Bayesian bridge regression for ordinal models with a practical application

Dhiyaa Hazem Qassem Aljabri and Rahim Alhamzawi

Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Iraq

Email: thyaa223@gmail.com

Abstract. In this paper, we introduce a Bayesian Bridge regression for ordinal data. The simulation study and real data analyses show that our proposed method performs very well compared with other methods. We consider a fully Bayesian approach that yields a new algorithm with tractable full conditional posteriors.

1- Introduction

Ordinal data is a type of statistical data where the covariates in form of naturally arranged categories. In this data, the difference between the two categories is unknown (Zhou, 2006). The ordinal regression is one of the types of regression analysis that predicts through using the ordinal variable (Winship and Mare, 1984). These data are found in many different fields such as social sciences, economic, medicine, political economy, climatology, psychology and many of sciences (Rahman, 2016; Alhamzawi, 2016b). An example of this data is the economy level for a country (low, medium, high). The levels take ordered values as following: the category low =1, the category medium =2 and the category high = 3. We cannot say that the medium level is a multiplier the low level. The response variable $y_i$ takes one of the categories $s$, where $s = 1, \ldots, S$, and $x_i = (x_{i1}, \ldots, x_{ip})^T$ predictors, and the $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{in})^T$ is a terms of normal error $(0, 1)$. The ordinal regression is given by:

$$pr(y_i = s) = F(x_i^T \beta), \; i = 1, \ldots, n, \; s = 1, \ldots, S,$$

where $F$ is a link function (Zhou, 2006). Analysis of ordinal data in the frequentist approach is characterized by relative simplicity. Although, the approximate theory for ordinal regression model has been well studied, a Bayesian method enables exact estimation even when $p$ is greater than $n$. In the Bayesian approaches, we need to define the model and determine the distribution of data. As for the analysis methods of regularization here, there are some complexities that face the researcher: First, we must specify a prior distribution. Second, it is difficult to calculate the posterior distribution (PD) and its summaries (Agresti, 2010). Although there are some Complexities in the Bayesian methods, these methods are better in terms of providing standard errors and credible intervals.

One of the ordinal regression problems when the number of covariates is increasing. Many of researches focus on a variable selection (vs) in the regression model, to obtain the appropriate model.
One of the most popular methods that used for (vs) is the Akaike Information Criteria (AIC) which can be written is:

\[ \text{AIC} = -2 \log m + 2p, \]

where \((m)\) is the probability function that estimated in (MLE), and \((p)\) it's the number of model parameters (Mallick, 2015). But the model be is inconsistent, especially when \(p\) size is large (Javed and Mantalos, 2013). Therefore, to get rid of this problem, the researcher tends to use another criterion such as the Bayesian Information Criterion (BIC),

\[ \text{BIC} = -2 \log m + p \log n, \]

where \(n\) is the sample size. Which was proposed to address the problem in the AIC, and the model selected will be consistent for select the appropriate model with probability 1 (Javed and Mantalos, 2013; Mallick, 2015).

Recently, researches have shown the regularization approaches which using estimation and (vs) simultaneously are effective. These methods improve prediction accuracy in linear regression (Zhou, 2006; Tibshirani, 1996). These methods assumed a penalty on size of regression parameters. This feature enables estimation of parameters even if a number of covariates are large and the sample size is relatively small (Alhamzawi and Ali, 2018b).

Suppose we have covariates \(x_i\), and the dependent variable is ordinal variable take categories \(y_i = s\), where \(s = 1, ..., S\). The methods of regularization have a general form:

\[ \min_{\beta} (y - X\beta)^T(y - X\beta) + \lambda \sum_{j=1}^{p} |\beta_j|^\delta, \quad \lambda, \delta \geq 0, \]

where \(\lambda\) is a positive tuning parameter, and \(\beta = (\beta_0, \beta_1, ..., \beta_{p-1})^T\) are regression coefficients. Here \(\delta\) is the parameter concavity which controls the amount concavity of the penalty function. (Mallick and Yi, 2018). The Formula (3) contains three special cases: the subset selection be the best if \((\delta = 0)\), the Lasso estimate when \((\delta = 1)\) and ridge regression (RR) when \((\delta = 2)\). The method of Bayesian bridge regression (BBR) has the desired good properties such as sparsity, oracle, as well as the unbiasedness when \((0 < \delta < 1)\) (Xu et al., 2010; Polson et al., 2014). In our current paper, we propose a Bayesian frame with the penalty \((0 < \delta < 1)\) for the (BBR) for ordinal data. The penalty \((0 < \delta < 1)\) provide a solution that is better than the Lasso estimator and adds oracle properties.

We organize this paper as follows. We take about the (BBR) in Section 2, we describe Bayesian bridge regression for ordinal data in section 3, in section 4 The Bayesian inferences, and in section 5 we describe Gibbs sampler (GS), in section 6 we implement the simulation study in order to check the performance of a proposed method. In section 7 the analyse to the Real data. The conclusion and Appendix in section 8.

2- Bayesian Bridge regression

Frank and Friedman introduced in (1993) the bridge regression (BR) classic which was specifically proposed for parameters, and then determine parameters with the following criterion (Mallick, 2015):

\[ \min_{\beta} (y - X\beta)^T(y - X\beta) + \lambda \sum_{j=1}^{p} |\beta_j|^\delta. \]

In (BR) we get the (vs) if \(0 < \delta < 1\), and the parameters shrinkage occurs if \((\delta > 0)\) (Park and Yoon, 2011). The (BR) is characterized by the characteristics of oracle (Knight and Fu, 2000; Xu et al., 2010). Although this estimate is good in terms of (vs) and estimating parameters, the (BR) produces incorrect standard errors. (Kyung et al., 2010). This reducing the practical use to the estimator. Many methods of Bayesian regularization are based on a scale mixture normal (SMN). Unfortunately, this
representation not properly available to the Bayesian bridge (BB) prior if the value of $\delta$ is $0 < \delta < 1$ (Mallick and Yi, 2018). Therefore, we using a scale mixture uniform (SMU) to represent the generalized Gaussian (GG) prior. Because this representation facilitates the work of Monte Carlo (MCMC) algorithm, which has good computational efficiency. If the ordinal data include more than three categories, here must determine a (prior distribution) for one or more categories (Johnson and Albert, 2006). In this paper, following (Mallick and Yi, 2018) we consider the conditional (GG) prior specification (the mean of GG distribution is 0, scale parameter (SP) is $\lambda^{-\frac{1}{2}}$, and shape parameter is $\delta$) of the model:

$$\pi(\beta) \propto \prod_{j=1}^{p} \exp\{-\lambda(\beta_j)^\delta\}. \quad (4)$$

We do not minimize the formula (3), we solve this problem when using (GS) that confirms construction of a Markov chain, that has a joint posterior for $\beta$ as a constant distribution. The statistical inference for (BB) is the reverse of the frequentist approach because it is simple. Moreover, can estimate the tuning parameter simply as a byproduct automatically to procedure MCMC (Mallick, 2015).

3- Bayesian bridge regression for ordinal data

Mallick and Yi (2018) proposed BBR using (SMU), which was characterized by good estimates compared with Bayesian and non-Bayesian methods. If possible responses for the response variable contains more than two categories and are ordinal in nature. The idea of "success" can be imagined in many different ways. When applying ordinal regression models, must assuming a simple imposition from data. The assumption is the odds must be parallel and proportional. This assumption indicates that all covariates have an equal effect on odds. If there (S) from the possible ordinal results, the predictions of the model does it will be $(S - 1)$ (O’Connell, 2006). Normal linear regression model:

$$y_i = x_i^T \beta + \varepsilon_i, \quad i = 1, \ldots, n,$$

$$\varepsilon_i \sim N(0, 1). \quad (5)$$

In a simple univariate case, the response variable $(y_i)$ takes one of $(S)$ ordered values (Jeliazkov et al., 2008). $X = (x_1, \ldots, x_n)^T$ is an $n \times p$ design matrix, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$ a vector of i.i.d. normal errors $(0, 1)$ (Park and Casella, 2008), where $E(\varepsilon_i|x_i) = 0$. Assuming that if there is a normal distribution (ND) with the cumulative distribution function (cdf) $F$, we can write the probability ($y$) to be equal to $s$ category through (Bürkner and Vuorre, 2018).

$$Pr(y = s) = F(y_{S+1}) - F(y_S). \quad (6)$$

In ordinal data $y_0, y_1, \ldots, y_S$ are cutpoints whose coordinates satisfy $-\infty = y_0 < y_1 < \cdots < y_{S-1} < y_S = \infty$. Here $y_{S-1}$ is the lower bound and the $y_S$ is the upper bound of the interval corresponding to response $s$ (Alhamzawi and Ali, 2018a). The latent variable $w_i$ depends on a $p$-vector for covariates $(x_i)$ through the model $w_i = x_i^T \beta + \varepsilon_i$, and the response variable

$$y_i = s \text{ if } y_{S-1} < w_i < y_S. \quad (7)$$

The cutpoint parameters are $-\infty = y_0 < y_1 < \cdots < y_{S-1} < y_S = \infty$ which determines the classification of data in categories $S$ (Feng et al., 2017). The (cdf) for $\varepsilon_i$, $F(\varepsilon_i)$ is $p_r(\varepsilon_i = s|\beta, y) = p_r(\{y_{S-1} < w_i\} \cap \{w_i \leq y_S\}) = \left[p_r\left(\{y_{S-1} < x_i^T \beta + \varepsilon_i\} \cap \{x_i^T \beta + \varepsilon_i \leq y_S\}\right)\right]$, we obtain that:
Given we can write the likelihood function for the model as:

$$p_r(y_i = s|\beta,\gamma) = F(y_i - x_i^T\beta) - F(y_{i-1} - x_i^T\beta).$$

(8)

Given \( y = (y_1, ..., y_n)^T \) we can write the likelihood function for the model as:

$$f(y|\beta,\gamma) = \prod_{i=1}^{n} \prod_{s=1}^{\mathbb{Y}} [F(y_s - x_i^T\beta) - F(y_{s-1} - x_i^T\beta)]^{I(y_i = s)},$$

(9)

where the indicator function for the event \( y_i = s \) is the \( 1\{y_i = s\} \), (Jeliazkov et al., 2008) (Alhamzawi, 2016b). Here if \( y_1 = 0 \) lead to removes the possibility of changing the distribution, and no shifting the probability of observing \( y_i \).

The identification problem is typically corrected simply by fixing the cutpoint, in adding to \( y_0 = -\infty \) and \( y_s = \infty \). In Bayesian inference, the most important steps are to select (prior distribution) for the parameters (Alhamzawi and Ali, 2018a; Walli, 2010). The prior selection plays the most important role (Alhamzawi and Yu, 2012). Must select a prior with carefully, because some problems can occur if the prior distribution is used without care. (Kinney and Dunson, 2007).

4- Bayesian inference

4.1- Prior elicitation:

To proceed a Bayesian analysis, we assume the generalized Gaussian distribution (GGD) which can be written such a scale mixture of uniform distribution (UD)

$$\frac{1}{2I(\frac{1}{\delta} + 1)} \exp\{ -\lambda|\xi|^{\delta} \} = \int^{1}_{x>0} \frac{1}{2I(\frac{1}{\delta} + 1)} z^{\frac{1}{\delta} + 1} \exp\{ -\lambda z \} dz$$

(10)

Equivalently, we may write (10) as follows:

$$\frac{1}{2I(\frac{1}{\delta} + 1)} \exp\{ \lambda|\xi|^{\delta} \} = \frac{1}{2I(\frac{1}{\delta} + 1)} \int^{1}_{x>\lambda|\xi|^{\delta}} \frac{x}{2} \exp\{ -\lambda x \} dx,$$

$$\int^{\frac{1}{\delta} + 1}_{-\frac{1}{\delta}} < x < \frac{1}{\delta} \frac{1}{2} \frac{1}{2I(\frac{1}{\delta} + 1)} v^{\frac{1}{\delta} + 1} \exp\{ -\lambda v \} dv .$$

(11)

The proofs are included in the Appendix 1.

In practice, we have found the mixture representation (11) perform better than (10) in sampling the regression coefficients in terms of prediction accuracy.

4.2- Hierarchical Representation:

We can write the hierarchical representation for Bayesian regression as follows

$$y_i = s \text{ if } y_{i-1} < w_i \leq y_s,$$

$$w_i|\beta \sim N(x_i^T\beta, 1).$$
4.3- Full Conditional Distributions

Under the hierarchical representation in (12), the full conditional distribution (FCD) of $\beta$ is given by:

$$\beta | y, X, v, \lambda \sim \mathcal{N}_p (\hat{\beta}_{OLS}, (X^T X)^{-1}) \prod_{j=1}^{p} I \left\{ |\beta_j| < \frac{v_j}{\lambda} \right\},$$  

(13)

The FCD of $v$ is given by:

$$v | y, X, \beta, \lambda \sim \prod_{j=1}^{p} \exp \left\{ \lambda \left( v_j - \frac{\lambda |\beta_j|^\delta}{\lambda} \right) \right\}.$$  

(14)

The proofs for (13 and 14) are included in the Appendix 2, and the (FCD) of ($y$) in formula (9).

Where $I(.)$ is an indicator function. To update ($\lambda$) the tuning parameter, must be we work directly with the density of the (GG), we can update ($\lambda$) when generating samples from the conditional posterior distribution (cpd). That is showing by (15), only if ($\lambda$) has distribution Gamma prior with parameters ($c, d$) (Mallick, 2015).

$$\pi (\lambda | y, X, \beta, v, \delta) \propto \lambda^{(c+p+\frac{p}{\delta})-1} \exp \left\{ -\lambda \left( d + \sum_{j=1}^{p} |\beta_j|^\delta \right) \right\}.$$  

(15)

The concavity parameter ($\delta$) is often prefixed beforehand (Mallick, 2015). (Xu et al., 2010) argued that the value of concavity parameter is $\delta = 0.5$ can be taken as a presentative of $\delta$, $0 < \delta < 1$. But in this article we don’t determine the value of $\delta$. However, $\delta$ can be assigned a beta distribution. If we assign a beta prior ($w, r$) for $\delta$, then the (PD) for $\delta$ is given by:

$$\pi (\delta | y, X, \beta, v, \lambda) \propto \delta^{w-1} (1-\delta)^{r-1} \frac{\Gamma(\frac{w}{\delta}+1)}{\Gamma(\frac{w}{\delta})} \exp \left\{ -\lambda \sum_{j=1}^{p} |\beta_j|^\delta \right\}.$$  

(16)

5 - Gibbs sampler

We using (GS) to simulate the unknowns from the (PD). Specifically, the algorithm of Gibbs sampling for the (BBR) is constructed through sampling coefficients from their (FCD) (Alhamzawi, 2016a). The (GS) algorithm is a special case of MCMC, especially useful when the (cpd) is known for each parameter (Yan and Su, 2009). To implement the (GS) algorithm, we must generate samples from conditional distributions (Czado, 1994). The (GS) is used to look for promising models without
computing the entire posterior (George and McCulloch, 1997). From (12) and (13) we can construct an efficient Gibbs sampler:

a- Sample $v_j$ from the exponential distribution $\{\exp(\lambda)I\{v_i > \lambda | \beta_j|\} \}$ and set $v_j = v_j^* + (|\beta_j|)^{\delta}$.

b- Sample $\beta$ from a truncated multivariate (ND) where

$$\beta | \gamma, X, \nu, \lambda \sim N_n(\bar{\beta}_{OLS}, (X^TX)^{-1}) \prod_{j=1}^{p} I(|\beta_j|^{\delta} < \frac{v_j}{\lambda}).$$

c- For $i = 1, ..., n$, Sample $w_i | y, \beta, \gamma \sim N_T(x_i^T \bar{\beta}, 1)$, where support is determined for the truncated normal distribution (NT) by the cutpoints $(y_{s-1}, y_s)$, that be associated with $y_i = s$.

d- Sampling $y_s$ from a (UD) determined between the max $w_i$ in category $s$ and min $w_i$ in category $s + 1$:

$$y_s | \gamma \sim U(\max \{w_i : y_i = s\}, \min \{w_i : y_i = s + 1\}), \ s = 2, ..., S - 1.$$  

The efficient (GS) based on this full conditions to extract samples from every full (cpd). Until all chains converge, the process will continue.

6- Simulation study

A simulation study was carried out to examine the performance of the proposed method and compare model fit with the Bayesian median Regression for Ordinal Model reported in (Rahman, 2016) and the ordinal probit model reported in Jeliazkov et al. (2008). In this simulation study, we simulate 1000 observations from the model $w_i = x_i^T \beta + \epsilon_i$. Here, $\beta = (-5, -10, 15)$, the predictors were sampled independently from a (UD) on the [-0.1, 0.1] and $\epsilon_i$ were sampled independently from a logistic distribution with location parameter $\mu = 0$ and (SP) $s = 1$. The cut-points used were $\delta_1 = -1.0, \delta_2 = -0.25, \delta_3 = 0.25$ and $\delta_4 = 0.84$. The number of observations corresponding to the five category of $y$ were 332, 123, 108, 141 and 296, respectively. The results are summarized in Table 1.

Table 1: Posterior mean and inefficiency factor (IF)

| Parameters | BBR (IF) | BMR (IF) | Ord (IF) | Propit (IF) |
|------------|----------|----------|----------|-------------|
| $\beta_1$  | -5.21    | -3.28    | -4.28    |             |
|            | (1.28)   | (1.93)   | (1.62)   |             |
| $\beta_2$  | -9.89    | -10.79   | -11.17   |             |
|            | (1.37)   | (1.78)   | (2.03)   |             |
| $\beta_3$  | 14.73    | 13.48    | 16.22    |             |
|            | (1.33)   | (2.04)   | (1.98)   |             |
| $\delta_1$ | -1.03    | -1.28    | -0.84    |             |
|            | (1.89)   | (2.39)   | (2.08)   |             |
| $\delta_2$ | -0.26    | -0.27    | -0.20    |             |
|            | (2.03)   | (2.19)   | (2.03)   |             |
| $\delta_3$ | 0.25     | 0.26     | 0.33     |             |
|            | (1.70)   | (1.83)   | (1.89)   |             |
| $\delta_4$ | 1.07     | 1.16     | 1.08     |             |
|            | (1.45)   | (1.89)   | (1.37)   |             |
From table 1, it can be seen that our proposed method performs better than the other methods in terms of prediction accuracy and inefficiency factor. The Deviance Information Criteria DIC was computed for the three models (BBR, BMR and Ord Propit) and the values were 438.27, 489.12, and 503.19, respectively. This also indicates that our proposed method performs better than the other methods.

7- Real data application
In this section, we use the Educational Attainment (NLSY) data (Jeliazkov et al., 2008) to analyse the performance of the proposed method. This NLSY data was started in 1979 by using 12,000 youths to conduct annual interviews on demographic questions. The outcome of interest is education degrees which has 4 levels: (1) less than high school, (2) high school degree, (3) college degree, and (4) graduate degree.

Table 2: Posterior mean for the Educational Attainment data

| Parameters          | BBR    | BMR    | Ord Propit |
|---------------------|--------|--------|------------|
| intercept           | -2.17  | -3.11  | -1.33      |
| family income       | 0.34   | 0.46   | 0.12       |
| mother’s education  | 0.22   | 0.17   | 0.06       |
| father’s education  | 0.19   | 0.27   | 0.09       |
| mother worked       | 0.05   | 0.10   | 0.04       |
| female              | 0.46   | 0.39   | 0.05       |
| black               | 0.39   | 0.31   | 0.05       |
| urban               | -0.08  | -0.11  | 0.06       |
| south               | 0.10   | 0.13   | 0.05       |
| age cohort 2        | -0.09  | -0.07  | 0.03       |
| age cohort 3        | -0.07  | -0.09  | 0.05       |
| age cohort 4        | 0.63   | 0.75   | 0.05       |
| $\delta_1$          | 0.97   | 0.80   | 0.01       |
| $\delta_2$          | 0.04   | 0.06   | 0.04       |

The DIC was computed for the three models (BBR, BMR and Ord Propit) and the values were 9776.24, 9811.22, and 9801.13, respectively. The results of DIC show that the proposed method performs better than the other methods.

8- Conclusion
In this paper, we have considered a Bayesian Bridge regression for univariate ordinal data, and proposes a method that can be extensively used in a wide class of applications. Simulation study shows that the proposed method is effective in regression. The real data analysis also show that our proposed method performs very well compared to other methods.

Appendix 1:

\[
\int_{z=|x|^{\delta}}^{z=1} \frac{1}{\lambda \Gamma(x+1)} \frac{1}{2z^{\delta}} e^{-\lambda z} \, dz
\]

Let \( v = \lambda z \Rightarrow z = \frac{v}{\lambda} \)

\( dv = \lambda \, dz \Rightarrow dz = \frac{dv}{\lambda} \), and \( z = \frac{v}{\lambda} \)

\( z > |x|^{\delta} \Rightarrow \frac{v}{\lambda} > |x|^{\delta} \Rightarrow vj > \lambda |x|^{\delta} \)
\[ z > |\beta| \frac{v}{\lambda} \Rightarrow v > |\beta| \frac{v}{\lambda} \Rightarrow v > \lambda|\beta| \frac{v}{\lambda} \]

\[
\int_{v > \lambda|\beta| \frac{v}{\lambda}} \frac{\lambda^\frac{1}{\delta}}{2v^\frac{1}{\delta}} \Gamma\left(\frac{1}{\delta} + 1\right) v^{\frac{1}{\delta} + 1 - 1} e^{-v} dv
\]

\[ \beta_j \sim \text{uniform} \left( -\frac{1}{\lambda}, \frac{1}{\lambda} \right) I\{|\beta_j| \leq \frac{v_j}{\lambda}\} \]

\[ v_j \sim \text{Gamma} \left( \frac{1}{\delta} + 1, 1 \right) I\{|v_j| > \lambda|\beta_j| \frac{v}{\lambda}\} \]

\[ \lambda \sim \text{Gamma}(c, d) I\{\lambda < \frac{v_j}{\beta_j^\delta}\} \]

**Appendix 2:**

The full conditional distribution

\[ \pi(\beta, v, \lambda, y, X) \propto \pi(y|X, \beta, \lambda) \pi(\beta|\lambda) \pi(v) \pi(\lambda) \]

The posterior distribution of \( \beta \) is:

\[ \pi(\beta, v, \lambda, y, X) \propto \pi(y|X, \beta, \lambda) \pi(\beta|\lambda). \]

\[ \propto \exp\left\{ -\frac{1}{2} (y - X\beta)^T(y - X\beta) \right\} \prod_{j=1}^{p} I\{|\beta_j| \leq \frac{v_j}{\lambda}\}, \]

\[ \propto \exp\left\{ -\frac{1}{2} (-2y^T X' \beta + \beta^T X'y \beta) \right\} \prod_{j=1}^{p} I\{|\beta_j| \leq \frac{v_j}{\lambda}\}, \]

\[ \propto \exp\left\{ -\frac{1}{2} (-2y^T X(X'T)^{-1}(X'T)X\beta + \beta^T X'y \beta) \right\} \prod_{j=1}^{p} I\{|\beta_j| \leq \frac{v_j}{\lambda}\}, \]

\[ \propto \exp\left\{ -\frac{1}{2} (-2\bar{\beta}^T X'T\beta + \beta^T X'y \beta) \right\} \prod_{j=1}^{p} I\{|\beta_j| \leq \frac{v_j}{\lambda}\}, \]

\[ \bar{\beta} \sim N_p(\bar{\beta}_{OLS}(X'T)^{-1}) \prod_{j=1}^{p} I\{|\beta_j| \leq \frac{v_j}{\lambda}\} \]

\[ \propto \exp\left\{ -\frac{1}{2} (-2y^T X(X'T + \eta_p)^{-1}(X'T + \eta_p) \beta + \beta^T (X'T + \eta_p) \beta) \right\} \prod_{j=1}^{p} I\{|\beta_j| \leq \frac{v_j}{\lambda}\}, \]

\[ \propto \exp\left\{ -\frac{1}{2} (-2\beta^T (X'T + \eta_p) \beta + \beta^T (X'T + \eta_p) \beta) \right\} \prod_{j=1}^{p} I\{|\beta_j| \leq \frac{v_j}{\lambda}\}. \]
The posterior distribution of $v$ is:

$$
\pi(v|\beta, \lambda) \propto \pi(\beta|v, \lambda)\pi(v),
$$

$$
\pi(v|y, X, \beta, \lambda) \propto \pi(v)I\{v_i > \lambda | \beta_j|^{\delta}\},
$$

$$
x \prod_{j=1}^{p} \exp(-v_j) I\{v_i > \lambda | \beta_j|^{\delta}\},
$$

$$
v \sim \prod_{j=1}^{p} \text{Exponential}(\lambda)I\{v_i > \lambda | \beta_j|^{\delta}\},
$$

(18)

The posterior distribution of $\lambda$ is:

$$
\pi(\lambda|\beta) \propto \pi(\beta|\lambda)\pi(\lambda)
$$

$$
\pi(\lambda|\beta) \propto \pi(\lambda)I\left\{\frac{v_j}{|\beta_j|^{\delta}}\right\}
$$

$$
\propto \lambda^{(c+1)-1} \exp\{-d\lambda\} I\left\{\frac{v_j}{|\beta_j|^{\delta}}\right\}
$$

$$
\propto \lambda \sim \text{Gamma}(c+1, d)I\left\{\lambda < \frac{v_j}{|\beta_j|^{\delta}}\right\},
$$

(19)

References:

[1] Agresti, A. (2010). *Analysis of ordinal categorical data* (Vol. 656). John Wiley and Sons.
[2] Alhamzawi, R. (2016a). Bayesian elastic net Tobit quantile regression. *Communications in Statistics-Simulation and Computation, 45*(7), 2409–2427.
[3] Alhamzawi, R. (2016b). Bayesian model selection in ordinal quantile regression. *Computational Statistics and Data Analysis, 103*, 68–78.
[4] Alhamzawi, R., and Ali, H. T. M. (2018a). Bayesian quantile regression for ordinal longitudinal data. *Journal of Applied Statistics, 45*(5), 815–828.
[5] Alhamzawi, R., and Ali, H. T. M. (2018b). Bayesian tobit quantile regression with penalty. *Communications in Statistics-Simulation and Computation, 47*(6), 1739–1750.
[6] Alhamzawi, R., and Yu, K. (2012). Variable selection in quantile regression via Gibbs sampling. *Journal of Applied Statistics, 39*(4), 799–813.
[7] Bürkner, P.-C., and Vuorre, M. (2018). Ordinal Regression Models in Psychology: A Tutorial. *PsyArXiv. September, 15.*
[8] Czado, C. (1994). Bayesian inference of binary regression models with parametric link. *Journal of Statistical Planning and Inference, 41*(2), 121–140.
[9] Feng X, Wu H, X. S. X. (2017). Bayesian adaptive lasso for ordinal regression with latent variables. *Sociological Methods and Research, 46*(4), 926–953.
[10] George, E. I., and McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica, 339–373.*
[11] Javed, F., and Mantalos, P. (2013). GARCH-type models and performance of information
criteria. Communications in Statistics–Simulation and Computation, 42(8), 1917–1933.
[12] Jeliazkov, I., Graves, J., and Kutzbach, M. (2008). Fitting and comparison of models for multivariate ordinal outcomes. In Bayesian econometrics (pp. 115–156). Emerald Group Publishing Limited.
[13] Johnson, V. E., and Albert, J. H. (2006). Ordinal data modeling. Springer Science and Business Media.
[14] Kinney, S. K., and Dunson, D. B. (2007). Fixed and random effects selection in linear and logistic models. Biometrics, 63(3), 690–698.
[15] Knight, K., and Fu, W. (2000). Asymptotics for lasso-type estimators. Annals of Statistics, 1356–1378.
[16] Kyung, M., Gill, J., Ghosh, M., and Casella, G. (2010). Penalized regression, standard errors, and Bayesian lassos. Bayesian Analysis, 5(2), 369–411.
[17] Mallick, H. (2015). Some contributions to Bayesian regularization methods with applications to genetics and clinical trials. The University of Alabama at Birmingham.
[18] Mallick, H., and Yi, N. (2018). Bayesian bridge regression. Journal of Applied Statistics, 45(6), 988–1008.
[19] O’Connell, A. A. (2006). Logistic regression models for ordinal response variables. Sage.
[20] Park, and Yoon, Y. J. (2011). Bridge regression: adaptivity and group selection. Journal of Statistical Planning and Inference, 141(11), 3506–3519.
[21] Park, and Casella, G. (2008). The bayesian lasso. Journal of the American Statistical Association, 103(482), 681–686.
[22] Polson, N. G., Scott, J. G., and Windle, J. (2014). The Bayesian bridge. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76(4), 713–733.
[23] Rahman, M. A. (2016). Bayesian quantile regression for ordinal models. Bayesian Analysis, 11(1), 1–24.
[24] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267-288.
[25] Walli, G. M. (2010). Bayesian variable selection in Normal regression models. na.
[26] Winship, C., and Mare, R. D. (1984). Regression models with ordinal variables. American Sociological Review, 512–525.
[27] Xu, Z., Zhang, H., Wang, Y., Chang, X., and Liang, Y. (2010). L 1/2 regularization. Science China Information Sciences, 53(6), 1159–1169.
[28] Yan, X., and Su, X. (2009). Linear regression analysis: theory and computing. World Scientific.
[29] Zhou, X. (2006). Bayesian inference for ordinal data. Rice University.