Bayesian Multinomial Model for Ordinal Data

Overview

This example illustrates how to use the MCMC procedure to fit a Bayesian multinomial model for categorical response data measured on an ordinal scale. The example also demonstrates how use the MCMC procedure to compute posterior means, credible intervals, and posterior distributions of the parameters and odds ratios for the multinomial model.

The SAS source code for this example is available as a text file attachment. In Adobe Acrobat, right-click the icon in the margin and select Save Embedded File to Disk. You can also double-click the icon to open the file immediately.

Analysis

Researchers study the results of a taste test on three different brands of ice cream. They want to assess the testers’ preference of the three brands. The taste of each brand is rated on a five-point scale from very good to very bad. The five points correspond to response variables Y1 through Y5, where Y1 represents very good and Y5 represents very bad. Response variables contain the number of taste testers who rate each brand in each category. The very bad taste level (Y5) is used as the reference response level. Two dummy variables, BRAND1 and BRAND2, are created to indicate Brands 1 and 2, respectively. Brand 3 is set as the reference level in this example and is represented in the data set when both of the dummy variables equal zero.

The following statements create the ICECREAM data set:

```sas
data icecream;
  input y1-y5 brand;
  if brand = 1 then brand1 = 1;
  else brand1 = 0;
  if brand = 2 then brand2 = 1;
  else brand2 = 0;
  keep y1-y5 brand1 brand2;
datalines;
70 71 151 30 46 1
20 36 130 74 70 2
50 55 140 52 50 3
;
```
Bayesian Multinomial Model

Multinomial ordinal models occur frequently in applications such as food testing, survey response, or anywhere order matters in the categorical response. Categorical data with an ordinal response correspond to multinomial models based on cumulative response probabilities (McCullagh and Nelder 1989). In this data set, the ordered response variable is the taste tester’s rating for a brand of ice cream.

Let the random variable $Y_{ij} = (Y_{i1}, \ldots, Y_{iJ})$ for brand $i = 1, 2, 3$, and let response level $j = 1, \ldots, 5$, be from an multinomial ordinal model with mutually exclusive, discrete response levels and probability mass function

$$f(Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ}) = \frac{n_i!}{\prod_{j=1}^{J} y_{ij}!} \prod_{j=1}^{J} \pi_{ij}^{y_{ij}}$$

where $y_{ij}$ represents the number of people from $i$th brand in the $j$th response level. For the grouped data, let $n_i = \sum_{j=1}^{5} y_{ij}$ denote the number of testers who taste the $i$th brand of ice cream and let $n = \sum_{i=1}^{3} n_i$.

Let $\pi_{ij} = P_r(Y_{ij} = j)$ denote the probability that the response of brand $i$ falls into the $j$th response level, and let $\sum_{j=1}^{J} \pi_{ij} = 1$. Let $\gamma_{ij} = P_r(Y_{ij} \leq j)$ denote the corresponding cumulative probability that the response falls in the $j$th level or below, so $\gamma_{ij} = \pi_{i1} + \ldots + \pi_{ij}$. The transformed cumulative probabilities are linear functions of the covariates written as $g(\gamma_{ij}) = \theta_j + X_i \beta$, where $g(\cdot)$ refers to the logit link function, $\beta$ represents the effects for the covariates, and $X_i = \{\text{BRAND}1_i, \text{BRAND}2_i\}$. Let $\theta_j$ represent the baseline value of the transformed cumulative probability for category $j$ such that the constraint $\theta_j < \theta_{j-1}$ holds for all $j$ (Albert and Chib 1993). Then

$$\gamma_{ij} = g^{-1}(X_i, \beta, \theta_j) = \text{logistic}(\theta_j + X_i \beta)$$

and the group probabilities for the $j$th levels are as follows:

$$\pi_{i1} = \text{logistic}(\theta_1 + X_i \beta)$$

$$\pi_{ij} = \gamma_{i1}$$

$$\pi_{ij} = \gamma_{ij} - \gamma_{i(j-1)} \quad \text{for } 1 \leq j \leq J$$

$$\pi_{iJ} = 1 - \sum_{j=1}^{J-1} \pi_{ij}$$

The likelihood function for the counts and corresponding covariates is

$$p(Y_{i1}, ..., Y_{i5} | \theta_1, ..., \theta_4, \beta_1, \beta_2, \text{BRAND}1_i, \text{BRAND}2_i) = \text{Multinomial}(\pi_{i1}, \ldots, \pi_{i5})$$

where $p(\cdot | \cdot)$ denotes a conditional probability density. The multinomial density is evaluated at the specified value of $Y_i$ and the corresponding probabilities $\pi_{ij}$, which are defined in Equation 12 through 14.

There are six parameters in the likelihood: the intercepts $\theta_1$ through $\theta_4$ and the regression parameters $\beta_1$ and $\beta_2$ that correspond to the relative Brand 1 and 2 effects, respectively.
Bayesian Multinomial Model

Suppose the following prior distributions are placed on the six parameters, where \( \pi(\cdot) \) indicates a prior distribution and \( \pi(\cdot|\cdot) \) indicates a conditional prior distribution:

\[
\begin{align*}
\pi(\theta_1) &= \text{normal}(0, \sigma^2 = 100) \\
\pi(\theta_2|\theta_1) &= \text{normal}(0, \sigma^2 = 100, \text{lower} = \theta_1) \\
\pi(\theta_3|\theta_2) &= \text{normal}(0, \sigma^2 = 100, \text{lower} = \theta_2) \\
\pi(\theta_4|\theta_3) &= \text{normal}(0, \sigma^2 = 100, \text{lower} = \theta_3) \\
\pi(\beta_1), \pi(\beta_2) &= \text{normal}(0, \sigma^2 = 1000)
\end{align*}
\] (16) through (20)

The joint prior distribution of \( \theta_1 \) through \( \theta_4 \) is the product of Equation 16 through 19. The prior distributions in Equation 17 through 19 represent truncated normal distributions with mean 0, variance 100, and the designated lower bound. The lower bound ensures that the order restriction on \( \theta \) is sustained.

Using Bayes’ theorem, the likelihood function and prior distributions determine the posterior distribution of the parameters as follows:

\[
\pi(\theta_1, \ldots, \theta_4, \beta_1, \beta_2|Y_1 = y_{i1}, \ldots, Y_5 = y_{i5}, \text{BRAND1}_i, \text{BRAND2}_i) \propto p(Y_1 = y_{i1}, \ldots, Y_5 = y_{i5}|\theta_1, \ldots, \theta_4, \beta_1, \beta_2, \text{BRAND1}_i, \text{BRAND2}_i) \pi(\beta_1) \pi(\beta_2) \prod_{j=2}^{4} \pi(\theta_j|\theta_{j-1})
\]

PROC MCMC obtains samples from the desired posterior distribution. You do not need to specify the exact form of the posterior distribution.

The odds ratio for comparing one brand to another can be written as

\[
\text{OR}_{rs} = \exp(\beta_r - \beta_s)
\]

for \( r, s \in \{1, 2, 3\} \). The odds ratio is useful for interpreting how the taste preference for the different brands of ice cream compares. For this example, Brand 3 is set as the reference level, which implies that \( \beta_3 = 0 \).

The following SAS statements fit the Bayesian multinomial ordinal model. The PROC MCMC statement invokes the procedure and specifies the input data set. The NBI= option specifies the number of burn-in iterations. The NMC= option specifies the number of posterior simulation iterations. The THIN=10 option specifies that one of every 10 samples is kept. The SEED= option specifies a seed for the random number generator (the seed guarantees the reproducibility of the random stream). The PROPCOV=QUANEW option uses the estimated inverse Hessian matrix as the initial proposal covariance matrix. The MONITOR= option outputs analysis on selected symbols of interest in the program.

```sas
ods graphics on;
proc mcmc data=icecream nbi=10000 nmc=25000 thin=10 seed=1181
  propcov=quanew monitor=(beta1 beta2 or12 or13 or23);
array theta[4];
array gamma[4];
parms theta1-theta4 beta1 beta2;
prior beta: ~ normal(0, var=1000);  
prior theta1 ~ normal(0, var=100);  
prior theta2 ~ normal(0, var=100, lower=theta1);  
prior theta3 ~ normal(0, var=100, lower=theta2);
```
prior theta4 ~ normal(0, var=100, lower=theta3);

mu = beta1*brand1 + beta2*brand2;
do j = 1 to 4;
    gamma[j] = logistic(theta[j] + mu);
end;
pi1 = gamma1;
pi2 = gamma2 - gamma1;
pi3 = gamma3 - gamma2;
pi4 = gamma4 - gamma3;
pi5 = 1 - sum(of pi1-pi4);

llike = logmpdfmultinom(of y1-y5, of pi1-pi5);
model dgeneral(llike);

beginnodedata;;
or12 = exp(beta1-beta2);
or13 = exp(beta1);
or23 = exp(beta2);
endnodedata;;
run;
ods graphics off;

Each of the two ARRAY statements associates a name with a list of variables and constants. The first
ARRAY statement specifies names for the intercept parameters. The second ARRAY statement contains the
y_{ij} parameters.

The PARMS statement puts all six parameters in a single block. The PRIOR statements specify priors for
the parameters as given in Equations 16 through 20.

The MU assignment statement calculates X_i\beta. The DO loop and coinciding GAMMA assignment state-
ments calculate y_{ij} for j = 1, ..., 4 as in Equation 11. The five ETA assignment statements calculate the
individual probabilities that an observation falls into the jth response level as in Equation 12 to Equation
14.

The LLIKE statement uses the multivariate LOGMPDFMULTIM function to calculate the value of the
log likelihood function as in Equation 15. The multinomial density is not listed in the section “Standard
Distributions” in the PROC MCMC documentation, so you use the DGENERAL function in the MODEL
statement. The letter "D" stands for discrete. The log likelihood is the input parameter because the DGEN-
ERAL function must be specified on the logarithm scale. Since the multivariate log likelihood function
takes the dependent variable into account, you do not need to explicitly state the dependent variable in the
MODEL statement.

The statements within the BEGINNODATA and ENDNODATA statements calculate the three odds ratios
for pairwise comparisons of ice cream brands according to Equation 21. The statements are enclosed within
the BEGINNODATA and ENDNODATA block to reduce unnecessary observation-level computations.

Figure 8 displays diagnostic plots to assess whether the Markov chains have converged.
The trace plot in Figure 8 indicates that the chain appears to have reached a stationary distribution. It also has good mixing and is dense. The autocorrelation plot indicates low autocorrelation and efficient sampling. Finally, the kernel density plot shows the smooth, unimodal shape of posterior marginal distribution for $\beta_1$. The remaining diagnostic plots (not shown here) similarly indicate good convergence in the other parameters.

Figure 9 displays a number of convergence diagnostics, including Monte Carlo standard errors, autocorrelations at selected lags, Geweke diagnostics, and the effective sample sizes.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MCSE</th>
<th>Standard Deviation</th>
<th>MCSE/SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta1</td>
<td>0.00379</td>
<td>0.1337</td>
<td>0.0283</td>
</tr>
<tr>
<td>beta2</td>
<td>0.00449</td>
<td>0.1403</td>
<td>0.0320</td>
</tr>
<tr>
<td>or12</td>
<td>0.0119</td>
<td>0.4054</td>
<td>0.0294</td>
</tr>
<tr>
<td>or13</td>
<td>0.00562</td>
<td>0.1981</td>
<td>0.0284</td>
</tr>
<tr>
<td>or23</td>
<td>0.00236</td>
<td>0.0743</td>
<td>0.0317</td>
</tr>
</tbody>
</table>
Figure 9 continued

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Lag 1</th>
<th>Lag 5</th>
<th>Lag 10</th>
<th>Lag 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta1</td>
<td>0.3284</td>
<td>-0.0011</td>
<td>0.0618</td>
<td>-0.0120</td>
</tr>
<tr>
<td>beta2</td>
<td>0.4089</td>
<td>0.0287</td>
<td>0.0232</td>
<td>0.0142</td>
</tr>
<tr>
<td>or12</td>
<td>0.3465</td>
<td>0.0059</td>
<td>0.0297</td>
<td>-0.0029</td>
</tr>
<tr>
<td>or13</td>
<td>0.3260</td>
<td>0.0014</td>
<td>0.0607</td>
<td>-0.0130</td>
</tr>
<tr>
<td>or23</td>
<td>0.3967</td>
<td>0.0338</td>
<td>0.0266</td>
<td>0.0144</td>
</tr>
</tbody>
</table>

Geweke Diagnostics

| Parameter | z      | Pr > |z| |
|-----------|--------|------|---|
| beta1     | -0.1114| 0.9113|
| beta2     | -1.5302| 0.1260|
| or12      | 1.4711 | 0.1413|
| or13      | -0.1486| 0.8819|
| or23      | -1.5676| 0.1170|

Effective Sample Sizes

<table>
<thead>
<tr>
<th>Parameter</th>
<th>ESS</th>
<th>Autocorrelation</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta1</td>
<td>1246.3</td>
<td>2.0060</td>
<td>0.4985</td>
</tr>
<tr>
<td>beta2</td>
<td>978.1</td>
<td>2.5560</td>
<td>0.3912</td>
</tr>
<tr>
<td>or12</td>
<td>1157.6</td>
<td>2.1596</td>
<td>0.4631</td>
</tr>
<tr>
<td>or13</td>
<td>1243.9</td>
<td>2.0098</td>
<td>0.4976</td>
</tr>
<tr>
<td>or23</td>
<td>992.7</td>
<td>2.5184</td>
<td>0.3971</td>
</tr>
</tbody>
</table>

Figure 10 reports summary and interval statistics for the regression parameters and odds ratios. The odds ratios provide the relative difference in one brand with respect to another and indicate whether there is a significant brand effect. The odds ratio for Brand 1 and Brand 2 is the multiplicative change in the odds of a taste tester preferring Brand 1 compared to the odds of the tester preferring Brand 2. The estimated odds ratio (OR) value is 2.8366 with a corresponding 95% equal-tail credible interval of (2.1197, 3.6740). Similarly, the odds ratio for Brand 1 and Brand 3 is 1.4787 with a 95% equal-tail credible interval of (1.1202, 1.9046). Finally, the odds ratio for Brand 2 compared to Brand 3 is 0.5271 with a 95% equal-tail credible interval of (0.3947, 0.6882). The lower categories indicate the favorable taste results; so Brand 1 scored significantly better when compared to Brand 2 or 3. Brand 2 scored less favorably when compared to Brand 3.
### Figure 10 Multinomial Model Summary and Interval Statistics

**The MCMC Procedure**

**Posterior Summaries**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>Deviation</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta1</td>
<td>2500</td>
<td>0.3822</td>
<td>0.1337</td>
<td>0.2957</td>
<td>0.3810</td>
<td>0.4718</td>
</tr>
<tr>
<td>beta2</td>
<td>2500</td>
<td>-0.6502</td>
<td>0.1403</td>
<td>-0.7456</td>
<td>-0.6488</td>
<td>-0.5590</td>
</tr>
<tr>
<td>or12</td>
<td>2500</td>
<td>2.8366</td>
<td>0.4054</td>
<td>2.5518</td>
<td>2.8079</td>
<td>3.1088</td>
</tr>
<tr>
<td>or13</td>
<td>2500</td>
<td>1.4787</td>
<td>0.1981</td>
<td>1.3440</td>
<td>1.4638</td>
<td>1.6029</td>
</tr>
<tr>
<td>or23</td>
<td>2500</td>
<td>0.5271</td>
<td>0.0743</td>
<td>0.4745</td>
<td>0.5227</td>
<td>0.5718</td>
</tr>
</tbody>
</table>

**Posterior Intervals**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Alpha</th>
<th>Equal-Tail Interval</th>
<th>HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta1</td>
<td>0.050</td>
<td>0.1135</td>
<td>0.6443</td>
</tr>
<tr>
<td>beta2</td>
<td>0.050</td>
<td>-0.9297</td>
<td>-0.3736</td>
</tr>
<tr>
<td>or12</td>
<td>0.050</td>
<td>2.1197</td>
<td>3.6740</td>
</tr>
<tr>
<td>or13</td>
<td>0.050</td>
<td>1.1202</td>
<td>1.9046</td>
</tr>
<tr>
<td>or23</td>
<td>0.050</td>
<td>0.3947</td>
<td>0.6882</td>
</tr>
</tbody>
</table>

### References
