Bayesian Analysis Principles

Bert Steece  
(berste@almaak.usc.edu)  
Dean of Faculty  
Marshall School of Business  
University of Southern California  

CSE’s Annual Research Review  
March 9, 1998

Outline

- Bayesian Analysis Theory
  - Bayes Theorem
- Different Types of Prior Information
  - Natural Conjugate Prior
  - g-Prior
  - Extension of g-Prior
- A Simple Cost Model
  - Noninformative Prior
  - Informative Prior
  - g-Prior
- Conclusions
A Simple Cost Model

- Suppose we have a simple software cost model

\[ \text{Effort} = \alpha \cdot \text{Size}^\beta \delta \]

- Linearizing, we get

\[
\begin{align*}
\ln(\text{Effort}) &= \ln(\alpha) + \beta \cdot \ln(\text{Size}) + \ln(\delta) \\
\ln(\text{Effort}) &= \beta y + \beta \cdot \ln(\text{Size}) + \epsilon \\
y &= X\beta + \epsilon
\end{align*}
\]

- How do we combine prior (expert) information with sample information?
  - Use Bayesian Approach

Bayes Theorem

\[
g(\beta \mid y) = \frac{f(y \mid \beta)g(\beta)}{f(y)}
\]

\[
g(\beta \mid y) \propto l(\beta \mid y) g(\beta)
\]

posterior information \(\propto\) sample information \(\times\) prior information

\[
\begin{array}{c}
\text{A - Priori} \\
\text{information}
\end{array} + \begin{array}{c}
\text{Sampling} \\
\text{Data:}
\end{array} = \begin{array}{c}
\text{A - Posterior} \\
\text{Model}
\end{array}
\]
Outline

- Different Types of Prior Information
  - Natural Conjugate Prior
  - g-Prior
  - Extension of g-Prior
- A Simple Cost Model
  - Noninformative Prior
  - Informative Prior
  - g-Prior
- Conclusions

**Posterior Mean with Natural Conjugate Prior**

- **Natural Conjugate Prior**
  \[ f(\beta, \sigma) = f_N(\beta, \sigma) f_{IG}(\sigma) \]
  where \( f_N(\beta, \sigma) \)
  
  is multivariate normal with prior mean \( \beta \)
  and prior covariance matrix \( \sigma^2 A^{-1} \)

- **Posterior Mean**
  \[ \beta = (A + X'X)^{-1}(A\beta + X'X\hat{\beta}) \]
  where \( A \) and \( \beta \) are specified by prior information
  and \( \hat{\beta} \) is the ordinary least squares estimate
Posterior Mean with g-Prior

Posterior Mean
\[ \bar{\beta} = (A + XX')^{-1}(A\bar{\beta} + XX'\hat{\beta}) \]
where \( A \) and \( \beta \) are specified by prior information and \( \hat{\beta} \) is the ordinary least squares estimate
But \( A \) is difficult to specify
- we assume that the structure of the prior covariance is the same as the structure of the sample covariance
- that is \( A = gXX' \)

Posterior Mean
\[ \bar{\beta} = g\bar{\beta} + \hat{\beta} \]
i.e. the weighted average of \( \bar{\beta} \) and \( \hat{\beta} \)

Posterior Variance
\[ V(\beta|\sigma, y, X) = \frac{(X'X)^{-1}\sigma^2}{1+g} \]

How do we interpret g?

When \( g = 0 \)
- our estimates of \( \beta \) depend only on sample information

When \( g = 1 \)
- we are equally weighting prior and sample information

When \( g > 1 \)
- we are giving greater weight to prior information
Extension of g-Prior

\[ \hat{\beta} = (gWX'XW + X'X)^{-1} (gWX'XW\hat{\beta} + X'X\hat{\beta}) \]

- W is a diagonal matrix
- When \( W = I \) i.e. identity matrix, we have used g-prior result described on previous slides

∠ Consider the model \( \text{Effort} = \alpha \cdot \text{Size}^{\beta_1} \delta \)
- We have different states of knowledge of \( \alpha \) and \( \beta_1 \)
- By assigning different values to \( W_{11} \) and \( W_{22} \), we can allow prior information to impact posterior distribution of \( \alpha \) and \( \beta_1 \) differently
- For example, if \( W_{11} = 0 \) and \( W_{22} > 1 \), then prior information will impact the posterior estimate of \( \beta_1 \) but the posterior estimate of \( \alpha \) will be the least squares

Outline

- A Simple Cost Model
  - Noninformative Prior
  - Informative Prior
  - g-Prior
- Conclusions
Bayesian Analysis on a Simple Cost Model

- Suppose we have a simple software cost model
  \[ \text{Effort} = \alpha \cdot \text{Size}^\beta \delta \]
- Linearizing, we get
  \[
  \ln(\text{Effort}) = \ln(\alpha) + \beta_1 \cdot \ln(\text{Size}) + \ln(\delta)
  \]
  \[
  \ln(\text{Effort}) = \beta_0 + \beta_1 \cdot \ln(\text{Size}) + \varepsilon
  \]
- We need to determine the values of \( \beta_0 \) and \( \beta_1 \)

Noninformative Prior

\( g = 0; \) posterior dependent on sample information only

\[
\begin{align*}
  f(\beta_0) &= 1 & -\infty < \beta_0 < +\infty \\
  f(\beta_1) &= 1 & -\infty < \beta_1 < +\infty
\end{align*}
\]
Modeling Under Complete Prior Uncertainty - Statistical Analysis

<table>
<thead>
<tr>
<th>Coefficient Estimates</th>
<th>Label</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_0$</td>
<td>1.206</td>
<td>0.199</td>
<td>6.062</td>
</tr>
<tr>
<td></td>
<td>$\beta_1$</td>
<td>1.03</td>
<td>0.407</td>
<td>21.93</td>
</tr>
</tbody>
</table>

$$\ln(\text{Effort}) = 1.206 + 1.03 \ln(\text{Size})$$

$$\text{Effort} = 33 \times \text{Size}^{1.206}$$

where $3.3 = e^{1.206}$

**Posterior Density Functions - Noninformative Prior**

$\beta_0$

$\beta_1$

But, can $B<1$? (i.e. economies of scale?)
Some Prior Knowledge on $\beta_1$

- Experience indicates that software exhibits diseconomies of scale
  \[ \text{[Banker94, Gulledge93]} \]
- Suppose we believe that $\beta_1 \geq 1$
  \[ f(\beta_1) = 1 \text{ if } \beta_1 > 1.0 \]
  \[ = 0 \text{ if } \beta_1 \leq 1.0 \]

Post Sample Density Functions - Inclusion of Prior Information

- 50% of area under curve

Inference of $\beta_1$ given data

$\text{Post Sample Density Functions}$

- Inclusion of Prior Information
Prior $2.5^{*} \text{Size}^{1.10}$

Data $3.3^{*} \text{Size}^{1.03}$

Posterior $2.9^{*} \text{Size}^{1.06}$

**Conclusions**

- Bayesian Approach formally incorporates experience-based prior information to sampling data
- Prior information can be specified in many different ways
- Well-suited when not enough data is available