

BAYESIAN METHODS FOR DISCRETE EVENT SIMULATION

Stephen E. Chick¹

¹Technology and Operations Management
INSEAD
Fontainebleau, France

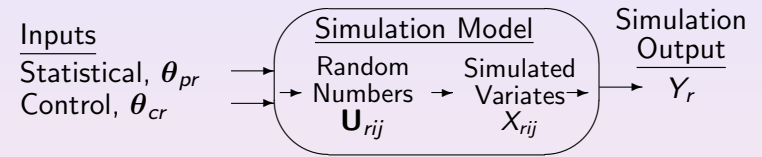
2004 Winter Simulation Conference

INSEAD

Chick

Bayesian Methods for Simulation

Why Bayesian methods in Simulation?



$$Y_r = g(\theta_p, \theta_c; \mathbf{U}_r)$$

- Example: Single Server Queue (M/M/1): $\theta_p = (\lambda, \mu_i) =$ arrival and service rates (server $i = 1, 2$)
- Output: $Y \approx \lambda / (\mu_i - \lambda) + \text{noise}$
- Simulation: Analyze stochastic processes via sample path generation. Inform decisions: pick design parameter $\theta_c = i$, value $h(E[Y | \theta, \theta_c])$
- Bayesian as alternative to frequentist

INSEAD

Chick

Bayesian Methods for Simulation

Why Bayesian methods?

- Glynn (1986): *Uncertainty analysis*. Not $\alpha = h(E[Y])$, but

$$\alpha(\theta) = h(E[Y | \theta])$$

- Unknown parameters, $p(\theta)$, data from modeled system to update
 - 1 Mean $E[\alpha(\Theta)]$
 - 2 Distribution of $\alpha(\Theta)$ induced by Θ
 - 3 Credible set: θ_{lo}, θ_{hi} so $p([h(\theta_{lo}), h(\theta_{hi})]) = 95\%$

- Chick (1997): Reviewed work to that date.
- Suggested broader range of application.
 - 1 *Ranking and selection*
 - 2 *Response surface modeling*
 - 3 *Experimental design*

INSEAD

Chick

Bayesian Methods for Simulation

The Point of Today

- Review some basic concepts of subjective probability, Bayesian statistics, decision theory.
- Identify several applications to simulation experiments.
- Summarize some implementation issues.
- Identify some areas for future work.

INSEAD

Chick

Bayesian Methods for Simulation

Related work

See the WSC paper for a long (but incomplete) citation list for work over the last 8 years on:

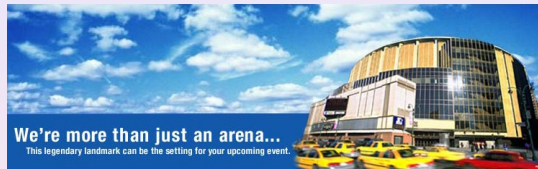
- Formal Bayes or decision theoretic theory
- Applications: scheduling, insurance, finance, traffic modeling, public health, waterway safety, supply chain and other areas
- Bayes and deterministic simulations
- Favorite books on subjective and Bayesian probability and decision theory

Health Economics: increasingly uses simulation (in addition to decision trees, Markov chains), and *increasingly* requires *probabilistic sensitivity analysis*.

Outline

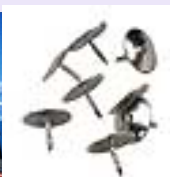
- 1 Getting down to brass tacks
 - Subjective and Bayesian methods
 - Assessing prior probability
 - Asymptotic Theorems
 - Decisions, loss, and value of information
 - Entropy and Kullback-Leibler Discrepancy
- 2 Applications
 - Uncertainty Analysis
 - Selecting from Multiple Candidate Distributions
 - Selecting the Best System
 - Metamodels
- 3 Implementation
- 4 Summary

Getting down to brass tacks



Probability of 7 heads in the first 10 flips?
How to approach the problem...

Getting down to brass tacks

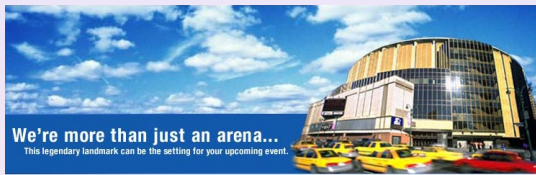


Probability of 7 heads in the first 10 flips?
Comte d'Alembert (18th cent.)

- Indifference says maybe $1/11$?
- But wait, for one flip, probability of heads is $1/2$?

See Savage (1972) and Kreps (1988).

Getting down to brass tacks



Probability of 7 heads in the first 10 flips?

Dwight (an unreconstructed frequentist)

- $\frac{10!}{7!3!}\theta^7(1-\theta)^3$, where $\theta = \lim_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n}$ (a.e.).
- If we rent Madison Square garden and flip the tack repeatedly, I can estimate θ for you. What confidence and how accurately do you need to know θ ?

Hmmm...

Getting down to brass tacks



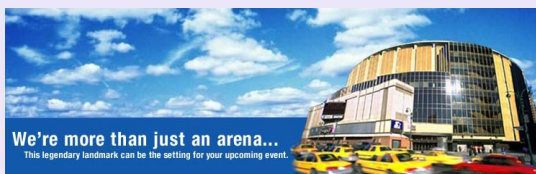
Probability of 7 heads in the first 10 flips?

Dwight (an unreconstructed frequentist)

- $\frac{10!}{7!3!}\theta^7(1-\theta)^3$, where $\theta = \lim_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n}$ (a.e.).
- If we rent Madison Square garden and flip the tack repeatedly, I can estimate θ for you. What confidence and how accurately do you need to know θ ?

Let's reformulate the question ...

Why am I a Bayesian?



Will you accept the following bet *now*? You get \$100 if there are 7 heads, but you pay \$5 if not.

more from Dwight

- I can't answer until I have a good idea of what θ is.
- Guessing wouldn't be scientific.

Why am I a Bayesian?



Ralph

Probability of 7 heads in the first 10 flips?

- I'm willing to use probability for personal judgments
- $\int_0^1 \frac{10!}{7!3!}\theta^7(1-\theta)^3\pi(\theta)d\theta$, where $\pi(\theta)$ is a *prior probability*.
- I'll update with *Bayes' rule*, to get *posterior probability*

$$p(\theta | \mathbf{x}_n) = \frac{\pi(\theta)p(\mathbf{x}_n | \theta)}{p(\mathbf{x}_n)} = \frac{\pi(\theta) \prod_{i=1}^n p(x_i | \theta)}{\int p(\mathbf{x}_n | \theta)d\pi(\theta)}$$

Why am I a Bayesian?

Lenny

Probability of 7 heads in the first 10 flips?

- Fair bets: I set $p(E_1) > p(E_2)$ if I prefer the first bet:



- Exchangeability (weaker than i.i.d.)

$$p(x_1, x_2, \dots, x_n) = p(x_{s_1}, x_{s_2}, \dots, x_{s_n})$$

for permutations s on $\{1, 2, \dots, n\}$ for arbitrary n .

Chick

Bayesian Methods for Simulation

Why am I a Bayesian?

Lenny

- Exchangeability plus conceptually infinite N imply

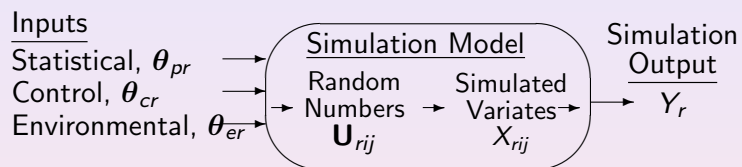
$$\lim_{N \rightarrow \infty} p(7 \text{ heads in first 10 flips}) = \int_0^1 \frac{10!}{7!3!} \theta^7 (1-\theta)^3 dF(\theta)$$

- de Finetti (1990)-like representation
- Ralph assumed conditional i.i.d., Lenny *derives* formula from exchangeability
- Probability defined by bet preferences, not repeated outcomes

Chick

Bayesian Methods for Simulation

Implication for Simulation: $Y_r = g(\theta_p, \theta_e, \theta_c; \mathbf{U}_r)$



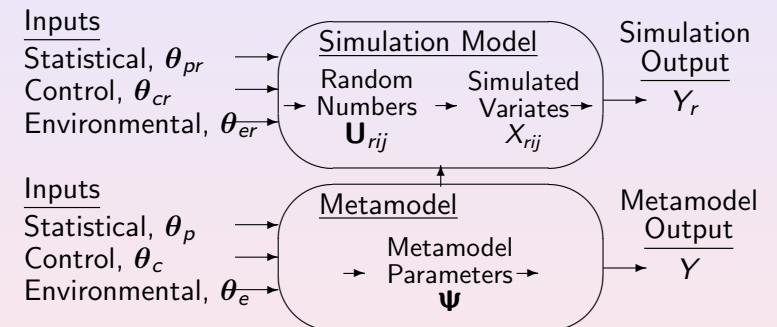
- Input selection: Infinite exchangeable sequence X_{ij} from modeled system to infer Θ_i

INSEAD

Chick

Bayesian Methods for Simulation

Implication for Simulation: $Y_r = g(\theta_p, \theta_e, \theta_c; \mathbf{U}_r)$



- Input selection: Infinite exchangeable sequence X_{ij} from modeled system to infer Θ_i
- Metamodeling: Infinite exchangeable Y_r to infer Ψ .

INSEAD

Chick

Bayesian Methods for Simulation

Part of the process

- Establish exchangeability arguments, posit potential likelihood functions
- Assess prior distributions
- Relevant asymptotic theorems
- Decisions, loss, and value of information
 - ranking and selection,
 - other experimental design issues

Outline

- 1 Getting down to brass tacks
 - Subjective and Bayesian methods
 - Assessing prior probability
 - Asymptotic Theorems
 - Decisions, loss, and value of information
 - Entropy and Kullback-Leibler Discrepancy
- 2 Applications
 - Uncertainty Analysis
 - Selecting from Multiple Candidate Distributions
 - Selecting the Best System
 - Metamodels
- 3 Implementation
- 4 Summary

Subjective Methods

We need a prior distribution for unknown parameters.
For a Bernoulli outcome ... de Finetti (1990), Savage (1972) require *You* to assess your personal belief, $\pi(\theta)$ to describe $p(\Theta \leq \theta)$

- Important gain in flexibility
- Consistent with expected value decision theory
- Kahneman, Slovic, and Tversky (1982) describe difficulties with elicitation ...

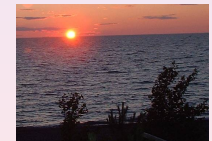
Some nonetheless seek 'automated' methods ...

Principle of Indifference

For finite exchangeable sequence, set

$$\theta_N = \frac{X_1 + \dots + X_N}{N} \in \{0/N, 1/N, \dots, (N-1)/N, 1\}$$

- Indifference: discrete uniform for finite N
- Limit: $\lim_{N \rightarrow \infty} p(\theta_N) \xrightarrow{D} \text{uniform}[0, 1]$
- Laplace (1812) used $\text{uniform}[0, 1]$ for his prior probability that the sun would come up tomorrow
- Coordinate dependence for continuous r.v. (X_i versus $\log X_i$)



Jeffrey's invariant prior

- $\pi(\theta) \propto |H(\theta)|^{1/2} d\theta$, where H is the expected information in one observation,

$$H(\theta) = E_X \left[-\frac{\partial^2 \log p(X | \theta)}{\partial \theta^2} \Big|_{\theta} \right], \quad (1)$$

'uniform' with respect to the natural metric induced by the likelihood function (Kass 1989)

- *Jeffreys' prior* for Bernoulli sampling is $\text{beta}(1/2, 1/2)$
- For some likelihoods, Jeffreys' prior is improper (doesn't integrate to 1). Might be used formally

INSEAD

Conjugate prior distribution

Bernoulli sampling

- Set $s_n = \sum_{i=1}^n x_i$
- Likelihood:
 $p(\mathbf{x}_n | \theta) \propto \theta^{s_n} (1 - \theta)^{n - s_n}$
- Prior:
 $\pi(\theta) \propto \theta^{\alpha-1} (1 - \theta)^{\beta-1}$, a $\text{beta}(\alpha, \beta)$ distribution
- Posterior:
 $\propto \theta^{\alpha+s_n-1} (1 - \theta)^{\beta+n-s_n-1}$, a $\text{beta}(\alpha + s_n, \beta + n - s_n)$ (conjugate) distribution

Exponential family

- Likelihood: $p(x | \theta) = a(x) h_0(\theta) \exp \left[\sum_{j=1}^d c_j \phi_j(\theta) h_j(x) \right]$
- Canonical conjugate prior: $p(\theta) = K(\mathbf{t}) [h_0(\theta)]^{n_0} \exp \left[\sum_{j=1}^d c_j \phi_j(\theta) t_j \right]$
- Posterior, given n data points: has parameters $n_0 + n$ and sum of $\mathbf{t} = (t_1, t_2, \dots, t_d)$ and sufficient statistics (Bernardo and Smith 1994)

INSEAD

Conjugate prior distribution

Bernoulli sampling

- Set $s_n = \sum_{i=1}^n x_i$
- Likelihood:
 $p(\mathbf{x}_n | \theta) \propto \theta^{s_n} (1 - \theta)^{n - s_n}$
- Prior:
 $\pi(\theta) \propto \theta^{\alpha-1} (1 - \theta)^{\beta-1}$, a $\text{beta}(\alpha, \beta)$ distribution
- Posterior:
 $\propto \theta^{\alpha+s_n-1} (1 - \theta)^{\beta+n-s_n-1}$, a $\text{beta}(\alpha + s_n, \beta + n - s_n)$ (conjugate) distribution

Exponential family

- Likelihood: $p(x | \theta) = a(x) h_0(\theta) \exp \left[\sum_{j=1}^d c_j \phi_j(\theta) h_j(x) \right]$
- Canonical conjugate prior: $p(\theta) = K(\mathbf{t}) [h_0(\theta)]^{n_0} \exp \left[\sum_{j=1}^d c_j \phi_j(\theta) t_j \right]$
- Posterior, given n data points: has parameters $n_0 + n$ and sum of $\mathbf{t} = (t_1, t_2, \dots, t_d)$ and sufficient statistics (Bernardo and Smith 1994)

INSEAD

Conjugate prior distribution

Bernoulli sampling

- Set $s_n = \sum_{i=1}^n x_i$
- Likelihood:
 $p(\mathbf{x}_n | \theta) \propto \theta^{s_n} (1 - \theta)^{n - s_n}$
- Prior:
 $\pi(\theta) \propto \theta^{\alpha-1} (1 - \theta)^{\beta-1}$, a $\text{beta}(\alpha, \beta)$ distribution
- Posterior:
 $\propto \theta^{\alpha+s_n-1} (1 - \theta)^{\beta+n-s_n-1}$, a $\text{beta}(\alpha + s_n, \beta + n - s_n)$ (conjugate) distribution

Exponential family

- Likelihood: $p(x | \theta) = a(x) h_0(\theta) \exp \left[\sum_{j=1}^d c_j \phi_j(\theta) h_j(x) \right]$
- Canonical conjugate prior: $p(\theta) = K(\mathbf{t}) [h_0(\theta)]^{n_0} \exp \left[\sum_{j=1}^d c_j \phi_j(\theta) t_j \right]$
- Posterior, given n data points: has parameters $n_0 + n$ and sum of $\mathbf{t} = (t_1, t_2, \dots, t_d)$ and sufficient statistics (Bernardo and Smith 1994)

INSEAD

'Noninformative'

- The uniform[0, 1] distribution is conjugate for Bernoulli sampling—a beta(1, 1) distribution.
- 'Noninformative' means 'evenly spread'—a heuristic term
- For canonical conjugate prior (for exponential family), the posterior has parameter $n_0 + n$
- Some think of n_0 as an 'effective' number of data points
- Noninformative can imply a small n_0

Others

- Jaynes (1983): maximum entropy methods
- Berger (1994), Kass and Wasserman (1996): Default rules

Useful? Actually informative?

Outline

- 1 Getting down to brass tacks
 - Subjective and Bayesian methods
 - Assessing prior probability
 - Asymptotic Theorems
 - Decisions, loss, and value of information
 - Entropy and Kullback-Leibler Discrepancy
- 2 Applications
 - Uncertainty Analysis
 - Selecting from Multiple Candidate Distributions
 - Selecting the Best System
 - Metamodels
- 3 Implementation
- 4 Summary

Classic Analogs for Infinite Exchangeable Sequences

Classical asymptotic theorems (e.g. Billingsley 1986) ...

- laws of large numbers (LLN)
- central limit theorem (CLT)
- law of iterated logarithm (LIL)

... have Bayesian interpretations if considered conditional on mean and standard deviation of an infinite exchangeable sequence.

Bayesian LLN

A Bayesian extension of the LLN allows for sample averages to converge to random variables rather than to 'true' means.

Theorem (Bayesian LLN)

If \bar{X}_n and \bar{Y}_m are respectively the averages of n and m exchangeable random quantities X_i (the two averages may or may not have some terms in common), the probability that

$$|\bar{X}_n - \bar{Y}_m| > \epsilon$$

may be made arbitrarily small by taking n and m sufficiently large (de Finetti 1990, p. 216 assumes a finite variance).

Theorem (Posterior Normality)

For each n , let $p_n(\cdot)$ be the posterior pdf of the d -dimensional parameter θ_n given $\mathbf{x}_n = (x_1, \dots, x_n)$, let $\tilde{\theta}_n$ be its mode (MAP), and define the $d \times d$ Bayesian observed information matrix Σ_n^{-1} by

$$\Sigma_n^{-1} = - \frac{\partial^2 \log p_n(\theta | \mathbf{x}_n)}{\partial \theta^2} \Big|_{\tilde{\theta}_n}. \quad (2)$$

Then $\phi_n = \Sigma_n^{-1/2}(\theta_n - \tilde{\theta}_n)$ converges in distribution to a standard (multivariate) normal random variable (Bernardo and Smith 1994, Prop 5.14 needs regularity conditions).

Frequentist analog: asserts that MLE is asymptotically normally distributed about a 'true' θ_0 (Law and Kelton 2000), as opposed to describing uncertainty about θ .

Outline

- 1 Getting down to brass tacks
 - Subjective and Bayesian methods
 - Assessing prior probability
 - Asymptotic Theorems
 - Decisions, loss, and value of information
 - Entropy and Kullback-Leibler Discrepancy
- 2 Applications
 - Uncertainty Analysis
 - Selecting from Multiple Candidate Distributions
 - Selecting the Best System
 - Metamodels
- 3 Implementation
- 4 Summary

Decisions under uncertainty

- Uncertainty described by probability \Rightarrow modeler can assess *expected value of information* (EVI) of additional data.
- EVI is useful in experimental design.
- EVI : value of resolving uncertainty with respect to a *loss function* $\mathcal{L}(d, \omega)$ that describes losses when a *decision* d is chosen when the state of nature is ω .
- Data from experiment can reduce uncertainty about ω , reduce expected loss.

Example: What is the mean? Setup

Example adapted from de Groot (1970) illustrates key concepts used for VIP procedures (Chick and Inoue 2001a)

- Decide if unknown mean W of normal distribution (known σ^2) is smaller (decision $d = 1$) or larger ($d = 2$) than w_0 .
- Exchangeable samples $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$, with $p(X_i) \sim \text{Normal}(w, \sigma^2)$, given $W = w$, are available
- Goal: design experiment (choose n) to balance sampling cost (cn) and expected opportunity cost if wrong answer chosen

$$\mathcal{L}(1, w) = \begin{cases} 0 & \text{if } w \leq w_0 \\ w - w_0 & \text{if } w > w_0, \end{cases}$$

$$\mathcal{L}(2, w) = \begin{cases} w_0 - w & \text{if } w \leq w_0 \\ 0 & \text{if } w > w_0. \end{cases}$$

Example: What is the mean? If we knew. . .

- Prior: $W \sim \text{Normal}(\mu, 1/\tau)$ is conjugate. NOTE: τ is the *precision* in our uncertainty about unknown mean, W .
- Posterior: Observing $\mathbf{X}_n = \mathbf{x}_n$ would result in

$$p(w | \mathbf{x}_n) \sim \text{Normal}(z, \tau_n^{-1})$$

$$z = \text{posterior mean of } W = E[W | \mathbf{x}_n] = \frac{\tau\mu + \frac{n}{\sigma^2}\bar{X}_n}{\tau + \frac{n}{\sigma^2}}$$

$$\tau_n = \text{posterior precision of } W = \tau + n/\sigma^2.$$

τ_n^{-1} equals the asymptotic posterior variance approximation Σ_n from theorem

INSEAD

Example: What is the mean? How much to know. . .

- Posterior mean z for unknown w influences the decision, but depends upon n , which is selected before \mathbf{X}_n is seen.
- Conditional distribution of \bar{X}_n given w is $\text{Normal}(w, \sigma^2/n)$
- *Predictive distribution* $p(z)$ of the posterior mean $Z = E[W | \mathbf{X}_n] = (\tau\mu + n\bar{X}_n)/\tau_n$
- Mixing over prior $\pi(w)$ implies a predictive distribution

$$Z \sim \text{Normal}(\mu, \tau_z^{-1})$$

$$\tau_z = \tau(\tau + n/\sigma^2)/(n/\sigma^2)$$

Note: $\tau_z^{-1} \rightarrow 0$ when $n \rightarrow 0$ (no new information).
If $n \rightarrow \infty$, then $\text{Var}[Z] \rightarrow \sigma^2$.

INSEAD

Example: What is the mean? What is the risk. . .

To minimize risk (sampling cost + expected loss from potentially incorrect decision), pick n to minimize a nested expectation

$$\rho(n) = cn + E_{\mathbf{X}_n}[E_W[\mathcal{L}(d(\mathbf{X}_n), W) | \mathbf{X}_n]].$$

- General technique: set $\mathcal{L}^*(d, w) = \mathcal{L}(d, w) - \mathcal{L}(1, w)$, which is 0 if $d = 1$ and is $w_0 - w$ if $d = 2$. Then

$$E_W[\mathcal{L}^*(d(\mathbf{X}_n), W) | \mathbf{X}_n] = \begin{cases} 0 & \text{if } d = 1 \\ w_0 - Z & \text{if } d = 2. \end{cases} \quad (3)$$

- To minimize loss in Eq. 3, assert $d(\mathbf{X}_n) = 2$ ('bigger') if posterior mean exceeds threshold, $Z > w_0$, and assert $d(\mathbf{X}_n) = 1$ ('smaller') if $Z \leq w_0$.

INSEAD

Example: What is the mean? Expected loss

- Decision depends on \mathbf{X}_n via Z ; Z has normal distribution
- Expected loss found with standard normal loss for newsboy

$$L_N[s] = \int_s^\infty (t - s)\phi(t)dt = \phi(s) - s(1 - \Phi(s))$$

$$E[\mathcal{L}^*(d(\mathbf{X}_n), W)] = E_{\mathbf{X}_n}[E_W[\mathcal{L}^*(d(\mathbf{X}_n), W) | \mathbf{X}_n]]$$

$$= -\tau_z^{-1} L_N[\tau_z^{-1/2}(w_0 - \mu)]$$

- Add back $E[\mathcal{L}(1, W)]$, use prior for W

$$E[\mathcal{L}(d(\mathbf{X}_n), W)] = \tau^{-1/2} L_N[\tau^{1/2}(w_0 - \mu)] - \tau_z^{-1/2} L_N[\tau_z^{1/2}(w_0 - \mu)]$$

- EVI and EVPI

INSEAD

Example: What is the mean? What is the risk...

- First-order optimality condition

$$\frac{\partial \rho}{\partial n} = \frac{1}{2} \tau_z^{-\frac{3}{2}} \phi[\tau_z^{\frac{1}{2}}(w_0 - \mu)] \cdot \frac{-\tau_z^2 \sigma^2}{n} + c = 0$$

- For small costs $c \rightarrow 0$, the sample size is large. Since $\tau_z \rightarrow \tau$ as $n \rightarrow \infty$, the optimal sample size n is asymptotically

$$n^* \approx \left(\tau^{\frac{1}{2}} \sigma^2 \phi[\tau^{\frac{1}{2}}(w_0 - \mu)] / (2c) \right)^{1/2}.$$

- Asymptotic approximations are a second useful tool to identify criteria-based sampling plans.
- Extensions of these ideas used to derive VIP procedures (Chick and Inoue 2001a).

INSEAD

Chick

Bayesian Methods for Simulation

Alternate approximation

- Regular exponential family: asymptotic variance approximation Σ_n from theorem simplifies to

$$H^{-1}(\theta) / (n_0 + n),$$

where H is the expected information from one observation (Eq. 1), if canonical conjugate prior distribution is used

- Approximate effect of m new samples is to change posterior to

$$\text{Normal} \left(\tilde{\theta}_n, \Sigma_n \frac{n_0 + n}{n_0 + n + m} \right).$$

- Used for OCBA (Chen 1996), sampling plans for field data (Ng and Chick 2001). Frequentist result to obtain CI of desired size (Law and Kelton 2000)

INSEAD

Chick

Bayesian Methods for Simulation

Outline

- 1 Getting down to brass tacks
 - Subjective and Bayesian methods
 - Assessing prior probability
 - Asymptotic Theorems
 - Decisions, loss, and value of information
 - Entropy and Kullback-Leibler Discrepancy

- 2 Applications

- Uncertainty Analysis
- Selecting from Multiple Candidate Distributions
- Selecting the Best System
- Metamodels

- 3 Implementation

- 4 Summary

INSEAD

Chick

Bayesian Methods for Simulation

Another loss function: discrepancy

- Kullback-Leibler discrepancy: difference between distributions
- Discrete distributions \tilde{p} and p ,

$$\delta(p \parallel \tilde{p}) = \sum \tilde{p}_i \log(\tilde{p}_i / q_i).$$

- Continuous r.v. X with densities \tilde{f} and $f_\theta = f(x \mid \theta)$,

$$\delta(f_\theta \parallel \tilde{f}) = \int \tilde{f}(x) \log \frac{\tilde{f}(x)}{f(x \mid \theta)} dx.$$

- One use: loss function for eliciting probability. If you believe the distribution is \tilde{f} , and you lose $\delta(f \parallel \tilde{f})$ if you provide a distribution f , then you should honestly report \tilde{f} (Bernardo 1979)

INSEAD

Chick

Bayesian Methods for Simulation

Discrepancy: Other uses

- Select design matrix \mathbf{d}_Θ of r vectors of inputs $(\theta_{pi}, \theta_{ei}, \theta_{ci})$ for $i = 1, 2, \dots, r$ with output \mathbf{Y} in order to best differentiate the posterior distribution of the response parameters ψ from the prior distribution for ψ (Bayesian D-optimal, Bernardo 1979; Smith and Verdinelli 1980; Ng and Chick 2002)

$$\int p(\mathbf{Y} | \mathbf{d}_\Theta) \left(\int p(\psi | \mathbf{Y}) \log \frac{p(\psi | \mathbf{Y})}{p(\psi)} d\psi \right) d\mathbf{Y}$$

- Select maximum entropy prior distribution (Jaynes 1983)
- More later: input distribution selection ...

Outline

- 1 Getting down to brass tacks
 - Subjective and Bayesian methods
 - Assessing prior probability
 - Asymptotic Theorems
 - Decisions, loss, and value of information
 - Entropy and Kullback-Leibler Discrepancy
- 2 Applications
 - Uncertainty Analysis
 - Selecting from Multiple Candidate Distributions
 - Selecting the Best System
 - Metamodels
- 3 Implementation
- 4 Summary

Why Uncertainty Analysis?

- Simple Example

$$\mu_j \rightarrow \text{Simulation} \rightarrow Y_j = 2\mu_j + e_j$$

- Input model: $X_\ell \sim \text{Normal}(\mu, \sigma_x^2)$, known σ_x^2
- Data: X_ℓ observed ($\ell = 1, 2, \dots, n_0$)
- Run r replications with \bar{X}_{n_0} input for μ
- Construct 90% CI:

$$\bar{Y}_r \pm z_{0.95} \frac{\hat{\sigma}_y}{\sqrt{r}}$$

Coverage?

If μ is known, expect coverage to be 90%. But μ is *not* known.

Why Uncertainty Analysis?

- Simple Example

$$\mu_j \rightarrow \text{Simulation} \rightarrow Y_j = 2\mu_j + e_j$$

- Input model: $X_\ell \sim \text{Normal}(\mu, \sigma_x^2)$, known σ_x^2
- Data: X_ℓ observed ($\ell = 1, 2, \dots, n_0$)
- Run r replications with \bar{X}_{n_0} input for μ
- Construct 90% CI:

$$\bar{Y}_r \pm z_{0.95} \frac{\hat{\sigma}_y}{\sqrt{r}}$$

Coverage?

If μ is known, expect coverage to be 90%. But μ is *not* known. CI can be meaningless (several authors. . .)

Why Uncertainty Analysis?

- Simple Example

$$\mu_j \rightarrow \text{Simulation} \rightarrow Y_j = 2\mu_j + e_j$$

- Input model: $X_\ell \sim \text{Normal}(\mu, \sigma_x^2)$, known σ_x^2
- Data: X_ℓ observed ($\ell = 1, 2, \dots, n_0$)
- Run r replications with \bar{X}_{n_0} input for μ
- Construct 90% CI:

$$\bar{Y}_r \pm z_{0.95} \frac{\bar{Y}_r}{\sqrt{V_{tot}}} \sqrt{V_{tot}}$$

Account for parameter uncertainty

$$V_{tot} = \frac{\sigma_y^2}{r} + \frac{4\sigma_x^2}{n_0}, \text{ so try } \bar{Y}_r \pm z_{0.95} \sqrt{V_{tot}}$$

Why Uncertainty Analysis?

- Simple Example

$$\mu_j \rightarrow \text{Simulation} \rightarrow Y_j = 2\mu_j + e_j$$

- Input model: $X_\ell \sim \text{Normal}(\mu, \sigma_x^2)$, known σ_x^2
- Data: X_ℓ observed ($\ell = 1, 2, \dots, n_0$)
- Run r replications with \bar{X}_{n_0} input for μ
- Construct 90% CI:

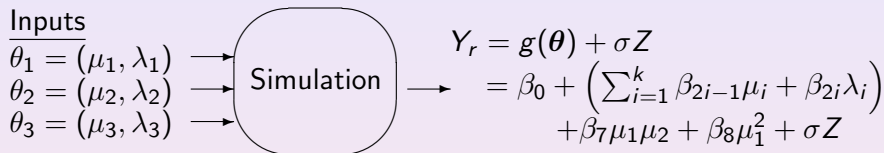
$$\bar{Y}_r \pm z_{0.95} \frac{\bar{Y}_r}{\sqrt{V_{tot}}} \sqrt{V_{tot}}$$

For known mean response g , input parameters $\theta = (\theta_1, \dots, \theta_k)$

$$V_{tot} \approx \frac{\sigma_y^2}{r} + \sum_{i=1}^k \frac{\beta_i H(\tilde{\theta}_i)^{-1} \beta_i^T}{n_i} \text{ where gradient } \beta_i = \left. \frac{\partial g(\tilde{\theta})^T}{\partial \theta_i} \right|_{\tilde{\theta}}$$

(asymptotics. Cheng & Holland; Ng & Chick; Wilson & Zouaoui)

Why Uncertainty Analysis? Estimate response



- Assess $X_{j\ell} \sim \text{Normal}(\mu_j, \lambda_j^{-1})$
- Observe data $X_{j\ell}$ are observed ($j = 1, 2, k; \ell = 1, 2, \dots, n_0$)
- Estimate unknown β with r_0 runs (e.g. CCD from DOE)

$$V_{tot} = \frac{\hat{\sigma}_y^2}{r_0} + \sum_{i=1}^k \underbrace{\frac{\partial g(\tilde{\theta}_n, \tilde{\beta}_{r_0})}{\partial \theta_i} \Sigma_{n_i} \frac{\partial g(\tilde{\theta}_n, \tilde{\beta}_{r_0})^T}{\partial \theta_i}}_{V_{par} = O(n_i^{-1})} + \underbrace{\Sigma_{\beta} \otimes \Sigma_{\theta}}_{V_{resp} = O(n_i^{-1} r)}$$

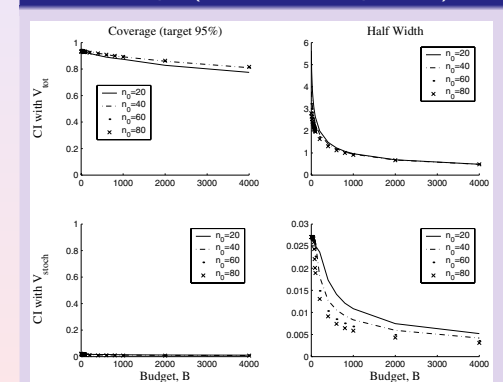
Uncertainty Analysis: Uncertainty Reduction

- Goal: Reduce uncertainty, not just quantify
 - r more replications
 - m_i more samples from source of randomness i
- Optimization:

$$\min_{r, m_i} \frac{\hat{\sigma}_y^2}{r_0 + r} + \sum_{i=1}^k \frac{\xi_i}{n_i + m_i} + \frac{\zeta_i}{(r_0 + r)(n_i + m_i)}$$

$$r, m_i \geq 0$$

CI Coverage (Chick and Ng 2004)



Uncertainty Analysis

- Sensitivity analysis: $E[g(\theta)]$ as a function of θ (average out stochastic uncertainty from \mathbf{u})
- Uncertainty analysis: $E[Y | \mathcal{E}]$, with both stochastic and parameter uncertainty, given all information \mathcal{E}
- Bayesian Model Average (BMA) estimates $E_Y[Y | \mathcal{E}]$

for $r = 1, \dots, R$ replications

sample parameter $\theta_r \sim p(\theta | \mathcal{E})$

for $i = 1, 2, \dots, n$

generate output y_{ri} given input θ_r

end loop

end loop

Generate estimate $\bar{y} = \sum_{r=1}^R \frac{1}{R} \sum_{i=1}^n \frac{y_{ri}}{n}$.

Variations on Estimating $E[Y | \mathcal{E}]$

- Zouaoui and Wilson (2003): decouple stochastic, parameter uncertainty; update estimate as new data becomes available with variations on BMA
- Andradóttir and Glynn (2004): biased estimates of $E[Y | \theta]$; quasi-random numbers; quadrature to select inputs θ_i
- Estimate distribution of conditional expectation $E[Y | \Theta, \mathcal{E}]$. Steckley and Henderson (2003) derive asymptotically optimal ways selecting r and n in BMA to produce a kernel density estimator (some conditions apply)

Outline

- 1 Getting down to brass tacks
 - Subjective and Bayesian methods
 - Assessing prior probability
 - Asymptotic Theorems
 - Decisions, loss, and value of information
 - Entropy and Kullback-Leibler Discrepancy
- 2 Applications
 - Uncertainty Analysis
 - Selecting from Multiple Candidate Distributions
 - Selecting the Best System
 - Metamodels
- 3 Implementation
- 4 Summary

Basic Problem

- Which distribution/parameter to pick?
- 'Usual':
 - 1 Pick q candidate input distributions (e.g. exponential, gamma, Weibull, lognormal)
 - 2 Find MLE $\hat{\theta}_i$ for candidates $i = 1, \dots, q$
 - 3 Goodness-of-fit (χ -squared, K-S, A-D) tests
- Concerns:
 - 1 CI coverage if MLE/best distribution selected
 - 2 How to select among unrejected models? ... (Lindley 1957; Berger and Pericchi 1996; Chick 2001)

Bayesian Input Selection

- BMA applies without change for q candidates
 - 1 Put prior on $(M = m, \theta_m)$, where $m \in \{1, 2, \dots, q\}$
 - 2 Compute posterior $p(m, \Theta_m | \mathcal{E})$, sample from it in BMA
- Chick (2001): stochastic process simulation context; moment matching method for commensurate prior distributions
- Zouaoui and Wilson (2001): decouple stochastic uncertainty from two types of structural uncertainty (candidate model & parameters); variance reduction for BMA; numerical analysis
- Model selected closest to the true model in sense of Kullback-Leibler divergence (Berk 1966; Bernardo and Smith 1994; Dmochowski 1999).

Outline

- 1 Getting down to brass tacks
 - Subjective and Bayesian methods
 - Assessing prior probability
 - Asymptotic Theorems
 - Decisions, loss, and value of information
 - Entropy and Kullback-Leibler Discrepancy
- 2 Applications
 - Uncertainty Analysis
 - Selecting from Multiple Candidate Distributions
 - Selecting the Best System
 - Metamodels
- 3 Implementation
- 4 Summary

Ranking and selection procedures differ?

- What is it for?
 - Select the 'best' of a finite set
 - Best identified by mean simulation response
- What are the approaches?
 - IZ, Indifference zone: $P(CS) \geq 1 - \alpha, \delta^*$, repeated applications of procedure
 - VIP, Bayesian value of information procedures: Like selection of mean bigger or smaller than threshold, Bayesian inference, loss, EVI
 - OCBA, Chen et al.: Heuristic, allocates samples to improve Bayesian PCS using 'thought experiment' $\sigma^2/r_0 \rightarrow \sigma^2/(r_0 + r)$

VIP procedures

- Motivated by 'statistical conservativeness' of IZ approaches
- Two-stage: Unknown means of several systems.
 - Opportunity cost and 0-1 loss ($P(CS)$)
 - Variances also unknown, different (conjugate prior, student marginal for mean)
 - Optimal solution unknown except special cases
 - Asymptotic approximation, Bonferroni bound for loss
- Sequential:
 - In theory, should improve things
 - Behrens-Fisher
 - Seems to work quite well

Procedure $\mathcal{LL}(\mathcal{B})$, for opportunity cost (linear loss)

- Specify the first-stage sample size r_0 . Take independent replications y_{i1}, \dots, y_{ir_0} , for each system, $i = 1, \dots, k$
- Compute all first-stage sample means $\bar{x}_i = \sum_{j=1}^{r_0} y_{ij} / r_0$ and sample variances $\hat{\sigma}_i^2 = \frac{\sum_{j=1}^{r_0} (y_{ij} - \bar{x}_i)^2}{r_0 - 1}$, order statistics $\bar{x}_{[1]} \leq \dots \leq \bar{x}_{[k]}$, and $\lambda_{i,k} = r_0 / (\hat{\sigma}_{[k]}^2 + \hat{\sigma}_{[i]}^2)$
- If sampling budget is B , run r_i more independent replications,

$$r_{[i]} = \frac{B + \sum_{j \in \mathcal{S}} r_0 c_j}{\sum_{j \in \mathcal{S}} \left(\frac{c_j c_{[j]} \hat{\sigma}_j^2 \eta_j}{\hat{\sigma}_{[i]}^2 \eta_{[i]}} \right)^{1/2}} - r_0$$

$$\eta_{[i]} = (\lambda_{i,k})^{1/2} \frac{(r_0 - 1) + \lambda_{i,k} (\bar{x}_{[k]} - \bar{x}_{[i]})^2}{(r_0 - 1) - 1} \phi_{r_0 - 1} [(\lambda_{i,k})^{1/2} (\bar{x}_{[k]} - \bar{x}_{[i]})]$$

if $[i] \neq [k]$ and $\eta_{[k]} = \sum_{j=1}^{k-1} \eta_{[j]}$

- Select system with largest $\bar{x}_{\bar{i}} = \sum_{j=1}^{r_0 + r_i} y_{ij} / (r_0 + r_i)$

INSEAD

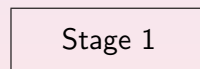
Sample Comparison with Combined Procedure, \mathcal{C} (MDM)

Figure of merit	Proc.	Number of systems, k			
		2	5	10	100
ANR	All	738	3,429	8,784	42,862
Empirical P(CS)	\mathcal{C}	0.8363	0.9140	0.9323	0.9763
	0-1(\mathcal{B}) $\mathcal{LL}(\mathcal{B})$	0.8527*	0.9117	0.9480*	0.9937*
Empirical frac. 'good' selections	\mathcal{C}	1.0000	0.9943	0.9990	1.0000
	0-1(\mathcal{B}) $\mathcal{LL}(\mathcal{B})$	1.0000	0.9953	0.9977	0.9997
Expected posterior PCS	\mathcal{C}	0.8336	0.8379	0.8649	0.9318
	0-1(\mathcal{B}) $\mathcal{LL}(\mathcal{B})$	0.8446*	0.8339	0.8821*	0.9717*
Expected bound, opp. cost	\mathcal{C}	0.0176	0.0138	0.0104	0.0037
	0-1(\mathcal{B}) $\mathcal{LL}(\mathcal{B})$	0.0157*	0.0128*	0.0075*	0.0012*
		0.0154*	0.0110*	0.0055*	0.0005*

INSEAD

VIP: Common Random Numbers

- Common random numbers (CRN) can sharpen contrasts between systems (e.g. same simulated demand pattern)
- Two-stage with screening (Chick and Inoue 2001b)
 - Run subset of systems in stage 2
 - Use 'missing data' formulas to update
 - Select from even screened systems
- Matrix intensive, heuristic provided

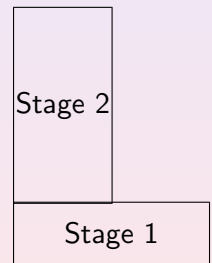


1 2 3 4 5 ... k

INSEAD

VIP: Common Random Numbers

- Common random numbers (CRN) can sharpen contrasts between systems (e.g. same simulated demand pattern)
- Two-stage with screening (Chick and Inoue 2001b)
 - Run subset of systems in stage 2
 - Use 'missing data' formulas to update
 - Select from even screened systems
- Matrix intensive, heuristic provided

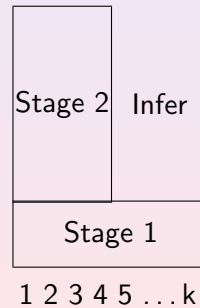


1 2 3 4 5 ... k

INSEAD

VIP: Common Random Numbers

- Common random numbers (CRN) can sharpen contrasts between systems (e.g. same simulated demand pattern)
- Two-stage with screening (Chick and Inoue 2001b)
 - Run subset of systems in stage 2
 - Use 'missing data' formulas to update
 - Select from even screened systems
- Matrix intensive, heuristic provided



INSEAD

Summaries

- VIP procedures have solid basis, perform numerically quite well (Chick and Inoue 2001a; Chick and Inoue 2002; Chick and Inoue 2001a; Inoue et al. 1999).
 - Matlab: All procedures (0-1 or opportunity cost loss; two-stage or independent sequential; two-stage CRN)
 - C: Independent replications, two-stage or sequential, both loss functions, *plus* variants (work in progress) to achieve Bayesian predictive targets for PCS and EOC
- OCBA has seen strong empirical performance. Connection?
- Have started to look at comparison of VIP with KN++ from Kim and Nelson (2001) (results not yet done)

INSEAD

Outline

- 1 Getting down to brass tacks
 - Subjective and Bayesian methods
 - Assessing prior probability
 - Asymptotic Theorems
 - Decisions, loss, and value of information
 - Entropy and Kullback-Leibler Discrepancy
- 2 Applications
 - Uncertainty Analysis
 - Selecting from Multiple Candidate Distributions
 - Selecting the Best System
 - Metamodels
- 3 Implementation
- 4 Summary

INSEAD

Metamodels

- Normal linear model

$$Y = \sum_{\ell=1}^p g_{\ell}(\boldsymbol{\theta})\beta_{\ell} + Z(\boldsymbol{\theta}; \mathbf{U}) = \mathbf{g}^T(\boldsymbol{\theta})\boldsymbol{\beta} + Z(\boldsymbol{\theta}; \mathbf{U}), \quad (4)$$

- Gaussian random function (GRF) - kriging

INSEAD

Normal linear model

$$Y = \sum_{\ell=1}^P g_{\ell}(\boldsymbol{\theta})\beta_{\ell} + Z(\boldsymbol{\theta}; \mathbf{U}) = \mathbf{g}^T(\boldsymbol{\theta})\boldsymbol{\beta} + Z(\boldsymbol{\theta}; \mathbf{U}), \quad (5)$$

- Model
 - Known: regression functions g_1, \dots, g_p
 - Unknown: coefficients $\boldsymbol{\beta}$, variance of zero-mean noise $Z(\cdot)$
- Conjugate prior $p(\boldsymbol{\beta}, \sigma^2)$ (if all factors active)
 - Inverted gamma distribution for unknown variance σ^2
 - Multivariate normal distribution for $\boldsymbol{\beta}$ given σ^2 ,
 - Raftery, Madigan, and Hoeting (1997) describe a relatively 'uninformative' prior distribution, 'good' results
- Identifying important factors like input distribution selection
 - 2^P candidate models
 - KL discrepancy-based design criterion balances factor identification and parameter estimation (Ng and Chick 2004a)



Normal linear model

$$Y = \sum_{\ell=1}^P g_{\ell}(\boldsymbol{\theta})\beta_{\ell} + Z(\boldsymbol{\theta}; \mathbf{U}) = \mathbf{g}^T(\boldsymbol{\theta})\boldsymbol{\beta} + Z(\boldsymbol{\theta}; \mathbf{U}), \quad (5)$$

- Model
 - Known: regression functions g_1, \dots, g_p
 - Unknown: coefficients $\boldsymbol{\beta}$, variance of zero-mean noise $Z(\cdot)$
- Conjugate prior $p(\boldsymbol{\beta}, \sigma^2)$ (if all factors active)
 - Inverted gamma distribution for unknown variance σ^2
 - Multivariate normal distribution for $\boldsymbol{\beta}$ given σ^2 ,
 - Raftery, Madigan, and Hoeting (1997) describe a relatively 'uninformative' prior distribution, 'good' results
- Identifying important factors like input distribution selection
 - 2^P candidate models
 - KL discrepancy-based design criterion balances factor identification and parameter estimation (Ng and Chick 2004a)



Normal linear model

$$Y = \sum_{\ell=1}^P g_{\ell}(\boldsymbol{\theta})\beta_{\ell} + Z(\boldsymbol{\theta}; \mathbf{U}) = \mathbf{g}^T(\boldsymbol{\theta})\boldsymbol{\beta} + Z(\boldsymbol{\theta}; \mathbf{U}), \quad (5)$$

- Model
 - Known: regression functions g_1, \dots, g_p
 - Unknown: coefficients $\boldsymbol{\beta}$, variance of zero-mean noise $Z(\cdot)$
- Conjugate prior $p(\boldsymbol{\beta}, \sigma^2)$ (if all factors active)
 - Inverted gamma distribution for unknown variance σ^2
 - Multivariate normal distribution for $\boldsymbol{\beta}$ given σ^2 ,
 - Raftery, Madigan, and Hoeting (1997) describe a relatively 'uninformative' prior distribution, 'good' results
- Identifying important factors like input distribution selection
 - 2^P candidate models
 - KL discrepancy-based design criterion balances factor identification and parameter estimation (Ng and Chick 2004a)



Gaussian random functions (GRF)

- Well-known in deterministic simulations, particularly in geostatistics (Cressie 1993; Santner et al. 2003)
- Provide flexibility that the linear model does not, and are useful when g takes a long time to compute.
- GRF for unknown nonstochastic g (no random numbers \mathbf{u}) is

$$Y(\boldsymbol{\theta}) = \sum_{\ell=1}^P g_{\ell}(\boldsymbol{\theta})\beta_{\ell} + Z(\boldsymbol{\theta}) = \mathbf{g}^T(\boldsymbol{\theta})\boldsymbol{\beta} + Z(\boldsymbol{\theta}) \quad (6)$$

for *known* regression functions g_1, \dots, g_p of \mathbb{R}^d , unknown regression coefficients $\boldsymbol{\beta}$, and a *zero-mean random second-order process* so that for any distinct inputs $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_m$, the vector (Y_1, \dots, Y_m) has (nonindependent) multivariate normal distribution, conditional on $\boldsymbol{\beta}$.



Inference with GRFs

- GRF: Determined by mean $\mathbf{g}^T(\boldsymbol{\theta})\boldsymbol{\beta}$ and covariance function

$$C^*(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \text{Cov}(Y(\boldsymbol{\theta}_1), Y(\boldsymbol{\theta}_2))$$

- Common to assume strong stationarity, so

$$C^*(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = C(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)$$

- Inference for $g(\boldsymbol{\theta})$ at $\boldsymbol{\theta}_{r+1}$ not yet input to simulation model with correlation function $R(\mathbf{h}) = C(\mathbf{h})/C(0)$ for $\mathbf{h} \in \mathbb{R}^d$.
Example: power exponential $R(\mathbf{h}) = \prod \exp[-|h_i/\gamma_i|^{p_i}]$ for $p_i \in [0, 2]$.
- Kriging (geostatistics) is best linear unbiased prediction (BLUP) for $g(\boldsymbol{\theta}_{r+1})$

INSEAD

More GRF

- Assessment of uncertainty in $g(\boldsymbol{\theta}_{r+1}) \Rightarrow$ experimental design technique to choose inputs to reduce response uncertainty (Santner et al. 2003)
- Also see tutorial by van Beers and Kleijnen (2004) in this year's WSC
- Deterministic simulation: (Sacks et al. 1989; O'Hagan et al. 1999; Kennedy and O'Hagan 2001; Santner et al. 2003; van Beers and Kleijnen 2003).
- More work is needed for GRFs in the stochastic simulation context.

INSEAD

Implementation

Issues:

- Maximize (MLE $\hat{\theta}$ or MAP $\tilde{\theta}$)
- Integrate (marginal distribution $p(\theta_1 | \mathbf{x}_n)$ from $p(\theta_1, \theta_2 | \mathbf{x}_n)$, or proportionality constant $c^{-1} = \int f(\mathbf{x}_n | \theta) d\pi(\theta)$)
- Simulate (sample from $p(\theta | \mathbf{x}_n)$ to estimate $E[\alpha(\Theta)]$)

Tools:

- Newton-Raphson, Nelder-Mead, expectation-maximization (EM) algorithm, ...
- Quadrature, normal approximation, data augmentation (IP algorithm), importance sampling (IS)
- Inversion, importance sampling (IS), Markov Chain Monte Carlo (MCMC)

Evans and Swartz 1995; Tanner 1996; Gilks et al. 1996; Devroye 2004, ...

INSEAD

Metropolis-Hastings: An MCMC Algorithm

Target: Sample from $p(\theta | \mathcal{E})$

Capable: Easily sample from $q(\cdot | \theta_{t-1})$

Initialize $t = 0, \theta_0$

for $t = 1, 2, \dots$

sample a candidate $\theta \sim q(\cdot | \theta_{t-1})$

compute acceptance probability

$$\alpha(\theta_{t-1}, \theta) = \min \left\{ 1, \frac{p(\theta | \mathcal{E}) \cdot q(\theta_{t-1} | \theta)}{p(\theta_{t-1} | \mathcal{E}) \cdot q(\theta | \theta_{t-1})} \right\}$$

generate an independent $u \sim \text{uniform}[0, 1]$

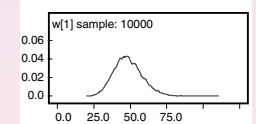
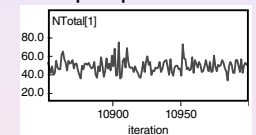
if $u \leq \alpha(\theta_{t-1}, \theta)$ then set $\theta_t \leftarrow \theta$

otherwise set $\theta_t \leftarrow \theta_{t-1}$

set $t \leftarrow t + 1$

end loop

Sample path



Density Estimate

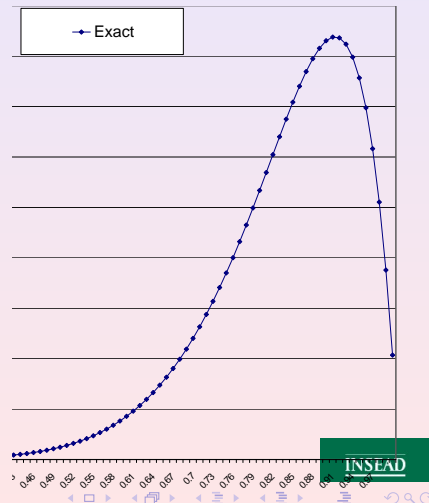
INSEAD

Approximating Posterior Distributions: Pros and Cons

Exact Posterior

- Exact
- Good if simple closed form known
- May be hard in general (mixtures, missing data, marginal distribution, curse of dimensionality?)

Here, gene linkage example (Tanner 1996)



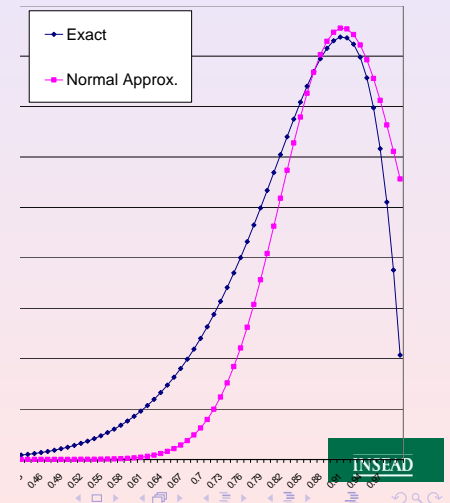
Chick

Bayesian Methods for Simulation

Approximating Posterior Distributions: Pros and Cons

Asymptotic normality (from theorem)

- Normal $(\tilde{\theta}_n, \Sigma_n)$
- Relatively easy to compute
- Requires many data points ($n > 20d$)
- Does not model skew, etc.



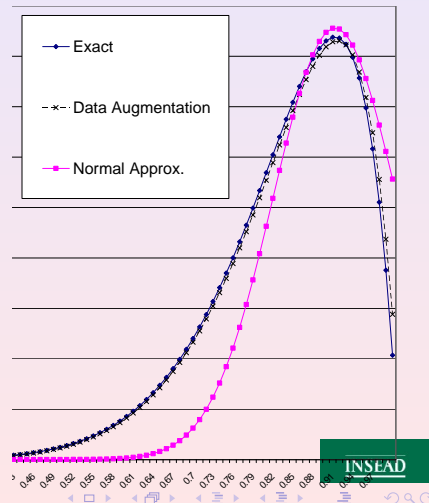
Chick

Bayesian Methods for Simulation

Approximating Posterior Distributions: Pros and Cons

Data Augmentation (IP algorithm)

- $p(\theta|\mathcal{E}) = \int p(\theta|Z, \mathcal{E})p(Z|\mathcal{E})$, average over $Z =$ 'missing data'
- 1 Set $i = 0$; $g_0(\theta) =$ current estimate of $p(\theta|\mathcal{E})$
- 2 Generate z_1, z_2, \dots, z_m from $g_i(\theta)$ by
 - 1 Sample $\theta_1, \dots, \theta_m$ from $g_i(\theta)$
 - 2 Sample z_1, \dots, z_m from $p(Z|\theta_i, \mathcal{E})$
- 3 $g_{i+1}(\theta) = \sum_{j=1}^m p(\theta|z_j, \mathcal{E})$



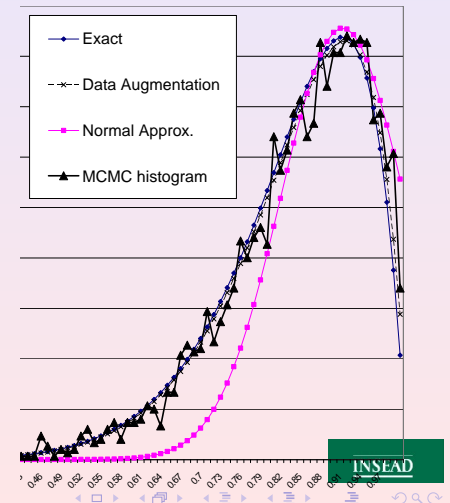
Chick

Bayesian Methods for Simulation

Approximating Posterior Distributions: Pros and Cons

MCMC with histogram

- General tool to sample (approximately) from posterior
- Complicated models possible
- Output analysis issues (convergence)
- Time-average (IP was average of distributions)
- MCMC can average distributions, too

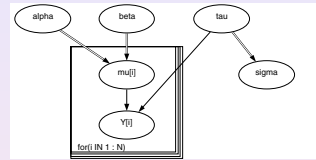


Chick

Bayesian Methods for Simulation

Some Tools

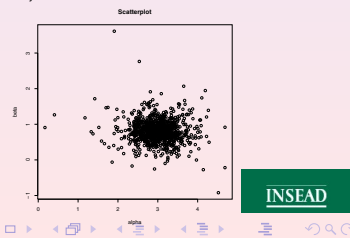
- Handcode: Matlab, C, Gauss
- WinBUGS (Spiegelhalter et al. 1996) (<http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml>)
- R (<http://www.r-project.org/>), S-PLUS packages, BOA add-on (<http://www.public-health.uiowa.edu/boa/>)
- Uncertainty analysis in spreadsheet Monte Carlo applications are available (e.g. Winston 2000).
- Most DEDS Commercial Tools: cumbersome to implement BMA



```

model
{
  for( i in 1 : N ) {
    Y[i] ~ dnorm(mu[i],tau)
    mu[i] <- alpha + beta * (x[i] - xbar)
  }
  tau ~ dgamma(0.001,0.001) sigma <- 1 / sqrt(tau)
  alpha ~ dnorm(0.0,1.0E-6)
  beta ~ dnorm(0.0,1.0E-6)
}

```

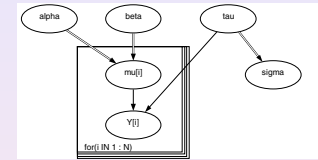


Chick

Bayesian Methods for Simulation

Some Tools

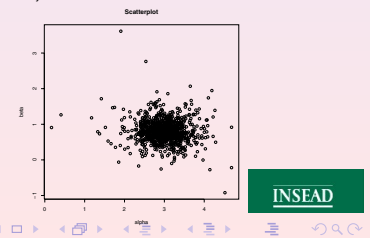
- Handcode: Matlab, C, Gauss
- WinBUGS (Spiegelhalter et al. 1996) (<http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml>)
- R (<http://www.r-project.org/>), S-PLUS packages, BOA add-on (<http://www.public-health.uiowa.edu/boa/>)
- Uncertainty analysis in spreadsheet Monte Carlo applications are available (e.g. Winston 2000).
- Most DEDS Commercial Tools: cumbersome to implement BMA



```

model
{
  for( i in 1 : N ) {
    Y[i] ~ dnorm(mu[i],tau)
    mu[i] <- alpha + beta * (x[i] - xbar)
  }
  tau ~ dgamma(0.001,0.001) sigma <- 1 / sqrt(tau)
  alpha ~ dnorm(0.0,1.0E-6)
  beta ~ dnorm(0.0,1.0E-6)
}

```



Chick

Bayesian Methods for Simulation

Bayesian Methods for Stochastic Process Simulation

- Themes:
 - Represent *all* uncertainty with probability, update with Bayes' rule, expected value of information for sampling decisions
 - Use simulation to efficiently estimate quantities of interest for a Bayesian analysis
- Applications: input distribution selection, uncertainty analysis, experimental design, ranking and selection
- Asymptotic approximations helpful when exact optimal solutions are hard to find

INSEAD

Chick

Bayesian Methods for Simulation

Research opportunities include

- Input modeling and uncertainty analysis (kernel estimation of conditional means; the effect of different candidate distributions on uncertainty; prior distributions elicitation; calibration/inverse problem)
- Response modeling (extend the Gaussian random field work for stochastic simulation; nonasymptotic sampling plans for input parameter inference to optimally reduce output uncertainty; reasoning about models; calibration. . .)
- Experimental design (quantiles, non-expected value goals; CRN for unknown input parameters for ranking and selection; non-Gaussian output for ranking and selection, GRFs)
- Improved computational tools (e.g. MCMC, software interop)

INSEAD

Chick

Bayesian Methods for Simulation