Wiley StatsRef:
Statistics Reference Online

Bayesian Ordinal Regression Modeling

By *James H. Albert*

Keywords: *latent data, Markov chain Monte Carlo, conditional means prior*

Abstract: This article provides an introduction to Bayesian modeling of an ordinal regression model, including the latent variable representation, choice of priors, and fitting by a Markov chain Monte Carlo algorithm. This article describes several extensions of the model, including hierarchical modeling, multivariate responses, and alternate latent data representations.

Ordinal data are a frequently encountered type of data in the social sciences. Survey data, in which respondents are asked to characterize their opinions on scales ranging from "strongly disagree" to "strongly agree," provide perhaps the most common example. Data recorded on scales ordered from one extreme of a quality to its opposite are often referred to as *Likert data* and the corresponding scale as a *Likert scale*.

One way to view ordinal data assumes the existence of an underlying latent (unobserved) variable associated with each response. Such a variable is often assumed to be drawn from a continuous distribution with a mean value that is modeled as a linear function of the respondent's covariate vector. This concept can be illustrated by a letter grade assigned to the students of a hypothetical statistics class. One assumes that there exists an underlying latent variable associated with each response. There are grade-cutoffs that divide the latent variable into categories, and the assigned grades, say A, B, C, D, and F, are observed for all students.

The latent variable formulation of the problem provides a model for the probability that a student receives a particular grade in the course, or in the more general case, that a response is recorded in a particular category. If we further assume that the responses or grades for a sample of n individuals are independent of one another given these probabilities, the sampling distribution for the observed data is given by a multinomial distribution.

To specify this multinomial distribution, suppose that there are C possible grades, which are denoted by $1, \ldots, C$. We observe the grades $y_1, \ldots, y_n$, where y_i denotes the observed grade for the ith individual. Associated with the ith individual's response, define a continuous latent variable Z_i. This variable is represented as $Z_i = x_i'\beta + \epsilon_i$, where x_i is a vector of covariates associated with the ith observation, and ϵ_i is distributed from the known distribution F. The grade $y_i = c$ is observed for the ith individual if the corresponding latent variable Z_i falls in the interval (γ_{c-1}, γ_c). If p_{ic} denotes the probability that a single response from the ith respondent falls into category c, we can represent the probability as

$$\begin{aligned} p_{ic} &= \Pr(\gamma_{c-1} < Z_i < \gamma_c) \\ &= F(\gamma_c - x_i'\beta) - F(\gamma_{c-1} - x_i'\beta) \end{aligned}$$

Bowling Green State University, Bowling Green, OH, USA

DOI: 10.1002/9781118445112.stat07867

1 Example

To illustrate an ordinal regression model, Long[1] describes a study to learn about variables helpful in predicting the article production of graduate student biochemistry doctoral programs. The response variable, the count of articles, is strongly right skewed and one approach for handling this skewness is to categorize the article count response into the ordinal categories "0 papers," "1–2 papers," "3–5 papers," and "more than 5 papers." One possible ordinal regression model is

$$Z_i = \beta_0 + \beta_1\, Female_i + Kid5_i + Ment_i + \epsilon_i$$

where Z_i is a continuous "publishing" latent variable, $Female_i$ indicates if the researcher is female, $Kid5_i$ is the number of children of age 5 or younger, and $Ment_i$ is the count of articles produced by the doctoral mentor. If one is fitting an ordinal probit regression, then one assumes that the latent data distribution F is the standard normal cdf.

2 Likelihood

Let p_i denote the corresponding vector of probabilities of the categories $1, \ldots, C$ for the ith individual $(p_{i1}, \ldots, p_{iC})$. Let $\mathbf{y} = (y_1, \ldots, y_n)$ denote the observed vector of responses for all individuals. The probability of observing these data, for a fixed value of the probability vectors $\{\mathbf{p}_i\}$, is given by the multinomial density proportional to

$$\Pr[y|\{p_i\}] \propto \prod_{i=1}^{n} p_{iy_i}$$

The unknown parameters are the regression vector β and the vector of cutoffs $\gamma = (\gamma_1, \ldots, \gamma_{C-1})$. Substituting the value for p_{ic}, the likelihood function of the unknown parameters is given by

$$L(\beta, \gamma) = \prod_{i=1}^{n} [F(\gamma_{y_i} - x_i'\beta) - F(\gamma_{y_i - 1} - x_i'\beta)]$$

Suppose that the vector of latent variables $Z = (Z_1, \ldots, Z_n)$ is introduced into the estimation problem. The likelihood function of this latent data Z and parameters β, γ is expressed as

$$L(\beta, \gamma, Z) = \prod_{i=1}^{n} f(Z_i - x_i'\beta) I(\gamma_{y_i - 1} \le Z_i < \gamma_{y_i})$$

where $I(\cdot)$ indicates the indicator function.

Ordinal regression models are often specified in terms of cumulative probabilities rather than individual category probabilities. Let $\theta_{ic} \sum_{j=1}^{c} p_{ij}$ denote the probability that the ith individual falls in category c or below. The ordinal model may be rewritten

$$\theta_{ic} = F(\gamma_c - x_i'\beta) \tag{1}$$

For example, if a logistic link function is assumed, this model becomes

$$\log\left(\frac{\theta_{ic}}{1 - \theta_{ic}}\right) = \gamma_c - x_i'\beta \tag{2}$$

See Ref. 2 and Chapter 3 of Ref. 3 for an overview of ordinal regresson models from a classical perspective. The MCMC fitting of the Bayesian form of this class of model was introduced by Albert and Chib[4] and

DOI: 10.1002/9781118445112.stat07867

described in detail in Ref. 5. General descriptions of MCMC algorithms can be found in Refs 6, 7, and 8. Congdon[9] provides a good survey of papers on Bayesian ordinal regression.

3 Priors

The ordinal regression model with C categories and $C-1$ unknown cutoff parameters $\gamma_1, \ldots, \gamma_{C-1}$ is overparameterized. To see this, note that if we add a constant to every cutoff value and subtract the same constant from the intercept regression term, the model is preserved. In the situation where little prior information is available, one approach to constructing a prior distribution is to first assume one cutoff parameter, say γ_1, is equal to zero—let γ^* denote the vector of cutoff parameters that are not fixed. Next assume that the vectors γ^* and β are independent. A uniform prior can be assigned on the components of γ^*, subject to the constraint

$$\gamma_2 \leq \cdots \leq \gamma_{C-1}$$

Similarly, a uniform prior can be selected for β

Instead of using a noninformative uniform distribution, one can construct a proper prior distribution that reflects one's prior beliefs about the location of the parameters. It is generally difficult to specify an informative prior distribution directly on the regression vector β and the vector of cutoffs γ. As the cutoff values and the regression coefficients are related nonlinearly to the probabilities of the categories $\{p_{ic}\}$, it can be hard to directly specify prior estimates of the components of β and γ. Instead, it is more convenient to construct an informative prior on (β, γ) indirectly by specifying distributions on the category probabilities. Johnson and Albert[5] generalize the conditional means approach of Ref. 10 for the binary regression situation.

4 MCMC Sampling

Suppose that the vector of regression parameters β and category cutoffs γ are assigned the prior density $g(\beta, \gamma)$. Then, the posterior density is given, up to a proportionality constant, by

$$g(\beta, \gamma|\mathbf{y}) \propto L(\beta, \gamma)g(\beta, \gamma)$$

where the likelihood function is given in Section 2 and the cutoffs satisfy the order restriction $\gamma_1 < \cdots < \gamma_{c-1}$.

The generic Metropolis–Hastings algorithms with multivariate normal proposal densities are not well suited for generating candidate vectors in the ordinal data setting. Instead, hybrid Metropolis–Hastings/Gibbs algorithms are generally used to sample from the posterior distribution of ordinal regression parameters. Several such algorithms have been proposed for the case of a probit link; among the more notable are those of Refs 4, 11, and 12. Each of these algorithms exploits the latent data formulation of ordinal response models. Johnson and Albert[5] describe the algorithm of Cowles that is relatively simple to implement, the simulated parameter values display good mixing, and the algorithm can be extended to models with arbitrary constraints on the category cutoffs.

5 Fitting of the Ordinal Regression for the Example

Martin *et al.*[13] have an R function `MCMCoprobit` in their `MCMCpack` package that generates a sample from the posterior distribution for a Bayesian ordered probit regression model using the algorithm of Ref. 11. The data set `bioChemists` is taken from the `pscl` package of Jackman[14]. The ordinal response

DOI: 10.1002/9781118445112.stat07867

variable is `ordinal_art` and the explanatory variables are `fem`, `kid5`, and `ment`. The default prior specifications are to assign uniform (improper) priors on both the regression vector β and the vector of cutpoints γ. The Bayesian ordinal probit regression is fit by the command

```
fit <- MCMCoprobit(ordinal_art ~ fem + kid5 + ment,
                   data=ord_data)
```

When one asks for a `summary` of this fit, the following output is given, which provides posterior means and standard deviations for each of the components of β and γ. The "Time-series SE" column provides simulation standard errors for the computed posterior means. When one inspects trace plots of the simulated values of the parameters, one notes that there is good mixing for the regression coefficients β_j, but there is some autocorrelation structure in the trace plots for the cut-point parameters γ_2 and γ_3.

```
1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                Mean       SD  Naive SE Time-series SE
(Intercept)  1.85704 0.093566 9.357e-04      1.016e-02
femWomen     0.12854 0.078027 7.803e-04      9.530e-04
kid5         0.09773 0.051430 5.143e-04      6.352e-04
ment        -0.02231 0.003863 3.863e-05      4.577e-05
gamma2       1.77464 0.072086 7.209e-04      1.184e-02
gamma3       2.29744 0.071816 7.182e-04      1.047e-02
```

6 Extensions of the Ordinal Response Model

There has been substantial research in Bayesian ordinal regression modeling and we provide an overview of different extensions of the basic ordinal regression model.

6.1 Alternative Latent Models

The general Bayesian formulation of the ordinal regression model is based on the normal latent data. Albert and Chib[15], following Ref. 3, describe Bayesian modeling using alternative latent data representations.

6.2 Hierarchical Models for Ordinal Data

A popular extension of Bayesian ordinal regression are hierarchical or multilevel regression models, where the data can be separated into groups, regression models are assigned to each group, and a second-level distribution modes the relationship between the group-specific regressions. These types of models are especially useful for studies where repeated ordinal measurements are made for each individual. Random effects models are developed in Refs 16–23, and 24.

6.3 Choice of Link Function

One assumption in the Bayesian model is the probit link that connects the cumulative probabilities with the linear predictor. Lang[25] illustrates Bayesian ordinal regression modeling with the use of a parametric family of mixture links.

DOI: 10.1002/9781118445112.stat07867

6.4 Multivariate Ordinal Data

Another extension, related to random effects modeling, assumes that the individuals have a general multivariate ordinal response. Bayesian analyes of multivariate ordinal data are described in Refs 26–29, and 30.

Related Articles

Gibbs Sampling; **Markov Chain Monte Carlo Algorithms**; **Ordinal Data**; **Probit Analysis**; **Probits**; **Probit Model**; **Multivariate Probit**; **Multinomial Probit and Logit**

References

[1] Long, J.S. (1990) The origins of sex differences in science. *Soc. Forces* **68** (3), 1297–1316.
[2] McCullagh, P. (1980) Regression models for ordinal data. *J. R. Stat. Soc. B* **42**, 109–142.
[3] Fahrmeier, L. and Tutz, G. (1994) *Multivariate Statistical Modeling Based on Generalized Linear Models*, Springer-Verlag, New York.
[4] Albert, J.H. and Chib, S. (1993) Bayesian analysis of binary and polychotomous response data. *J. Am. Stat. Assoc.* **88**, 669–679.
[5] Johnson, V.E. and Albert, J.H. (1999) *Ordinal Data Modeling*, Springer Science and Business Media, New York.
[6] Gelfand, A.E. and Smith, A.F.M. (1990) Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* **85**, 398–409.
[7] Chib, S. and Greenberg, E. (1995) Understanding the Metropolis—Hastings algorithm. *Am. Stat.* **49**, 327–335.
[8] Tierney, L. (1994) Markov chains for exploring posterior distributions. *Ann. Stat.* **22**, 1701–1762. (With discussion.)
[9] Congdon, P. (2005) *Bayesian Models for Categorical Data*, John Wiley & Sons, Inc., New York.
[10] Bedrick, E.J., Christensen, R., and Johnson, W. (1996) A new perspective on priors for generalized linear models. *J. Am. Stat. Assoc.* **91**, 1450–1460.
[11] Cowles, M.K. (1996) Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Stat. Comput.* **6**, 101–111.
[12] Nandram, B. and Chen, M.-H. (1996) Accelerating Gibbs sampler convergence in the generalized linear models via a reparameterization. *J. Stat. Comput. Simul.* **54**, 129–144.
[13] Martin, A.D., Quinn, K.M., and Park, J.H. (2011) MCMCpack: Markov chain Monte Carlo in R. *J. Stat. Softw.* **42** (9), 1–21. R package version 1.3-5.
[14] Jackman, S. (2015) *pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory*, Stanford University. Department of Political Science, Stanford University, Stanford, CA. R package version 1.4.9. http://pscl.stanford.edu/
[15] Albert, J.H. and Chib, S. (2001) Sequential ordinal modeling with applications to survival data. *Biometrics* **57**, 829–836.
[16] Gajewski, B.J., Hart, S., Bergquist-Beringer, S., and Dunton, N. (2007) Inter-rater reliability of pressure ulcer staging: ordinal probit Bayesian hierarchical model that allows for uncertain rater response. *Stat. Med.* **26**, 4602–4618.
[17] Gill, J. and Casella, G. (2009) Nonparametric priors for ordinal Bayesian social science models: specification and estimation. *J. Am. Stat. Assoc.* **104**, 453–454.
[18] Ishwaran, H. (2000) Univariate and multirater ordinal cumulative link regression with covariate specific cutpoints. *Can. J. Stat.* **28**, 715–730.
[19] Ishwaran, H. and Gatsonis, C.A. (2000) A general class of hierarchical ordinal regression models with applications to correlated ROC analysis. *Can. J. Stat.* **28**, 731–750.
[20] Manda, S.O. (2002) A Bayesian ordinal model for heterogeneity in a multi-centre myocardial infarction clinical trial. *Stat. Med.* **21**, 3011–3022.
[21] Mwalili, S.M., Lesaffre, E., and Declerck, D. (2005) A Bayesian ordinal logistic regression model to correct for interobserver measurement error in a geographical oral health study. *J. R. Stat. Soc., Ser. C: Appl. Stat.* **54**, 77–93.
[22] Qiu, Z., Song, P.X.K., and Tan, M. (2002) Bayesian hierarchical models for multi-level repeated ordinal data using WinBUGS. *J. Biopharm. Stat.* **12**, 121–135.
[23] Tan, M., Qu, Y., Mascha, E.D., and Schubert, A. (1999) A Bayesian hierarchical model for multi-level repeated ordinal data: analysis of oral practice examinations in a large anaesthesiology training programme. *Stat. Med.* **18**, 1983–1992.
[24] Xie, M., Simpson, D.G., and Carroll, R.J. (2000) Random effects in censored ordinal regression: latent structure and Bayesian approach. *Biometrics* **56**, 376–383.

DOI: 10.1002/9781118445112.stat07867

[25] Lang, J.B. (1999) Bayesian ordinal and binary regression models with a parametric family of mixture links. *Comput. Stat. Data Anal.* **28**, 59–87.

[26] Biswas, A. and Das, K. (2002) A Bayesian analysis of bivariate ordinal data: Wisconsin epidemiologic study of diabetic retinopathy revisited. *Stat. Med.* **21**, 549–559.

[27] DeYoreo, M. and Kottas, A. (2014) Bayesian Nonparametric Modeling for Multivariate Ordinal Regression, arXiv:1408.1027.

[28] Kottas, A., Müller, P., and Quintana, F. (2005) Nonparametric Bayesian modeling for multivariate ordinal data. *J. Comput. Graph. Stat.* **14**, 610–625.

[29] Lawrence, E., Bingham, D., Liu, C., and Nair, V.N. (2008) Bayesian inference for multivariate ordinal data using parameter expansion. *Technometrics* **50**, 182–191.

[30] Webb, E.L. and Forster, J.J. (2008) Bayesian model determination for multivariate ordinal and binary data. *Comput. Stat. Data Anal.* **52**, 2632–2649.

DOI: 10.1002/9781118445112.stat07867