

Introduction to Bayesian Statistics

Shin-ichi Mayekawa

mayekawa@hum.titech.ac.jp

www.ms.hum.titech.ac.jp

Outline

1. Introduction to Bayes Theorem
2. Binomial Model and Poisson Model
3. Normal Model
4. Normal Regression Model
5. Bayesian Numerical Calculations
(Markov Chain Monte Carlo)
6. Application of MCMC
7. Model Selection
8. miscellaneous topics

2

Numerical Evaluation of Post.Dist.

- Calculation of posterior mean, var etc by numerical integration.

Simplest method is the Grid Method.

- In this ppt presentation, roman letters X is used to denote the parameters such as θ , π , μ or σ .

3

Grid Method

- When the kernel of the posterior pdf is known:

$$f(x) = c \times g(x)$$

use the grid

$$x_{(1)}, x_{(2)}, \dots, x_{(n)}$$

to approximate the probability at each grid point by

$$f^*(x_{(j)}) = \frac{1}{d} g(x_{(j)})$$

where

$$d = \sum_{j=1}^n g(x_{(j)})$$

Normalized so that the sum is equal to unity. \rightarrow probability

4

Grid Method

- posterior mean

$$\mathcal{E}(x) \simeq \sum_j^n x_{(j)} f^*(x_{(j)})$$

sum of value x prob

- posterior mean of any function

$$\mathcal{E}(h(x)) \simeq \sum_j^n h(x_{(j)}) f^*(x_{(j)})$$

- Examples: posterior variance

$$h(x) = (x - \mathcal{E}(x))^2$$

Grid Method

- Example: Be(4,6)

exact mean = 0.4 and exact variance = 0.0218

grid: n=10 $x_{(1)} = 0.05, x_{(2)} = 0.15, \dots, x_{(10)} = 0.95$

post pdf kernel: $g(x) = 1000 \times x^{a-1} (1-x)^{b-1}$

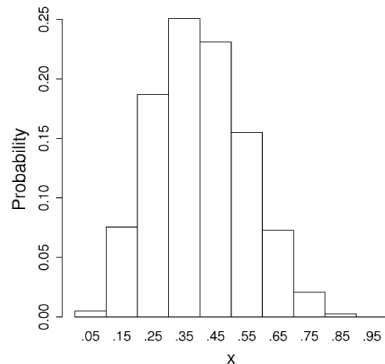
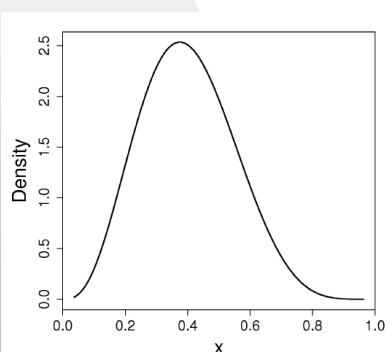
	1	2	3	4	5	6	7	8	9	10	合計
$x_{(j)}$	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95	
$g(x_{(j)})$	0.097	1.498	3.708	4.975	4.586	3.070	1.442	0.412	0.047	0.000	19.834
$f^*(x_{(j)})$	0.005	0.076	0.187	0.251	0.231	0.155	0.073	0.021	0.002	0.000	1

Grid Method

- Example: Be(4,6)

exact mean = 0.4 and exact variance = 0.0218

	1	2	3	4	5	6	7	8	9	10	合計
$x_{(j)}$	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95	
$g(x_{(j)})$	0.097	1.498	3.708	4.975	4.586	3.070	1.442	0.412	0.047	0.000	19.834
$f^*(x_{(j)})$	0.005	0.076	0.187	0.251	0.231	0.155	0.073	0.021	0.002	0.000	1



Grid Method

- Multivariate posterior pdf

$$f(x_1, x_2, \dots, x_p) = c \times g(x_1, x_2, \dots, x_p)$$

grid: $n_1 \times n_2 \times \dots \times n_p$

$x_{i(j_i)}$ ($j_i = 1, 2, \dots, n_i, i = 1, 2, \dots, p$)

prob:

$$f^*(x_{1(j_1)}, x_{2(j_2)}, \dots, x_{p(j_p)}) = \frac{1}{d} \times g(x_{1(j_1)}, x_{2(j_2)}, \dots, x_{p(j_p)})$$

$$d = \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \dots \sum_{j_p=1}^{n_p} g(x_{1(j_1)}, x_{2(j_2)}, \dots, x_{p(j_p)})$$

Normalized so that the sum is equal to unity. → probability 8

Grid Method

- Posterior expectation of function h:

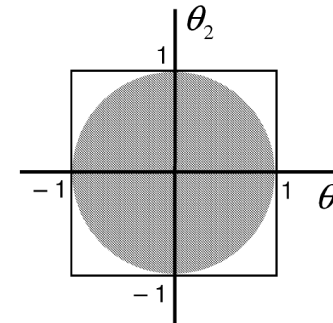
$$\mathcal{E}(h(x_i)) \simeq \sum_{j_1}^{n_1} \sum_{j_2}^{n_2} \dots \sum_{j_p}^{n_p} h(x_{i(j_i)}) f^*(x_{1(j_1)}, x_{2(j_2)}, \dots, x_{p(j_p)})$$

$$\mathcal{E}(h(x_1, x_2, \dots, x_p)) \simeq \sum_{j_1}^{n_1} \sum_{j_2}^{n_2} \dots \sum_{j_p}^{n_p} h(x_{1(j_1)}, x_{2(j_2)}, \dots, x_{p(j_p)}) \times f^*(x_{1(j_1)}, x_{2(j_2)}, \dots, x_{p(j_p)})$$

9

Grid Method

- effective area where pdf > 0



cube whole domain $(2)^n$
 circle area of concentration $V^n = \frac{\pi^{\frac{n}{2}}}{\frac{n}{2} \Gamma(\frac{n}{2})}$

dim n	1	2	3	4	6	8	10
circle	2	3.14	4.19	4.93	5.17	4.06	2.55
cube	2	4	8	16	64	256	1024
ratio	100%	78.5%	52.4%	30.8%	8.07%	1.59%	0.25%

10

Numerical Evaluation of Post.Dist.

- Posterior Simulation: Monte Carlo Method

Generate many random numbers from the posterior distribution and calculate sample statistics such as mean, std, etc to evaluate the properties of the posterior distribution.

Can be used when the posterior distribution cannot be identified with a known distribution.

11

Monte Carlo Method

- Posterior Simulation: Monte Carlo Method

Let

$$X^{(t)} \quad (t = 0, 1, 2, \dots, N)$$

be a set of iid random variables from pdf f:

Monte Carlo Sample.

iid = identically and independently distributed

Then,

$$\mathcal{E}(X) \simeq \bar{X} = \frac{1}{N} \sum_{t=1}^N X^{(t)} \quad \text{var}(X) \simeq \frac{1}{N} \sum_{t=1}^N (X^{(t)} - \bar{X})^2$$

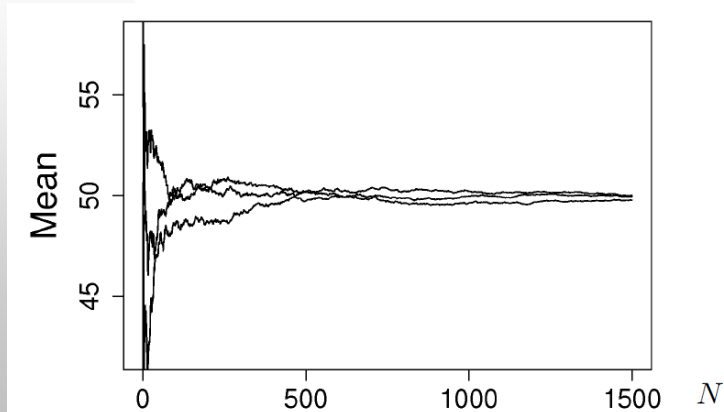
$$\mathcal{E}(h(X)) = \int_X h(x) f(x) dx = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N h(X^{(t)})$$

12

Numerical Evaluation of Post.Dist.

- Example: mean of $N(50, 10)$

Three sequences of the means up to N-th obs.



13

Monte Carlo Method

- Multivariate posterior pdf

$$f(x_1, x_2, \dots, x_p)$$

Monte Carlo Sample

$$(x_1^{(1)}, x_2^{(1)}, \dots, x_p^{(1)}), (x_1^{(2)}, x_2^{(2)}, \dots, x_p^{(2)}), \dots, (x_1^{(N)}, x_2^{(N)}, \dots, x_p^{(N)})$$

$X_i^{(t)}$ is from the marginal distribution.

$$\mathcal{E}(X_i) \simeq \frac{1}{N} \sum_{t=1}^N X_i^{(t)}$$

$$\text{var}(X_i) \simeq \frac{1}{N} \sum_{t=1}^N (X_i^{(t)} - \bar{X}_i)^2$$

14

Monte Carlo Method

- Monte Carlo Method can be used whenever we can generate random variables from the posterior distribution.
- Recent statistical packages and computer languages have random number generators from **known** distributions.
- However, if the posterior distribution is identified with the known distribution, there are no need to use Monte Carlo method.
- Therefore, Monte Carlo method is only useful when the posterior distribution is **not** of the standard form.
- What should we do, then?

15

Normal Model with Semi Natural Conjugate Prior

- prior

$$\mu \sim N(\gamma_0, \phi_0^2)$$

$$\sigma \sim \chi^{-1}(\kappa_0^2, \nu_0)$$

$$h(\mu, \sigma) = h(\mu)h(\sigma)$$

$$\propto \frac{1}{\sqrt{\phi_0^2}} \exp\left(-\frac{1}{2\phi_0^2}(\mu - \gamma_0)^2\right) \times \sigma^{-\nu_0-1} \exp\left(-\frac{\kappa_0^2}{2\sigma^2}\right)$$

- posterior

$$h(\mu, \sigma | \mathbf{y}) \propto f(\mathbf{y} | \mu, \sigma) h(\mu, \sigma)$$

$$\propto \exp\left[-\frac{1}{2} \left(\frac{\text{RSS}_y + \kappa_0^2}{\sigma^2} + \frac{N(\gamma_0 - \bar{y})^2}{\sigma^2 + N\phi_0^2} + \left(\frac{N}{\sigma^2} + \frac{1}{\phi_0^2}\right) \left(\mu - \frac{\gamma_0\sigma^2 + N\phi_0^2\bar{y}}{\sigma^2 + N\phi_0^2}\right)^2 \right)\right]$$

$$\times \sigma^{-N-\nu_0-1} \phi_0^{-1}$$

cannot be identified.

16

Normal Model with Semi Natural Conjugate Prior

- full conditional posterior

$$h(\mu, |\sigma, \mathbf{y}) \propto \exp \left[-\frac{1}{2} \left(\frac{N}{\sigma^2} + \frac{1}{\phi_0^2} \right) \left(\mu - \frac{\gamma_0 \sigma^2 + N \phi_0^2 \bar{y}}{\sigma^2 + N \phi_0^2} \right)^2 \right]$$

$$\mu | \sigma, \mathbf{y} \sim N \left[\frac{1}{\frac{N}{\sigma^2} + \frac{1}{\phi_0^2}} \left(\frac{1}{\phi_0^2} \gamma_0 + \frac{N}{\sigma^2} \bar{y} \right), \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\phi_0^2}} \right]$$

$$h(\sigma | \mu, \mathbf{y}) \propto \sigma^{-(N+\nu_0)-1} \times \exp \left[-\frac{1}{2} \left(\frac{\text{RSS}_y + N(\mu - \bar{y})^2 + \kappa_0^2}{\sigma^2} \right) \right]$$

$$\sigma | \mu, \mathbf{y} \sim \chi^{-1}(N + \nu_0, N(\mu - \bar{y})^2 + \text{RSS}_y + \kappa_0^2)$$

17

Markov Chain Monte Carlo: MCMC

- Metropolis-Hastings
- Gibbs Sampler

18