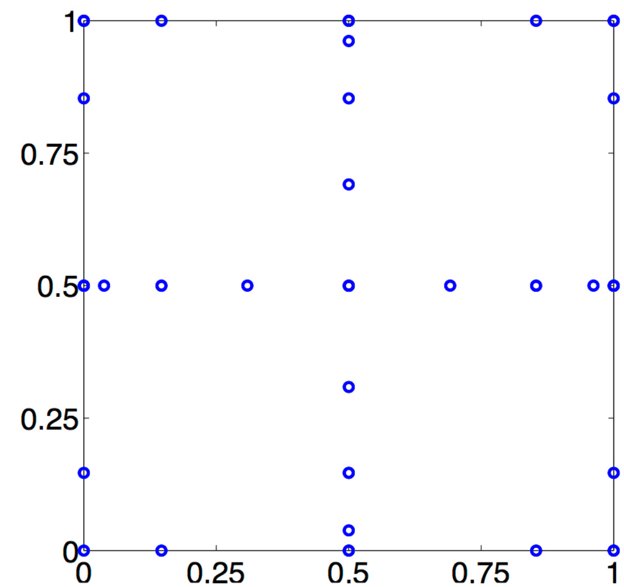
Questions:

- How do we reduce required number of points while maintaining accuracy?

Tensored Grids: Exponential growth



Sparse Grids: Same accuracy



p	R_ℓ	Sparse Grid $\mathcal{R}$	Tensored Grid $R = (R_\ell)^p$
2	9	29	81
5	9	241	59,049
10	9	1581	$> 3 \times 10^9$
50	9	171,901	$> 5 \times 10^{47}$
100	9	1,353,801	$> 2 \times 10^{95}$

Numerical Quadrature

Problem:

- Accuracy of methods diminishes as parameter dimension p increases
- Suppose $f \in C^\alpha([0,1]^p)$
- Tensor products: Take R_ℓ points in each dimension so $R = (R_\ell)^p$ total points
- Quadrature errors:

Newton-Cotes: $E \sim \mathcal{O}(R_\ell^{-\alpha}) = \mathcal{O}(R^{-\underline{\alpha/p}})$

Gaussian: $E \sim \mathcal{O}(e^{-\beta R_\ell}) = \mathcal{O}\left(e^{-\beta \underline{\sqrt[p]{R}}}\right)$

Sparse Grid: $E \sim \mathcal{O}\left(\mathcal{R}^{-\alpha} \log(\mathcal{R})^{\underline{(p-1)(\alpha+1)}}\right)$

Numerical Quadrature

Problem:

- Accuracy of methods diminishes as parameter dimension p increases
- Suppose $f \in C^\alpha([0,1]^p)$
- Tensor products: Take R_ℓ points in each dimension so $R = (R_\ell)^p$ total points
- Quadrature errors:

Newton-Cotes: $E \sim \mathcal{O}(R_\ell^{-\alpha}) = \mathcal{O}(R^{-\underline{\alpha/p}})$

Gaussian: $E \sim \mathcal{O}(e^{-\beta R_\ell}) = \mathcal{O}\left(e^{-\beta \underline{\sqrt[p]{R}}}\right)$

Sparse Grid: $E \sim \mathcal{O}\left(\mathcal{R}^{-\alpha} \log(\mathcal{R})^{\underline{(p-1)(\alpha+1)}}\right)$

- Alternative: Monte Carlo quadrature

$$\int_{\mathbb{R}^p} f(q)\rho(q)dq \approx \frac{1}{R}\sum_{r=1}^{R} f(q^r) \quad , \quad E \sim \left(\frac{1}{\sqrt{R}}\right)$$

- Advantage: Errors independent of dimension p
- Disadvantage: Convergence is very slow!

Numerical Quadrature

Problem:

- Accuracy of methods diminishes as parameter dimension p increases
- Suppose $f \in C^\alpha([0,1]^p)$
- Tensor products: Take R_ℓ points in each dimension so $R = (R_\ell)^p$ total points
- Quadrature errors:

Newton-Cotes: $E \sim \mathcal{O}(R_\ell^{-\alpha}) = \mathcal{O}(R^{-\underline{\alpha/p}})$

Gaussian: $E \sim \mathcal{O}(e^{-\beta R_\ell}) = \mathcal{O}\left(e^{-\beta \underline{\sqrt[p]{R}}}\right)$

Sparse Grid: $E \sim \mathcal{O}\left(\mathcal{R}^{-\alpha} \log(\mathcal{R})^{\underline{(p-1)(\alpha+1)}}\right)$

- Alternative: Monte Carlo quadrature

$$\int_{\mathbb{R}^p} f(q)\rho(q)dq \approx \frac{1}{R}\sum_{r=1}^{R} f(q^r) \quad , \quad E \sim \left(\frac{1}{\sqrt{R}}\right)$$

- Advantage: Errors independent of dimension p
- Disadvantage: Convergence is very slow!

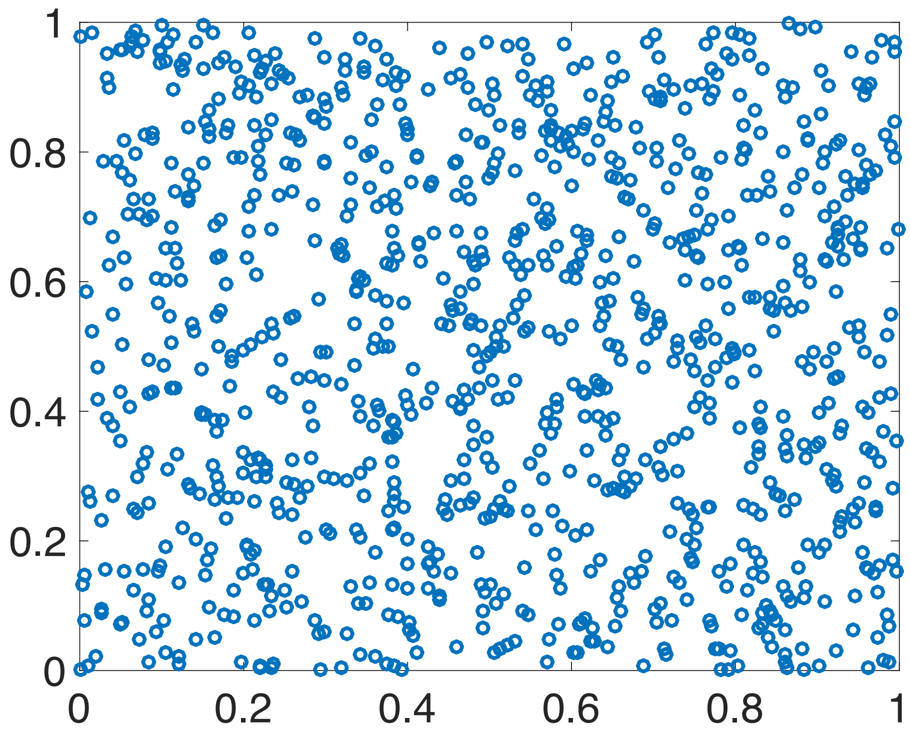
Conclusion: For high enough dimension p, monkeys throwing darts will beat Gaussian and sparse grid techniques!

Monte Carlo Sampling Techniques

Issues:

- Very low accuracy and slow convergence
- Random sampling may not "randomly" cover space …

Samples from Uniform Distribution

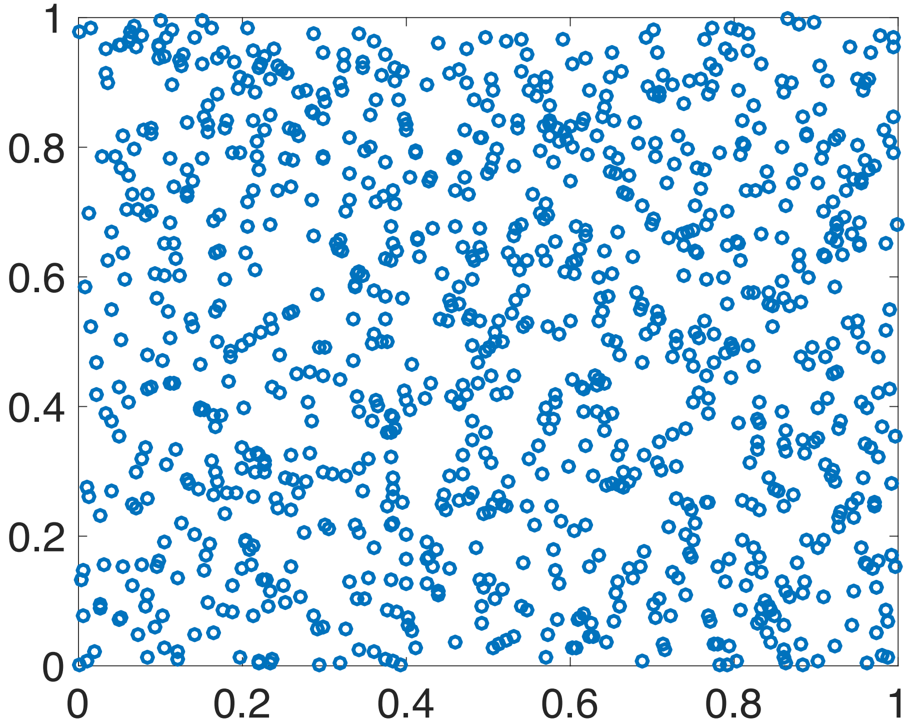


Monte Carlo Sampling Techniques

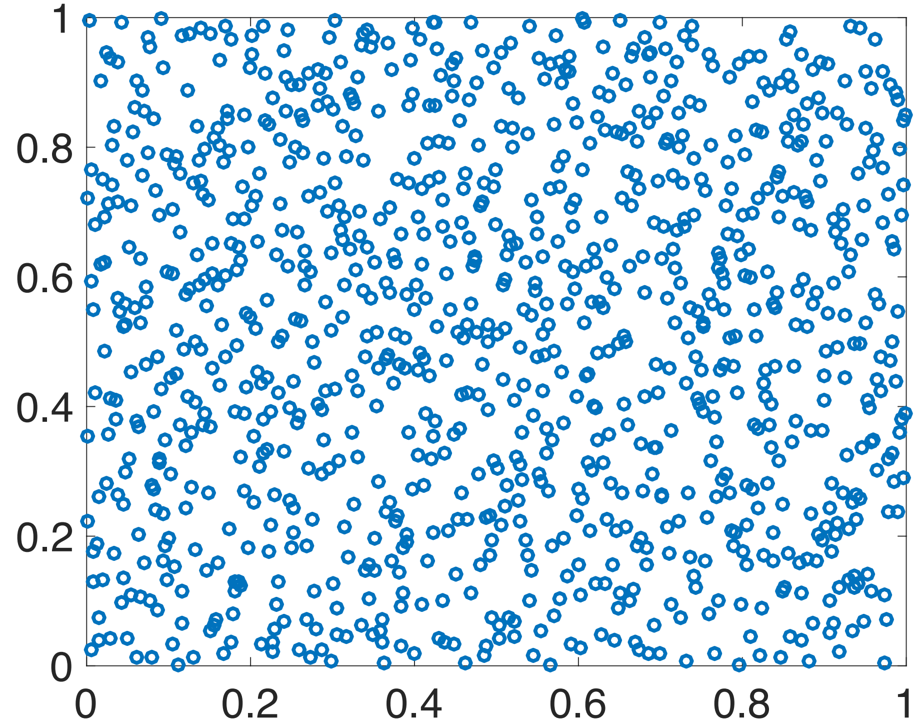
Issues:

- Very low accuracy and slow convergence
- Random sampling may not “randomly” cover space …

Samples from Uniform Distribution



Sobol’ Points



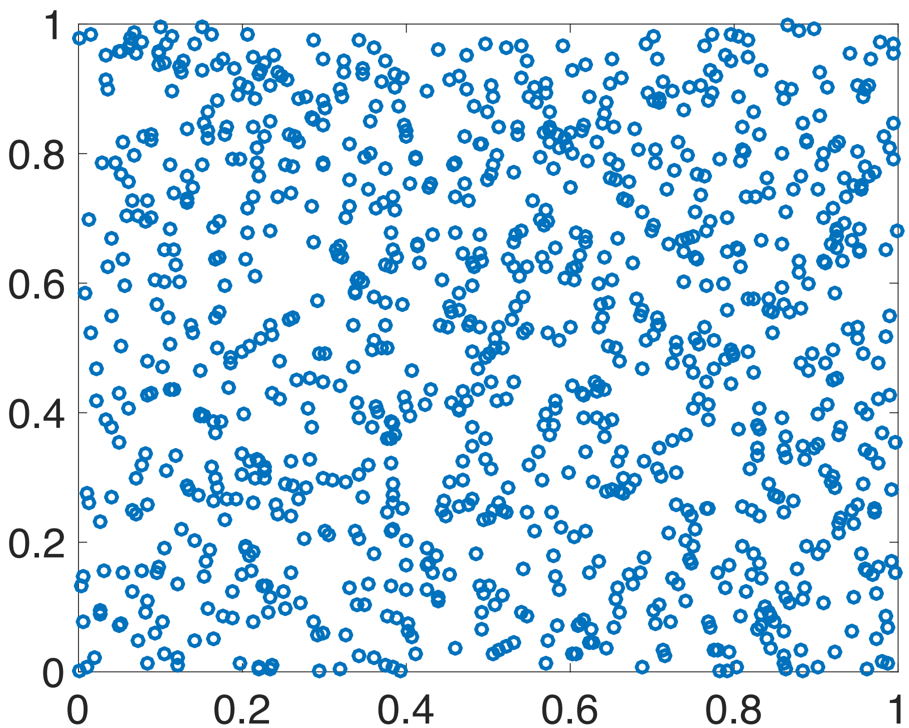
Sobol’ Sequence: Use a base of two to form successively finer uniform partitions of unit interval and reorder coordinates in each dimension.

Quasi-Monte Carlo Sampling Techniques

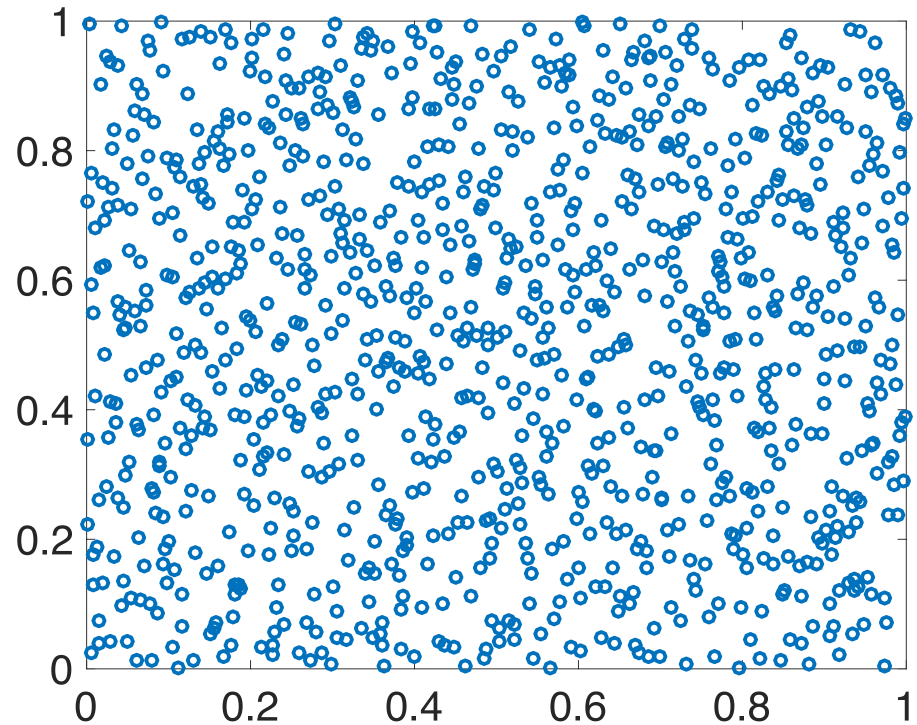
Issues:

- Very low accuracy and slow convergence
- Random sampling may not "randomly" cover space …

Samples from Uniform Distribution



Sobol' Points



$$\int_{\mathbb{R}^p} f(q)\rho(q)dq \approx \frac{1}{R}\sum_{r=1}^{R} f(q^r) \quad , \quad E \sim \left(\frac{1}{\sqrt{R}}\right)$$

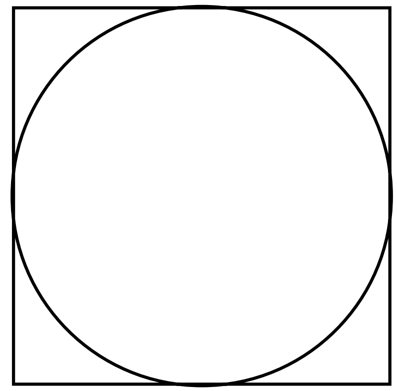
$$E \sim \mathcal{O}\left(\frac{(\log R)^p}{R}\right)$$

Monte Carlo Sampling Techniques

Example: Use Monte Carlo sampling to approximate area of circle

$$\frac{A_c}{A_s} = \frac{\pi r^2}{4r^2} = \frac{\pi}{4}$$

$$\Rightarrow A_c = \frac{\pi}{4} A_s$$



Strategy:

- Randomly sample N points in square $\Rightarrow$ approximately $N\frac{\pi}{4}$ in circle
- Count M points in circle

$$\Rightarrow \pi \approx \frac{4M}{N}$$

Quasi-Monte Carlo:

- SAMSI Program on *Quasi-Monte Carlo and High Dimensional Sampling Methods in Applied Math* in 2017-18

MATLAB Example

Monte Carlo Quadrature:

- Run rand_points.m to observe uniformly sampled and Sobol' points.
- Run pi_approx.m with different values of N to see if you observe convergence rate of $1/\sqrt{N}$

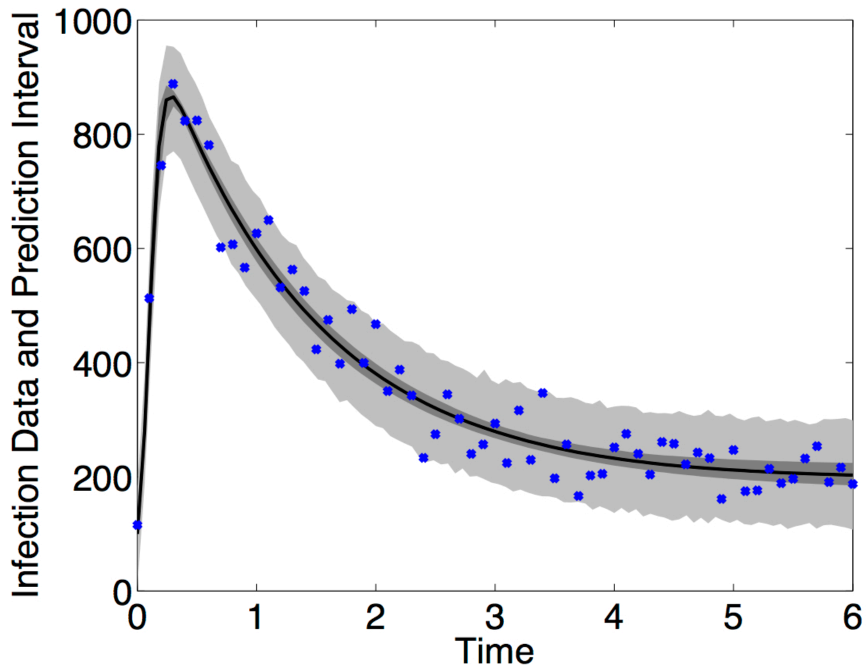
Website:

- http://www4.ncsu.edu/~rsmith/DATAWORKS18/

Confidence, Credible and Prediction Intervals

Note:

- We now know how to compute the mean response for the QoI.
- How do we compute appropriate intervals?



SIR Model:

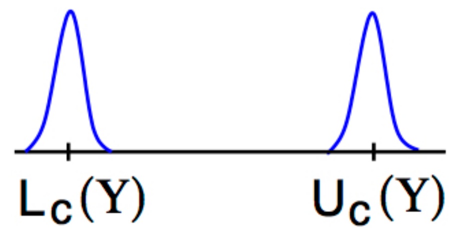
$$\frac{dS}{dt} = \delta N - \delta S - \gamma IS \quad , \; S(0) = S_0$$ Susceptible

$$\frac{dI}{dt} = \gamma IS - (r + \delta)I \quad , \; I(0) = I_0$$ Infectious

$$\frac{dR}{dt} = rI - \delta R \quad , \; R(0) = R_0$$ Recovered

Confidence, Credible and Prediction Intervals

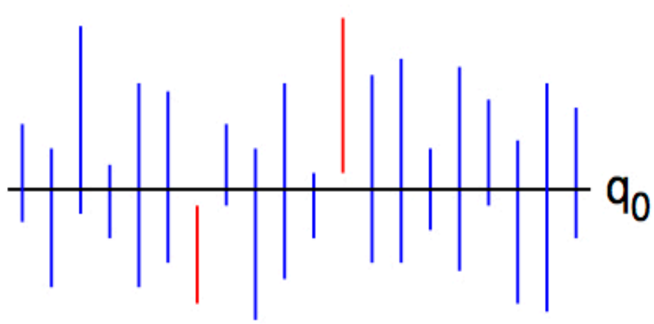
Data: $\Upsilon = [\Upsilon_1, \cdots, \Upsilon_n]$ of iid random observations



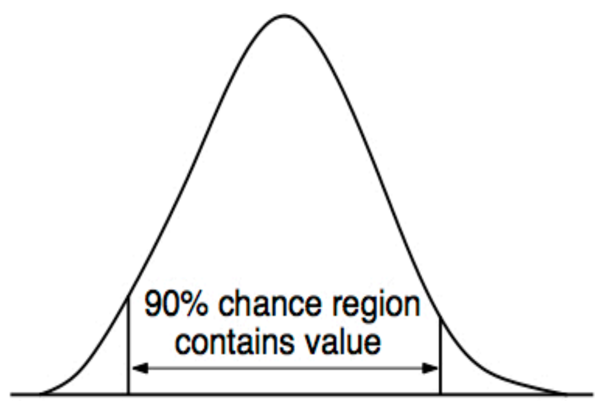
Confidence Interval (Frequentist): A $100 \times (1-\alpha)$% confidence interval for a fixed, unknown parameter q_0 is a random interval $[L_c(\Upsilon), U_c(\Upsilon)]$, having probability at least $1-\alpha$ of covering q_0 under the joint distribution of Υ.

Credible Interval (Bayesian): A $100 \times (1-\alpha)$ % credible interval is that having probability at least $1-\alpha$ of containing q.

Strategy: Sample out of parameter density $\rho_Q(q)$



90% Confidence Intervals



90% Credible Interval

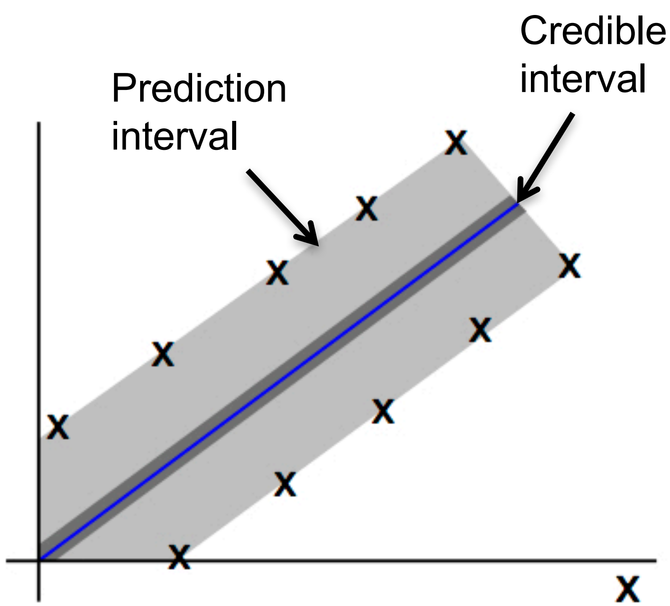
Confidence, Credible and Prediction Intervals

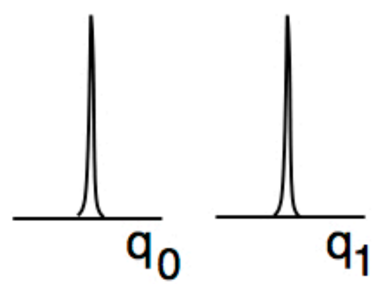
Data: $\Upsilon = [\Upsilon_1, \cdots, \Upsilon_n]$ of iid random observations

Prediction Interval: A $100 \times (1-\alpha)$% prediction interval for a future observable Υ_{n+1} is a random interval $[L_c(\Upsilon), U_c(\Upsilon)]$ having probability at least $1-\alpha$ of of containing Υ_{n+1} under the joint distribution of $(\Upsilon, \Upsilon_{n+1})$.

Example: Consider linear model

$$\Upsilon_i = q_0 + q_1 x_i + \varepsilon_i \ , \ i = 1, \cdots, n$$





Example: HIV Model

Model:

$$\dot{T}_1 = \underline{\lambda_1} - \underline{d_1} T_1 - (1-\varepsilon) k_1 V T_1$$

$$\dot{T}_2 = \lambda_2 - d_2 T_2 - (1 - f\varepsilon) k_2 V T_2$$

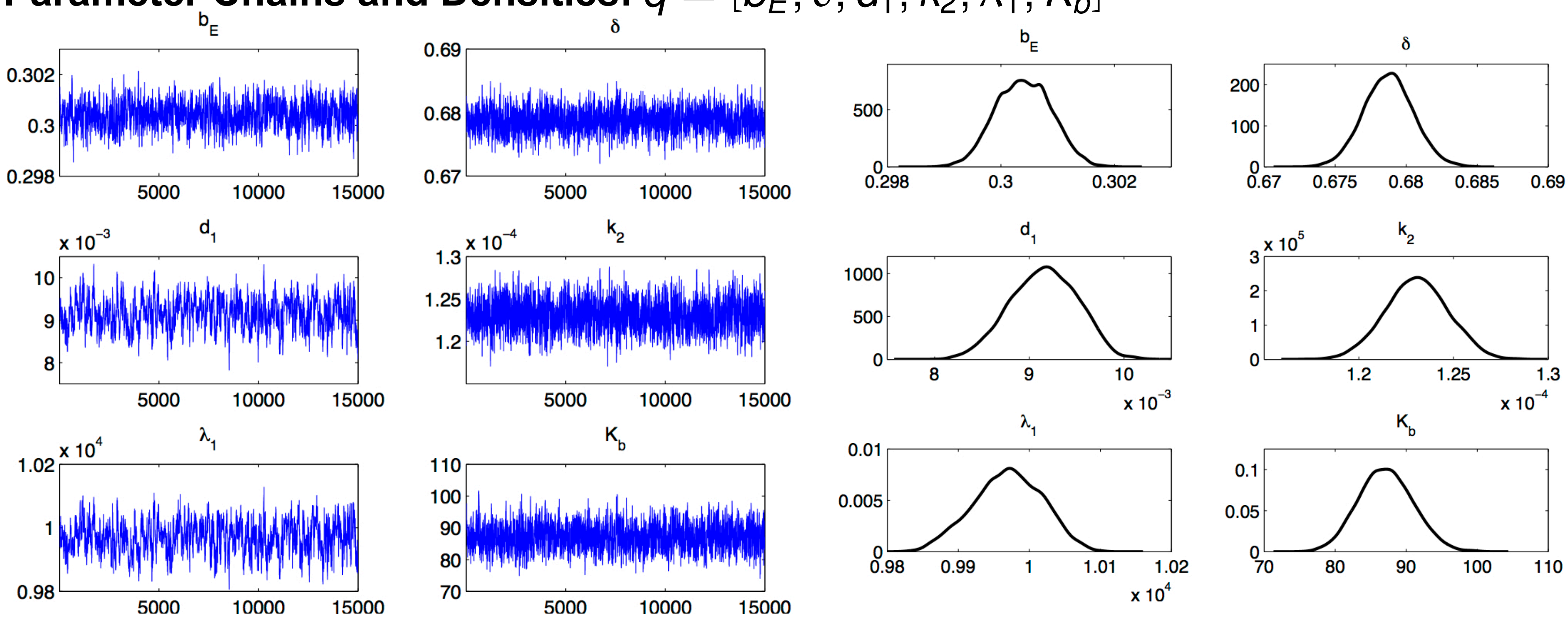
$$\dot{T}_1^* = (1-\varepsilon) k_1 V T_1 - \underline{\delta} T_1^* - m_1 E T_1^*$$

$$\dot{T}_2^* = (1 - f\varepsilon) \underline{k_2} V T_2 - \underline{\delta} T_2^* - m_2 E T_2^*$$

$$\dot{V} = N_T \underline{\delta} (T_1^* + T_2^*) - cV - [(1-\varepsilon)\rho_1 k_1 T_1 + (1 - f\varepsilon)\rho_2 k_2 T_2] V$$

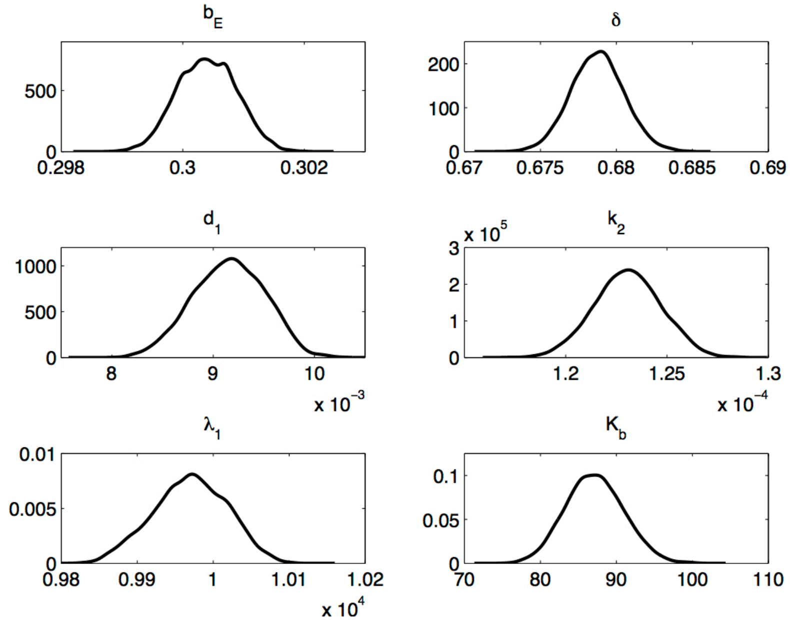
$$\dot{E} = \lambda_E + \frac{\underline{b_E}(T_1^* + T_2^*)}{T_1^* + T_2^* + \underline{K_b}} E - \frac{d_E(T_1^* + T_2^*)}{T_1^* + T_2^* + K_d} E - \delta_E E$$

Parameter Chains and Densities: $q = [b_E, \delta, d_1, k_2, \lambda_1, K_b]$



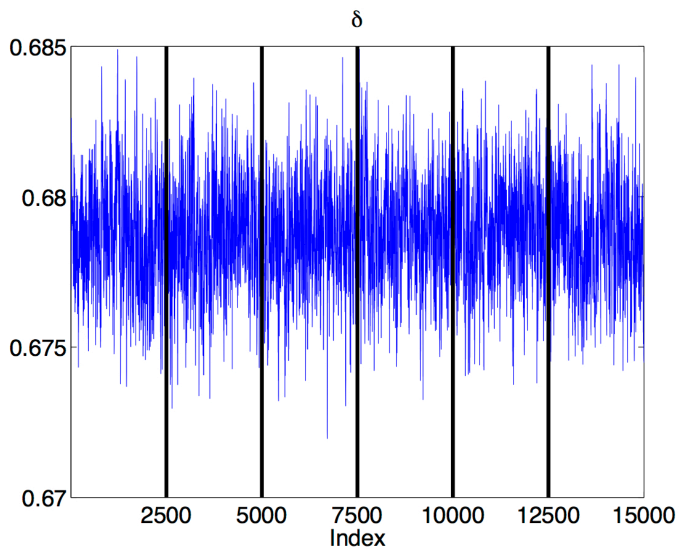
Propagation of Uncertainty – HIV Example

Parameter Densities:

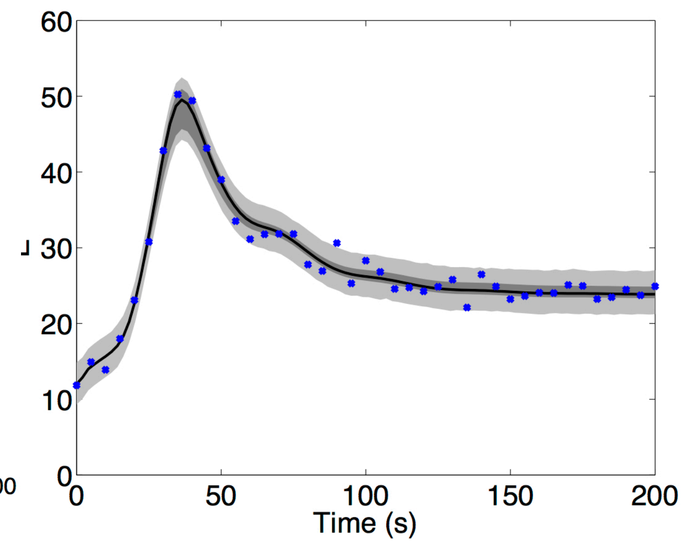


Techniques:

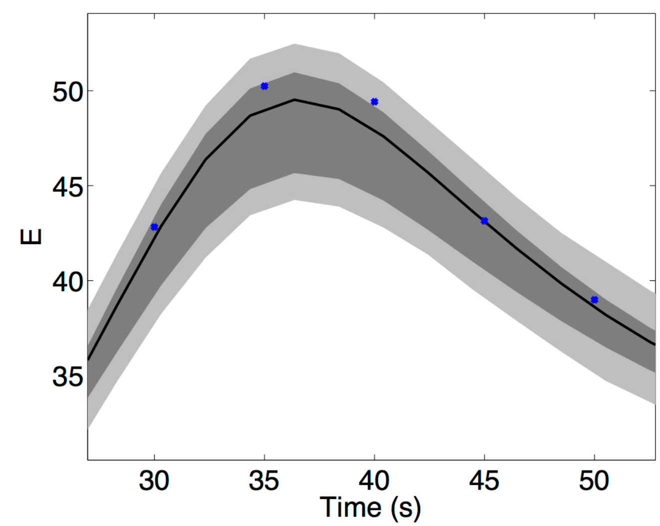
- Sample from parameter and observation error densities to construct mean response, credible intervals, and prediction intervals for QoI.
- Slow convergence rate $\mathcal{O}(1/\sqrt{M})$



Samples from Chain



Data, Credible Intervals and Prediction Intervals



Non-Gaussian Credible and Prediction Intervals

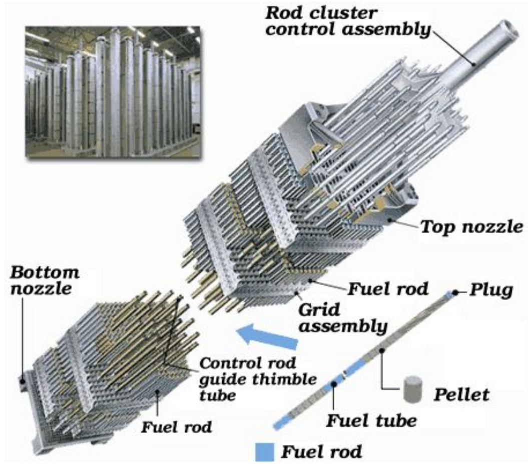
Use of Prediction Intervals: Nuclear Power Plant Design

Subchannel Code (COBRA-TF): numerous closure relations, ~70 parameters

e.g., Dittus—Boelter Relation

$$Nu = {\color{red}0.023} Re^{0.8} Pr^{0.4}$$

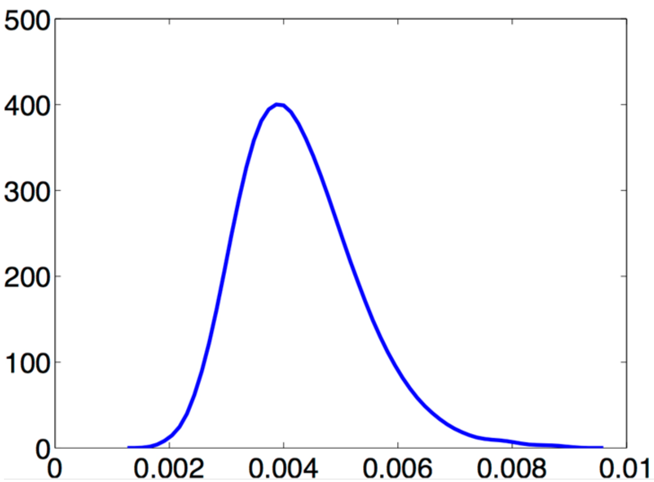
Nu: Nusselt number
Re: Reynolds number
Pr: Prandtl number



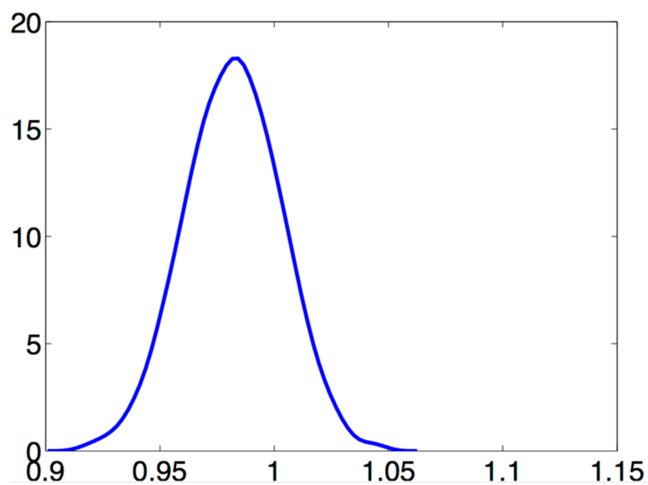
Industry Standard: Employ conservative, uniform, bounds

i.e., [0, 0.046], [0, 1.6], [0,0.8]

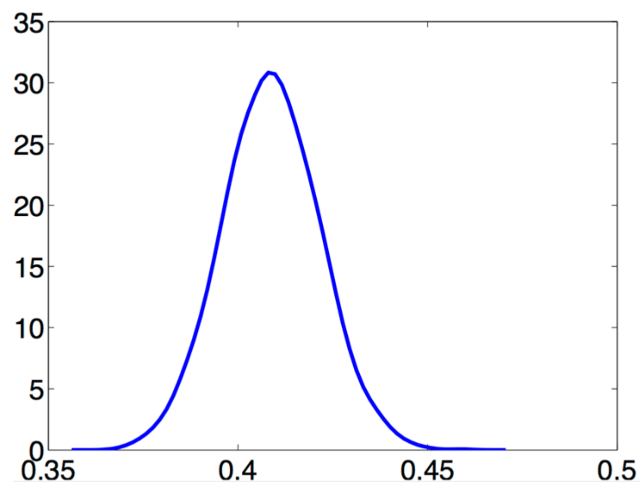
Bayesian Analysis: Employ conservative bounds as priors



$2\sigma \approx 0.0035$



$2\sigma \approx 0.06$

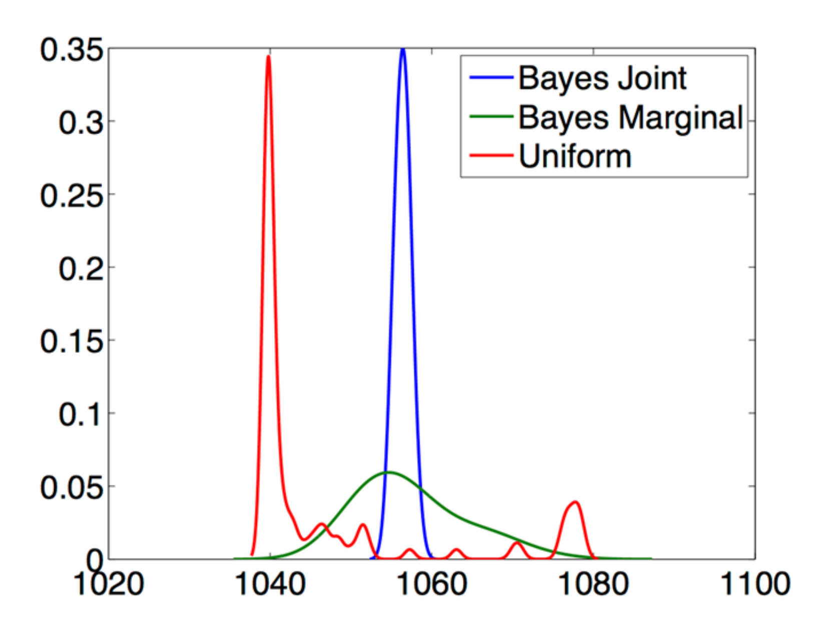


$2\sigma \approx 0.03$

Note: Substantial reduction in parameter uncertainty

Use of Prediction Intervals: Nuclear Power Plant Design

Strategy: Propagate parameter uncertainties through COBRA-TF to determine uncertainty in maximum fuel temperature



Notes:

- Temperature uncertainty reduced from 40 degrees to 5 degrees
- Can run plant 20 degrees hotter, which significantly improves efficiency

Ramification: Savings of 10 billion dollars per year for US power plants

Issues:

- We considered only one of many closure relations
- Nuclear regulatory commission takes years to change requirements and codes

Good News: We are now working with Westinghouse to reduce uncertainties.

MATLAB Example

SIR Model:

$$\frac{dS}{dt} = \delta N - \delta S - \gamma IS \quad , \; S(0) = S_0 \qquad \text{Susceptible}$$

$$\frac{dI}{dt} = \gamma IS - (r + \delta)I \quad , \; I(0) = I_0 \qquad \text{Infectious}$$

$$\frac{dR}{dt} = rI - \delta R \quad , \; R(0) = R_0 \qquad \text{Recovered}$$

Note:

- Run either the 3 or 4 parameter model and compute the prediction intervals.

Website:

- http://www4.ncsu.edu/~rsmith/DATAWORKS18/