Monte Carlo Statistical Methods

George Casella
University of Florida
July 27, 2006
casella@stat.ufl.edu

Based on

- Monte Carlo Statistical Methods,
  Christian Robert and George Casella,
  2004, Springer-Verlag
- Programming in R (available as a free download from
  http://www.r-project.org
- Also WinBugs, available free from
  http://www.mrc-bsu.cam.ac.uk/bugs/
- R programs for the course available at
  http://www.stat.ufl.edu/ casella/mcsm/

The Multi-Stage Gibbs Sampler

- Suppose that for some $p > 1$, the random variable $\mathbf{X} \in \mathcal{X}$ can be written as $\mathbf{X} = (X_1, \ldots, X_p)$, where the X_i's are either uni- or multidimensional.
- Moreover, suppose that we can simulate from the corresponding univariate conditional densities $f_1, \ldots, f_p$, that is, we can simulate

  $$X_i | x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p \sim f_i(x_i | x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p)$$

  for $i = 1, 2, \ldots, p$.

The Multi-Stage Gibbs Sampler

Given $\mathbf{x}^{(t)} = (x_1^{(t)}, \ldots, x_p^{(t)})$, generate

1. $X_1^{(t+1)} \sim f_1(x_1|x_2^{(t)}, \ldots, x_p^{(t)})$;

2. $X_2^{(t+1)} \sim f_2(x_2|x_1^{(t+1)}, x_3^{(t)}, \ldots, x_p^{(t)})$,

$\vdots$

p. $X_p^{(t+1)} \sim f_p(x_p|x_1^{(t+1)}, \ldots, x_{p-1}^{(t+1)})$.

- The densities $f_1, \ldots, f_p$ are called the *full conditionals*
- These are the only densities used for simulation, even in a high-dimensional problem.

Hierarchical Models - Introduction

- A hierarchical model is of the form

$$\begin{aligned} \mathbf{X} &\sim f(\mathbf{x}|\theta) \\ \theta &\sim g(\theta|\beta) \\ \beta &\sim h(\beta|\lambda) \\ \lambda &\sim k(\lambda) \end{aligned}$$

  - All hyperparameters specified at deepest level
  - Effect of deeper hyperparameters is lower

- Easy to get joint distribution
- Easy to pick off full conditionals

Hierarchical Models - Introduction - 2

- Hierarchical Model

$$\begin{aligned} \mathbf{X} &\sim f(\mathbf{x}|\theta) \\ \theta &\sim \pi(\theta|\beta) \\ \beta &\sim \pi(\beta|\lambda) \\ \lambda &\sim \pi(\lambda) \end{aligned}$$

- Joint distribution

$$f(\mathbf{x}|\theta) \times \pi(\theta|\beta) \times \pi(\beta|\lambda) \times \pi(\lambda)$$

- Full Conditionals

$$\pi(\theta|\mathbf{x}, \beta, \lambda) \propto \text{ terms in joint involving } \theta$$

etc...

Hierarchical Models - Introduction - 3

- Normal Hierarchical Model (Conjugate)

$$\begin{aligned} \mathbf{X} &\sim N(\theta, \sigma^2) \\ \theta &\sim N(\theta_0, \tau^2\sigma^2) \\ \sigma^2 &\sim \text{Inverted Gamma}(a, b) \end{aligned}$$

- Here θ_0, τ^2, a, b are specified
  - Usual to take $\tau^2 \approx 10$ (variance ratio)
  - Choose a, b to give prior a big variance

Normal Hierarchical Models

- Normal Hierarchical Model

$$\begin{aligned} X_i &\sim N(\theta, \sigma^2), \quad i = 1, \ldots, n \\ \theta &\sim N(\theta_0, \tau^2\sigma^2) \\ \sigma^2 &\sim \text{Inverted Gamma}(a, b) \end{aligned}$$

- Joint Distribution

$$f(\mathbf{x}, \theta, \sigma^2) \propto \left[\frac{1}{\sigma} e^{-\sum_i (x_i - \theta)^2/(2\sigma^2)}\right] \times \left[\frac{1}{\tau\sigma} e^{-(\theta-\theta_0)^2/(2\tau^2\sigma^2)}\right] \times \left[\frac{1}{(\sigma^2)^{a+1}} e^{1/b\sigma^2}\right]$$

Normal Hierarchical Models -2

- Joint Distribution

$$f(\mathbf{x}, \theta, \sigma^2) \propto \left[\frac{1}{\sigma} e^{-\sum_i (x_i - \theta)^2/(2\sigma^2)}\right] \times \left[\frac{1}{\tau\sigma} e^{-(\theta-\theta_0)^2/(2\tau^2\sigma^2)}\right] \times \left[\frac{1}{(\sigma^2)^{a+1}} e^{1/b\sigma^2}\right]$$

- θ full conditional

$$\pi(\theta|\mathbf{x}, \sigma^2) \propto \left[\frac{1}{\sigma} e^{-\sum_i (x_i - \theta)^2/(2\sigma^2)}\right] \times \left[\frac{1}{\tau\sigma} e^{-(\theta-\theta_0)^2/(2\tau^2\sigma^2)}\right] \times \left[\frac{1}{(\sigma^2)^{a+1}} e^{1/b\sigma^2}\right]$$

$$= \text{Normal}$$

- σ^2 full conditional

$$\pi(\sigma^2|\mathbf{x}, \theta) \propto \left[\frac{1}{\sigma} e^{-\sum_i (x_i - \theta)^2/(2\sigma^2)}\right] \times \left[\frac{1}{\tau\sigma} e^{-(\theta-\theta_0)^2/(2\tau^2\sigma^2)}\right] \times \left[\frac{1}{(\sigma^2)^{a+1}} e^{1/b\sigma^2}\right]$$

$$= \text{Inverted Gamma}$$

Normal Hierarchical Models -3

- To estimate θ and σ^2

$$\begin{aligned} X_i &\sim N(\theta, \sigma^2), \quad i = 1, \ldots, n \\ \theta &\sim N(\theta_0, \tau^2\sigma^2) \\ \sigma^2 &\sim \text{Inverted Gamma}(a, b) \end{aligned}$$

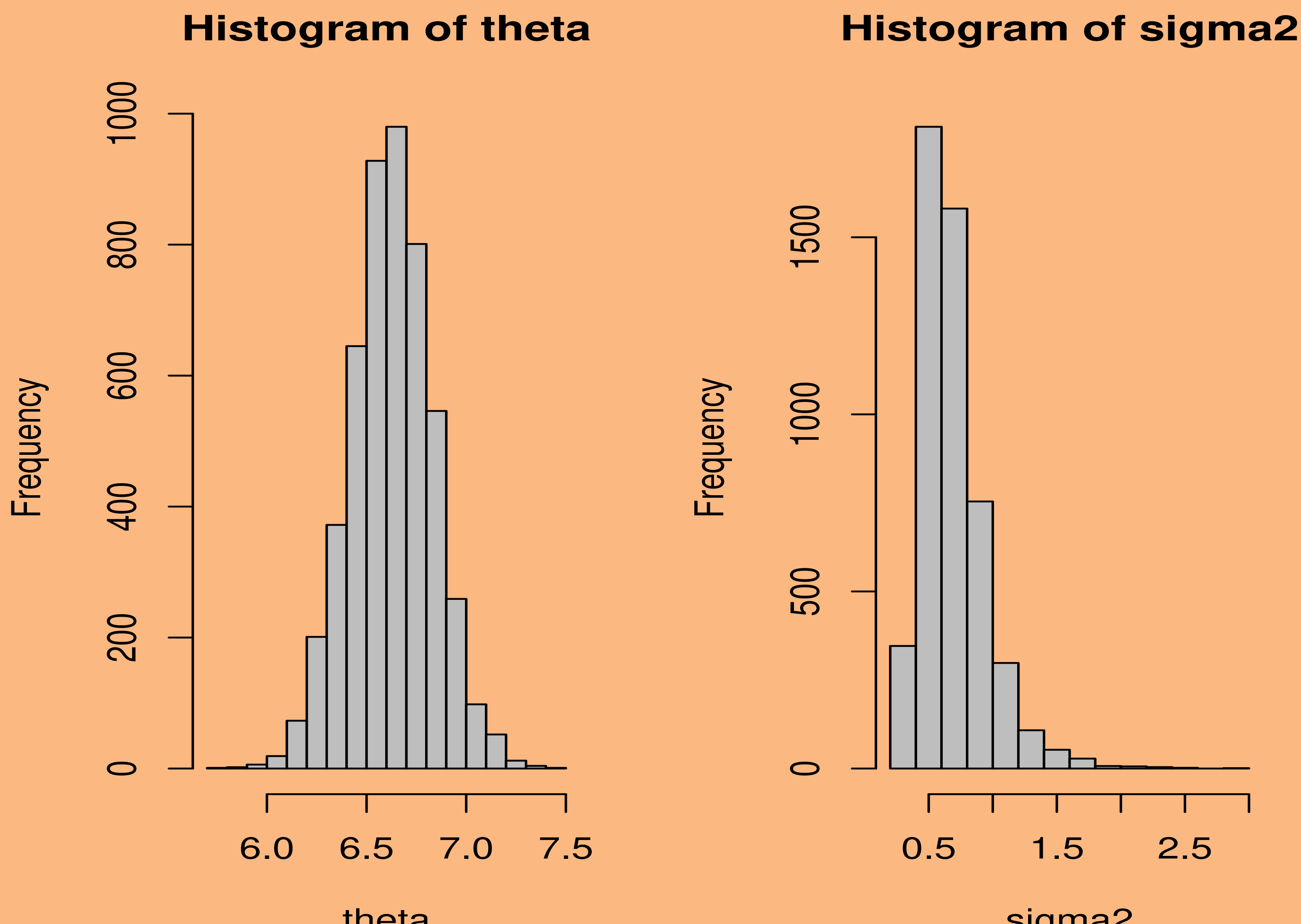
- Use a Gibbs sampler with

$$\theta \sim \mathrm{N}\left(\frac{1}{1+n\tau^2}\,\theta_0 + \frac{n\tau^2}{1+n\tau^2}\,\bar{x},\ \frac{\sigma^2\tau^2}{1+n\tau^2}\right)$$

$$\frac{1}{\sigma^2} \sim \text{Gamma}\left(\frac{n+1}{2} + a,\ \frac{1}{\frac{\sum_i (x_i-\theta)^2}{2} + \frac{(\theta-\theta_0)^2}{2} + \frac{1}{b}}\right)$$

Example

- Energy Intake (Megajoules) over 24 hours, 15 year old females
- R program NormalHierarchy-1



Normal Hierarchical Models -3a

- To avoid specifying θ_0 use the hierarchy

$$\begin{aligned} X_i &\sim N(\theta, \sigma^2), \quad i = 1, \ldots, n \\ \theta &\sim \text{Uniform}(-\infty, \infty) \\ \sigma^2 &\sim \text{Inverted Gamma}(a, b) \end{aligned}$$

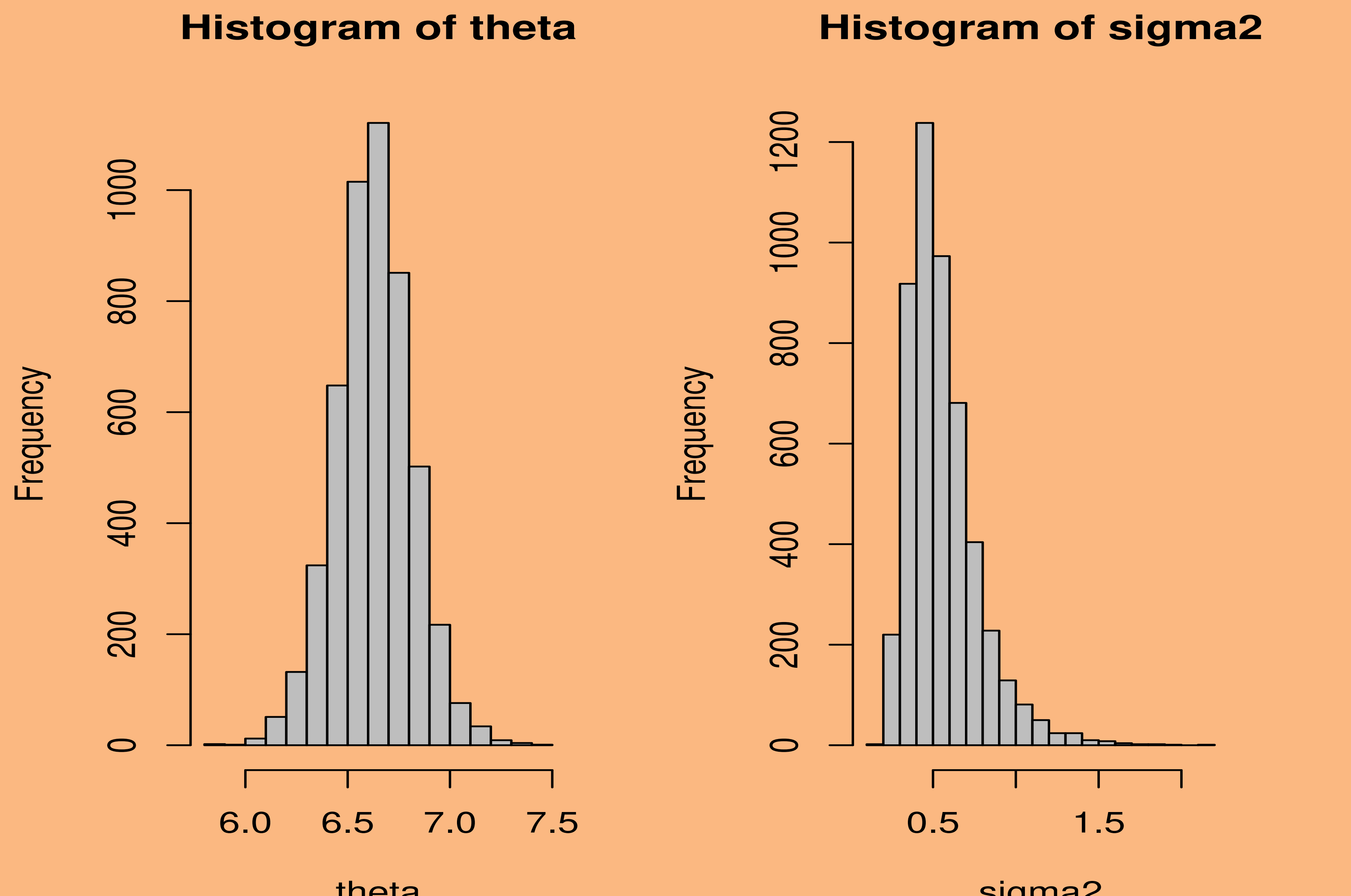
- which gives a Gibbs sampler with

$$\theta \sim \mathrm{N}\left(\bar{x},\ \sigma^2\right)$$

$$\frac{1}{\sigma^2} \sim \text{Gamma}\left(\frac{n}{2} + a,\ \frac{1}{\frac{\sum_i (x_i - \theta)^2}{2} + \frac{1}{b}}\right)$$

Example

- Energy Intake (Megajoules) over 24 hours, 15 year old females
- R program NormalHierarchy-2



Normal Hierarchical Models -4

- A bit more complicated - oneway anova: $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$
- A full hierarchical specification

$$\begin{aligned} Y_{ij} &\sim N(\mu + \alpha_i, \sigma^2), \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i \\ \mu &\sim \text{Uniform}(-\infty, \infty) \\ \alpha_i &\sim N(0, \tau^2), \quad i = 1, \ldots, k \\ \sigma^2 &\sim \text{Inverted Gamma}(a_1, b_1) \\ \tau^2 &\sim \text{Inverted Gamma}(a_2, b_2) \end{aligned}$$

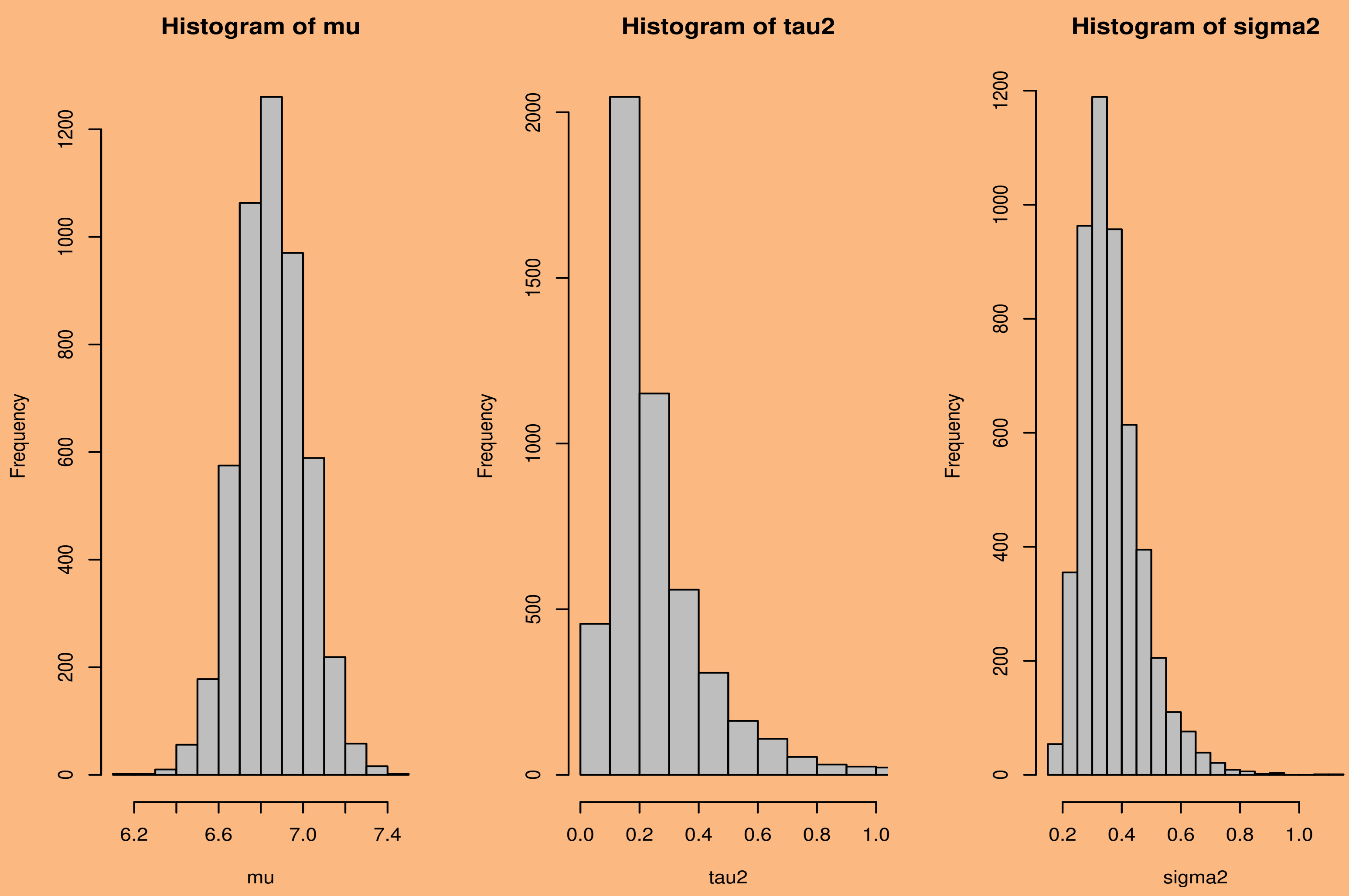
Normal Hierarchical Models -4a

- Oneway anova: $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$
- with Gibbs sampler

$$
\begin{aligned}
\mu &\sim N\left(\bar{y} - \bar{\alpha},\ \frac{\sigma^2}{\sum_i n_i}\right) \\
\alpha_i &\sim N\left(\frac{n_i \sigma^2 \tau^2}{\sigma^2 + n_i \tau^2}(\bar{y}_i - \mu),\ \frac{\sigma^2 \tau^2}{\sigma^2 + n_i \tau^2}\right) \\
\frac{1}{\sigma^2} &\sim \text{Gamma}\left(\frac{\sum_i n_i}{2} + a_1,\ \frac{1}{\frac{\sum_{ij}(y_{ij} - \alpha_i - \mu)^2}{2} + \frac{1}{b_1}}\right) \\
\frac{1}{\tau^2} &\sim \text{Gamma}\left(\frac{k}{2} + a_2,\ \frac{1}{\frac{\sum_i \alpha_i^2}{2} + \frac{1}{b_2}}\right)
\end{aligned}
$$

Example

- Energy Intake (Megajoules) over 24 hours, 15 year old females and 15 year old males
- R program NormalHierarchy-3



A lazy hierarchical specification

$$
\begin{aligned}
Y_{ij} &\sim N(\mu + \alpha_i, \sigma^2), \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i \\
\alpha_i &\sim N(0, \tau^2), \quad i = 1, \ldots, k
\end{aligned}
$$

- The classical random effects model
- We can set up a Gibbs sampler

Random Effects Model

$$\begin{aligned} Y_{ij} &\sim N(\mu + \alpha_i, \sigma^2), \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i \\ \alpha_i &\sim N(0, \tau^2), \quad i = 1, \ldots, k \end{aligned}$$

- with Gibbs sampler

$$\begin{aligned} \alpha_i &\sim \mathrm{N}\left(\frac{n_i\tau^2}{n_i\tau^2 + \sigma^2}\,(\bar{y}_i - \mu),\ \frac{n_i\tau^2\sigma^2}{n_i\tau^2 + \sigma^2}\right) \\ \mu &\sim \mathrm{N}\left(\bar{y} - \bar{\alpha},\ \frac{\sigma^2}{\sum_i n_i}\right) \\ \frac{1}{\sigma^2} &\sim \mathrm{Gamma}\left(\frac{\sum_i n_i}{2} - 1,\ \frac{2}{\sum_{ij}(y_{ij} - \mu - \alpha_i)^2}\right) \\ \frac{1}{\tau^2} &\sim \mathrm{Gamma}\left(\frac{k}{2} - 1,\ \frac{2}{\sum_i \alpha_i^2}\right) \end{aligned}$$

Problem!!

- This is not a Gibbs sampler
- Conditional distributions do not exist!
- Result of using improper priors
  - Improper priors sometimes OK
  - Sometimes: bad conditionals
  - Sometimes: good conditionals, bad posterior ← REAL BAD
  - Extremely hard to detect
- Moral: Best to use proper priors

Hierarchical Models: Animal epidemiology

- Research in animal epidemiology sometimes uses data from groups of animals, such as litters or herds.
- Such data may not follow some of the usual assumptions of independence, etc., and, as a result, variances of parameter estimates tend to be larger ("overdispersion")
- Data on the number of cases of clinical mastitis in dairy cattle herds over a one year period.

Hierarchical Models: Animal epidemiology

- $X_i \sim \mathcal{P}(\lambda_i)$, where λ_i is the underlying rate of infection in herd i
- To account for overdispersion, put a gamma prior distribution on the Poisson parameter. A complete hierarchical specification is

$$\begin{aligned} X_i &\sim \mathcal{P}(\lambda_i), \\ \lambda_i &\sim \mathcal{G}a(\alpha, \beta_i), \\ \beta_i &\sim \mathcal{G}a(a, b), \end{aligned}$$

  where α, a, and b are specified.

- The posterior density of λ_i, $\pi(\lambda_i|\mathbf{x}, \alpha)$, can now be simulated via the Gibbs sampler

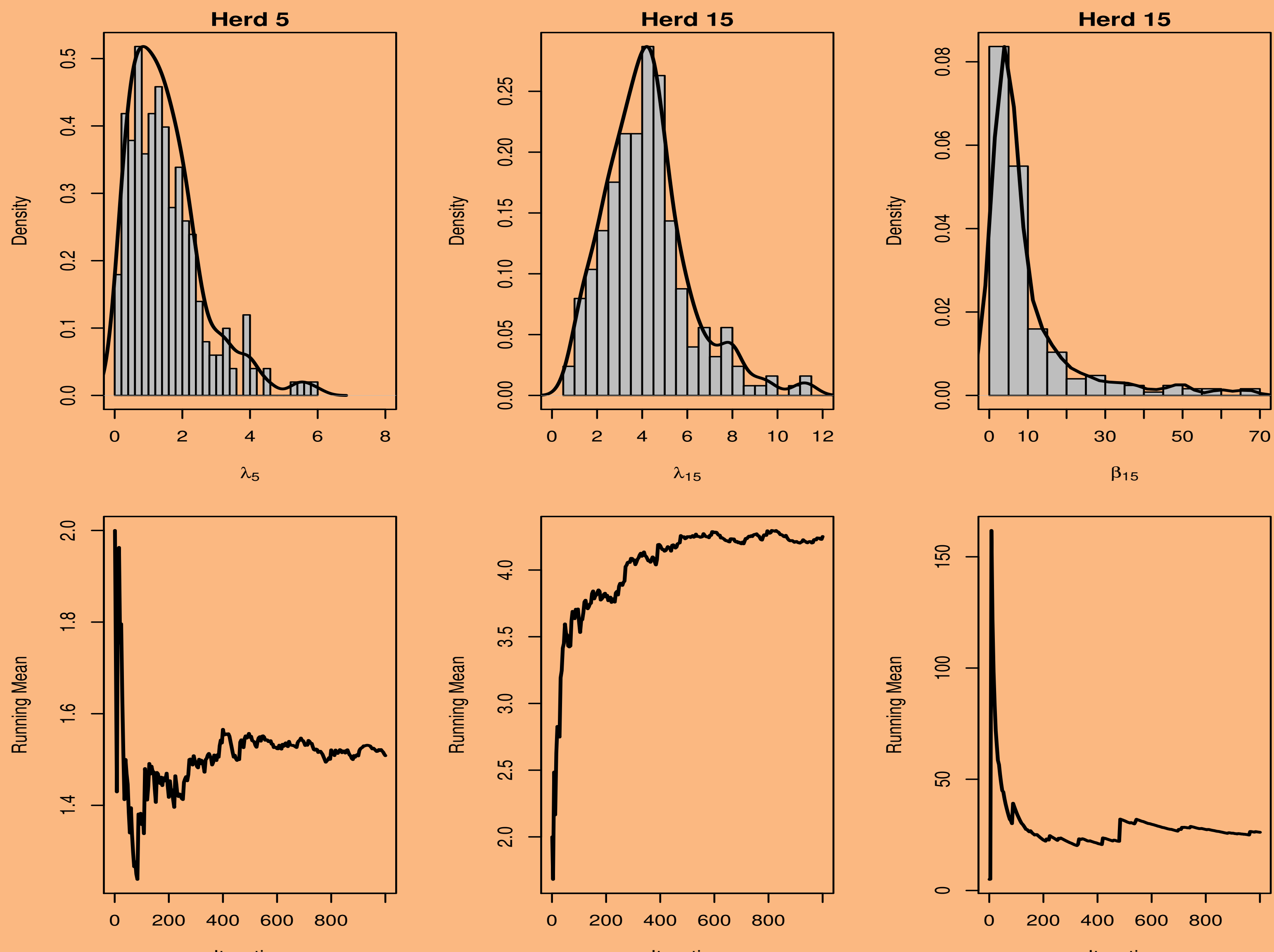
$$\begin{aligned} \lambda_i &\sim \pi(\lambda_i|\mathbf{x}, \alpha, \beta_i) = \mathcal{G}a(x_i + \alpha, 1 + \beta_i), \\ \beta_i &\sim \pi(\beta_i|\mathbf{x}, \alpha, a, b, \lambda_i) = \mathcal{G}a(\alpha + a, \lambda_i + b)\,. \end{aligned}$$

Animal Epidemiology R code

```
xdata <-c(0,0,1,1,2,2,2,2,2,2,4,4,4,5,5,5,5,5,5,6,6,8,8,8,9,9,9,
      10,10,12,12,13,13,13,13,18,18,19,19,19,19,20,20,22,22,22,23,25)
nx<-length(xdata)
nsim<-1000;
lambda<-array(2,dim=c(nsim,nx));beta<-array(5,dim=c(nsim,nx));
alpha<-.1;a<-1;b<-1;
for(i in 2:nsim){
for(j in 1:nx){
beta[i,j]<-1/rgamma(1,shape=alpha+a,scale=1/(lambda[i-1,j]+(1/b)));
lambda[i,j]<-rgamma(1,shape=xdata[j]+alpha,scale=1/(1+(1/beta[i,j])))
}
}
```

Gibbs sampler output

- Selected estimates of λ_i and β_i.



Prediction - Introduction

- For the simple model

$$\mathbf{X} \sim f(\mathbf{x}|\theta)$$
$$\theta \sim g(\theta)$$

- The predictive density of a new $\mathbf{X}$ is

$$\pi(x_{\text{new}}|\mathbf{x}_{\text{old}}) = \int f(x_{\text{new}}|\theta)\pi(\theta|\mathbf{x}_{\text{old}})d\theta$$

  - $\pi(\theta|\mathbf{x}_{\text{old}})$ is the posterior density
  - Averages over the parameter values

- If $\theta_1, \ldots, \theta_M \sim \pi(\theta|\mathbf{x}_{\text{old}})$

$$\pi(x_{\text{new}}|\mathbf{x}_{\text{old}}) \approx \frac{1}{M}\sum_i f(x_{\text{new}}|\theta_i)$$

Prediction - Introduction -2

- For the hierarchical model

$$\begin{aligned} \mathbf{X} &\sim f(\mathbf{x}|\theta) \\ \theta &\sim g(\theta|\beta) \\ \beta &\sim h(\beta|\lambda) \\ \lambda &\sim k(\lambda) \end{aligned}$$

- the Gibbs sampler give us a $(\theta_i, \beta_i, \lambda_i)$, $i = 1, \ldots, M$
  - A sample from the joint distribution.
- Using Monte Carlo sums

$$\pi(x_{\text{new}}|\mathbf{x}_{\text{old}}) \approx \frac{1}{M} \sum_i f(x_{\text{new}}|\theta_i)$$

- A Conditionally Independent Hierarchical Model

Oneway Anova Predictive Density

- Energy Intake - oneway anova: $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$
- A full hierarchical specification

$$\begin{aligned} Y_{ij} &\sim N(\mu+\alpha_i, \sigma^2), \quad i=1,\ldots,k, \quad j=1,\ldots,n_i \\ \mu &\sim \text{Uniform}(-\infty,\infty) \\ \alpha_i &\sim N(0,\tau^2), \quad i=1,\ldots,k \\ \sigma^2 &\sim \text{Inverted Gamma}(a_1,b_1) \\ \tau^2 &\sim \text{Inverted Gamma}(a_2,b_2) \end{aligned}$$

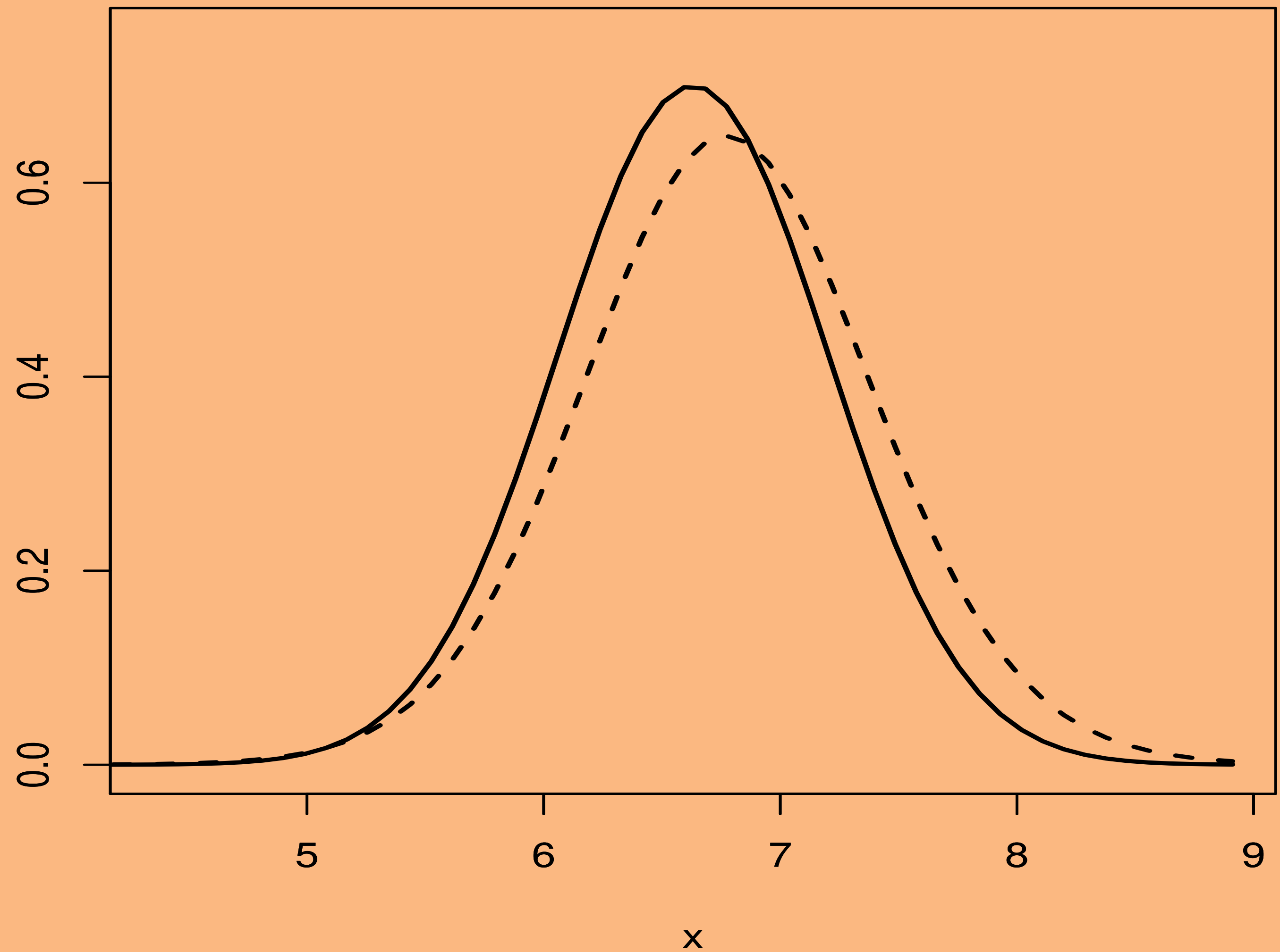
- Predictive Density for Group i

$$\pi(y_{\text{new}}|\mathbf{y}) = \frac{1}{M}\sum_{j=1}^{M} \frac{1}{\sqrt{2\pi\sigma_j^2}}\, e^{-.5(y_{\text{new}}-\mu_j-\alpha_{ij})^2/\sigma_j^2}$$

where $(\mu_j, \alpha_{ij}, \sigma_j^2)$ are a sample from the posterior distribution.

Energy Intake - Predictive density for females

- R program NormalPrediction-3



◦ solid = "naive" prediction ◦ dashed = predictive density

PKPD Medical Models

- ***Pharmacokinetics*** is the modeling of the relationship between the dosage of a drug and the resulting concentration in the blood.
- Gilks *et al.* (1993) approach:
  - Estimate pharmacokinetic parameters using mixed-effects model and nonlinear structure
  - Also robust to the outliers common to clinical trials
- For a given dose d_i administered at time 0 to patient i, the measured log concentration in the blood at time t_{ij}, X_{ij}, is assumed to follow a normal distribution

$$X_{ij} \sim N(\log g_{ij}(\lambda_i), \sigma^2),$$

PKPD Medical Models

- $X_{ij} \sim N(\log g_{ij}(\lambda_i), \sigma^2)$,
- $\lambda_i = (\log C_i, \log V_i)'$ are parameters for the ith individual, σ^2 is the measurement error variance, and g_{ij} is given by

$$g_{ij}(\lambda_i) = \frac{d_i}{V_i} \exp\left(-\frac{C_i \, t_{ij}}{V_i}\right).$$

  - C_i represents *clearance*
  - V_i represents *volume* for patient i.

- We complete the hierarchical specification with

$$\log C_i \sim \mathcal{N}(\mu_C, \sigma_C^2) \text{ and } \log V_i \sim \mathcal{N}(\mu_V, \sigma_V^2).$$

  with $\mu_C, \sigma_C^2, \mu_V, \sigma_V^2$ fixed.