

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)

$$f(y|\theta) \quad \theta \text{ is called the parameter of the model.}$$

- 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.

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

3

Posterior Distribution

- Normal

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

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

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

- Beta

$$h(\theta|y) \propto \theta^{y+\alpha_0-1} (1-\theta)^{N-y+\beta_0-1}$$

$$\theta | \sim \text{Beta}(y + \alpha_0, N - y + \beta_0)$$

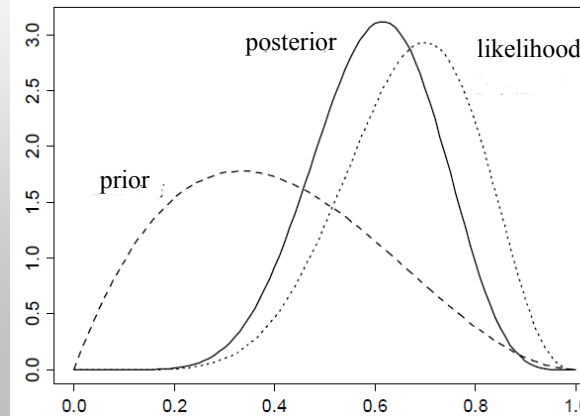
4

Posterior Distribution

- Posterior distribution of θ , $h(\theta|y)$, contains all the information about θ obtained from the observed data y .
- We can calculate the following:
 - posterior probability
 $\Pr(\theta_1 < \theta < \theta_2 | y) \rightarrow$ **credibility interval**
 hypothesis testing
- How close is the given value, θ_1 , to the center of the posterior distribution?
 $\Pr(h(\theta | y) > h(\theta_1 | y)) \rightarrow$ **Highest Posterior Density Region**
HPDR or HDR
 hypothesis testing

Example

- Ten tosses of a coin:
 H T H H T H H H H T $N=10, y=7$
- Prior: $B(2,3)$
- Posterior: $B(9,6) = B(2+7, 3+10-7)$



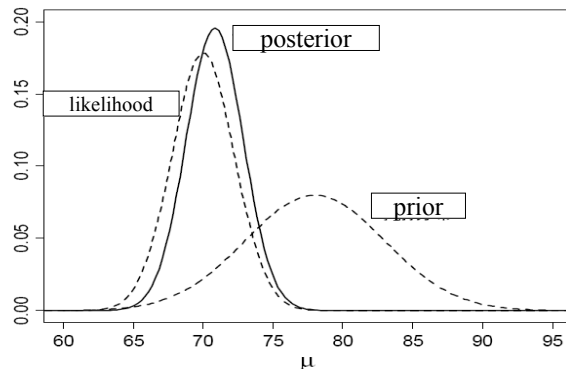
	prior	posterior
mean	0.40	0.60
mode	0.33	0.62
std	0.20	0.12

Posterior var(std) is smaller than prior var.

Normal Model (var. known) example

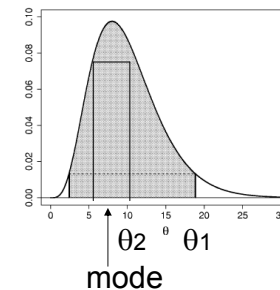
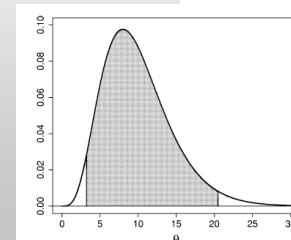
- data and prior
 $N = 20 \quad \bar{y} = 70 \quad \mu_0 = 75, \sigma_0 = 5 \quad c = 10$
- post mean and std

$$\text{mean} = \frac{70 \times \frac{20}{100} + 75 \times \frac{1}{25}}{\frac{20}{100} + \frac{1}{25}} \approx 70.83 \quad \text{std} = \sqrt{\frac{1}{\frac{20}{100} + \frac{1}{25}}} \approx 2.04$$



HPDR or HDR

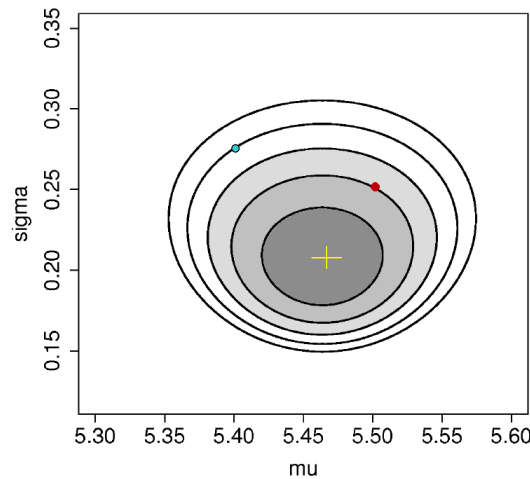
- How close is the given value, θ_1 , to the center (mode) of the posterior distribution?
 Think of the density $h(\theta | y)$ as a random variable and calculate
 $\Pr(h(\theta | y) > h(\theta_1 | y)) = \alpha.$ When h is t , **F** distribution
 Small α indicates closeness.



shortest-length interval of the size $1 - \alpha$

HPDR or HDR

- Bivariate case:
How plausible is the value (5.50, 0.25)?



shortest-area region of the size $1 - \alpha$

Comparison of Two Proportions

- Distribution of the difference of two independent beta random variables

$$\pi_1 \sim \text{Be}(\alpha_1, \beta_1), \quad \pi_2 \sim \text{Be}(\alpha_2, \beta_2),$$

For example, we would like to calculate

$$\Pr(\pi_1 > \pi_2 | y) = \Pr(\pi_1 - \pi_2 > 0 | y).$$

Let $d = \pi_1 - \pi_2$ and $d_2 = \pi_2$ and use **convolution**.

change of variable technique

$$\begin{aligned} h(d|n_1, x_1, n_2, x_2) &= \int_{\Pi_2} h_1(d + \pi_2 | n_1, x_1) h_2(\pi_2 | n_2, x_2) d\pi_2 \\ &= \begin{cases} \int_{-d}^1 h_1(d + \pi_2 | n_1, x_1) h_2(\pi_2 | n_2, x_2) d\pi_2 & \text{if } -1 < d < 0 \\ \int_0^{1-d} h_1(d + \pi_2 | n_1, x_1) h_2(\pi_2 | n_2, x_2) d\pi_2 & \text{if } 0 < d < 1 \end{cases} \end{aligned}$$

Comparison of Two Proportions

- Example:

$$\pi_1 \sim \text{Be}(\alpha_1, \beta_1), \quad \pi_2 \sim \text{Be}(\alpha_2, \beta_2), \quad \alpha_1=6, \beta_1=3, \alpha_2=3, \beta_2=5$$

$$h(d|n_1, x_1, n_2, x_2)$$

$$= \begin{cases} \frac{7}{143}(d+1)^{10}(d(d(5d(3d-14)+116)-70)+15) & \text{if } -1 < d \leq 0 \\ -\frac{7}{143}(d-1)^5(d+1)(d(d(d(d(d(5d(3d+28)+576)+1344)+1856)+1344)+576)+140)+15) & \text{if } 0 < d < 1 \end{cases}$$

$$\Pr(\pi_1 > \pi_2 | n_1, x_1, n_2, x_2) = \Pr(d > 0 | n_1, x_1, n_2, x_2) =$$

$$= \int_0^1 h(d|n_1, x_1, n_2, x_2) dd = \frac{386}{429} \approx 0.900$$

Comparison of Two Proportions

- Numerical evaluation: **Monte Carlo** integration

In order to calculate $\Pr(\pi_1 > \pi_2 | y)$,

Generate many pairs of random variables:

$$\Pi_1 \sim \text{Be}(\alpha_1, \beta_1), \quad \Pi_2 \sim \text{Be}(\alpha_2, \beta_2).$$

For each pair, calculate the difference

$$D = \Pi_1 - \Pi_2$$

Calculate the proportion of $D > 0$.

More generally, calculate the frequency distribution of D .

Table of D_p : $\Pr(D < D_p) = p$

D_p percentiles

p	0.025	0.05	0.10	0.20	0.25	0.50	0.75	0.80	0.90	0.95	0.975
D_p	-0.162	-0.088	-0.000	0.104	0.144	0.302	0.500	0.484	0.571	0.636	0.688

Comparison of Two Proportions

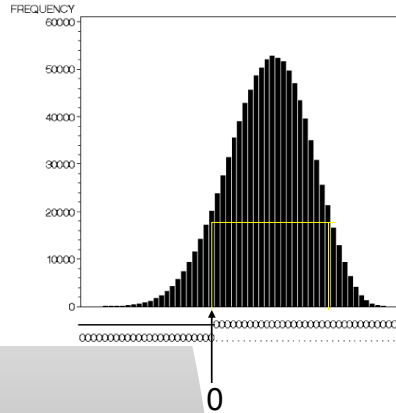


Table of D_p : $\Pr(D < D_p) = p$

p	0.025	0.05	0.10	0.20	0.25	0.50	0.75	0.80	0.90	0.95	0.975
D_p	-0.162	-0.088	-0.000	0.104	0.144	0.302	0.500	0.484	0.571	0.636	0.688

Comparison of Two Normal Means

- Distribution of the difference of two independent Normal random variables (variance known)

$$\Pr(\mu_1 - \mu_2 > 0 \mid y) \leftarrow \text{Normal Distribution}$$

- Distribution of the difference of two independent t random variables (variance unknown)

$$\Pr(\mu_1 - \mu_2 > 0 \mid y) \leftarrow \text{Behrens-Fisher Distribution}$$

Comparison of Two Normal Means

- Numerical evaluation by Monte Carlo method
- Example:

data

$$n_1 = 20, \bar{y}_1 = 50, S_1^2 = 11.4,$$

$$n_2 = 12, \bar{y}_2 = 55, S_2^2 = 36.667$$

posterior t hyper params

$$\gamma_1^* = 50, \nu_1^* = 19, \kappa_1^* = 0.6,$$

$$\gamma_2^* = 55, \nu_2^* = 11, \kappa_2^* = 3.333$$

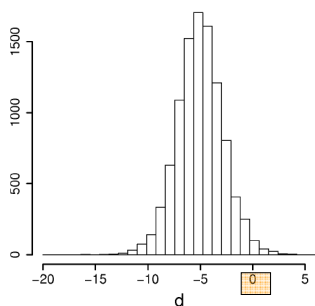


Table of D_p : $\Pr(D < D_p) = p$

0.025	0.05	0.10	0.20	0.25	0.50
-9.310	-8.536	-7.696	-6.745	-6.391	-4.998
0.50	0.75	0.80	0.90	0.95	0.975
-4.998	-3.604	-3.250	-2.285	-1.453	-0.682

Hypothesis Testing

- testing hypotheses such as

$$H_1: \mu_1 = \mu_2 \quad \text{v.s.} \quad H_2: \mu_1 \neq \mu_2$$

Derive the posterior distribution of $\mu_1 - \mu_2$ and calculate the probability such as

$$\Pr(\mu_1 - \mu_2 > 0 \mid y) .$$

Calculate the **HDR** which just contains 0 and check if α is small enough.

Bayes Factor, BIC, DIC etc

Normal Regression Model

- model

The observation y depends on the regressor variables X which are treated as known constant.

$$y_i = \sum_{j=1}^q x_{ij}\beta_j + e_i, \quad e_i \sim N(0, \sigma^2)$$

$$y_i \sim N\left(\sum_{j=1}^q x_{ij}\beta_j, \sigma^2\right)$$

In matrix notation:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad \mathbf{e} \sim MVN(\mathbf{0}, \sigma^2 \mathbf{I})$$

17

Normal Regression Model

- normal model

$$f(y_i | \boldsymbol{\beta}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \sum_{j=1}^q x_{ij}\beta_j)^2}{2\sigma^2}\right)$$

- likelihood

$$\begin{aligned} f(\mathbf{y} | \boldsymbol{\beta}, \sigma^2) &= \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \sum_{j=1}^q x_{ij}\beta_j)^2}{2\sigma^2}\right) \\ &\propto \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})/\sigma^2\right) \end{aligned}$$

ANOVA decomposition

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'(\mathbf{X}'\mathbf{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad R_{ss\boldsymbol{\beta}} = (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'(\mathbf{X}'\mathbf{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

$$R_{ss_y} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

18

ANOVA decomposition

scalar

$$\sum_{i=1}^N (y_i - \mu)^2 = N(\mu - \bar{y})^2 + R_{ss_y}$$

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i \quad \leftarrow \text{LS or ML.}$$

$$R_{ss_y} = \sum_{i=1}^N (y_i - \bar{y})^2 \quad \leftarrow \text{does not contain } \mu.$$

vector

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'(\mathbf{X}'\mathbf{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad \leftarrow \text{LS or ML.}$$

$$R_{ss_y} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad \leftarrow \text{does not contain } \boldsymbol{\beta}.$$

$$R_{ss\boldsymbol{\beta}} = (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'(\mathbf{X}'\mathbf{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

19

Normal Regression Model

- likelihood (joint pdf of \mathbf{y} as a function of $(\boldsymbol{\beta}, \sigma)$)

$$\begin{aligned} f(\mathbf{y} | \boldsymbol{\beta}, \sigma^2) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\left(-\frac{1}{2\sigma^2}(R_{ss\boldsymbol{\beta}} + R_{ss_y})\right) \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^N (\sigma^2)^{-N/2} \exp\left(-\frac{1}{2\sigma^2}(R_{ss\boldsymbol{\beta}} + R_{ss_y})\right) \\ &\propto \sigma^{-N} \exp\left(-\frac{1}{2\sigma^2} \left[(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'(\mathbf{X}'\mathbf{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + R_{ss_y} \right]\right) \end{aligned}$$

- natural conjugate prior for $(\boldsymbol{\beta}, \sigma)$

$$\begin{pmatrix} \boldsymbol{\beta} \\ \sigma \end{pmatrix} \sim MVN\chi^{-1}(\boldsymbol{\gamma}_0, \boldsymbol{\Phi}_0, \nu_0, \kappa_0^2)$$

$$h(\boldsymbol{\beta}, \sigma) \propto \sigma^{-\nu_0 - q - 1} \exp\left[-\frac{1}{2\sigma^2} \left((\boldsymbol{\beta} - \boldsymbol{\gamma}_0)' \boldsymbol{\Phi}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\gamma}_0) + \kappa_0^2 \right)\right]$$

20

Normal Regression Model

- joint posterior pdf of (β, σ)

$$\begin{aligned}
 h(\beta, \sigma | \mathbf{y}) &\propto f(\mathbf{y} | \beta, \sigma) h(\beta, \sigma) \\
 &= \sigma^{-N} \exp \left[-\frac{1}{2\sigma^2} \left((\beta - \hat{\beta})' (\mathbf{X}' \mathbf{X}) (\beta - \hat{\beta}) + R_{ssy} \right) \right] \\
 &\quad \times \sigma^{-\nu_0 - p - 1} \exp \left[-\frac{1}{2\sigma^2} \left((\beta - \gamma_0)' \Phi_0^{-1} (\beta - \gamma_0) + \kappa_0^2 \right) \right] \\
 &= \exp \left[-\frac{1}{2\sigma^2} \left((\beta - \hat{\beta})' (\mathbf{X}' \mathbf{X}) (\beta - \hat{\beta}) + (\beta - \gamma_0)' \Phi_0^{-1} (\beta - \gamma_0) + R_{ssy} + \kappa_0^2 \right) \right] \\
 &\quad \times \sigma^{-N - \nu_0 - q - 1} \quad \text{completion of square: sum of two QFs} \\
 &\propto \exp \left[-\frac{1}{2\sigma^2} \left((\beta - \tilde{\beta})' (\mathbf{X}' \mathbf{X} + \Phi_0^{-1}) (\beta - \tilde{\beta}) \right. \right. \\
 &\quad \left. \left. + (\hat{\beta} - \gamma_0)' \left((\mathbf{X}' \mathbf{X})^{-1} + \Phi_0 \right)^{-1} (\hat{\beta} - \gamma_0) + R_{ssy} + \kappa_0^2 \right) \right] \\
 &\quad \times \sigma^{-N - \nu_0 - q - 1}
 \end{aligned}$$

21

Sum of Two Quadratic Forms

scalar	$ax^2 + bx + c = a \left(x + \frac{b}{2a} \right)^2 - \frac{b^2 - 4ac}{4a}$
vector	$\mathbf{x}' \mathbf{A} \mathbf{x} - 2\mathbf{b}' \mathbf{x} = (\mathbf{x} - \mathbf{d})' \mathbf{A} (\mathbf{x} - \mathbf{d}) - \mathbf{b}' \mathbf{A} \mathbf{b}$ where $\mathbf{d} = \mathbf{A}^{-1} \mathbf{b}$
sum of two quad forms	$(\mathbf{x} - \mathbf{v}_1)' \mathbf{W}_1 (\mathbf{x} - \mathbf{v}_1) + (\mathbf{x} - \mathbf{v}_2)' \mathbf{W}_2 (\mathbf{x} - \mathbf{v}_2)$ $= (\mathbf{x} - \tilde{\mathbf{x}})' (\mathbf{W}_1 + \mathbf{W}_2) (\mathbf{x} - \tilde{\mathbf{x}})$ $+ (\mathbf{v}_1 - \mathbf{v}_2)' \mathbf{W}_1 (\mathbf{W}_1 + \mathbf{W}_2)^{-1} \mathbf{W}_2 (\mathbf{v}_1 - \mathbf{v}_2)$ $= (\mathbf{x} - \tilde{\mathbf{x}})' (\mathbf{W}_1 + \mathbf{W}_2) (\mathbf{x} - \tilde{\mathbf{x}})$ $+ (\mathbf{v}_1 - \mathbf{v}_2)' (\mathbf{W}_1^{-1} + \mathbf{W}_2^{-1})^{-1} (\mathbf{v}_1 - \mathbf{v}_2)$ where $\tilde{\mathbf{x}} = (\mathbf{W}_1 + \mathbf{W}_2)^{-1} (\mathbf{W}_2 \mathbf{v}_2 + \mathbf{W}_1 \mathbf{v}_1)$

22

Normal Regression Model

- joint posterior pdf of (β, σ)

$$\begin{aligned}
 h(\beta, \sigma | \mathbf{y}) &\propto \\
 &\exp \left[-\frac{1}{2\sigma^2} \left((\beta - \tilde{\beta})' (\mathbf{X}' \mathbf{X} + \Phi_0^{-1}) (\beta - \tilde{\beta}) + s_1 + R_{ssy} + \kappa_0^2 \right) \right] \\
 &\quad \times \sigma^{-N - \nu_0 - q - 1}
 \end{aligned}$$

where

$$\begin{aligned}
 \tilde{\beta} &= (\mathbf{X}' \mathbf{X} + \Phi_0^{-1})^{-1} (\Phi_0^{-1} \gamma_0 + \mathbf{X}' \mathbf{X} \hat{\beta}) \\
 s_1 &= (\hat{\beta} - \gamma_0)' \left((\mathbf{X}' \mathbf{X})^{-1} + \Phi_0 \right)^{-1} (\hat{\beta} - \gamma_0) \\
 &= (\hat{\beta} - \gamma_0)' \mathbf{X}' \mathbf{X} (\mathbf{X}' \mathbf{X} + \Phi_0^{-1})^{-1} \Phi_0^{-1} (\hat{\beta} - \gamma_0)
 \end{aligned}$$

$$\begin{pmatrix} \beta \\ \sigma \end{pmatrix} \sim N\chi^{-1} \left(N + \nu_0, \tilde{\beta}, (\mathbf{X}' \mathbf{X} + \Phi_0^{-1})^{-1}, R_{ssy} + \kappa_0^2 + s_1 \right)$$

23

Normal Regression Model

- joint posterior distribution of (β, σ)

$$\begin{pmatrix} \beta \\ \sigma \end{pmatrix} \sim N\chi^{-1} \left(N + \nu_0, \tilde{\beta}, (\mathbf{X}' \mathbf{X} + \Phi_0^{-1})^{-1}, R_{ssy} + \kappa_0^2 + s_1 \right)$$

- marginal posterior distribution of (β, σ)

$$\beta | \mathbf{y} \sim MVt \left(N + \nu_0, \tilde{\beta}, \frac{R_{ssy} + \kappa_0^2 + s_1}{N + \nu_0} (\mathbf{X}' \mathbf{X} + \Phi_0^{-1})^{-1} \right)$$

$$\sigma | \mathbf{y} \sim \chi^{-1} \left(N + \nu_0, R_{ssy} + \kappa_0^2 + s_1 \right)$$

non-informative prior $\nu_0 = -q, \kappa_0^2 = 0, \Phi_0^{-1} = \mathbf{O}$

24

Normal Regression Model

- predictive distribution

$$f(\tilde{y}|\mathbf{y}) = \int_0^\infty \int_{-\infty}^\infty f(\tilde{y}|\beta, \sigma)h(\beta, \sigma|\mathbf{y})d\beta d\sigma$$

$$= \int_0^\infty \int_{-\infty}^\infty f(\tilde{y}|\beta, \sigma)h(\beta|\sigma, \mathbf{y})h(\sigma|\mathbf{y})d\beta d\sigma$$

$$\tilde{\mathbf{y}} \sim MVt(\nu_1, \mathbf{X}\gamma_1, \frac{\kappa_1^2}{\nu_1}(\mathbf{X}\Phi_1\mathbf{X}' + \mathbf{I}))$$

$$\tilde{y} \sim t(\nu_1, \mathbf{x}'\gamma_1, \frac{\kappa_1^2}{\nu_1}(\mathbf{x}'\Phi_1\mathbf{x} + 1))$$

25

Normal Regression Model

- univariate Normal model as a submodel

$$\mathbf{X} \rightarrow \mathbf{1} \quad \gamma_0 \rightarrow \gamma_0$$

$$\beta \rightarrow \mu \quad \Phi_0 \rightarrow \frac{1}{n_0}$$

$$\hat{\beta} = \frac{1}{N} \sum_{i=1}^N y_i = \bar{y}$$

$$\tilde{\beta} = \frac{1}{N + \phi^{-1}}(\phi^{-1}\gamma_0 + N\bar{y}) = \frac{1}{N + n_0^{-1}}(n_0\gamma_0 + N\bar{y})$$

$$s_1 = \frac{1}{\frac{1}{N} + \phi}(\bar{y} - \gamma_0)^2 = \frac{1}{\frac{1}{N} + \frac{1}{n_0}}(\bar{y} - \gamma_0)^2$$

$$Rss_y = \sum_{i=1}^N (y_i - \bar{y})^2$$

26

Normal Regression: Example

Willerman et al. (1991) IQ and the Brain Size

X0	IQ	Brain Size	Weight	Gender
1	99	92.8799	146	女 0
1	92	85.4258	175	女 0
1	89	90.4858	134	男 1
1	96	87.8897	146	女 0
1	83	86.5363	135	女 0
1	100	94.5088	178	男 1
1	101	80.8020	136	女 0
1	80	88.9083	180	男 1
1	97	90.5940	186	男 1
1	91	83.1772	114	女 0
1	85	79.8612	140	女 0
1	103	106.2462	187	男 1
1	77	79.3549	106	女 0
1	103	99.7925	192	男 1
1	90	87.9987	181	男 1
1	83	83.4344	143	女 0
1	88	89.3983	139	女 0
1	81	93.0016	148	男 1
1	89	93.5863	179	男 1
mean	90.9	89.15	155.0	

N=19, q=4

y : IQ (100,15)
X0: intercept
X1: brain size (pixel)
X2: body weight (Lb)
X3: gender

$$\beta = \begin{pmatrix} \beta_0 \text{ int} \\ \beta_1 \text{ b.s.} \\ \beta_2 \text{ b.w.} \\ \beta_3 \text{ gen} \end{pmatrix}$$

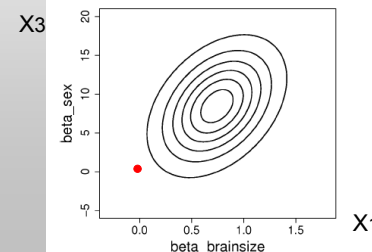
27

Normal Regression: Example

- non-informative analysis

$$\hat{\beta} = \begin{pmatrix} 0.027 \\ 0.742 \\ 0.131 \\ 8.442 \end{pmatrix}, \quad \mathbf{X}'\mathbf{X} = \begin{pmatrix} 19 & 1694 & 2945 & 10 \\ 1694 & 151843 & 264719 & 849 \\ 2945 & 264719 & 468755 & 1380 \\ 10 & 849 & 1380 & 10 \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \text{ int} \\ \beta_1 \text{ b.s.} \\ \beta_2 \text{ b.w.} \\ \beta_3 \text{ gen} \end{pmatrix}$$

$$\gamma = \begin{pmatrix} 0.027 \\ 0.742 \\ 0.131 \\ 8.442 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 917.595 & -9.079 & -0.345 & -99.379 \\ -9.079 & 0.119 & -0.012 & 0.638 \\ -0.345 & -0.012 & 0.009 & 0.198 \\ -99.379 & 0.638 & 0.198 & 22.427 \end{pmatrix}$$



How close is (0,0) to the mode?

HDR by F

28

One Way Analysis of Variance comparison of means

- q = 2 groups, independent observations

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{N_1,1} \\ y_{12} \\ y_{22} \\ \vdots \\ y_{N_2,2} \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

$e \sim MVN(0, \sigma^2 \mathbf{I})$
common error variance

29

One Way Analysis of Variance comparison of means

- prior $MVN\chi^{-1}(\gamma_0, \Phi_0, \nu_0, \kappa_0^2)$
- posterior

$$\boldsymbol{\beta} | \mathbf{y} \sim MVt \left(N + \nu_0, \tilde{\boldsymbol{\beta}}, \frac{Rss_y + \kappa_0^2 + s_1}{N + \nu} (\mathbf{X}'\mathbf{X} + \Phi_0^{-1})^{-1} \right)$$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} N_1 & 0 \\ 0 & N_2 \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix}$$

$$Rss_y = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \sum_{j=1}^2 \sum_{i=1}^{N_i} (y_{ij} - \bar{y}_j)^2$$

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X} + \Phi_0^{-1})^{-1} (\Phi_0^{-1} \gamma_0 + \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}})$$

$$s_1 = (\hat{\boldsymbol{\beta}} - \gamma_0)' ((\mathbf{X}'\mathbf{X})^{-1} + \Phi_0)^{-1} (\hat{\boldsymbol{\beta}} - \gamma_0)$$

30

One Way Analysis of Variance comparison of means

- linear contrast

$$d = \mu_1 - \mu_2 = \mathbf{a}'\boldsymbol{\beta} \quad \mathbf{a}' = (1 \ -1)$$

$$d | \mathbf{y} \sim t \left(N + \nu_0, \mathbf{a}'\tilde{\boldsymbol{\beta}}, \frac{Rss_y + \kappa_0^2 + s_1}{N_1 + N_2 + \nu_0} \mathbf{a}'(\mathbf{X}'\mathbf{X} + \Phi_0^{-1})^{-1} \mathbf{a} \right)$$

non-informative prior $\Phi_0^{-1} = \mathbf{O}, \kappa_0^2 = 0, \nu_0 = -2$

$$d | \mathbf{y} \sim t \left(N_1 + N_2 - 2, \bar{y}_1 - \bar{y}_2, \frac{Rss_y}{N_1 + N_2 - 2} \left(\frac{1}{N_1} + \frac{1}{N_2} \right) \right)$$

HDR by F

31

Two Way Analysis of Variance

- additive model

y_{kij} k-th observation at the cell i-j

$$y_{kij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{kij}, \quad e_{kij} \sim N(0, \sigma^2)$$

cell means

$$\mathcal{E}(y_{kij}) = \mu_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$$

zero-sum constraints

$$\sum_{i=1}^{n_A} \alpha_i = 0, \quad \sum_{j=1}^{n_B} \beta_j = 0 \quad \alpha_{n_A} = - \sum_{i=1}^{n_A-1} \alpha_i \quad \text{and} \quad \beta_{n_B} = - \sum_{j=1}^{n_B-1} \beta_j$$

$$\sum_{i=1}^{n_A} \alpha\beta_{ij} = \sum_{j=1}^{n_B} \alpha\beta_{ij} = 0$$

32

Two Way Analysis of Variance

- parameters in terms of the cell means

		factor B					means
		1	...	j	...	n_B	
factor A	1	μ_{11}	...	μ_{1j}	...	μ_{1n_B}	$\mu_{1.}$
	\vdots	\vdots					
	i	μ_{i1}		μ_{ij}		μ_{in_B}	$\mu_{i.}$
	\vdots	\vdots					
	n_A	$\mu_{n_A 1}$		$\mu_{n_A j}$		$\mu_{n_A n_B}$	$\mu_{n_A.}$
means		$\mu_{.1}$		$\mu_{.j}$		$\mu_{.n_B}$	μ

33

Two Way Analysis of Variance

- parameters in terms of the cell means

$$\mu = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \mu_{ij}$$

$$\alpha_i = \mu_{i.} - \mu$$

$$\beta_j = \mu_{.j} - \mu$$

$$\alpha\beta_{ij} = \mu_{ij} - (\mu + \alpha_i + \beta_j)$$

$$\mu_{i.} = \frac{1}{n_B} \sum_{j=1}^{n_B} \mu_{ij}$$

$$\mu_{.j} = \frac{1}{n_A} \sum_{i=1}^{n_A} \mu_{ij}$$

34

Two Way ANOVA: Design Matrix

$$\mathbf{y} = \begin{pmatrix} y_{111} \\ y_{211} \\ \vdots \\ y_{N_{111}} \\ y_{112} \\ y_{212} \\ \vdots \\ y_{N_{1212}} \\ y_{113} \\ y_{213} \\ \vdots \\ y_{N_{1313}} \\ y_{121} \\ y_{221} \\ \vdots \\ y_{N_{221}} \\ y_{122} \\ y_{222} \\ \vdots \\ y_{N_{222}} \\ y_{123} \\ y_{223} \\ \vdots \\ y_{N_{2323}} \end{pmatrix}
 \quad
 \mathbf{X} = \begin{pmatrix} \mu & \alpha_1 & \beta_1 & \beta_2 & \alpha\beta_{11} & \alpha\beta_{12} \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & -1 & -1 & -1 & 1 & 1 \end{pmatrix}
 \quad
 \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ \beta_2 \\ \alpha\beta_{11} \\ \alpha\beta_{12} \end{pmatrix}
 \quad
 \mathbf{X} = [\mathbf{1}, \mathbf{X}_\alpha, \mathbf{X}_\beta, \mathbf{X}_{\alpha\beta}]$$

35

Two Way ANOVA

- posterior distribution

$$N = \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} N_{ij}$$

$$R_{SS_y} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \sum_{k=1}^{N_{ij}} (y_{kij} - \bar{y}_{ij})^2$$

$$\hat{\mu} = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \bar{y}_{ij}$$

$$\hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..}$$

$$\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}$$

$$\hat{\alpha}\hat{\beta}_{ij} = \bar{y}_{ij} - (\bar{y}_{..} + \hat{\alpha}_i + \hat{\beta}_j) = \bar{y}_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$$

36

Two Way ANOVA: non-informative

$$C = (X'X + \Phi_0^{-1})^{-1}$$

$$\hat{\beta} = (X'X)^{-1} X'y$$

$$\beta = \begin{pmatrix} \mu \\ a \\ b \\ ab \end{pmatrix} \quad C = \begin{bmatrix} c_{\mu} & c'_{\mu,\alpha} & c'_{\mu,\beta} & c'_{\mu,\alpha\beta} \\ c_{\mu,\alpha} & C_{\alpha} & C_{\alpha,\beta} & C_{\alpha,\alpha\beta} \\ c_{\mu,\beta} & C_{\beta,\alpha} & C_{\beta} & C_{\beta,\alpha\beta} \\ c_{\mu,\alpha\beta} & C_{\alpha\beta,\alpha} & C_{\alpha\beta,\beta} & C_{\alpha\beta} \end{bmatrix} = \begin{bmatrix} \frac{1}{N} & 0' & 0' & 0' \\ 0 & (X'_a X_a)^{-1} & 0 & 0 \\ 0 & 0 & (X'_b X_b)^{-1} & 0 \\ 0 & 0 & 0 & (X'_{\alpha\beta} X_{\alpha\beta})^{-1} \end{bmatrix}$$

posterior distribution

$$\mu|y \sim t(\tilde{\mu}, c_{\mu}, N + \nu)$$

$$a|y \sim t(\tilde{a}, C_{\alpha}, N + \nu)$$

$$b|y \sim t(\tilde{b}, C_{\beta}, N + \nu)$$

$$ab|y \sim t(\tilde{ab}, C_{\alpha\beta}, N + \nu)$$

HDR by F

$$F_{\alpha} = \frac{1}{n_A - 1} \frac{N - n_A n_B}{Rss_y} \tilde{a}' C_{\alpha}^{-1} \tilde{a}$$

$$F_{\beta} = \frac{1}{n_B - 1} \frac{N - n_A n_B}{Rss_y} \tilde{b}' C_{\beta}^{-1} \tilde{b}$$

$$F_{\alpha\beta} = \frac{1}{(n_A - 1)(n_B - 1)} \frac{N - n_A n_B}{Rss_y} \tilde{ab}' C_{\alpha\beta}^{-1} \tilde{ab}$$