Bayesian Computation Using Design of Experiments-Based Interpolation Technique

V. Roshan Joseph

Technometrics, August, 2012
(with discussions)

Supported by
NSF CMMI-1030125

Georgia Institute of Technology

INDUSTRIAL & SYSTEMS ENGINEERING

Bayesian Computation

- Many intractable high-dimensional integrals
  - Posterior distribution
  - Posterior summaries
  - Marginal posterior distributions
  - Posterior predictive distributions
- Use approximation methods

Deterministic Methods

- Laplace’s Approximation (e.g. Tierney and Kadane 1986)
  - May not be accurate.
- Gaussian Quadrature (e.g. Naylor and Smith 1982)
  - Can produce accurate results.
  - Cannot be used for high dimensions.

Simulation-Based Methods

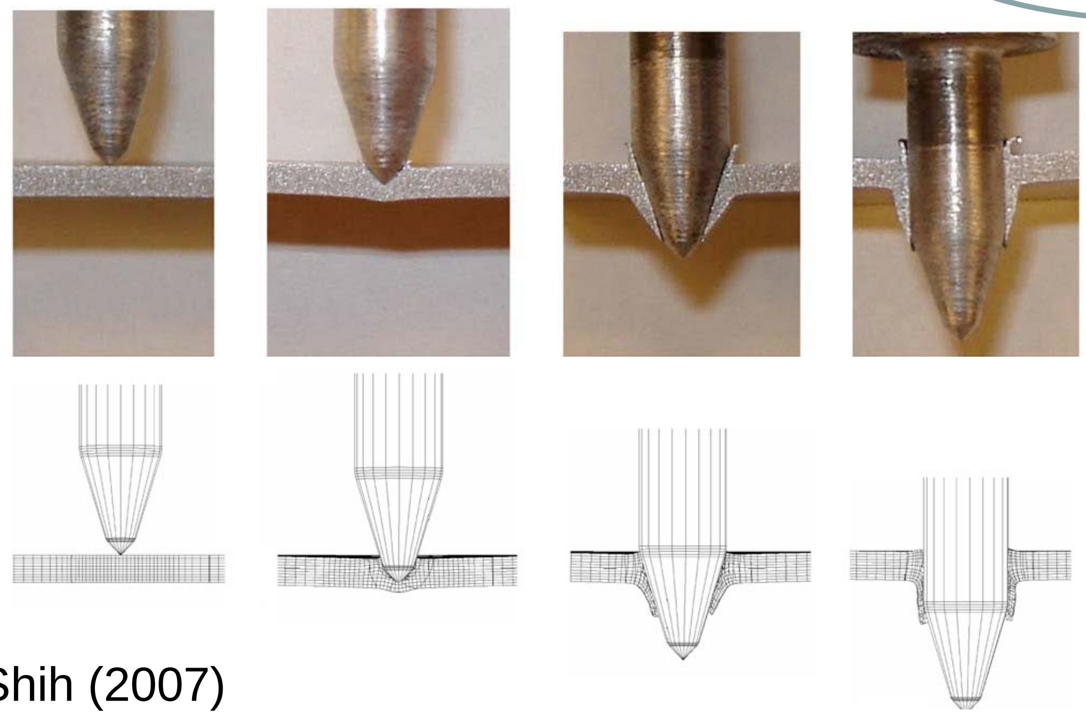
- MC/MCMC (Metropolis et al. 1953, Hastings 1970, Geman and Geman 1984, Gelfand and Smith 1990, …)
  - Can obtain results with arbitrary precision
  - Suffer less from the curse of dimensionality
  - Convergence issues
  - High computational cost when dealing with computationally expensive posteriors

Examples of computationally expensive posteriors

- Model calibration

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \{y_i - f(\boldsymbol{x}_i;\boldsymbol{\theta})\}^2$$

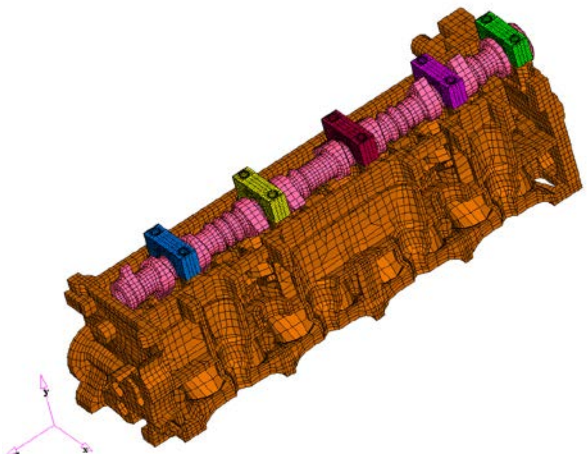
Expensive



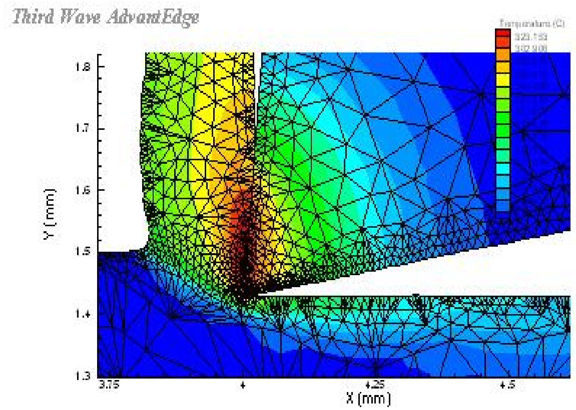
Miller and Shih (2007)

Examples-continued

- Geostatistics/Spatial statistics/Computer experiments



Joseph, Hung, Sudjianto (2008)



Hung, Joseph, Melkote (2009)

$$f(\boldsymbol{x})|\mu, \sigma^2, \boldsymbol{\theta} \sim GP(\mu, \sigma^2 R(.;\boldsymbol{\theta}))$$

$$p(\boldsymbol{y}|\mu, \sigma^2, \boldsymbol{\theta}) \propto \frac{1}{\sigma^n |\boldsymbol{R}(\boldsymbol{\theta})|^{1/2}} \exp\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - \mu\mathbf{1})'\boldsymbol{R}(\boldsymbol{\theta})^{-1}(\boldsymbol{y} - \mu\mathbf{1})\}$$

Expensive

Examples-continued

- Inexpensive likelihood, but appearing within some algorithms.

- Simulation-based nonlinear optimal design (Muller 1999)

$$\max_{\boldsymbol{D}} \frac{1}{N} \sum_{i=1}^{N} \log p(\boldsymbol{\theta}_i | \boldsymbol{D}, \boldsymbol{y}_i)$$

Expensive

Not Expensive

New Deterministic Method

- **D**esign **o**f Experiments-based **I**nterpolation **t**echnique (**DoIt**)
  - Design of experiments
  - Interpolation methods (e.g., kriging)
- Advantages
  - Can obtain results with arbitrary precision
  - Suffer less from the curse of dimensionality
  - Works much faster than MC/MCMC
- Disadvantages
  - Small to moderate number dimensions
  - Continuous parameters

Earlier Work

- Bayes-Hermite Quadrature
  - O'Hagan (1991)
  - Kennedy (1998)
  - Rasmussen and Ghahramani (2003)
- Hybrid Monte Carlo using Gaussian Process Models
  - Rasmussen (2003)
  - Bliznyuk et al. (2008)
  - Henedrson et al. (2008)
  - Fielding, Nott, and Liong (2011)

Laplace's Approximation

- Bayesian model:

$$\begin{aligned} \boldsymbol{y}|\boldsymbol{\theta} &\sim p(\boldsymbol{y}|\boldsymbol{\theta}) \\ \boldsymbol{\theta} &\sim p(\boldsymbol{\theta}) \end{aligned}$$

- Unnormalized posterior: $h(\boldsymbol{\theta}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$

$$\hat{\boldsymbol{\theta}} = argmax_{\boldsymbol{\theta}} h(\boldsymbol{\theta})$$

$$h(\boldsymbol{\theta}) \approx h(\hat{\boldsymbol{\theta}}) \exp\{-\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\}$$

$$\boldsymbol{\theta}|\boldsymbol{y} \sim^a N(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma})$$

$$\boldsymbol{\Sigma} = [-\nabla^2 \log(h(\hat{\boldsymbol{\theta}}))]^{-1}$$

DoIt

- Unnormalized normal density function:

$$g(\boldsymbol{\theta}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \exp\{-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu})\}$$

- Experimental design:

$$\boldsymbol{D} = \{\boldsymbol{\nu}_1, \cdots, \boldsymbol{\nu}_m\}$$

- DoIt:

$$h(\boldsymbol{\theta}) \approx \sum_{i=1}^{m} c_i g(\boldsymbol{\theta}; \boldsymbol{\nu}_i, \boldsymbol{\Sigma})$$

- Kriging, Radial Basis Function,…

DoIt-continued

- Evaluations: $\boldsymbol{h} = (h_1, \cdots, h_m)'$, where $h_i = h(\boldsymbol{\nu}_i)$
- To get interpolation:

$$\boldsymbol{G}\boldsymbol{c} = \boldsymbol{h},$$

where $\boldsymbol{c} = (c_1, \cdots, c_m)'$ and $\boldsymbol{G}_{ij} = g(\boldsymbol{\nu}_i; \boldsymbol{\nu}_j, \boldsymbol{\Sigma})$

$$\tilde{\boldsymbol{c}} = \boldsymbol{G}^{-1}\boldsymbol{h}$$

- Let $\boldsymbol{g}(\boldsymbol{\theta}) = (g(\boldsymbol{\theta}; \boldsymbol{\nu}_1, \boldsymbol{\Sigma}), \cdots, g(\boldsymbol{\theta}; \boldsymbol{\nu}_m, \boldsymbol{\Sigma}))'$

$$\tilde{h}(\boldsymbol{\theta}) = \tilde{\boldsymbol{c}}'\boldsymbol{g}(\boldsymbol{\theta})$$

DoIt-continued

- Marginal likelihood

$$\begin{aligned}\int \tilde{h}(\boldsymbol{\theta}) d\boldsymbol{\theta} &= \tilde{\boldsymbol{c}}' \int \boldsymbol{g}(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= (2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2} \tilde{\boldsymbol{c}}' \mathbf{1}\end{aligned}$$

- Posterior distribution

$$\tilde{p}(\boldsymbol{\theta}|\boldsymbol{y}) \approx \frac{\tilde{\boldsymbol{c}}' \boldsymbol{g}(\boldsymbol{\theta})}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2} \tilde{\boldsymbol{c}}' \mathbf{1}} = \frac{\tilde{\boldsymbol{c}}' \boldsymbol{\phi}(\boldsymbol{\theta})}{\tilde{\boldsymbol{c}}' \mathbf{1}}$$

$$\boxed{\tilde{p}(\boldsymbol{\theta}|\boldsymbol{y}) \approx \frac{\sum_{i=1}^{m} \tilde{c}_i \phi(\boldsymbol{\theta}; \boldsymbol{\nu}_i, \boldsymbol{\Sigma})}{\sum_{i=1}^{m} \tilde{c}_i}}$$

Example

- Bayesian model:

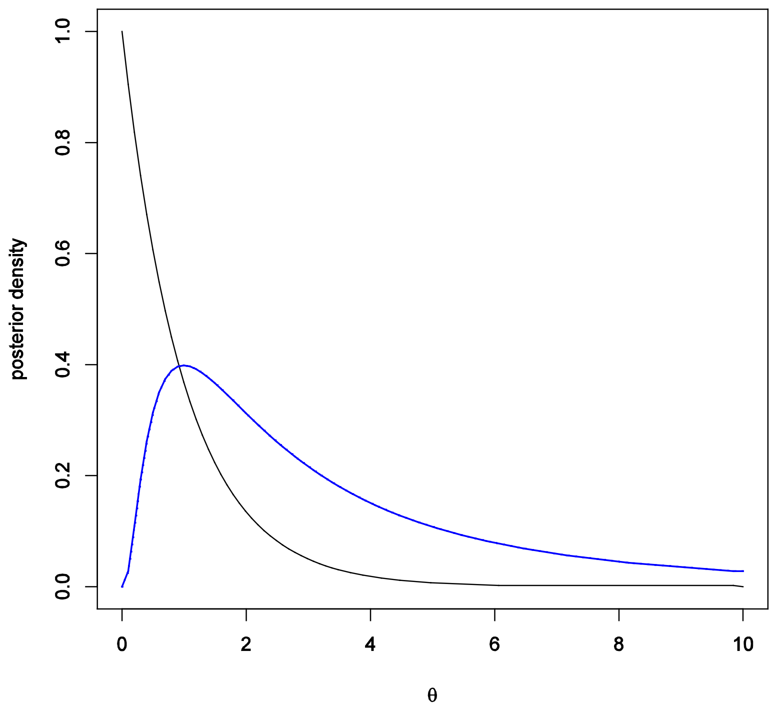
$$\begin{aligned} y|\theta &\sim Poisson(\theta), \\ p(\theta) &\propto 1. \end{aligned}$$

- For *y*=0, posterior distribution:

$$\theta|y \sim Exp(1).$$

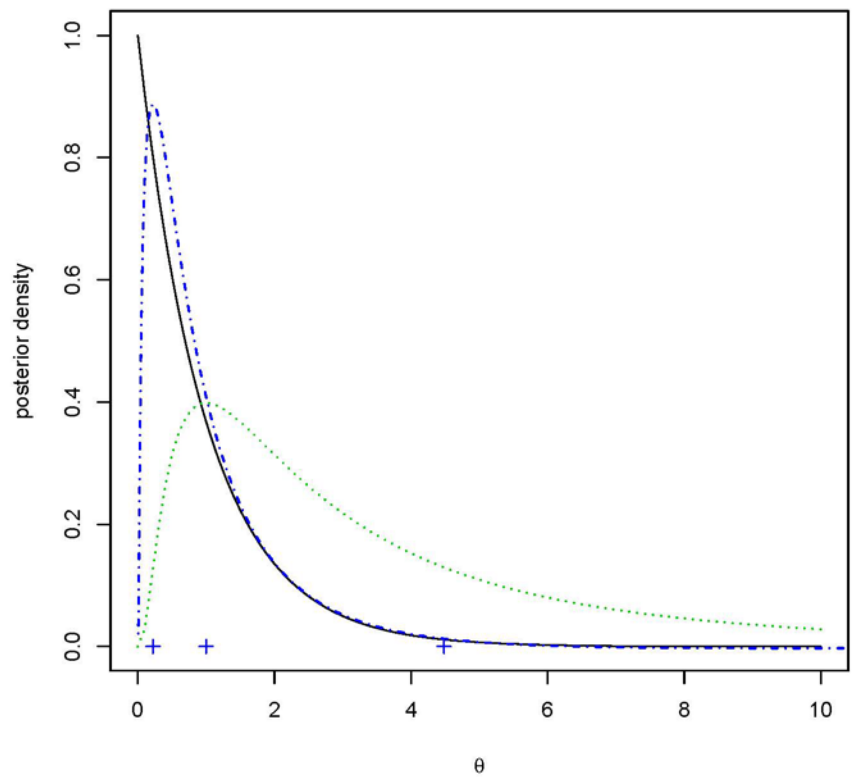
Example-continued

- Laplace's approximation: $\gamma = \log(\theta)$
- $\hat{\gamma} = 0 \text{ and } \hat{\sigma}^2 = 1$: $\theta|y \sim^a log - normal(0,1).$



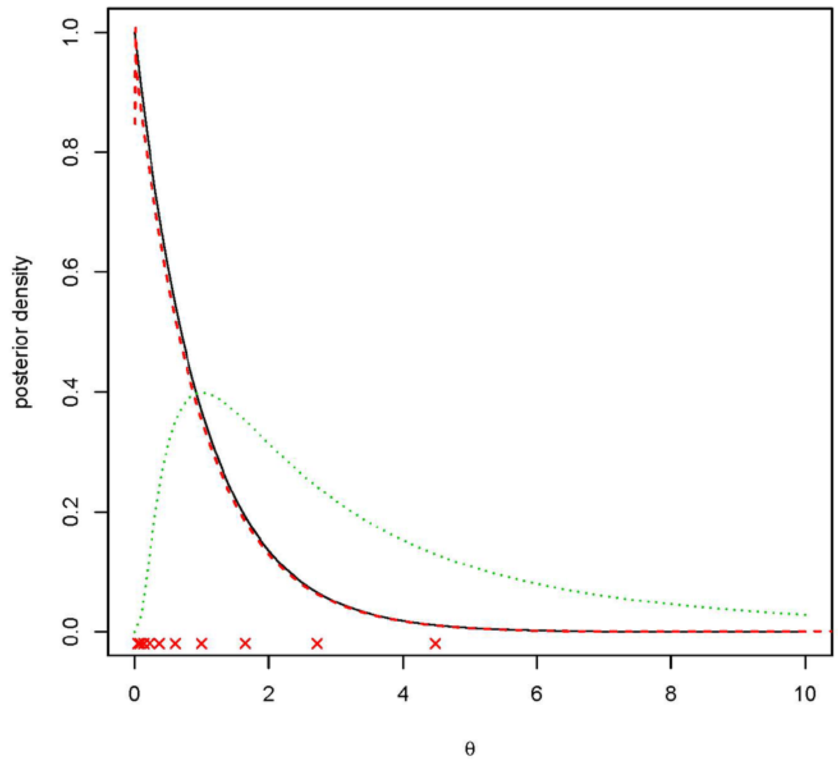
Example-continued

- **DoIt** : $\nu_1 = \hat{\gamma},\ \nu_2 = \hat{\gamma} - 1.5\sigma,\ \text{and}\ \nu_2 = \hat{\gamma} + 1.5\sigma$



Example-continued

- **DoIt** : 10 equally spaced points from $\hat{\gamma} - 3\sigma$ to $\hat{\gamma} + 1.5\sigma$



A Result

Theorem 1: If $h(\boldsymbol{\theta})$ is continuous, then for any $\alpha \in (0,1)$ and any $\epsilon > 0$, there exists a finite number of points $\boldsymbol{D} = \{\boldsymbol{\nu}_1, \cdots, \boldsymbol{\nu}_m\}$ in $\boldsymbol{\Theta}$ such that

$$\left|\frac{\hat{h}(\boldsymbol{\theta}) / \int_{\boldsymbol{\Theta}} \hat{h}(\boldsymbol{\theta}) d\boldsymbol{\theta}}{h(\boldsymbol{\theta}) / \int_{\boldsymbol{\Theta}} h(\boldsymbol{\theta}) d\boldsymbol{\theta}} - 1\right| < \epsilon \tag{6}$$

for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\hat{h}(\boldsymbol{\theta})$ is any continuous and uniformly convergent interpolator of $h(\boldsymbol{\theta})$ on $\boldsymbol{D}$ and $\boldsymbol{\Theta}$ is the $(1-\alpha)$ highest posterior density (HPD) credible set.

As $\alpha \to 0$,

$$\frac{\hat{h}(\boldsymbol{\theta}) / \int_{\boldsymbol{\Theta}} \hat{h}(\boldsymbol{\theta}) d\boldsymbol{\theta}}{h(\boldsymbol{\theta}) / \int_{\boldsymbol{\Theta}} h(\boldsymbol{\theta}) d\boldsymbol{\theta}} \to \tilde{p}(\boldsymbol{\theta}|\boldsymbol{y}) / p(\boldsymbol{\theta}|\boldsymbol{y}),$$

and as $\epsilon \to 0$,

$$\tilde{p}(\boldsymbol{\theta}|\boldsymbol{y}) / p(\boldsymbol{\theta}|\boldsymbol{y}) \to 1.$$

Unknown Posterior Mode

- Assume that $\boldsymbol{\Sigma} = diag\{\sigma_1^2, \cdots, \sigma_d^2\}$
- Leave-one-out cross validation: $e_i = h_i - \tilde{h}_{(i)i}$
- From the kriging literature

$$e_i = \frac{(\boldsymbol{G}^{-1})_i}{(\boldsymbol{G}^{-1})_{ii}} \boldsymbol{h}$$

- Minimize

$$MSCV = \frac{1}{m} \boldsymbol{e}' \boldsymbol{e}.$$

- or

$$WMSCV = \frac{1}{m} \boldsymbol{e}' diag(\boldsymbol{G}^{-1}) \boldsymbol{e}$$

Example

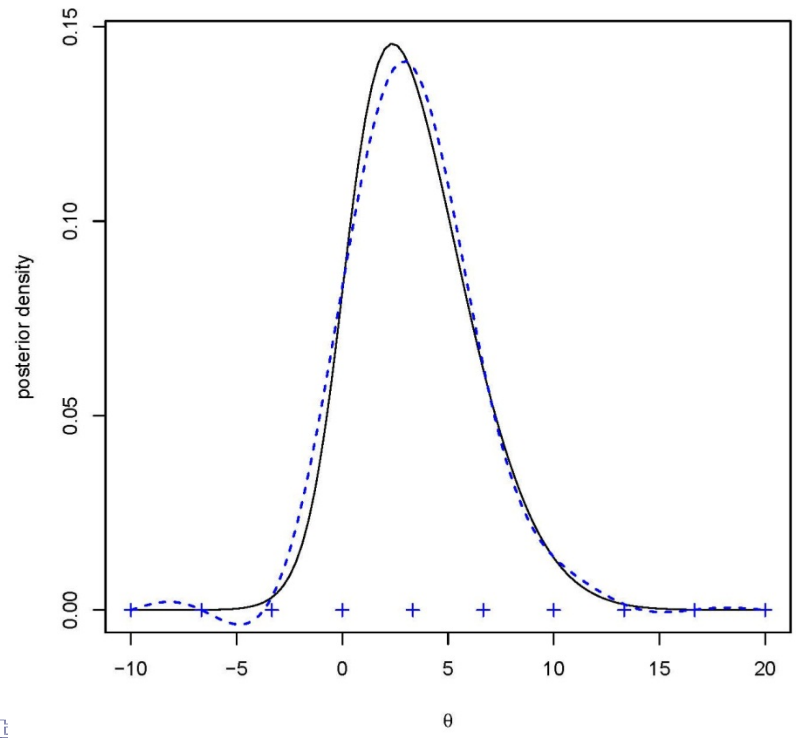
- Bayesian model:

$$y|\theta \sim Bernoulli(\{1+\exp(-\theta)\}^{-1}),$$
$$\theta \sim N(\mu, \tau^2).$$

- Suppose $y = 1$, $\mu = 1$ and $\tau = 4$.

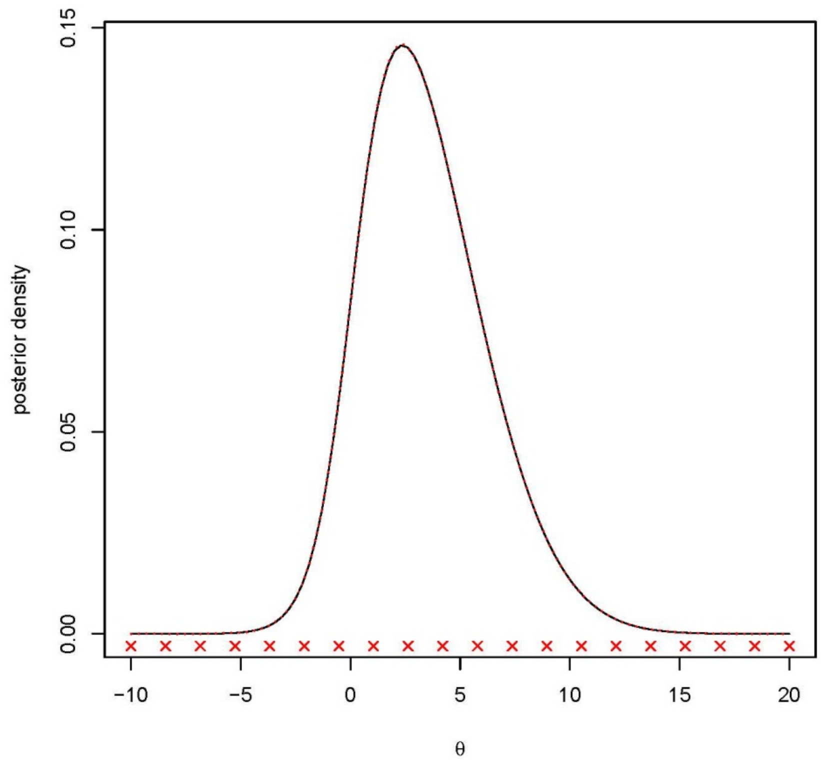
Example-continued

- Suppose we sample 10 equally spaced points from -10 to 20.
- Minimizing *WMSCV*: $\hat{\sigma}^2 = 9.30$



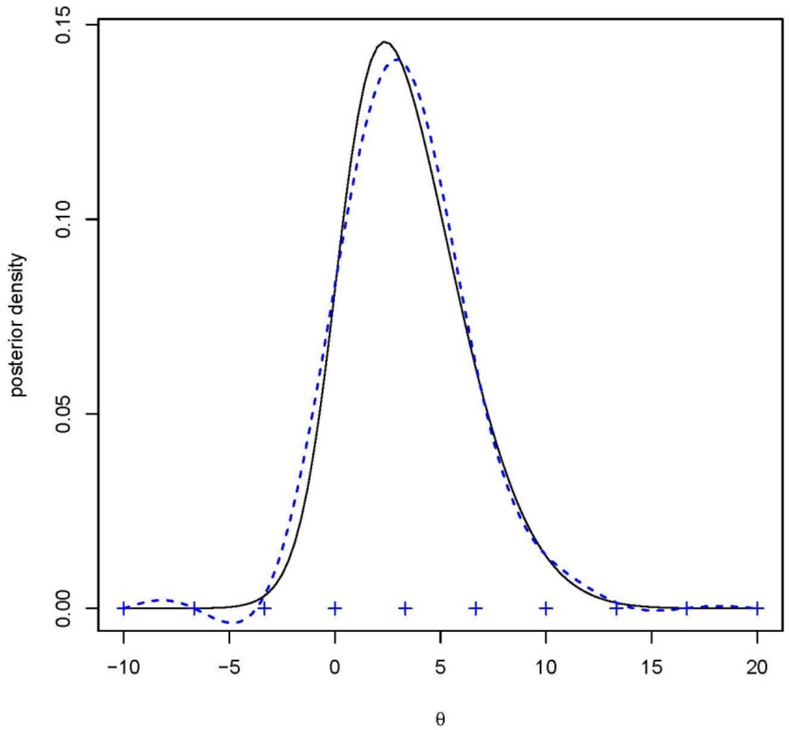
Example-continued

- m=20



Negativity Problem

- The coefficients $\tilde{c}_i$ can be negative and can lead to negative posterior density values.



Mixture Normal Approximation

- Restrict c_i's to be nonnegative:

$$\min_{\boldsymbol{c} \geq 0} (\boldsymbol{h} - \boldsymbol{G}\boldsymbol{c})'\boldsymbol{G}^{-1}(\boldsymbol{h} - \boldsymbol{G}\boldsymbol{c})$$

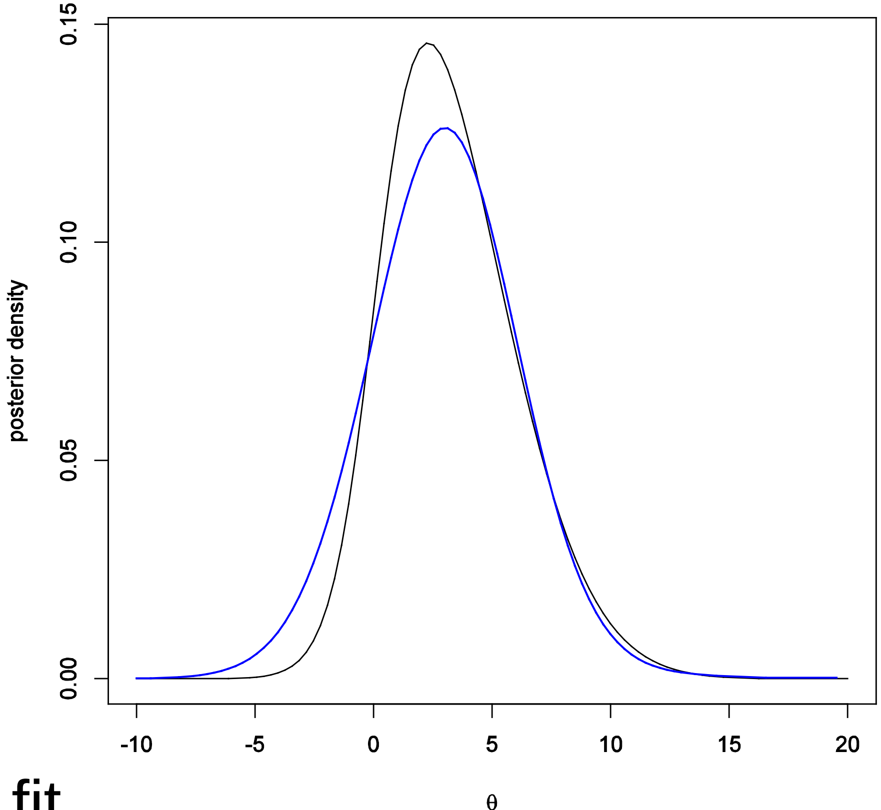
- Quadratic program.
- Then, **DoIt**

$$\hat{p}(\boldsymbol{\theta}|\boldsymbol{y}) \approx \frac{\hat{\boldsymbol{c}}'\phi(\boldsymbol{\theta};\boldsymbol{\Sigma})}{\hat{\boldsymbol{c}}'\boldsymbol{1}}$$

becomes a mixture normal approximation.

Mixture Normal Approximation

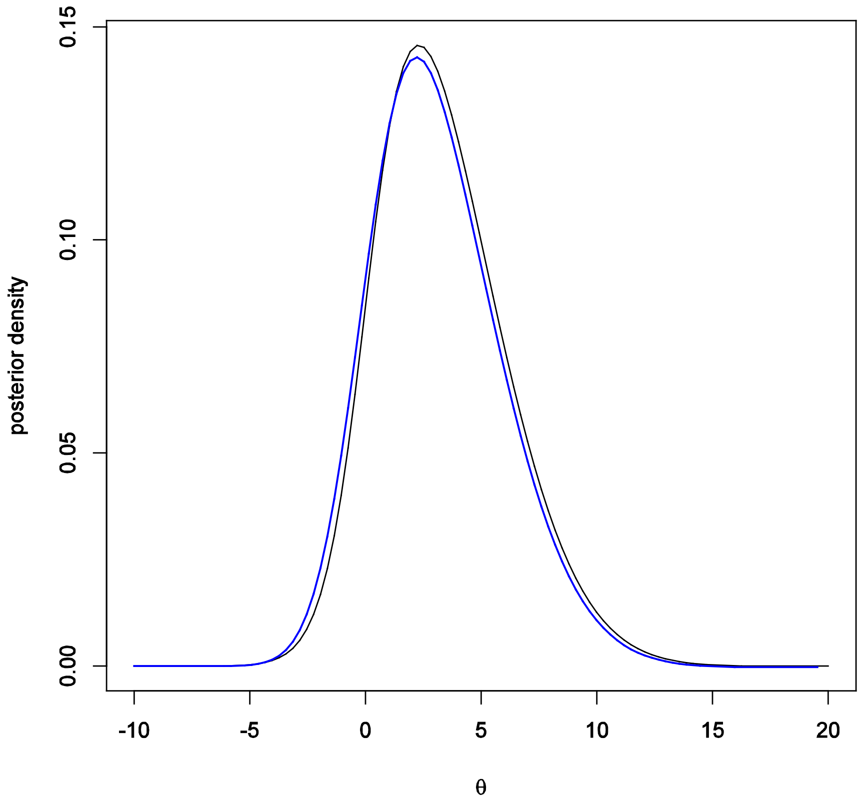
- m=10



- Not a good fit.

Mixture Normal Approximation

- m=20



- Better, but not good enough.

Improved DoIt

- DoIt :

$$h(\boldsymbol{\theta}) \approx \sum_{i=1}^{m} \hat{c}_i g(\boldsymbol{\theta}; \boldsymbol{\nu}_i, \boldsymbol{\Sigma}) \left\{ a + \sum_{i=1}^{m} b_i g(\boldsymbol{\theta}; \boldsymbol{\nu}_i, \boldsymbol{\Lambda}) \right\}$$

- Let $z_i = h(\boldsymbol{\nu}_i)/\hat{\boldsymbol{c}}'\boldsymbol{g}(\boldsymbol{\nu}_i; \boldsymbol{\Sigma})$ for $i = 1, \cdots, m.$

- Then, $\hat{\boldsymbol{b}} = \boldsymbol{G}(\boldsymbol{\Lambda})^{-1}(\boldsymbol{z} - a\boldsymbol{1})$

- New approximation:

$$\hat{h}(\boldsymbol{\theta}) = \hat{\boldsymbol{c}}'\boldsymbol{g}(\boldsymbol{\theta}; \boldsymbol{\Sigma})\{a + \hat{\boldsymbol{b}}'\boldsymbol{g}(\boldsymbol{\theta}; \boldsymbol{\Lambda})\}$$

Improved DoIt-continued

$$\begin{aligned}\int \hat{h}(\boldsymbol{\theta})d\boldsymbol{\theta} &= a\hat{\boldsymbol{c}}' \int \boldsymbol{g}(\boldsymbol{\theta};\boldsymbol{\Sigma})d\boldsymbol{\theta} + \hat{c}' \int \boldsymbol{g}(\boldsymbol{\theta};\boldsymbol{\Sigma})\boldsymbol{g}(\boldsymbol{\theta};\boldsymbol{\Lambda})'d\boldsymbol{\theta}\ \hat{\boldsymbol{b}} \\ &= a\hat{\boldsymbol{c}}'(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}\mathbf{1} + \hat{\boldsymbol{c}}'(2\pi)^{d/2}\frac{|\boldsymbol{\Sigma}\boldsymbol{\Lambda}|^{1/2}}{|\boldsymbol{\Sigma}+\boldsymbol{\Lambda}|^{1/2}}\boldsymbol{G}(\boldsymbol{\Sigma}+\boldsymbol{\Lambda})\hat{\boldsymbol{b}}\end{aligned}$$

$$a = \int \hat{z}(\boldsymbol{\theta})\frac{\hat{\boldsymbol{c}}'\boldsymbol{\phi}(\boldsymbol{\theta};\boldsymbol{\Sigma})}{\hat{\boldsymbol{c}}'\mathbf{1}}\,d\boldsymbol{\theta} = \frac{\int \hat{h}(\boldsymbol{\theta})d\boldsymbol{\theta}}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}\hat{\boldsymbol{c}}'\mathbf{1}}$$

$$a = \frac{\hat{\boldsymbol{c}}'\boldsymbol{G}(\boldsymbol{\Sigma}+\boldsymbol{\Lambda})\boldsymbol{G}(\boldsymbol{\Lambda})^{-1}\boldsymbol{z}}{\hat{\boldsymbol{c}}'\boldsymbol{G}(\boldsymbol{\Sigma}+\boldsymbol{\Lambda})\boldsymbol{G}(\boldsymbol{\Lambda})^{-1}\mathbf{1}}$$

Improved DoIt-continued

$$\int \hat{h}(\boldsymbol{\theta})d\boldsymbol{\theta} = a(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}\hat{\boldsymbol{c}}'\mathbf{1}$$

$$\boxed{\hat{p}(\boldsymbol{\theta}|\boldsymbol{y}) \approx \frac{\hat{\boldsymbol{c}}'\boldsymbol{\phi}(\boldsymbol{\theta};\boldsymbol{\Sigma})}{\hat{\boldsymbol{c}}'\mathbf{1}}\{1 + \hat{\boldsymbol{b}}'\boldsymbol{g}(\boldsymbol{\theta};\boldsymbol{\Lambda})/a\}}$$

- $\boldsymbol{\Lambda}$ can be obtained using cross validation.

Improved DoIt-continued

$$\text{Let } \boldsymbol{V} = \boldsymbol{\Sigma}(\boldsymbol{\Sigma}+\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda} \text{ and } \boldsymbol{\mu}_{ij} = \boldsymbol{V}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu}_i + \boldsymbol{\Lambda}^{-1}\boldsymbol{\nu}_j)$$

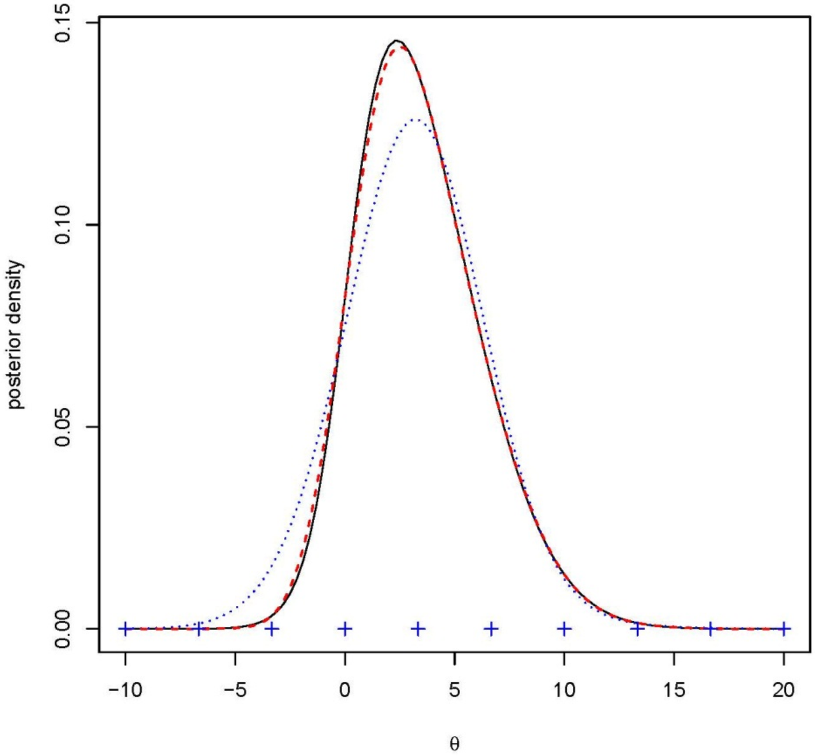
$$\hat{p}(\boldsymbol{\theta}|\boldsymbol{y}) \approx \frac{\sum_{i=1}^{m}\hat{c}_i\phi(\boldsymbol{\theta};\boldsymbol{\nu}_i,\boldsymbol{\Sigma}) + \sum_{i=1}^{m}\sum_{j=1}^{m} d_{ij}\phi(\boldsymbol{\theta};\boldsymbol{\mu}_{ij},\boldsymbol{V})}{\sum_{i=1}^{m}\hat{c}_i}$$

- where

$$d_{ij} = \frac{\hat{c}_i\hat{b}_j|\boldsymbol{\Lambda}|^{1/2}}{a|\boldsymbol{\Sigma}+\boldsymbol{\Lambda}|^{1/2}} g(\boldsymbol{\nu}_i;\boldsymbol{\nu}_j,\boldsymbol{\Sigma}+\boldsymbol{\Lambda}).$$

Binary Data Example

- m=10



Another Example

- Marin and Robert (2007)

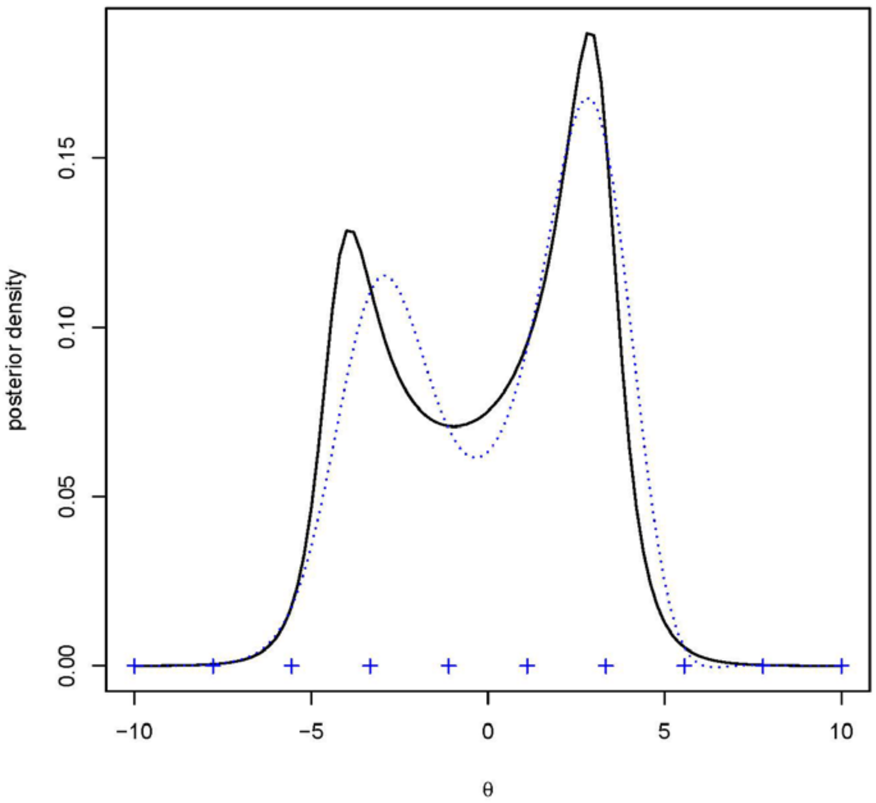
$$\begin{aligned} y|\theta &\sim Cauchy(\theta, 1), \\ \theta &\propto N(0, (\sqrt{10})^2). \end{aligned}$$

- Unnormalized posterior:

$$h(\theta) = \frac{\exp(-\theta^2/20)}{\prod_{i=1}^{2}(1 + (y_i - \theta)^2)}.$$

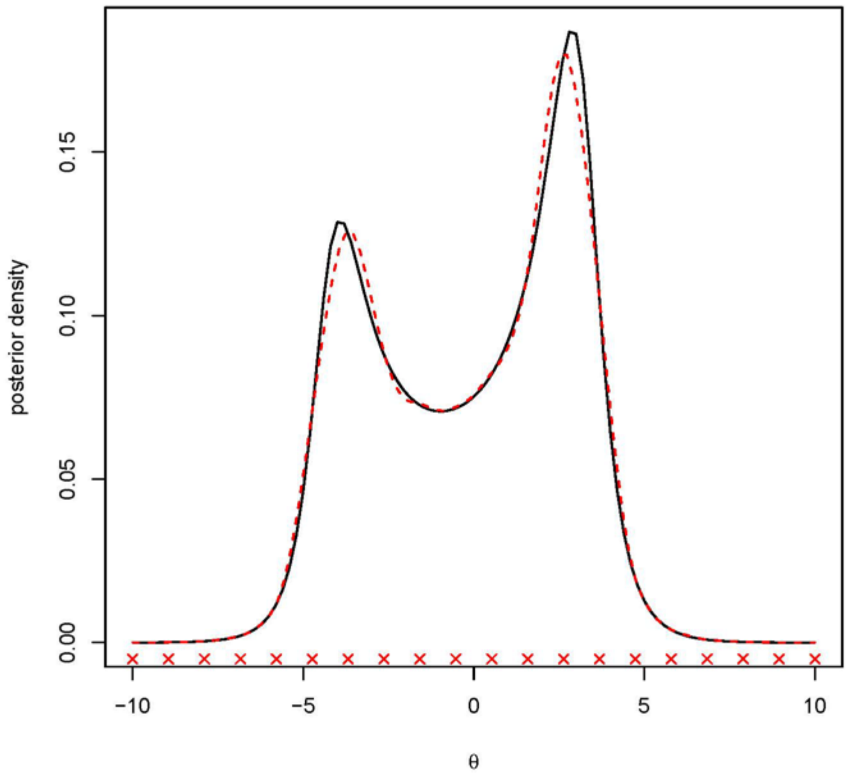
Example-continued

Suppose we sample 10 equally spaced points from -10 to 10.



Example-continued

- m=20



Marginal Posterior Distributions

- Using properties of multivariate normal distribution:

$$\hat{p}(\theta_k|\boldsymbol{y}) \approx \frac{\sum_{i=1}^{m} \hat{c}_i \phi(\theta_k; \nu_{ik}, \boldsymbol{\Sigma}_{kk}) + \sum_{i=1}^{m} \sum_{j=1}^{m} d_{ij} \phi(\theta_k; \mu_{ijk}, \boldsymbol{V}_{kk})}{\sum_{i=1}^{m} \hat{c}_i}$$

Posterior Summaries

$$E(\boldsymbol{\theta}|\boldsymbol{y}) = \bar{\boldsymbol{\theta}} \approx \frac{\sum_{i=1}^{m} \hat{c}_i \boldsymbol{\nu}_i + \sum_{i=1}^{m} \sum_{j=1}^{m} d_{ij} \boldsymbol{\mu}_{ij}}{\sum_{i=1}^{m} \hat{c}_i}$$

$$var(\boldsymbol{\theta}|\boldsymbol{y}) \approx \frac{\sum_{i=1}^{m} \hat{c}_i (\boldsymbol{\nu}_i \boldsymbol{\nu}_i' + \boldsymbol{\Sigma}) + \sum_{i=1}^{m} \sum_{j=1}^{m} d_{ij} (\boldsymbol{\mu}_{ij} \boldsymbol{\mu}_{ij}' + \boldsymbol{V})}{\sum_{i=1}^{m} \hat{c}_i} - \bar{\boldsymbol{\theta}} \bar{\boldsymbol{\theta}}'$$

Posterior Expectation

- More generally,

$$\begin{aligned}\xi &= E\{f(\boldsymbol{\theta})|\boldsymbol{y}\}\\ &\approx \int f(\boldsymbol{\theta})\frac{\hat{\mathbf{c}}'\phi(\boldsymbol{\theta};\boldsymbol{\Sigma})}{a\hat{\mathbf{c}}'\mathbf{1}}\{a+\hat{\boldsymbol{b}}'\boldsymbol{g}(\boldsymbol{\theta};\boldsymbol{\Lambda})\}\, d\boldsymbol{\theta}\end{aligned}$$

for some function $f(\boldsymbol{\theta})$.

- Use approximations.

Posterior Expectation-continued

- First, let $z(\boldsymbol{\theta}) = a + \hat{\boldsymbol{b}}'\boldsymbol{g}(\boldsymbol{\theta};\boldsymbol{\Lambda})$ and $f^*(\boldsymbol{\theta}) = f(\boldsymbol{\theta})z(\boldsymbol{\theta})$.

$$\xi \approx \frac{1}{a\hat{\boldsymbol{c}}'\boldsymbol{1}} \sum_{i=1}^{m} \hat{c}_i \int f^*(\boldsymbol{\theta})\phi(\boldsymbol{\theta};\boldsymbol{\nu}_i,\boldsymbol{\Sigma})\, d\boldsymbol{\theta}$$

- Let $\boldsymbol{f}^* = (f^*(\boldsymbol{\nu}_1),\cdots,f^*(\boldsymbol{\nu}_m))' = \boldsymbol{f}\odot\boldsymbol{z}$

- Approximate $f^*(\boldsymbol{\theta})$ using a kriging predictor:

$$f^*(\boldsymbol{\theta}) = \alpha z(\boldsymbol{\theta}) + \boldsymbol{g}(\boldsymbol{\theta};\boldsymbol{\Omega})'\boldsymbol{G}(\boldsymbol{\Omega})^{-1}(\boldsymbol{f}^* - \alpha\boldsymbol{z})$$

Posterior Expectation-continued

$$\xi \approx \alpha + \frac{|\boldsymbol{\Omega}|^{1/2}}{a\hat{\boldsymbol{c}}'\boldsymbol{1}|\boldsymbol{\Omega} + \boldsymbol{\Sigma}|^{1/2}} \hat{c}' \boldsymbol{G}(\boldsymbol{\Omega} + \boldsymbol{\Sigma}) \boldsymbol{G}(\boldsymbol{\Omega})^{-1} (\boldsymbol{f}^* - \alpha \boldsymbol{z})$$

- Choose $\alpha = \xi$ (Joseph 2006),

$$\xi \approx \frac{\hat{c}' \boldsymbol{G}(\boldsymbol{\Omega} + \boldsymbol{\Sigma}) \boldsymbol{G}(\boldsymbol{\Omega})^{-1} \boldsymbol{f}^*}{\hat{c}' \boldsymbol{G}(\boldsymbol{\Omega} + \boldsymbol{\Sigma}) \boldsymbol{G}(\boldsymbol{\Omega})^{-1} \boldsymbol{z}}$$

- Take $\boldsymbol{\Omega} = \boldsymbol{\Lambda}$,

$$\boxed{\xi \approx \frac{\hat{c}' \boldsymbol{G}(\boldsymbol{\Sigma} + \boldsymbol{\Lambda}) \boldsymbol{G}(\boldsymbol{\Lambda})^{-1} \boldsymbol{f}^*}{\hat{c}' \boldsymbol{G}(\boldsymbol{\Sigma} + \boldsymbol{\Lambda}) \boldsymbol{G}(\boldsymbol{\Lambda})^{-1} \boldsymbol{z}}}$$

Example

- Posterior predictive density in the binary data example:

$$y|\boldsymbol{y} \sim Bernoulli(\xi),$$

where

$$\xi = E(\{1+\exp(-\theta)\}^{-1}|\boldsymbol{y}).$$

- Numerical integration: $\xi = .8496$
- Kriging approximation: $\xi \approx .8478$
- First order approximation: $\xi \approx \{1 + \exp(-\hat{\theta})\}^{-1} = .914$

Experimental Design

- Initial space-filling design
- Sequential design

Space-Filling Design

- By Laplace's approximation

$$\boldsymbol{\theta}|\boldsymbol{y} \sim^a N(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma})$$

- Transform:

$$\boldsymbol{\alpha} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$$

- Then,

$$\boldsymbol{\alpha}|\boldsymbol{y} \sim^a N(\boldsymbol{0}, \boldsymbol{I})$$

First choose a design $\boldsymbol{D}^* = (\boldsymbol{\nu}_1^*, \cdots, \boldsymbol{\nu}_m^*)'$ from $(0,1)^d$

$$\boldsymbol{D} = (\hat{\boldsymbol{\theta}} + \boldsymbol{\Sigma}^{1/2}\Phi^{-1}(\boldsymbol{\nu}_1^*), \cdots, \hat{\boldsymbol{\theta}} + \boldsymbol{\Sigma}^{1/2}\Phi^{-1}(\boldsymbol{\nu}_m^*))'$$

Space-Filling Design-continued

- Latin Hypercube Design (LHD)
- Maximin LHD (Morris and Mitchell 1995)

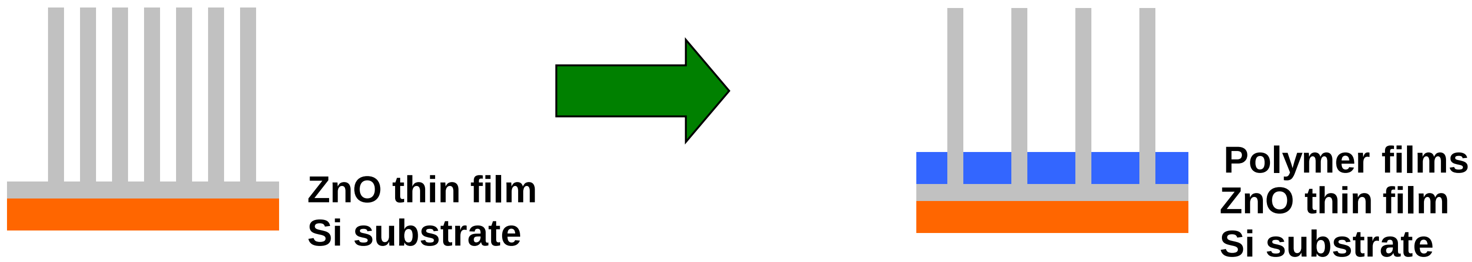
- Let $\boldsymbol{\nu}_1^* = \mathbf{.5} = (.5, \cdots, .5)'$

- Find the remaining m-1 points by minimizing

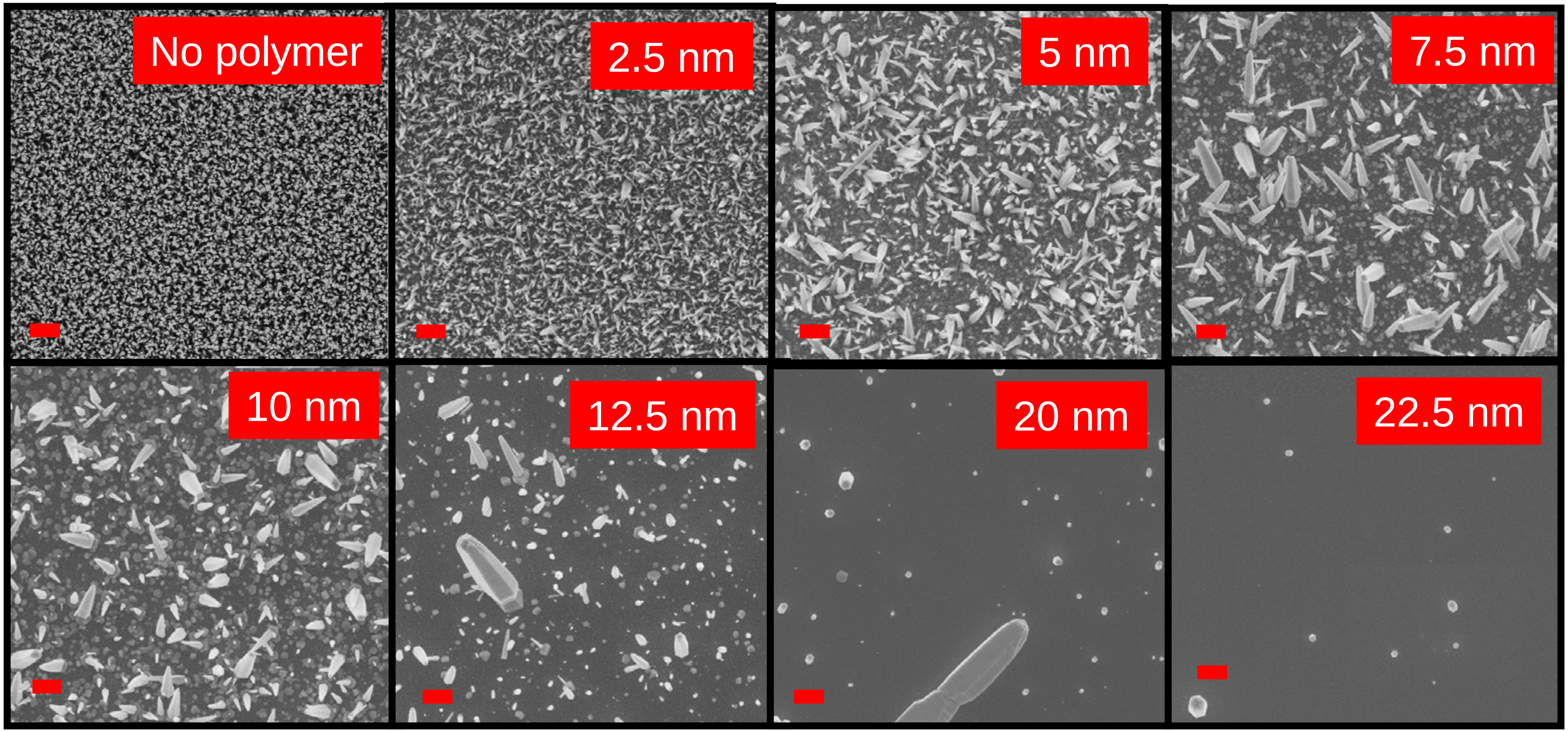
$$\left\{ \sum_{i=2}^{m} \sum_{j=2}^{m} 1/d^k(\boldsymbol{\nu}_i^*, \boldsymbol{\nu}_j^*) \right\}^{1/k}$$

Example: Density Control of Nanowires

- Dasgupta, Weintraub, and Joseph (2011)



Results (1st set of experiments)



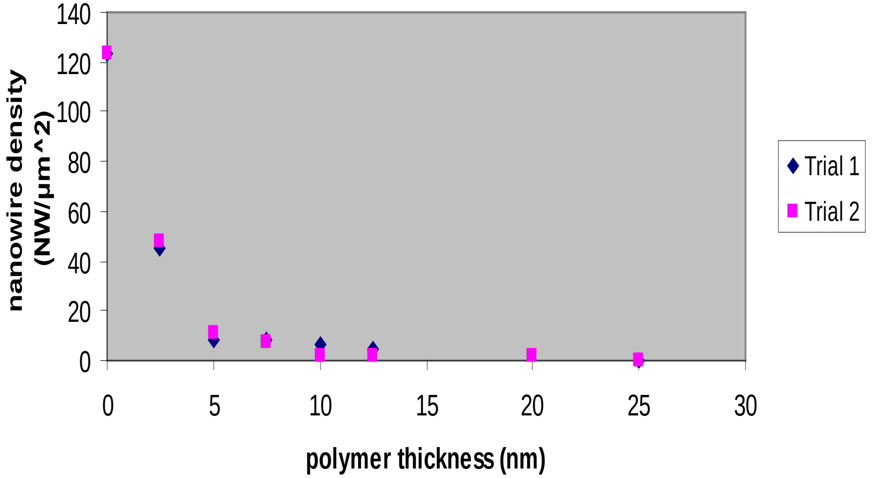
(All images at 5000x magnification; ▬ = 1 μm)

Georgia Institute of Technology

ISYE INDUSTRIAL & SYSTEMS ENGINEERING

Experimental Data

# Bi Layers	Thickness (nm)	Density (NW/μm^2)	
		Trial 1	Trial 2
0	0.0	123	123
1	2.5	46	48
2	5.0	8	11
3	7.5	8	7
4	10.0	7	2
5	12.5	5	1
8	20.0	2	-
10	25.0	0	0



Example-continued

- Density of nanowires (*y*)
- Thickness of polymer films (*x*)

$$\begin{aligned} y_{ij}|\boldsymbol{\theta}, \boldsymbol{u} &\sim Poisson(\mu(x_i)) \\ \mu(x_i) &= \left[\theta_1 \exp(-\theta_2 x_i^2) + \theta_3\{1 - \exp(-\theta_2 x_i^2)\}\Phi(-x_i/\theta_4)\right] u_i \end{aligned}$$

for $i = 1, \cdots, 8$ and $j = 1, 2.$

- Let $\gamma_i = \log(\theta_i)$ for $i = 1, \cdots, 4.$
  and $\alpha_i = \log(u_i)$ for $i = 1, \cdots, 8$

- Prior: $p(\boldsymbol{\gamma}) \propto 1,$ and $\alpha_i \sim^{iid} N(0, .1^2)$

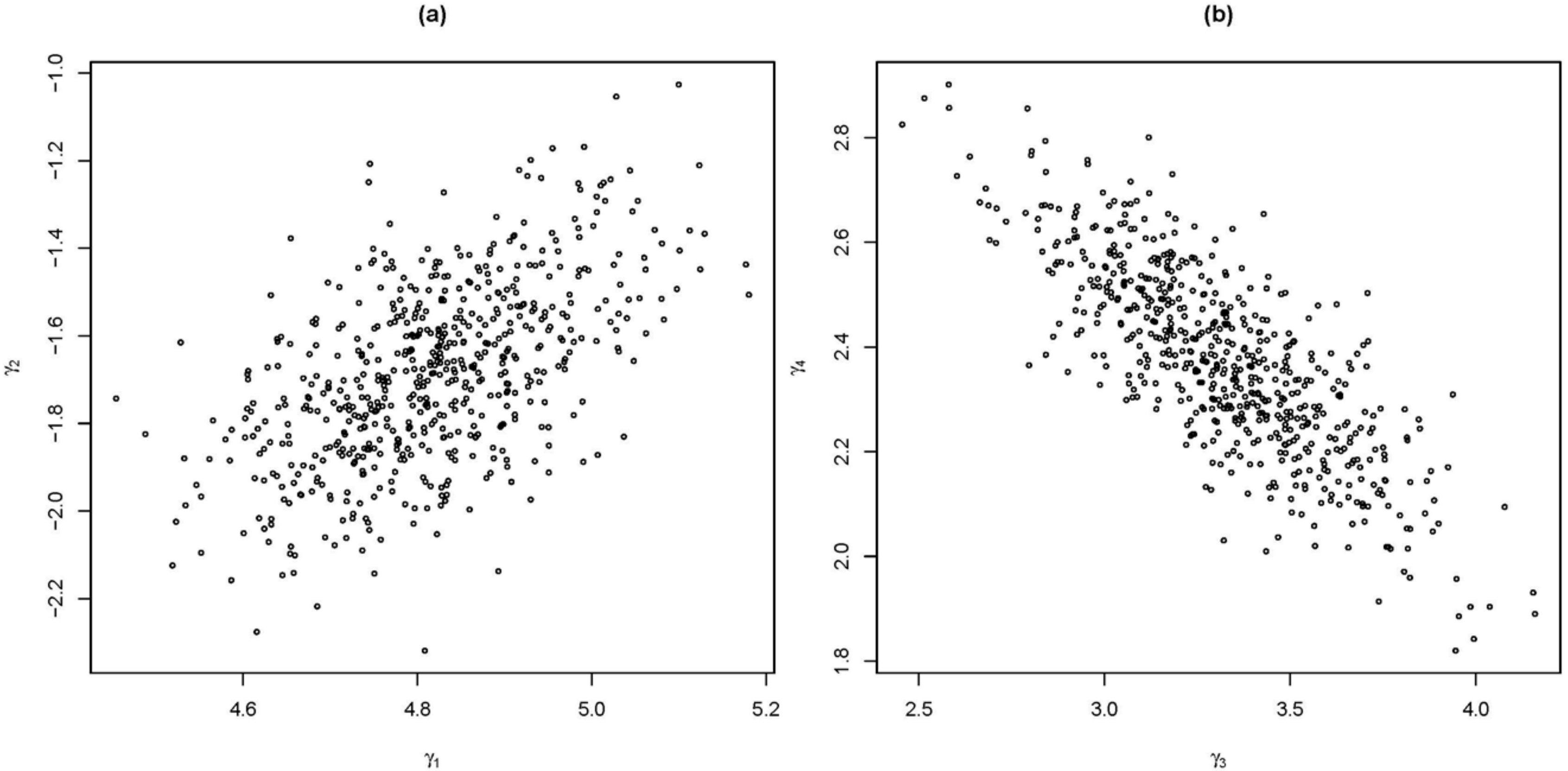
Example-continued

- Posterior mode: $\hat{\gamma} = (4.82, -1.69, 3.32, 2.37)'$

$$\hat{\boldsymbol{\alpha}} = (-0.003, 0.005, -0.008, 0.014, -0.007, -0.007, 0.011, -0.005)'$$

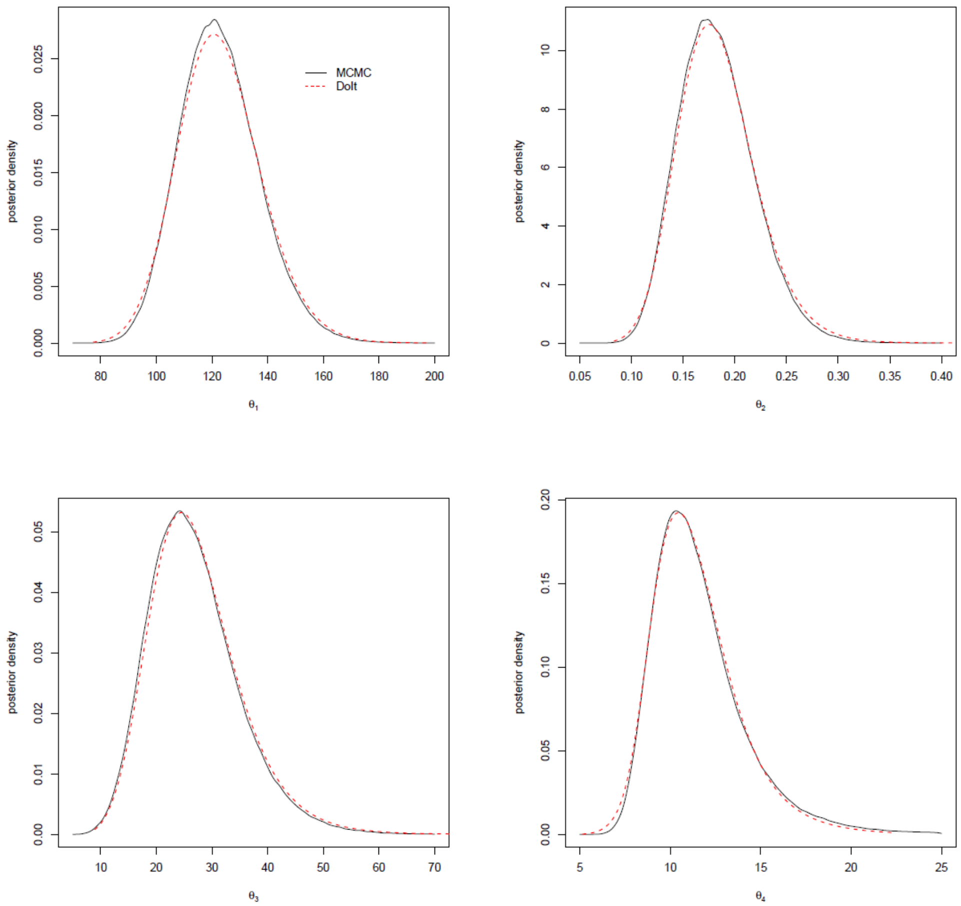
- $\boldsymbol{\Sigma}$ obtained through numerical differentiation.
- $m = 50 \times 12 = 600$
- MmLHD
- $$\boldsymbol{D} = (\hat{\boldsymbol{\theta}} + \boldsymbol{\Sigma}^{1/2}\Phi^{-1}(\boldsymbol{\nu}_1^*), \cdots, \hat{\boldsymbol{\theta}} + \boldsymbol{\Sigma}^{1/2}\Phi^{-1}(\boldsymbol{\nu}_m^*))'$$

Example-continued



Example-continued

- Metropolis algorithm with 3,000,000 samples (black line).

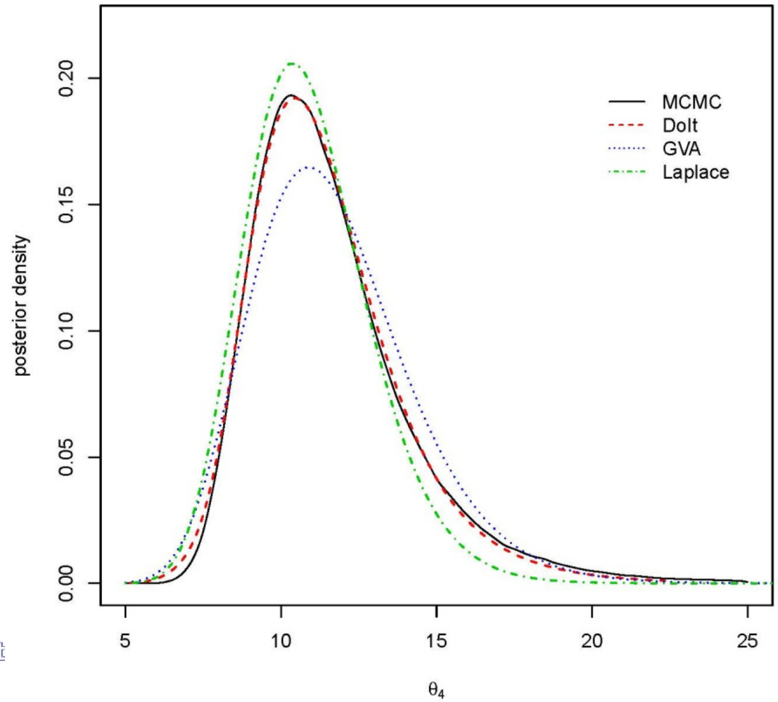


Georgia Institute of Technology

ISyE INDUSTRIAL & SYSTEMS ENGINEERING

Comparison

- Quadrature method: didn't converge!
- Gaussian Variational Approximation (GVA) (Ormerod and Wand 2012)



Sequential Design

- Add points one-by-one to improve approximation.
- Optimal design theory: add the new point at a location with the largest prediction uncertainty.
- **DoIt** can be viewed as a simple kriging predictor given $\hat{\boldsymbol{c}}'\boldsymbol{g}(\boldsymbol{\theta};\boldsymbol{\Sigma})$.
- Conditional prediction variance is

$$\left(\hat{\boldsymbol{c}}'\boldsymbol{g}(\boldsymbol{\theta};\boldsymbol{\Sigma})\right)^2\{1-\boldsymbol{g}(\boldsymbol{\theta};\boldsymbol{\Lambda})'\boldsymbol{G}^{-1}(\boldsymbol{\Lambda})\boldsymbol{g}(\boldsymbol{\theta};\boldsymbol{\Lambda})\}$$

$$\boldsymbol{\nu}_{m+1} = arg\max_{\boldsymbol{\theta}}\left(\hat{\boldsymbol{c}}'\boldsymbol{g}(\boldsymbol{\theta};\boldsymbol{\Sigma})\right)^2\{1-\boldsymbol{g}(\boldsymbol{\theta};\boldsymbol{\Lambda})'\boldsymbol{G}^{-1}(\boldsymbol{\Lambda})\boldsymbol{g}(\boldsymbol{\theta};\boldsymbol{\Lambda})\}$$

Sequential Design-continued

- Let $v_{(i)}$ be the leave-one-out estimate of the prediction variance

$$v_{(i)} \approx \left(h_i + l_i - \frac{\boldsymbol{G}_i^{-1}(\boldsymbol{\Sigma})}{\boldsymbol{G}_{ii}^{-1}(\boldsymbol{\Sigma})}(\boldsymbol{h} + \boldsymbol{l}) \right)^2 \frac{1}{\boldsymbol{G}_{ii}^{-1}(\boldsymbol{\Lambda})}$$

for $i = 1, \cdots, m$, where $\boldsymbol{l} = \boldsymbol{G}(\boldsymbol{\Sigma})\hat{\boldsymbol{c}} - \boldsymbol{h}$

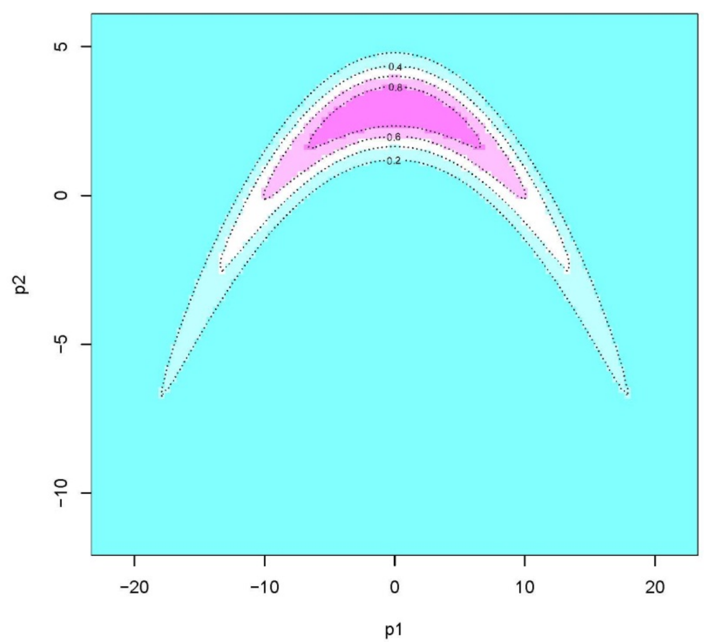
Let $i^* = arg\,\max_i v_{(i)}$.

- Optimize in the neighborhood of $\boldsymbol{\nu}_{i^*}$

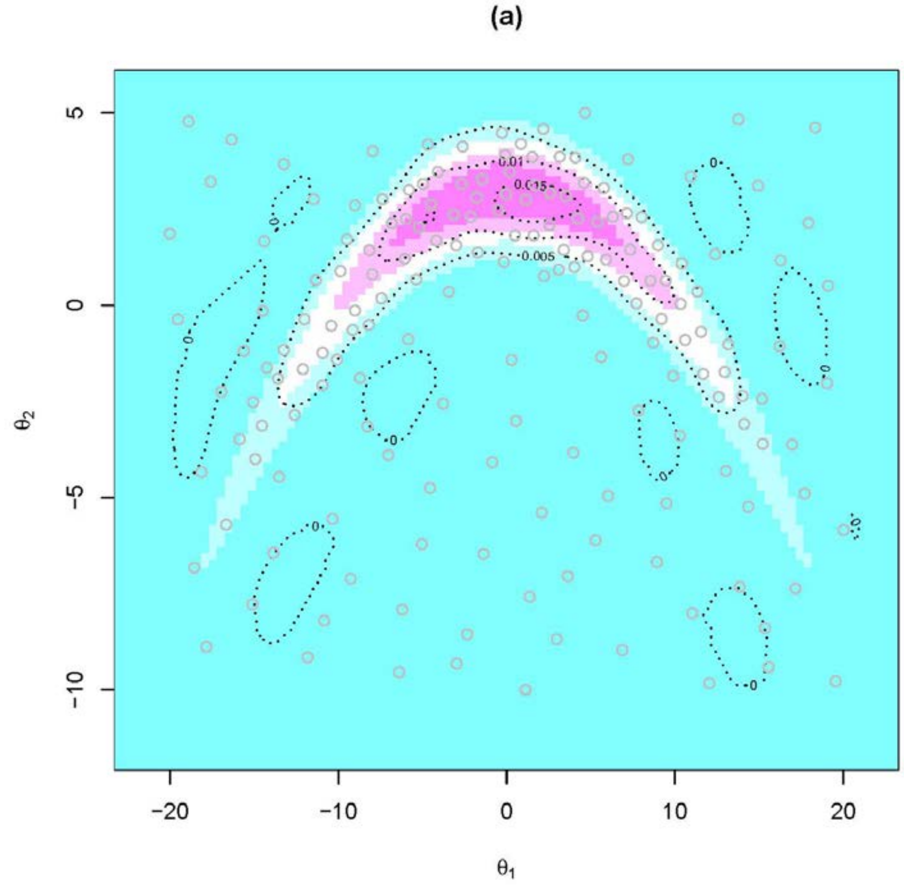
Example

- Haario, Saksman, and Tamminen (2001)

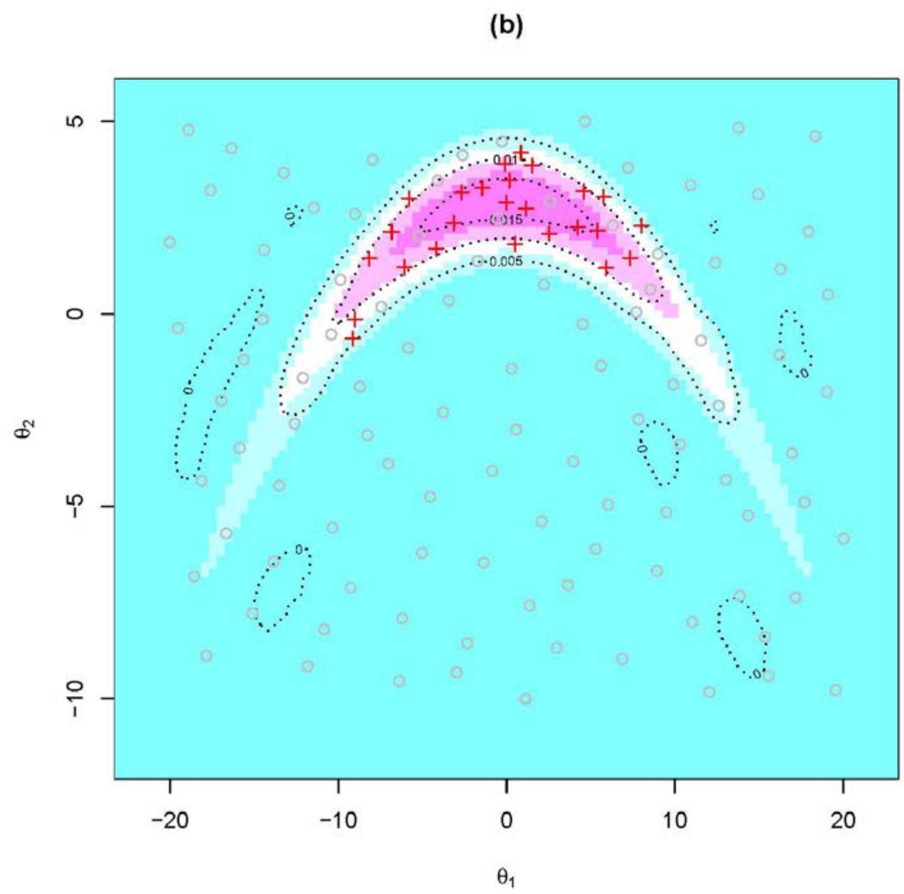
$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \phi\left((\theta_1, \theta_2 + .03\theta_1^2 - 3)'; (0,0)', diag\{100, 1\}\right)$$



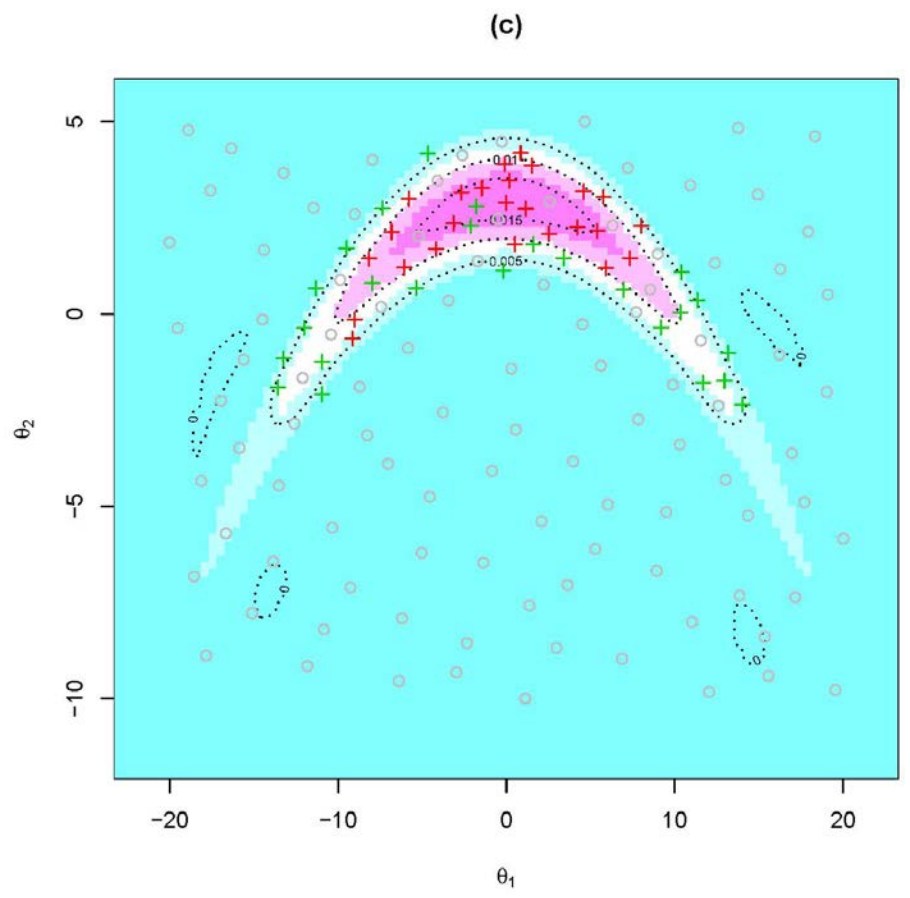
m=100



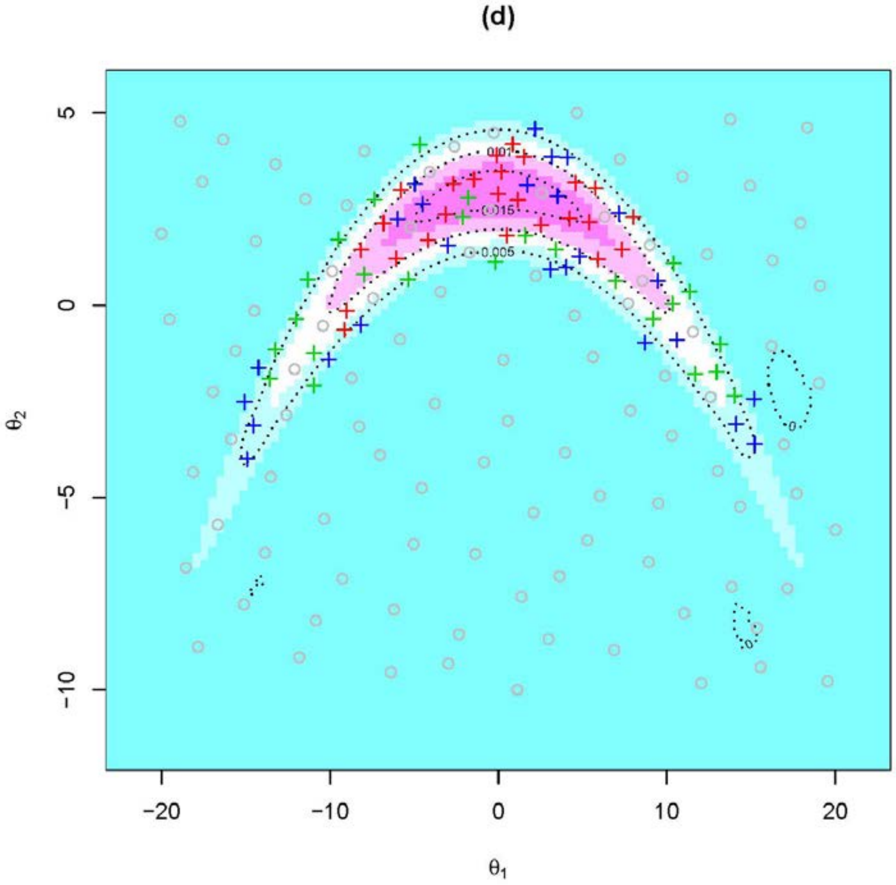
m=100+25



m=100+50



m=100+75

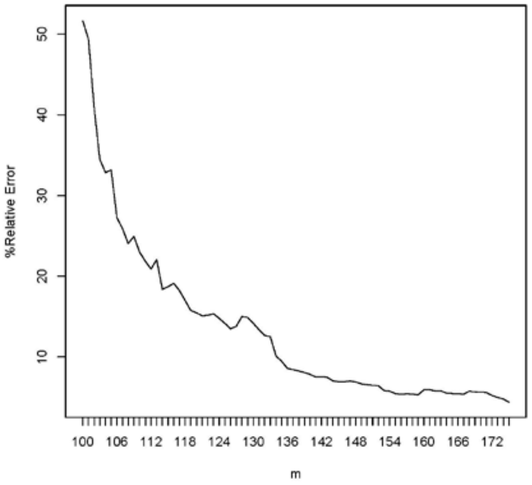


Example-continued

- % Relative Error: $\%RE = \dfrac{\overline{|cv|}}{\bar{h}} \times 100,$

where $\overline{|cv|} = E(|cv(\boldsymbol{\theta})|\ |\boldsymbol{y})$ and $\bar{h} = E(h(\boldsymbol{\theta})|\boldsymbol{y})$

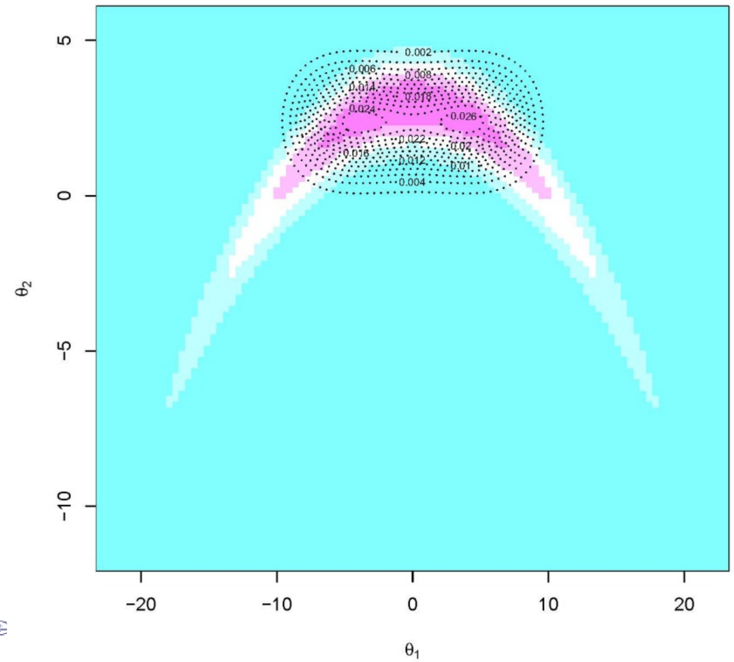
$$cv_i = h_i - \left(h_i + l_i - \frac{\boldsymbol{G}_i^{-1}(\boldsymbol{\Sigma})}{\boldsymbol{G}_{ii}^{-1}(\boldsymbol{\Sigma})}(\boldsymbol{h} + \boldsymbol{l})\right)\left(\frac{h_i}{h_i + l_i} - \frac{\boldsymbol{G}_i^{-1}(\boldsymbol{\Lambda})}{\boldsymbol{G}_{ii}^{-1}(\boldsymbol{\Lambda})}(\frac{\boldsymbol{h}}{\boldsymbol{h} + \boldsymbol{l}} - a\boldsymbol{1})\right)$$



Comparison: Variational Bayes

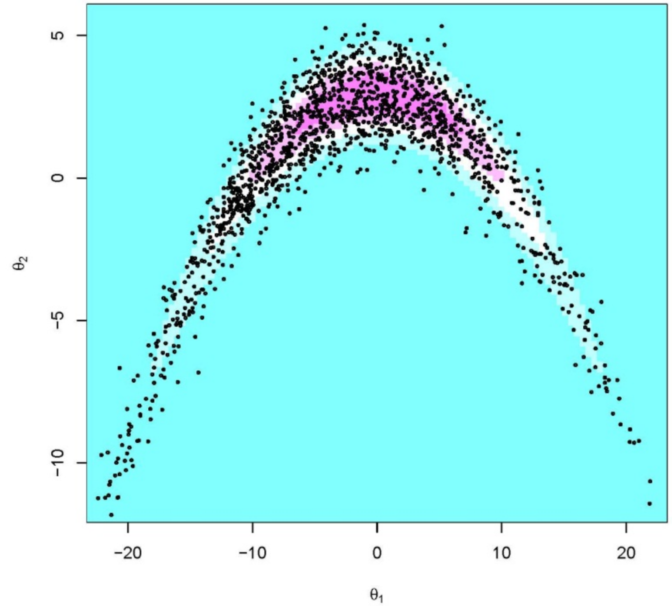
$$\hat{p}_{VB}(\boldsymbol{\theta}|\boldsymbol{y}) = q_1(\theta_1)\phi(\theta_2; \mu_2, \sigma_2^2),$$

$$\text{where } q_1(\theta_1) \propto \exp\{-.5[\theta_1^2/100 + (\mu_2 + .03\theta_1^2 - 3)^2]\},$$
$$\mu_2 = -.03(\mu_1^2 + \sigma_1^2) + 3, \text{ and } \sigma_2^2 = 1.$$



Comparison: Hybrid MCMC

- Fielding, Nott, and Liong (2011)



- CPU time: Hybrid MCMC=90 mins, DoIt=3 mins.

Discussion from Dagupta and Meng: connections to QMC

- Experimental design
  - QMC: low discrepancy sequence in $[0,1]^d$
  - DoIt: initial space-filling design+ sequential design
- Posterior summaries
  - QMC: Monte Carlo average
  - DoIt: smooth interpolation + analytical evaluation
- DoIt is likely to perform better than QMC when the posterior densities are smooth.

Numerical Comparison

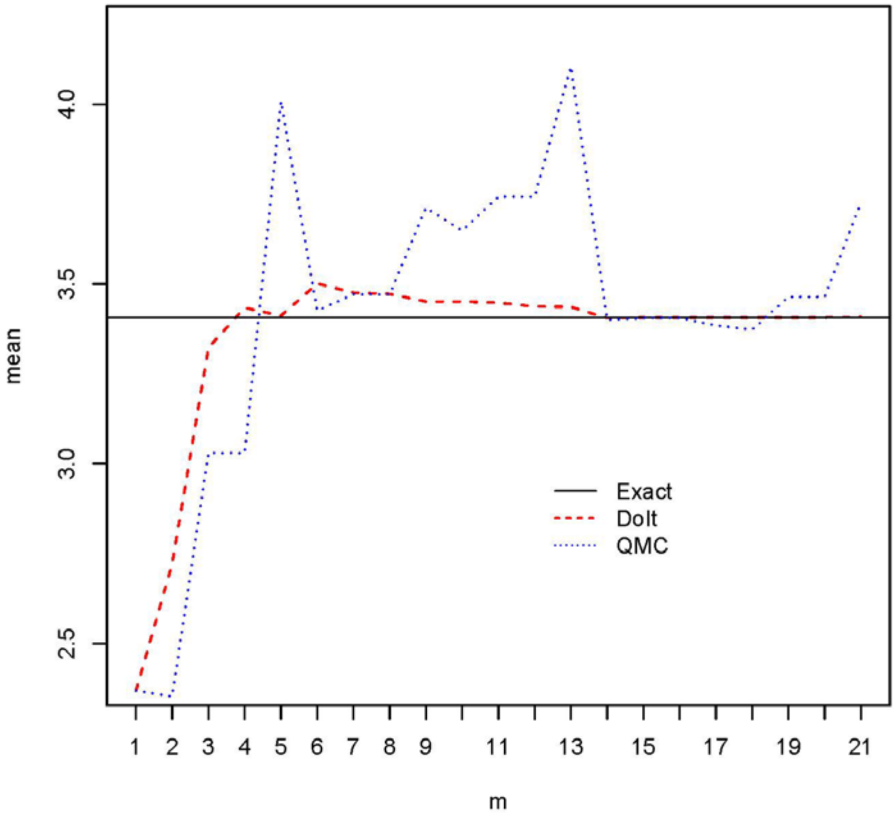
- Binary data example

$$y|\theta \sim Bernoulli(\{1+\exp(-\theta)\}^{-1}),$$
$$\theta \sim N(\mu, \tau^2).$$

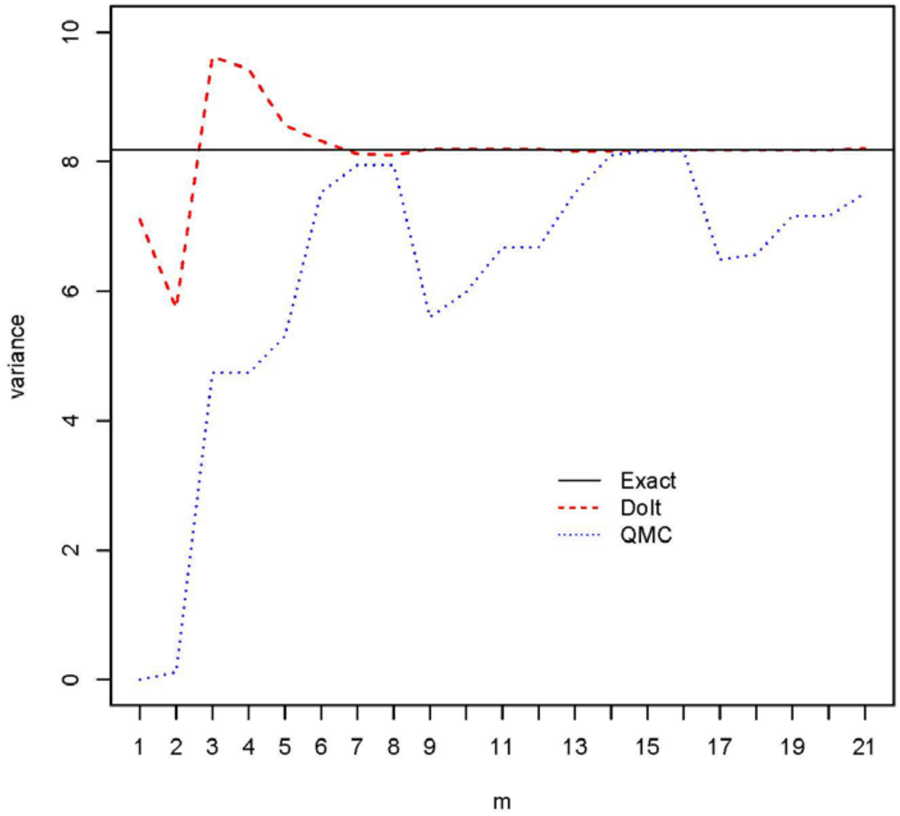
- Laplace approximation: $N(2.37, 2.67^2)$
- van der Corput sequence: ½, ¼, ¾, ….and re-scaled to [2.37-15,2.37+15].

$$\hat{\theta}_{QMC} = E(\theta|y)_{QMC} = \frac{\sum_{i=1}^{m} h_i \nu_i}{\sum_{i=1}^{m} h_i} \text{ and } var(\theta|y)_{QMC} = \frac{\sum_{i=1}^{m} h_i (\nu_i - \hat{\theta}_{QMC})^2}{\sum_{i=1}^{m} h_i}.$$

Posterior Mean



Posterior Variance



Georgia Institute of Technology

ISyE INDUSTRIAL & SYSTEMS ENGINEERING

Hierarchical Models

- May contain very large number of parameters
- Not easy to find a good space-filling design in high dimensions.
  - Make use of the hierarchical structure.

$$\boldsymbol{y}|\boldsymbol{\theta} \sim p(\boldsymbol{y}|\boldsymbol{\theta}),\ \boldsymbol{\theta}|\boldsymbol{\eta} \sim p(\boldsymbol{\theta}|\boldsymbol{\eta}),\ \text{and } \boldsymbol{\eta} \sim p(\boldsymbol{\eta}).$$

Hierarchical Models-continued

- Suppose we can obtain explicit expression of

$$p(\boldsymbol{y}|\boldsymbol{\eta}) = \int p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\eta})\ d\boldsymbol{\theta},$$

- And have the conditional distribution

$$p(\boldsymbol{\theta}|\boldsymbol{\eta},\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\eta})$$

- Use **DoIt** : $\hat{p}(\boldsymbol{\eta}|\boldsymbol{y}) \approx \dfrac{\hat{\boldsymbol{c}}'\phi(\boldsymbol{\eta};\boldsymbol{\Sigma})}{\hat{\boldsymbol{c}}'\mathbf{1}}\{1 + \hat{\boldsymbol{b}}'\boldsymbol{g}(\boldsymbol{\eta};\boldsymbol{\Lambda})/a\}$

- Then, $\hat{p}(\boldsymbol{\theta}|\boldsymbol{y}) \approx \dfrac{\hat{\boldsymbol{c}}'\boldsymbol{G}(\boldsymbol{\Sigma}+\boldsymbol{\Lambda})\boldsymbol{G}(\boldsymbol{\Lambda})^{-1}\boldsymbol{p}^*(\boldsymbol{\theta})}{\hat{\boldsymbol{c}}'\boldsymbol{G}(\boldsymbol{\Sigma}+\boldsymbol{\Lambda})\boldsymbol{G}(\boldsymbol{\Lambda})^{-1}\boldsymbol{z}}$

where $\boldsymbol{p}^*(\boldsymbol{\theta}) = (p(\boldsymbol{\theta}|\boldsymbol{\nu}_1,\boldsymbol{y}),\cdots,p(\boldsymbol{\theta}|\boldsymbol{\nu}_m,\boldsymbol{y}))' \odot \boldsymbol{z}$

Example 1: A Longitudinal Data Analysis

- Orthodontic measurements on 27 children (Phinheiro and Bates 2000)

$$
\begin{aligned}
y_{ij}|\boldsymbol{\beta}, u_i, \sigma_\epsilon^2 &\sim^{ind.} N(\beta_0 + u_i + \beta_1 age_{ij} + \beta_2 sex_i, \sigma_\epsilon^2), \\
u_i|\sigma_u^2 &\sim^{iid} N(0, \sigma_u^2), \\
\boldsymbol{\beta} &\sim N(\mathbf{0}, 10^8 \boldsymbol{I}_3), \\
\sigma_\epsilon^2, \sigma_u^2 &\propto^{ind.} IG(.01, .01),
\end{aligned}
$$

for $i = 1, \cdots, 27$ and $j = 1, \cdots, 4.$

- **32 parameters**

Georgia Institute of Technology

ISYE INDUSTRIAL & SYSTEMS ENGINEERING

Longitudinal Data-continued

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\epsilon}$$

- Integrating out u

$$\boldsymbol{y}|\boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_u^2 \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma_\epsilon^2 \boldsymbol{I}_{108} + \sigma_u^2 \boldsymbol{Z}\boldsymbol{Z}')$$

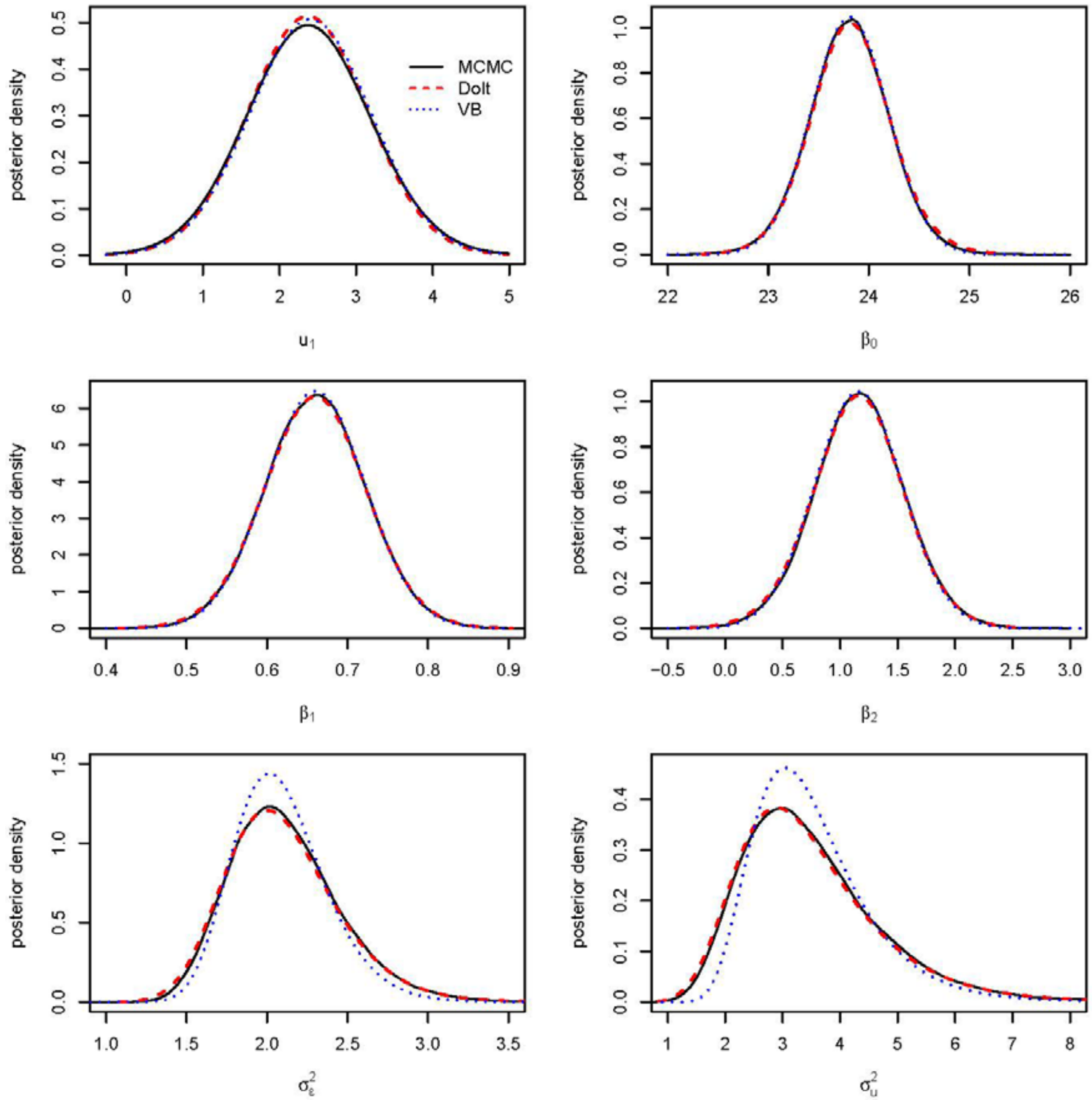
- Also,

$$\boldsymbol{u}|\beta, \sigma_\epsilon^2, \sigma_u^2, \boldsymbol{y} \sim N((\boldsymbol{Z}'\boldsymbol{Z} + \frac{\sigma_\epsilon^2}{\sigma_u^2}\boldsymbol{I}_{27})^{-1}\boldsymbol{Z}'(\boldsymbol{y} - \boldsymbol{X}\beta), (\boldsymbol{Z}'\boldsymbol{Z} + \frac{\sigma_\epsilon^2}{\sigma_u^2}\boldsymbol{I}_{27})^{-1})$$

- **DoIt** : 250-run MmLHD in **5 dimensions**

Georgia Institute of Technology

ISyE INDUSTRIAL & SYSTEMS ENGINEERING

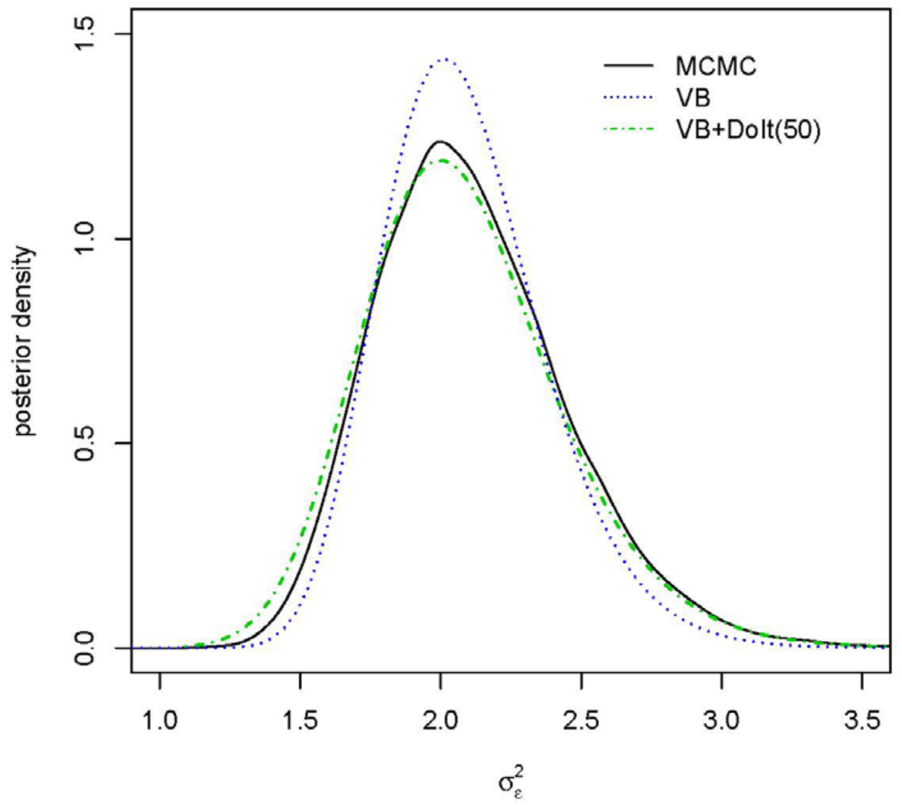


MCMC
DoIt
VB
posterior density
u_1
β_0
β_1
β_2
σ_ε^2
σ_u^2

Discussion from Ormerod & Wand

- VB approximation can be improved using the grid-based variational approximation method in Ormerod (2011).
- DoIt can be used for the same purpose!
  - Center the experimental design using VB estimates.

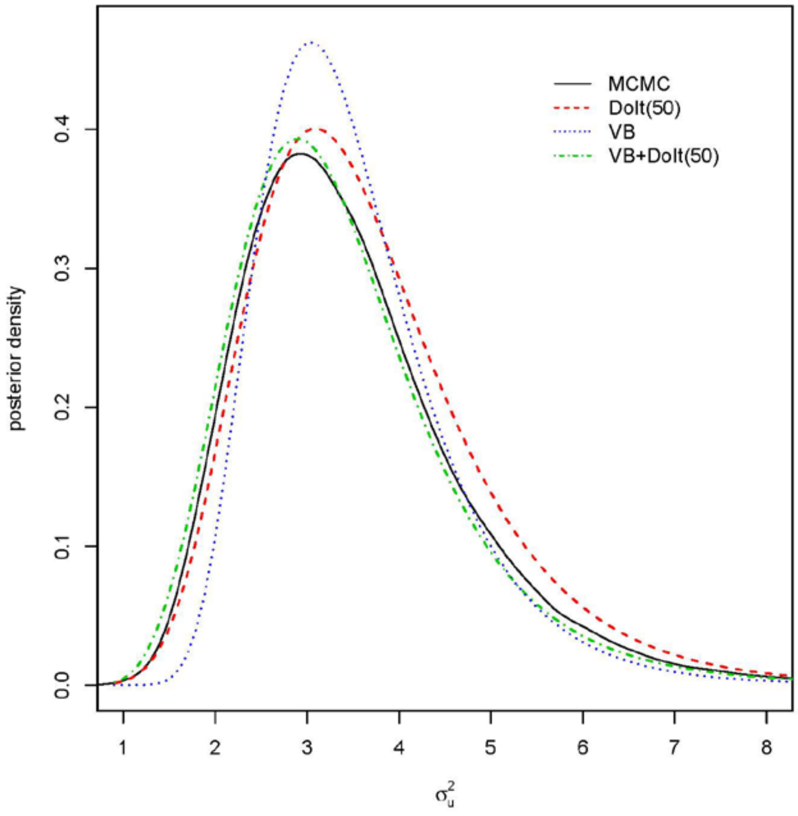
VB+DoIt



MCMC
VB
VB+DoIt(50)
posterior density
0.0
0.5
1.0
1.5
1.0
1.5
2.0
2.5
3.0
3.5
σ_ε^2

VB+DoIt

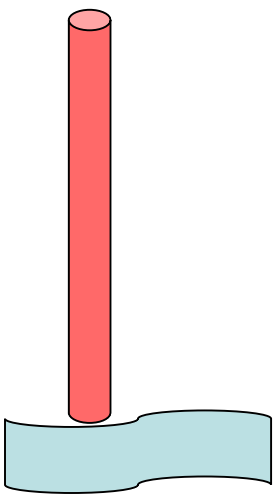
- VB can be used to improve DoIt, whenever VB implementation is readily available.



Example 2: LAMM

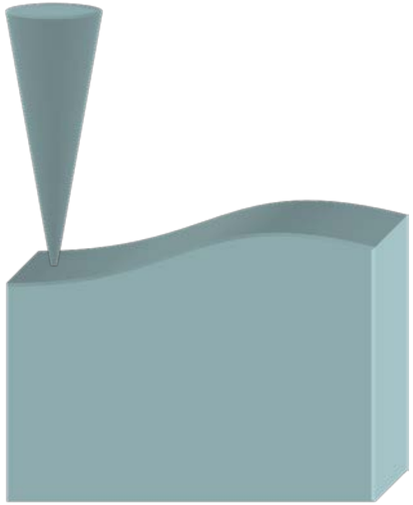
Laser assisted mechanical micromachining (LAMM) integrates *thermal softening* with *mechanical micro cutting*

(Singh, Joseph, and Melkote 2011)



Laser heating

+



Mechanical micromachining

= LAMM

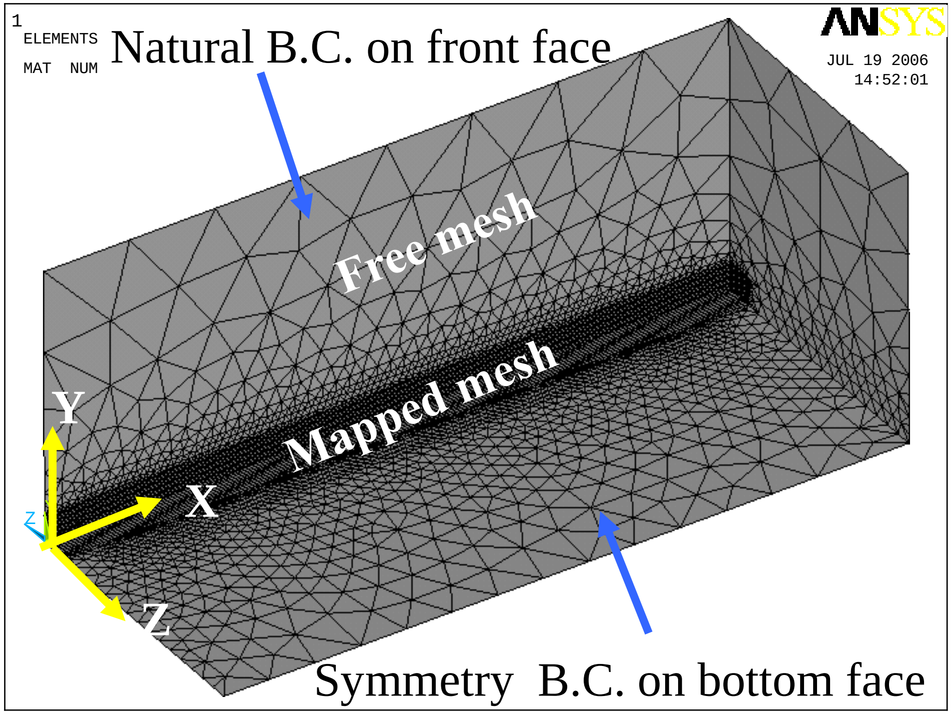
Georgia Institute of Technology

ISyE INDUSTRIAL & SYSTEMS ENGINEERING

Objective

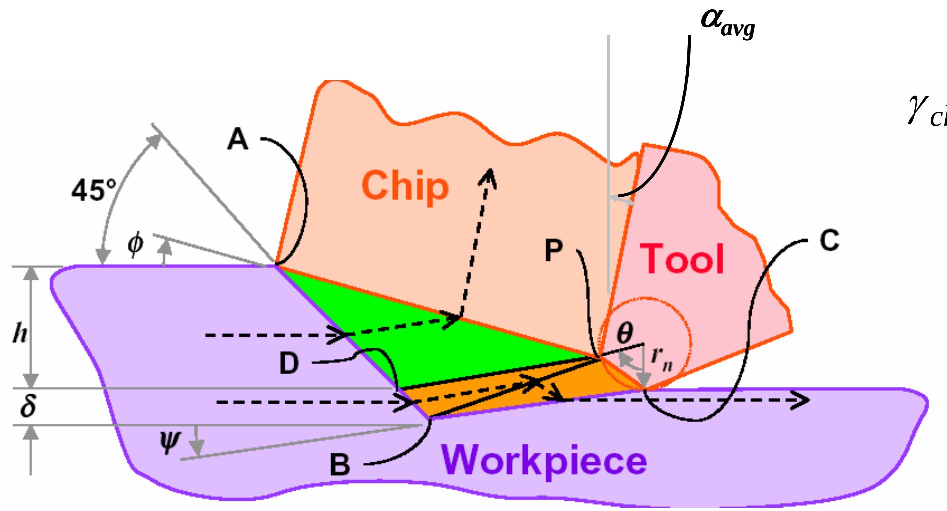
Find optimum processing conditions that minimize cutting/thrust forces and thermal damage.

Thermal Model



- Mapped dense mesh (25 μm x 12.5 μm x 20μm)
- An 8 noded 3-D thermal element (Solid70) is used
- Gaussian distribution of heat flux applied to a 5x5 element matrix which sweeps the mesh on the front face

Geometric Model



(Manjunathiah et. al, 2000)

$$\gamma_{chip} = \frac{\sqrt{2}\,\sin\theta_{PD}}{\sin(\pi/4+\theta_{PD})} + \frac{\cos(\alpha_{avg}+\theta_{PD})}{\cos(\alpha_{avg}-\phi)\sin(\phi+\theta_{PD})}$$

$$\gamma_{work} = \frac{\sqrt{2}\sin\theta_{PD}}{\sin(\pi/4+\theta_{PD})} + \frac{\sin(\theta_{PD}+\theta/2)}{\sin(\theta_{PB}+\theta/2)\sin(\theta_{PB}+\theta_{PD})} + \frac{\sin\theta/2}{\sin\psi\sin(\psi+\theta/2)}$$

$$\dot{\gamma}_{chip} = 2V\frac{\gamma_{chip}}{\sqrt{2}\,\sin(\pi/4+\theta_{PD})\overline{PD}}$$

$$\dot{\gamma}_{work} = 2V\frac{\gamma_{work}}{\sqrt{2}\sin(\pi/4+\theta_{PD})\overline{PD} + \frac{\sin(\psi+\theta/2)}{\sin\psi}\overline{PC}}$$

$$\gamma_{eff} = \frac{v_{chip}\gamma_{chip} + v_{work}\gamma_{work}}{v_{chip}+v_{work}}$$

$$\dot{\gamma}_{eff} = \frac{v_{chip}\dot{\gamma}_{chip} + v_{work}\dot{\gamma}_{work}}{v_{chip}+v_{work}}$$

For plane strain conditions,

$$\varepsilon = \gamma_{eff}/\sqrt{3}$$

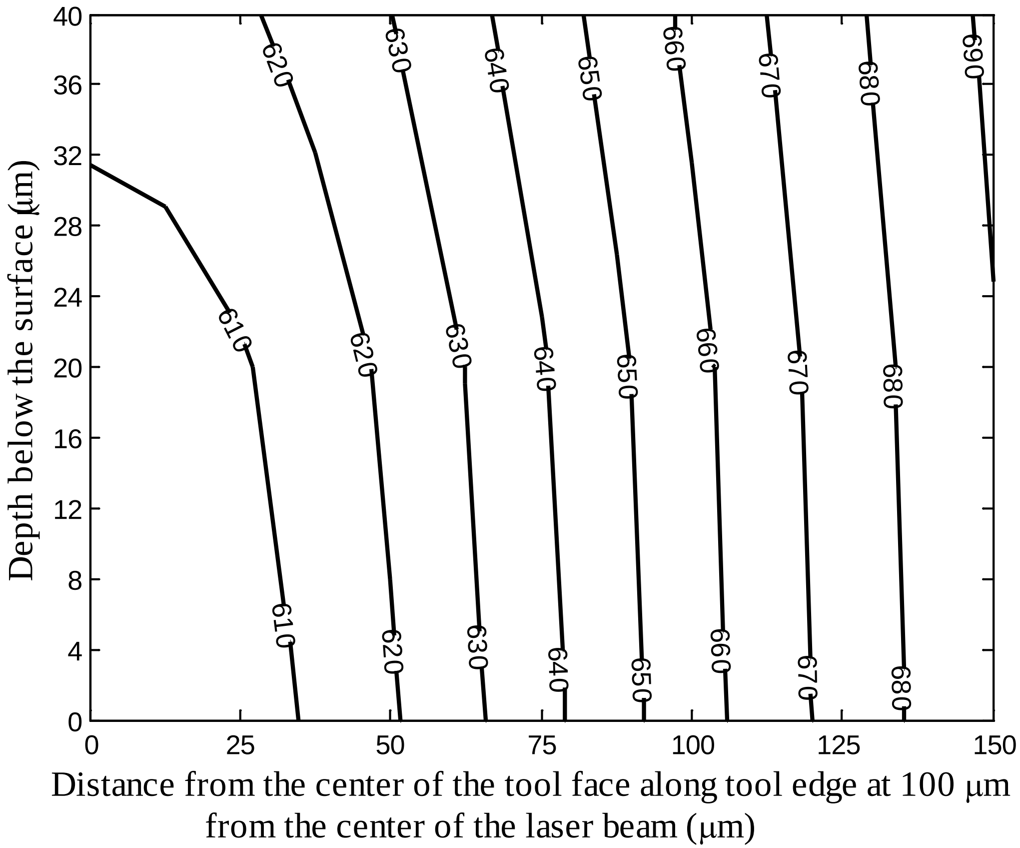
$$\dot{\varepsilon} = \dot{\gamma}_{eff}/\sqrt{3}$$

Shear Flow Strength

$$\sigma(\varepsilon,\dot{\varepsilon},T,HRC)=\left(A+B\varepsilon^{n}+C\,ln(\varepsilon+\varepsilon_{0})+D\right)\left(1+E\,ln\left(\frac{\dot{\varepsilon}}{\dot{\varepsilon}_{0}}\right)\right)\left(1-\left(T^{*}\right)^{m}\right)$$

Yan et al., 2007

$$S=\sigma/\sqrt{3}$$



10W laser power, 10 mm/min speed 100 μm laser-tool distance and 110 μm spot size

Georgia Institute of Technology

ISyE INDUSTRIAL & SYSTEMS ENGINEERING

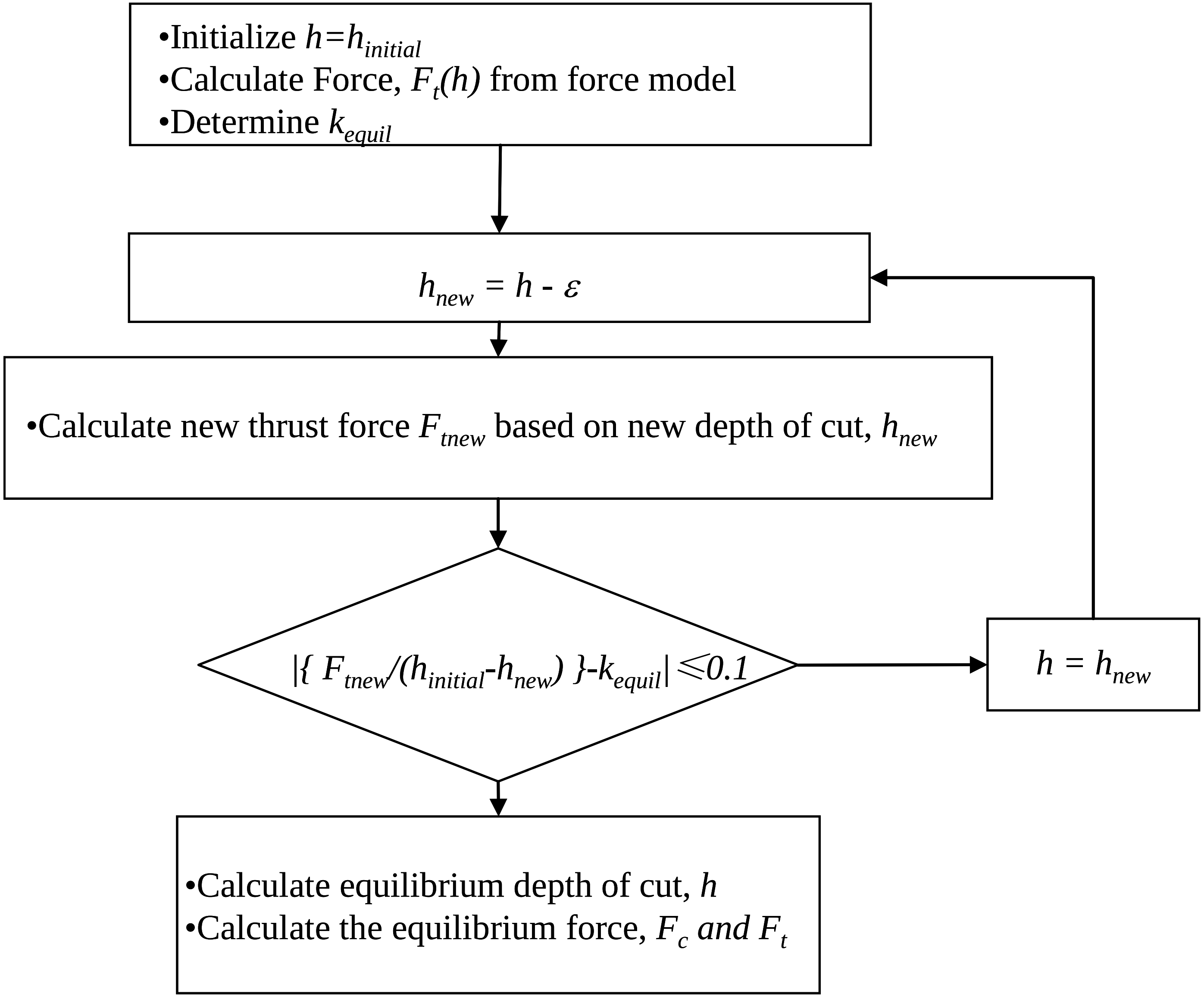
Forces

- Cutting and thrust forces,

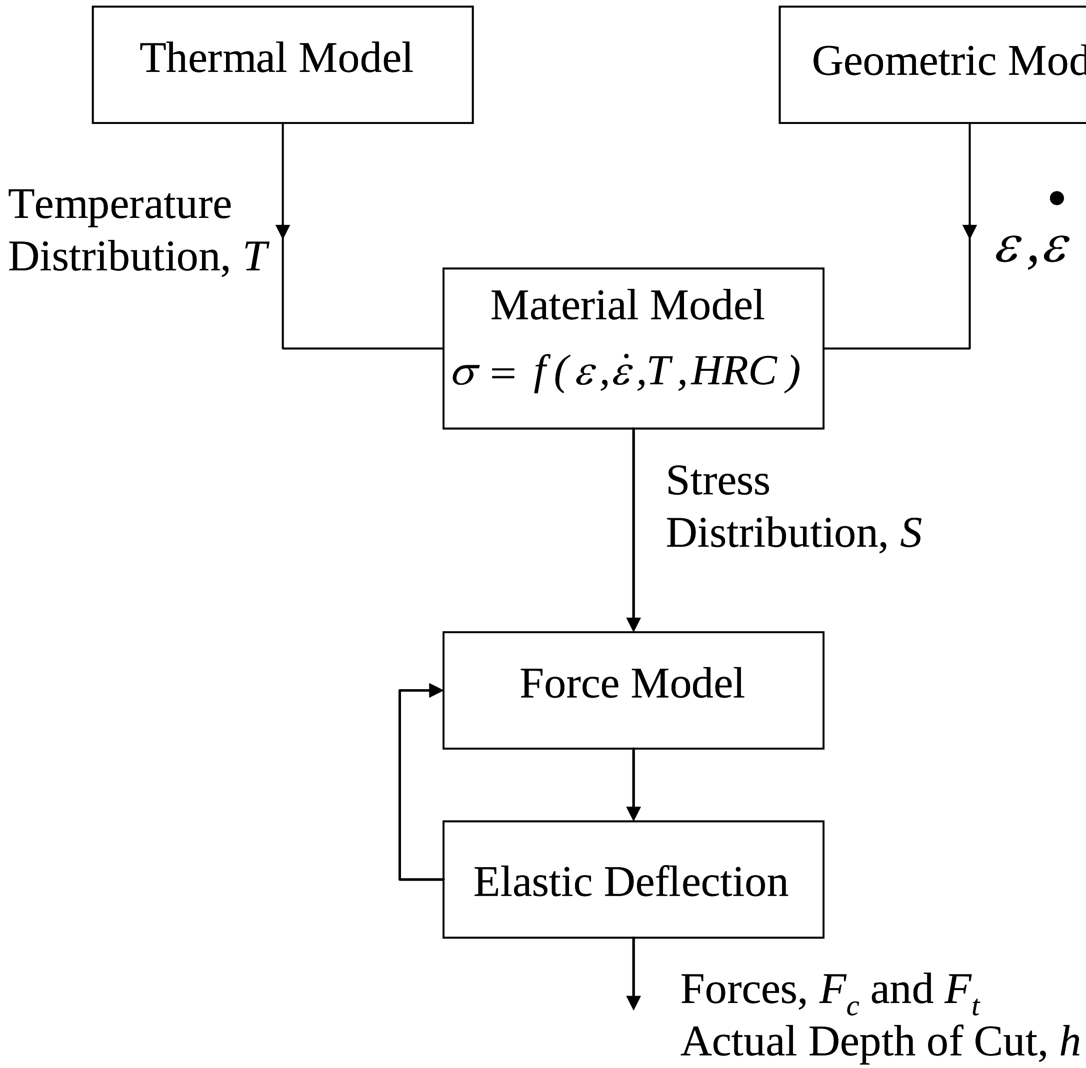
$$F_c = \{(h - p)\cot\phi + h + r_n \sin\theta - (k-1)\delta\}\sum_{i=1}^{n}\overline{S}(i)w(i)$$

$$F_t = \{(h - p)\cot\phi - h + r_n \sin\theta + (k-1)\delta\cot\psi\}\sum_{i=1}^{n}\overline{S}(i)w(i)$$

Equilibrium Forces/Deflection



Force model


Thermal Model
Geometric Model
Temperature Distribution, T
$\varepsilon, \dot{\varepsilon}$
Material Model
$\sigma = f(\varepsilon, \dot{\varepsilon}, T, HRC)$
Stress Distribution, S
Force Model
Elastic Deflection
Forces, F_c and F_t
Actual Depth of Cut, h

LAMM-continued

$$
\begin{aligned}
y_i &= \theta(\boldsymbol{x}_i) \\
\theta(\boldsymbol{x}) &\sim GP(\mu, \tau^2 r),\ r(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\{-\sum_{k=1}^{4} \alpha_k (x_{ik} - x_{jk})^2\} \\
p(\mu, \tau^2) &\propto 1/\tau^2 \\
\gamma_i &= \log(\alpha_i) \sim^{iid} N(0, 1).
\end{aligned}
$$

LAMM-continued

- Unnormalized posterior

$$
\begin{aligned}
h(\theta(\boldsymbol{x}), \mu, \tau^2, \gamma) &\propto p(\boldsymbol{y}|\theta(\boldsymbol{x}), \mu, \tau^2, \gamma)p(\theta(\boldsymbol{x})|\mu, \tau^2, \gamma)p(\mu, \tau^2)p(\boldsymbol{\gamma}), \\
&= p(\theta(\boldsymbol{x})|\mu, \tau^2, \boldsymbol{\gamma}, \boldsymbol{y})p(\boldsymbol{y}|\mu, \tau^2, \gamma)p(\mu, \tau^2)p(\boldsymbol{\gamma})
\end{aligned}
$$

- Integrating out $\theta(\boldsymbol{x}), \mu, \text{ and } \tau^2,$

$$
h(\boldsymbol{\gamma}) = |\boldsymbol{R}|^{-1/2}(\mathbf{1}'\boldsymbol{R}^{-1}\mathbf{1})^{-1/2}[(\boldsymbol{y} - \hat{\mu}\mathbf{1})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \hat{\mu}\mathbf{1})]^{-(n-1)/2}
$$

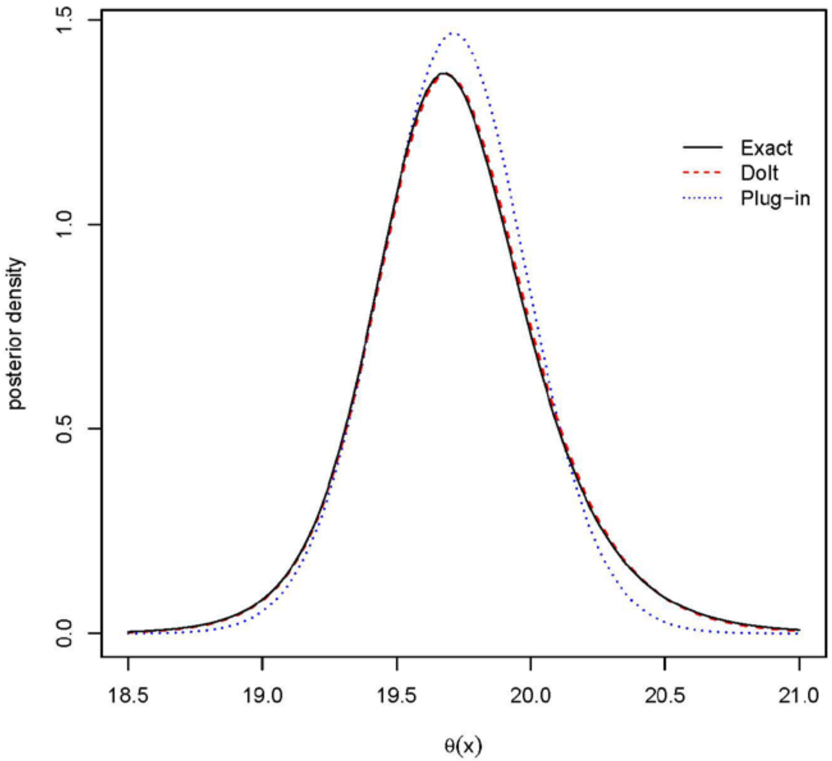
LAMM-continued

- Conditional distribution: $\dfrac{\theta(\boldsymbol{x}) - \hat{\theta}(\boldsymbol{x})}{\sqrt{V(\boldsymbol{x})}} | \gamma, \boldsymbol{y} \sim t_{n-1},$

$$\begin{aligned}
\hat{\theta}(\boldsymbol{x}) &= \hat{\mu} + \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \hat{\mu}\mathbf{1}), \\
V(\boldsymbol{x}) &= \hat{\tau}^2 \left(1 - \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x}) + \frac{\{1 - \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\mathbf{1}\}^2}{\mathbf{1}'\boldsymbol{R}^{-1}\mathbf{1}} \right), \\
\hat{\mu} &= \frac{\mathbf{1}'\boldsymbol{R}^{-1}\boldsymbol{y}}{\mathbf{1}'\boldsymbol{R}^{-1}\mathbf{1}}, \\
\hat{\tau}^2 &= \frac{1}{n-1}(\boldsymbol{y} - \hat{\mu}\mathbf{1})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \hat{\mu}\mathbf{1}).
\end{aligned}$$

LAMM-continued

- **DoIt** : m=100 (red), MCMC: m=100,000 (black)



LAMM-continued

- Computational time
  - **DoIt** = 3 seconds
  - MCMC=10 minutes

200 times faster!

- Computationally expensive likelihood.
  - Computational complexity: $O(n^3)$.
  - Here *n*=48.

Conclusions

- A new deterministic approximation method using **design of experiments** and **interpolation** techniques.
  - Very general.
  - Can obtain the results with arbitrary precision.
  - Suffer less from the curse of dimensionality compared to lattice-based quadrature methods.
  - Some of the very large hierarchical problems can be solved efficiently.
  - Very fast!
  - Almost a black box method (no extra programming or derivations are needed).

Conclusions

- Disadvantages
  - Not as flexible as MCMC.
  - Not easy to find good space-filling designs in high dimensions.
  - Can handle only continuous parameters.
  - Cannot ensure nonnegativity of the posterior density.

Nonnegative DoIt

- Joseph, *Technometrics*, 2013, February.

Let $\hat{\boldsymbol{c}} = \boldsymbol{b}$, $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}$, and $a = 0$.

$$h(\boldsymbol{\theta}) \approx \left\{ \sum_{i=1}^{m} b_i g(\boldsymbol{\theta}; \boldsymbol{\nu}_i, \boldsymbol{\Sigma}) \right\}^2,$$

which will always be nonnegative!

$$\sqrt{h(\boldsymbol{\theta})} \approx \sum_{i=1}^{m} b_i g(\boldsymbol{\theta}; \boldsymbol{\nu}_i, \boldsymbol{\Sigma})$$

Future Research

- Fast generation of design points at high probability regions.
  - Need a better design strategy.

- Fast approximation using local versions of Σ .
  - Need a better modeling strategy.

- Topics for future research!

Conclusions

If you have a Bayesian problem, then just

Do It !