Bayesian Statistics and R

Peng Ding,
School of Mathematical Sciences,
Peking Univ.

December 16, 2008
Introduction

- What is *Probability*?
- Frequentist: LLN, CLT
- Bayesian: Prior and Posterior
- What is *Statistics*?
• Moment Estimation (Karl Pearson)
• Maximum Likelihood Estimation (Gauss, R. A. Fisher)
• Bayesian Method (Bayes)
• Empirical Bayesian (Robbins)
• ...

Bayesian Statistics and R
Frequentist vs Bayesian

• Frequentist:
  · parameters are constant to be estimate
  · point estimation and interval estimation

• Bayesian:
  · parameters are random variables
  · Prior + Model → Posterior
  · all information are contained in posterior distribution
Bayes’ Formula and Bayesian Statistics

• Bayes’ Formula:
If $\theta \in \Theta$ has prior distribution $\pi(\theta)$, and the observed data $y$ comes from conditional distribution $p(y|\theta)$. Then the posterior distribution of $\theta$ given $y$ is

$$
\pi(\theta|y) = \frac{p(y|\theta)\pi(\theta)}{\int_{\Theta} p(y|\theta)\pi(\theta)\,d\theta}
$$

• Bayesian Statistics:
  · choose prior
  · model observed data
  · inference based on posterior distribution
How to choose *Prior*?

- Congugate Prior Distribution:
  - Prior and Posterior have the same form
  - Congugate Prior for Exponential Family

\[
p(y_i|\theta) = f(y_i)g(\theta) \exp \left\{ \phi(\theta)^T u(y_i) \right\}
\]

\[
p(y|\theta) = \left( \prod_{i=1}^{n} f(y_i) \right) g(\theta)^n \exp \left\{ \phi(\theta)^T \sum_{i=1}^{n} u(y_i) \right\}
\]

Choose Prior as:

\[
p(\theta) \propto g(\theta)^\eta \exp \{ \phi(\theta)^T \nu \}
\]

The Posterior is:

\[
p(\theta|y) \propto g(\theta)^{n+\eta} \exp \{ \phi(\theta)^T (\nu + \sum_{i=1}^{n} u(y_i)) \}
\]
How to choose *Prior*?

- Non-informative Prior Distribution:
  - Baysian Assumption

\[ p(\theta) \propto \text{constant}, \theta \in \Theta \]

- Jeffreys Prior

\[ \pi(\theta) \propto |I(\theta)|^{1/2} \]

where \( I(\theta) \) is the Fisher Information Matrix.
How to explore the *Posterior*?

- **Direct Calculation:**
  - The posterior has explicit and simple form!
- **Simulation the Posterior:**
  - Sampling from the posterior distribution
  - Markov Chain Monte Carlo (MCMC):
    - Gibbs Sampler and Metropolis-Hasting Algorithm
Gibbs Sampling

- $X \sim \pi(x), x = (x_1, ..., x_n)$
- Initial value $x^{(0)} = (x_1^{(0)}, ..., x_n^{(0)})$
- The $t$-th iteration:
  - Sample $x_1^{(t)} \sim \pi(x_1|x_2^{(t-1)}, ..., x_n^{(t-1)})$;
  - ...  
  - Sample $x_i^{(t)} \sim \pi(x_i|x_1^{(t)}, ..., x_{i-1}^{(t)}, x_{i+1}^{(t-1)}, ..., x_n^{(t-1)})$;
  - ...  
  - Sample $x_n^{(t)} \sim \pi(x_n|x_1^{(t)}, ..., x_{n-1}^{(t)})$.
- Under some regular conditions, the distribution of $x$ converges to the stationary distribution of the Markov Chain: $\pi(x)$. 

Peng Ding, School of Mathematical Sciences, Peking Univ.
M-H Algorithm

- Given an irreducible transition probability $q(\cdot, \cdot)$,
- Given a function

$$\alpha(\cdot, \cdot) = \min\{1, \frac{\pi(x')q(x', x)}{\pi(x)q(x, x')}\}, \quad 0 < \alpha \leq 1.$$ 

- At time $t$, $X^{(t)} = x$,
  - Generate a potential transition $x \rightarrow x'$ by $q(x, \cdot)$;
  - With probability $\alpha(x, x')$, accept $x'$; with probability $1 - \alpha(x, x')$ stay at $x$.
- Under some regular conditions, $\pi(x)$ is the stationary distribution of this Markov Chain.
• MCMCpack

library(lattice)
library(coda)
library(MASS)
library(MCMCpack)

• Famous software WinBUGS: Bayesian inference Using Gibbs Sampling.
Example 1: Binomial Distribution

- Model
  \[ p(y|\theta) \propto \theta^y (1 - \theta)^{n-y} \]

- Prior
  \[ p(\theta) \propto \theta^{\alpha-1} (1 - \theta)^{\beta-1} \]

- Posterior
  \[ p(\theta|y) \propto \theta^{\alpha+y-1} (1 - \theta)^{\beta+n-y-1} \]

- R code

```r
posterior <- MCbinomialbeta(y=3,n=12,alpha=1,beta=1,mc=5000)
summary(posterior)
plot(posterior)
```
1. Empirical mean and standard deviation for each variable, plus standard error of the mean:

<table>
<thead>
<tr>
<th>Mean</th>
<th>SD</th>
<th>Naive SE</th>
<th>Time-series SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.286442</td>
<td>0.116600</td>
<td>0.001649</td>
<td>0.001924</td>
</tr>
</tbody>
</table>

2. Quantiles for each variable:

<table>
<thead>
<tr>
<th>2.5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.08972</td>
<td>0.19926</td>
<td>0.27805</td>
<td>0.36167</td>
<td>0.53906</td>
</tr>
</tbody>
</table>
Posterior for Binomial Parameter

Trace of $\pi$

Density of $\pi$

Iterations

$N = 5000$  Bandwidth = 0.0225
Example 2: Poisson Distribution

- **Model**
  \[ p(y|\lambda) \propto \prod_{i=1}^{n} \lambda^{y_i} e^{-\lambda} \]

- **Prior**
  \[ p(\lambda) \propto e^{\beta \lambda} \lambda^{\alpha-1} \]

- **Posterior**
  \[ \lambda|y \sim \text{Gamma}(\alpha + n\bar{y}, \beta + n) \]

- **R code**

  ```r
  y<-rpois(1000,lamba=2)
  posterior <- MCpoissongamma(y, 15, 1, 5000)
  summary(posterior)
  plot(posterior)
  ```
Posterior for Poisson Parameter

**Trace of lambda**

- Iterations range from 0 to 4000, with values fluctuating between 1.80 and 2.00.

**Density of lambda**

- N = 5000
- Bandwidth = 0.008447

---

Peng Ding, School of Mathematical Sciences, Peking Univ.

Bayesian Statistics and R
Example 3: Normal Distribution with Variance known

• Model

\[ p(y|\mu) \propto \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2 \right\} \]

• Prior

\[ p(\mu) \propto \exp\left\{ -\frac{1}{2\tau_0^2} (\mu - \mu_0)^2 \right\} \]

• Posterior

\[ \mu|y \sim \mathcal{N}(\mu_1, \tau_1^2) \]

where

\[ \mu_1 = \frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2}, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i / n, \]

\[ \frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2}. \]
Example 3: Normal Distribution with Variance known

- R code

```r
y <- rnorm(1000, 5, 1)
posterior <- MCnormalnormal(y, sigma2=1, mu0=0, tau20=100, mc=5000)
summary(posterior)
plot(posterior)
```
Posterior for Normal Parameter: $\mu$

**Trace of $\mu$**

**Density of $\mu$**

- Iterations: 0 to 4000
- N = 5000
- Bandwidth = 0.006083
Example 4: Normal Distribution with unknown Variance

- **Model**

\[
p(y|\mu) \propto \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right\}
\]

- **Prior (Semi-Conjugate)**

\[
\mu \sim N(b_0, B_0^2) \\
\sigma^2 \sim \text{Inverse - } \chi^2(2c_0, 2d_0)
\]

- **R code**

```r
y <- rnorm(1000, 2, 1)
posterior <- MCMCregress(y ~ 1, b0 = 0, B0 = 0, c0 = 0.001, d0 = 0.001)
summary(posterior)
plot(posterior)
```
Posterior for Normal Parameter: $\mu$ and $\sigma$
Example 5: Multinomial Distribution

- Model

\[ p(y|\theta) \propto \prod_{i=1}^{n} \theta_i^{y_i} \]

- Prior

\[ p(\theta|\alpha) \propto \prod_{i=1}^{n} \theta_i^{\alpha_i - 1} \]

- R code

```r
posterior <- MCmultinomdirichlet(c(727, 583, 137), c(1, 1, 1), mc=10000)
summary(posterior)
plot(posterior)
```
Posterior for Multinomial Parameter

Trace of $\pi.1$

Density of $\pi.1$

$N = 10000$  Bandwidth = 0.002186

Trace of $\pi.2$

Density of $\pi.2$

$N = 10000$  Bandwidth = 0.002162

Trace of $\pi.3$

Density of $\pi.3$

$N = 10000$  Bandwidth = 0.0013
Generalized Linear Model

- $E(y|x) = g^{-1}(\beta^T x)$
- Different $g(\cdot)$–Link Function, different models
  - Linear Regression: $g(t) = t$: $E(y|x) = \beta^T x$
  - Logistic Regression: $g(t) = \log(\frac{t}{1-t})$: \[ \log\left(\frac{P(y=1|x)}{1-P(y=1|x)}\right) = \beta^T x \]
  - Probit Regression: $g(t) = \Phi^{-1}(t)$: \[ \Phi^{-1}(P(y = 1|x)) = \beta^T x \]
  - Poisson Regression: $g(t) = \log(t)$: \[ \log(E(y|x)) = \beta^T x \]
Example 6: Linear Regression

\[ X \leftarrow \text{rnorm}(100, 2, 1) \]
\[ Y \leftarrow 1 + 2 \times X + \text{rnorm}(100, 0, 1) \]
\[ \text{posterior} \leftarrow \text{MCMCregress}(Y \sim X, b0 = 0, B0 = 0, \]
\[ \quad c0 = 0.001, d0 = 0.001, \text{verbose=1000}) \]
\[ \text{plot(posterior)} \]
\[ \text{summary(posterior)} \]
Posterior for Linear Regression

**Trace of (Intercept)**

- Iterations: 2000, 4000, 6000, 8000, 10000

**Density of (Intercept)**

- N = 10000
- Bandwidth = 0.03855

**Trace of X**

- Iterations: 2000, 4000, 6000, 8000, 10000

**Density of X**

- N = 10000
- Bandwidth = 0.01715

**Trace of sigma2**

- Iterations: 2000, 4000, 6000, 8000, 10000

**Density of sigma2**

- N = 10000
- Bandwidth = 0.02661
Example 7: Logistic Regression

```r
x <- rnorm(1000)
y <- rbinom(1000, 1, exp(1 - x)/(1 + exp(1 - x)))
posterior <- MCMClogit(y ~ x, b0=0, B0=.001)
plot(posterior)
summary(posterior)
```
Posterior for Logistic Regression

Trace of (Intercept)

Density of (Intercept)

Trace of $x$

Density of $x$
Exmaple 8: Probit Regression

```r
y <- rbinom(1000, 1, pnorm(1 - x))
posterior <- MCMCprobit(y ~ x, b0 = 0, B0 = .001)
plot(posterior)
summary(posterior)
```
Posterior for Probit Regression

Trace of (Intercept)

Density of (Intercept)

Trace of $x$

Density of $x$

N = 10000  Bandwidth = 0.009996

N = 10000  Bandwidth = 0.01089
Example 9: Poisson Regression

```r
x <- rnorm(100)
y <- rpois(100, exp(1 + x))
posterior <- MCMCpoisson(y ~ x)
plot(posterior)
summary(posterior)
```
Other Models

- Gaussian Mixture Model
- Latent Class Analysis
- Hierarchical Models
- Perhaps any parametric models: Examples in WinBUGS
- Bayesians believes that all inference and more is Bayesian territory.—*Bayesian Nonparametrics*, J. K. Ghosh and R. V. Ramamoorthi, Springer(2003)
Reference and recommendatory books

• I would like to thank the organizers of the first R conference in China.
• Also I am grateful to my mentor Professor Zhi Geng for his introduction of GLM and Bayesian Methods.
Thank you!