

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

Bayesian Inference

- Specify the likelihood (joint pdf of data)

$f(y|\theta)$ θ is called the parameter of the model.

- Express your prior belief on the model parameter as the prior distribution. (prior pdf)

$h(\theta)$

- Update your prior belief on the model parameter by combining the prior pdf and the likelihood.

$$\begin{aligned} h(\theta|y) &= \frac{1}{\int_{\Theta} f(y|\theta)h(\theta)d\theta} f(y|\theta)h(\theta) \\ &\propto f(y|\theta)h(\theta) \end{aligned}$$

3

Natural Conjugate Prior

- Nice prior in the sense that the prior and the posterior distributions belong to the same class.

- prior $h(\theta)$
- posterior $h(\theta|y) \propto f(y|\theta)h(\theta)$

- natural conjugate prior depends on the likelihood.

4

How to Find Natural Conjugate Prior

1. Consider the likelihood as the function of θ .

Binomial likelihood $\theta^y (1 - \theta)^{N-y}$

2. Enrich the function by replacing y etc by the hyper-parameters.

Repalce y and N etc. $\theta^{\alpha_0-1} (1 - \theta)^{\beta_0-1}$

3. Identify the resulting distribution of θ .

Beta distribution

$$h(\theta) = \frac{1}{B(\alpha_0, \beta_0)} \theta^{\alpha_0-1} (1 - \theta)^{\beta_0-1}$$

5

Poisson

- pdf

$$f(y|\mu) = \frac{\mu^y}{y!} \exp(-\mu)$$

- joint pdf = likelihood

$$f(\mathbf{y}|\mu) = \prod_{i=1}^N \frac{\mu^{y_i}}{y_i!} \exp(-\mu) = \frac{\mu^{y_+}}{\left(\prod_{i=1}^N y_i!\right)} \exp(-N\mu)$$

$$y_+ = \sum_{i=1}^N y_i$$

- natural conjugate prior pdf

$$h(\mu) = \frac{1}{\Gamma(r_0)\theta_0^{r_0}} \times \exp\left(-\frac{\mu}{\theta_0}\right) \mu^{r_0-1}$$

$$\mu \sim \text{Gamma}(r_0, \theta_0) \quad r_0 = \frac{1}{2}, \frac{1}{\theta_0} = 0$$

6

Poisson (cont)

- likelihood

$$f(\mathbf{y}|\mu) = \prod_{i=1}^N \frac{\mu^{y_i}}{y_i!} \exp(-\mu) = \frac{\mu^{y_+}}{\left(\prod_{i=1}^N y_i!\right)} \exp(-N\mu)$$

- natural conjugate prior pdf

$$h(\mu) = \frac{1}{\Gamma(r_0)\theta_0^{r_0}} \times \exp\left(-\frac{\mu}{\theta_0}\right) \mu^{r_0-1}$$

$\mu \sim \text{Gamma}(r_0, \theta_0)$

- posterior pdf

$$h(\mu|\mathbf{y}) \propto f(\mathbf{y}|\mu) \times h(\mu) =$$

$$\frac{\mu^{y_+}}{\left(\prod_{i=1}^N y_i!\right)} \exp(-N\mu) \times \frac{1}{\Gamma(r_0)\theta_0^{r_0}} \times \exp\left(-\frac{\mu}{\theta_0}\right) \mu^{r_0-1}$$

$$\propto \mu^{y_+ + r_0 - 1} \exp\left(-\frac{\mu(1 + N\theta_0)}{\theta_0}\right) \cdot \text{Gamma}(y_+ + r_0, \frac{1}{\theta_0} + N)$$

7

Poisson Example

- data

number of typos per page
n=5, y_i=1, 8, 3, 0, 2 mean=2.8

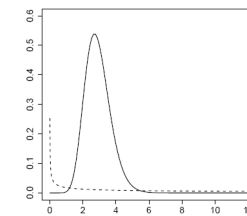
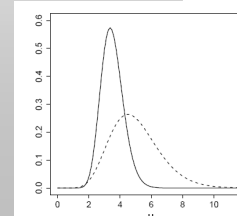
- prior

$$r_0 = 21/2, \theta_0 = 1/2$$

- posterior

$$\mu|\mathbf{y} \sim \text{Gamma}(49/2, 1/7)$$

posterior mean = 3.50, post std = 0.71



8

How to Choose Point Estimate

Expected loss

Choose d which minimizes:

$$\mathcal{E}L(d|\mathbf{y}) = \mathcal{E}_{\Theta}(L(d, \theta)|\mathbf{y}) = \int_{\Theta} L(d, \theta) h(\theta|\mathbf{y}) d\theta$$

Various types of expected loss

$$\text{Squared Error Loss} = (d - \theta)^2$$

minimized by:

mean

$$\text{Bilinear Error Loss} = \begin{cases} a(d - \theta) & \text{if } d \leq \theta \\ b(\theta - d) & \text{if } d > \theta \end{cases}$$

median

$$\text{0-1 Error Loss} = \begin{cases} 0 & \text{if } |d - \theta| \leq b \\ 1 & \text{if } |d - \theta| > b \end{cases}$$

mode

9

Point Estimate

Squared error loss

$$\begin{aligned} \mathcal{E}L(d|\mathbf{y}) &= \int_{\Theta} (d - \theta)^2 h(\theta|\mathbf{y}) d\theta = \int_{\Theta} (d^2 - 2d\theta + \theta^2) h(\theta|\mathbf{y}) d\theta \\ &= d^2 - 2d \int_{\Theta} \theta h(\theta|\mathbf{y}) d\theta + \int_{\Theta} \theta^2 h(\theta|\mathbf{y}) d\theta \end{aligned}$$

Minimum by differentiation

$$d = \int_{\Theta} \theta h(\theta|\mathbf{y}) d\theta = \mathcal{E}(\theta|\mathbf{y})$$

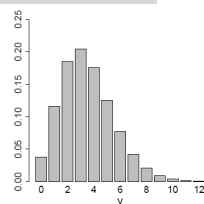
10

Poisson Example

posterior predictive pdf

$$\begin{aligned} h(\tilde{y}|\mathbf{y}) &= \int f(\tilde{y}|\mu) h(\mu|\mathbf{y}) d\mu \\ &= \int \frac{\mu^{\tilde{y}}}{\tilde{y}!} \exp(-\mu) \times \frac{1}{\Gamma(r_1)\theta_1^{r_1}} \times \exp\left(-\frac{\mu}{\theta_1}\right) \mu^{r_1-1} d\mu \\ &\propto \frac{1}{\tilde{y}!} \int \mu^{\tilde{y}+r_1-1} \exp\left[-\mu\left(\frac{\theta_1}{\theta_1+1}\right)^{-1}\right] d\mu = \frac{1}{\tilde{y}!} \times \Gamma(\tilde{y} + r_1) \left(\frac{\theta_1}{\theta_1+1}\right)^{\tilde{y}+r_1} \\ &\propto \frac{1}{\tilde{y}!} \times \Gamma(\tilde{y} + r_1) \left(\frac{\theta_1}{\theta_1+1}\right)^{\tilde{y}} = \frac{\Gamma(\tilde{y} + r_1)}{\Gamma(\tilde{y} + 1)} \left(1 - \frac{1}{\theta_1 + 1}\right)^{\tilde{y}} \end{aligned}$$

$$\tilde{y}|\mathbf{y} \sim \text{negbin}(r_1, \frac{1}{\theta_1 + 1}) \quad r_1 = 49/2, \theta_1 = 1/7$$



11

Normal Model: continuous obs (y)

voltage of battery

α	$\beta = 1.2$	$\beta = 1.5$
0.95	0.001	0.000
1.00	0.005	0.000
1.05	0.030	0.000
1.10	0.104	0.000
1.15	0.220	0.000
1.20	0.282	0.000
1.25	0.220	0.001
1.30	0.104	0.005
1.35	0.030	0.030
1.40	0.005	0.104
1.45	0.001	0.220
1.50	0.000	0.282
1.55	0.000	0.220
1.60	0.000	0.104
1.65	0.000	0.030
1.70	0.000	0.005
1.75	0.000	0.001

A digital voltage meter is used to measure the voltage of a battery whose true voltage is either 1.2v or 1.5v.

The digital voltage meter display the voltage with the step of 0.05 volt as shown in the CPT.

$$\Pr(B = \beta_j | A = \alpha_i) = \frac{\frac{1}{c_j} \exp(-\frac{1}{2}(\alpha_i - \beta_j)^2/0.05) \Pr(B = \beta_j)}{\sum_k^a \frac{1}{c_k} \exp(-\frac{1}{2}(\alpha_i - \beta_k)^2/0.05) \Pr(B = \beta_k)}$$

12

Normal Model (variance known)

pdf

$$Y \sim N(\mu, c^2) \quad f(y|\mu) = \frac{1}{\sqrt{2\pi c^2}} \times \exp\left(-\frac{1}{2}(y - \mu)^2/c^2\right)$$

joint pdf = likelihood

$$\begin{aligned} f(\mathbf{y}|\mu) &= \prod_{i=1}^N \left[\frac{1}{\sqrt{2\pi c^2}} \times \exp\left(-\frac{1}{2}(y_i - \mu)^2/c^2\right) \right] \\ &= \left[\prod_{i=1}^N \frac{1}{\sqrt{2\pi c^2}} \right] \times \exp\left(-\frac{1}{2c^2} \sum_{i=1}^N (y_i - \mu)^2\right) \quad \text{completion of square} \\ &= \left[\prod_{i=1}^N \frac{1}{\sqrt{2\pi c^2}} \right] \times \exp\left(-\frac{\text{RSS}_y}{2c^2} - \frac{1}{2} \frac{N(\mu - \bar{y})^2}{c^2}\right) \quad \text{ANOVA decomposition} \end{aligned}$$

$$\begin{aligned} \bar{y} &= \frac{1}{N} \sum_{i=1}^N y_i \quad \text{sufficient stat} \\ \text{RSS}_y &= \sum_{i=1}^N (y_i - \bar{y})^2 \end{aligned}$$

13

Normal Model (var. known)

likelihood as the function of μ

$$\begin{aligned} f(\mathbf{y}|\mu) &= \left[\prod_{i=1}^N \frac{1}{\sqrt{2\pi c^2}} \right] \times \exp\left(-\frac{\text{RSS}_y}{2c^2} - \frac{1}{2} \frac{N(\mu - \bar{y})^2}{c^2}\right) \\ &\propto \exp\left(-\frac{1}{2} \frac{N(\mu - \bar{y})^2}{c^2}\right) \end{aligned}$$

prior pdf

$$h(\mu) \propto \exp\left(-\frac{1}{2} \frac{(\mu - \mu_0)^2}{\sigma_0^2}\right) \quad \mu \sim N(\mu_0, \sigma_0^2)$$

14

Normal Model (var. known)

posterior

$$\begin{aligned} h(\mu|\mathbf{y}) &\propto f(\mathbf{y}|\mu)h(\mu) \\ &\propto \exp\left(-\frac{1}{2} \frac{N(\mu - \bar{y})^2}{c^2}\right) \times \exp\left(-\frac{1}{2} \frac{(\mu - \mu_0)^2}{\sigma_0^2}\right) \\ &\propto \exp\left[-\frac{1}{2} \left(\frac{N(\mu - \bar{y})^2}{c^2} + \frac{(\mu - \mu_0)^2}{\sigma_0^2} \right)\right] \\ &\propto \exp\left[-\frac{1}{2} \left(\left(\frac{N}{c^2} + \frac{1}{\sigma_0^2} \right) \mu^2 - 2 \left(\frac{N\bar{y}}{c^2} + \frac{\mu_0}{\sigma_0^2} \right) \mu \right)\right] \\ &\propto \exp\left[-\frac{1}{2} \left(\frac{1}{\frac{N}{c^2} + \frac{1}{\sigma_0^2}} \right)^{-1} \left(\mu - \frac{\frac{N\bar{y}}{c^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{N}{c^2} + \frac{1}{\sigma_0^2}} \right)^2\right] \end{aligned}$$

$$\sigma_1^2 = \frac{1}{\frac{N}{c^2} + \frac{1}{\sigma_0^2}}$$

note

$$h(\mu|\mathbf{y}) \sim N\left(\frac{\frac{N\bar{y}}{c^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{N}{c^2} + \frac{1}{\sigma_0^2}}, \frac{1}{\frac{N}{c^2} + \frac{1}{\sigma_0^2}}\right)$$

1/(post. var.) =
1/(prior. var.) + 1/(sampling var.)
post. mean = weighted average of
sample mean and prior mean

15

Normal Model (var. known)

posterior predictive pdf

$$\begin{aligned} h(\tilde{y}|\mathbf{y}) &= \int_{-\infty}^{\infty} f(\tilde{y}|\mu)h(\mu|\mathbf{y})d\mu \\ &\propto \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2c^2}(\tilde{y} - \mu)^2\right) \exp\left(-\frac{1}{2\sigma_1^2}(\mu - \mu_1)^2\right) d\mu \end{aligned}$$

$$\begin{aligned} \tilde{y}|\mu &\sim N(\mu, c^2) \\ \mu|\mathbf{y} &\sim N(\mu_1, \sigma_1^2) \end{aligned}$$

$$\tilde{y}|\mathbf{y} \sim N(\mu_1, c^2 + \sigma_1^2) \quad \sigma_1^2 = \frac{1}{\frac{N}{c^2} + \frac{1}{\sigma_0^2}}$$

note

posterior predictive var is larger.

16

Normal Model (var. known) example

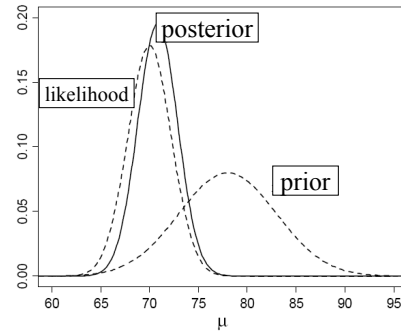
- data and prior

$$N = 20 \quad \bar{y} = 70 \quad \mu_0 = 75, \sigma_0 = 5 \quad c = 10$$

- post mean and std

$$\text{mean} = \frac{70 \times \frac{20}{100} + 75 \times \frac{1}{25}}{\frac{20}{100} + \frac{1}{25}} \approx 70.83$$

$$\text{std} = \sqrt{\frac{1}{\frac{20}{100} + \frac{1}{25}}} \approx 2.04$$



17

Normal Model (mean known)

- likelihood as the function of σ

$$\begin{aligned} f(\mathbf{y}|\sigma) &= \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \times \exp\left(-\frac{1}{2}(y_i - b)^2/\sigma^2\right) \\ &= \left[\prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \right] \times \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - b)^2\right) \\ &= \left[\prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \right] \times \exp\left(-\frac{RSS_b}{2\sigma^2}\right) \quad RSS_b = \sum_{i=1}^N (y_i - b)^2 \\ &\propto \sigma^{-N} \exp\left(-\frac{RSS_b}{2\sigma^2}\right) \end{aligned}$$

- prior

$$\begin{aligned} h(\sigma) &= \frac{(\kappa_0^2)^{\frac{\nu_0}{2}}}{2^{\frac{\nu_0}{2}-1} \Gamma(\frac{\nu_0}{2})} \times \sigma^{-\nu_0-1} \exp\left(-\frac{\kappa_0^2}{2\sigma^2}\right) \quad \sigma \sim \chi^{-1}(\nu_0, \kappa_0^2) \\ &\propto \sigma^{-\nu_0-1} \exp\left(-\frac{\kappa_0^2}{2\sigma^2}\right) \quad \kappa_0^2 = 0, \nu_0 = 0 \end{aligned}$$

18

Normal Model (mean known)

- posterior pdf

$$\begin{aligned} h(\sigma|\mathbf{y}) &\propto f(\mathbf{y}|\sigma)h(\sigma) \\ &\propto \sigma^{-N} \exp\left(-\frac{RSS_b}{2\sigma^2}\right) \times \sigma^{-\nu_0-1} \exp\left(-\frac{\kappa_0^2}{2\sigma^2}\right) \\ &\propto \sigma^{-(\nu_0+N)-1} \exp\left(-\frac{(RSS_b + \kappa_0^2)}{2\sigma^2}\right) \end{aligned}$$

$$\sigma|\mathbf{y} \sim \chi^{-1}(\nu_0 + N, RSS_b + \kappa_0^2)$$

19

Normal Model (mean known)

- posterior predictive

$$\begin{aligned} h(\tilde{y}|\mathbf{y}) &= \int_0^\infty f(\tilde{y}|\sigma) \times h(\sigma|\mathbf{y}) d\sigma \\ &\propto \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} \times \exp\left(-\frac{1}{2}(\tilde{y} - b)^2/\sigma^2\right) \times \sigma^{-\nu_1-1} \exp\left(-\frac{\kappa_1^2}{2\sigma^2}\right) d\sigma \\ &\propto \int_0^\infty \sigma^{-\nu_1-2} \exp\left[-\frac{1}{2\sigma^2}((\tilde{y} - b)^2 + \kappa_1^2)\right] d\sigma = \frac{2^{\frac{\nu_1-1}{2}} \times \Gamma(\frac{\nu_1+1}{2})}{[(\tilde{y} - b)^2 + \kappa_1^2]^{\frac{\nu_1+1}{2}}} \\ &\propto [(\tilde{y} - b)^2 + \kappa_1^2]^{-\frac{\nu_1+1}{2}} \end{aligned}$$

$$\tilde{y}|\mathbf{y} \sim t(\nu_1, b, \frac{\kappa_1^2}{\nu_1})$$

20

Normal Model (both unknown)

likelihood

$$f(\mathbf{y}|\mu, \sigma) \propto \left(\prod_{i=1}^N \frac{1}{\sigma} \right) \times \prod_{i=1}^n \exp \left(-\frac{1}{2} (y_i - \mu)^2 / \sigma^2 \right)$$

$$\propto \sigma^{-N} \times \exp \left(-\frac{1}{2} \frac{\text{RSS}_y}{\sigma^2} - \frac{1}{2} \frac{N(\mu - \bar{y})^2}{\sigma^2} \right)$$

prior

$$h(\mu, \sigma) \propto \sigma^{-\nu_0-2} \exp \left(-\frac{1}{2\sigma^2} (n_0(\mu - \gamma_0)^2 + \kappa_0^2) \right)$$

$$\begin{pmatrix} \mu \\ \sigma \end{pmatrix} \sim N\chi^{-1} \left(\gamma_0, \frac{1}{n_0}, \nu_0, \kappa_0^2 \right) \quad \begin{aligned} \mu|\sigma &\sim N(\gamma_0, \sigma^2 \frac{1}{n_0}) \\ \sigma &\sim \chi^{-1}(\kappa_0^2, \nu_0) \end{aligned}$$

$$n_0 = 0 \quad \nu_0 = -1$$

21

Normal Model (both unknown)

posterior pdf

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

$$\times \sigma^{-\nu_0-2} \exp \left(-\frac{1}{2\sigma^2} (n_0(\mu - \gamma_0)^2 + \kappa_0^2) \right)$$

$$= \sigma^{-N-\nu_0-2} \exp \left[-\frac{1}{2\sigma^2} (\text{RSS}_y + N(\mu - \bar{y})^2 + n_0(\mu - \gamma_0)^2 + \kappa_0^2) \right]$$

$$= \exp \left[-\frac{1}{2\sigma^2} ((\mu - \tilde{\mu})^2 (N + n_0) + s_1 + \text{RSS}_y + \kappa_0^2) \right] \sigma^{-N-\nu_0-2}$$

$$\tilde{\mu} = \frac{1}{N + n_0} (n_0 \gamma_0 + N \bar{y})$$

$$s_1 = \frac{N n_0}{N + n_0} (\bar{y} - \gamma_0)^2$$

$$\begin{pmatrix} \mu \\ \sigma \end{pmatrix} \sim N\chi^{-1} \left(\tilde{\mu}, \frac{1}{N + n_0}, N + \nu_0, \kappa_0^2 + s_1 + \text{RSS}_y \right)$$

22

Normal Model (both unknown)

posterior pdf

$$\begin{pmatrix} \mu \\ \sigma \end{pmatrix} \sim N\chi^{-1} \left(\tilde{\mu}, \frac{1}{N + n_0}, N + \nu_0, \kappa_0^2 + s_1 + \text{RSS}_y \right)$$

$$\mu|\mathbf{y} \sim t(N + \nu_0, \tilde{\mu}, \frac{\kappa_0^2 + s_1 + \text{RSS}_y}{(N + \nu_0)(N + n_0)})$$

$$\sigma|\mathbf{y} \sim \chi^{-1}(N + \nu_0, \kappa_0^2 + s_1 + \text{RSS}_y)$$

23

Normal Model (both unknown) example

data and prior

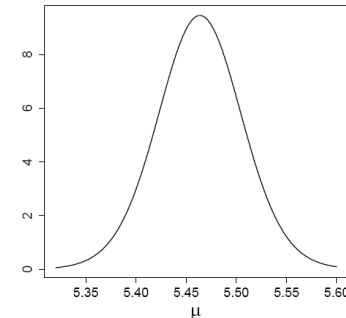
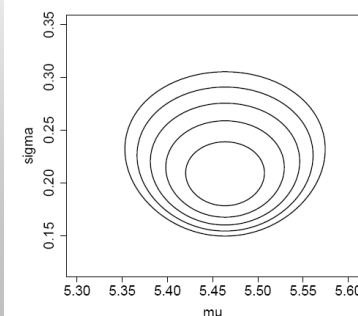
$$N = 23 \quad \bar{y} \approx 5.48, \quad \text{RSS}_y \approx 0.798 \quad \gamma_0 = 5.31, \quad n_0 = 3 \quad \nu_0 = 2 \quad \kappa_0^2 = 0.26$$

posterior

$$\tilde{\mu} \approx 5.46, \quad s_1 \approx 0.078$$

$$\begin{pmatrix} \mu \\ \sigma \end{pmatrix} \sim N\chi^{-1}(5.46, 1/26, 25, 1.13)$$

$$\mu|\mathbf{y} \sim t(25, 5.46, 0.00174)$$



24

Normal Model (both unknown)

- posterior predictive pdf

$$\begin{aligned}h(\tilde{y}|\mathbf{y}) &= \int_0^\infty \int_{-\infty}^\infty f(\tilde{y}|\mu, \sigma) h(\mu, \sigma|\mathbf{y}) d\mu d\sigma \\&= \int_0^\infty \int_{-\infty}^\infty f(\tilde{y}|\mu, \sigma) h(\mu|\sigma, \mathbf{y}) h(\sigma|\mathbf{y}) d\mu d\sigma\end{aligned}$$

$$\tilde{y}|\mu, \sigma \sim N(\mu, \sigma^2)$$

$$\mu|\sigma, \mathbf{y} \sim N(\gamma_1, \frac{\sigma^2}{n_1})$$

$$\sigma|\mathbf{y} \sim \chi^{-1}(\nu_1, \kappa_1^2)$$

$$\tilde{y}|\mathbf{y} \sim t(\nu_1, \gamma_1, \frac{\kappa_1^2(n_1 + 1)}{\nu_1 n_1})$$