Ordinal Regression Made Easy: A Bayesian Nonparametric Approach

Jizhou Kang [1] Athanasios Kottas [1]

[1]Department of Statistics, University of California, Santa Cruz

Introduction

Ordinal responses are widely encountered in many fields, including econometrics and the biomedical and social sciences, typically accompanied by covariate information. A univariate ordinal response Y with C categories can be encoded as a $C-$dimensional binary vector $\mathbf{Y}$. The modeling challenge for the ordinal regression problem involves capturing general regression relationships in the response probabilities (referred as the probability response curves), while appropriately accounting for the ordinal nature of the response distribution.

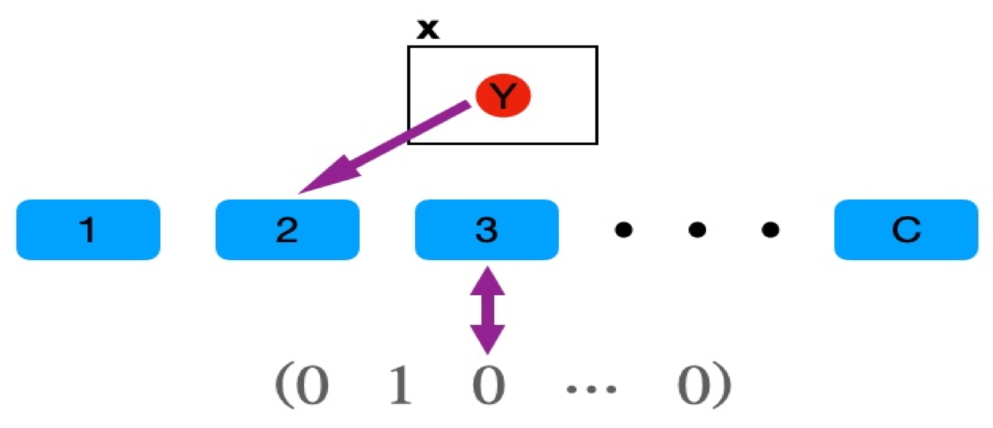


Figure 1. Illustration of data structure.

Traditional Approach

Traditionally, the ordinal regression problem is approached by treating the ordinal responses as a discretized version of latent continuous responses, which are usually assumed to be normally distributed, resulting in popular ordinal probit models.

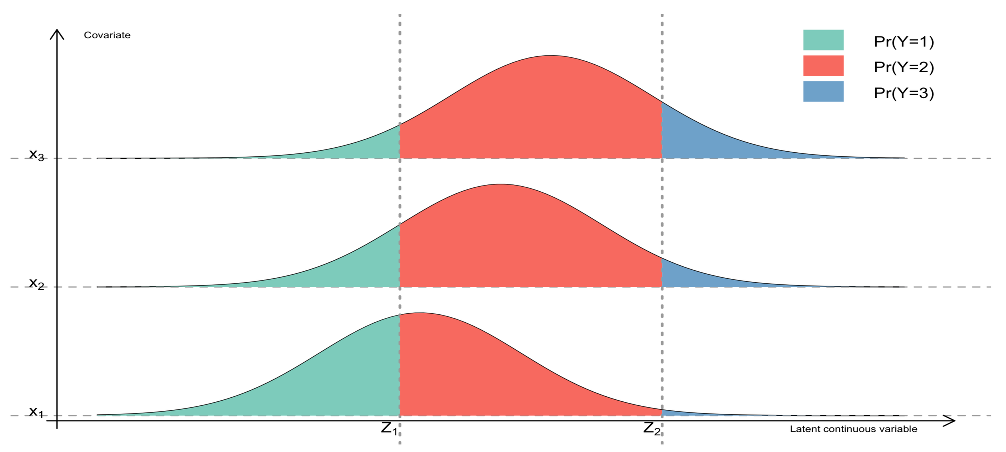


Figure 2. Illustration of probit regression.

The traditional approach is limited in the following ways:

- $Pr(Y=1|\mathbf{x})$ and $Pr(Y=C|\mathbf{x})$ are monotonically increasing or decreasing as a function of covariate $\mathbf{x}$, and they must have the opposite type of monotonicity;
- The direction of monotonicity changes exactly once in moving from category 1 to C (referred to as the single crossing property);
- The relative effect of one covariate to another is the same for every ordinal level and any covariate value.

Our Approach

- We directly model the discrete distribution of the ordinal responses, using the continuation-ratio logits representation of multinomial distribution.

$$\mathbf{Y} \sim Mult(\mathbf{Y}|m, \pi_1, \cdots, \pi_C) \Longleftrightarrow$$
$$\mathbf{Y} \sim Bin(Y_1|m_1, \varphi(\theta_1)) \cdots Bin(Y_{C-1}|m_{C-1}, \varphi(\theta_{C-1})) = K(\mathbf{Y}|\mathbf{m}, \boldsymbol{\theta})$$

- We generalize this parametric model via Bayesian nonparametric modeling. Placing a covariate-dependent nonparametric prior.

$$\mathbf{Y}|G_{\mathbf{x}} \sim \int K(\mathbf{Y}|\mathbf{m}, \boldsymbol{\theta}) dG_{\mathbf{x}}(\boldsymbol{\theta}) = \sum_{\ell=1}^{\infty} \omega_\ell(\mathbf{x}) K(\mathbf{Y}|\mathbf{m}, \boldsymbol{\theta}_\ell(\mathbf{x}))$$

Model Properties

The continuation-ratio logits structure offers a sequential mechanism to allocate the ordinal response Y.

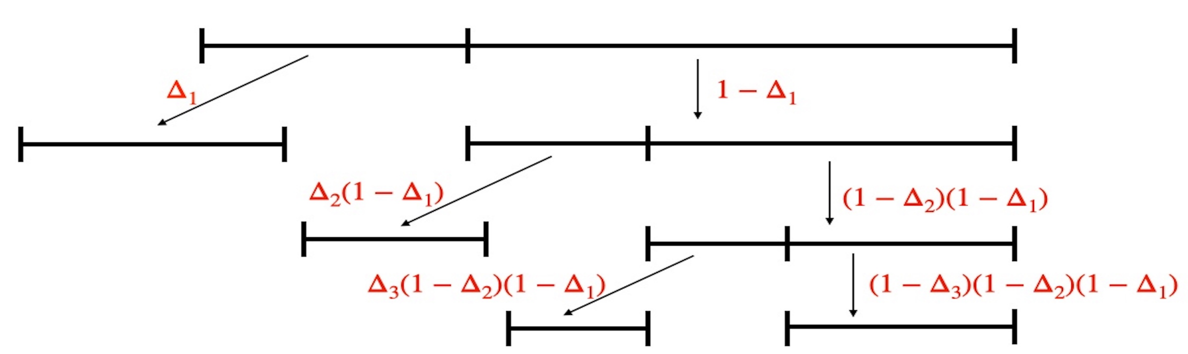
Generate $\mathcal{H}_1 \sim Bern(\Delta_1), \Delta_1 = \varphi(\mathbf{x}^T\boldsymbol{\beta}_1), \boldsymbol{\beta}_1 \sim N(\boldsymbol{\mu}_1, \Sigma_1)$

$\mathcal{H}_1 = 1$: $Y = 1$; $\mathcal{H}_1 = 0$: Generate $\mathcal{H}_2 \sim Bern(\Delta_2), \Delta_2 = \varphi(\mathbf{x}^T\boldsymbol{\beta}_2), \boldsymbol{\beta}_2 \sim N(\boldsymbol{\mu}_2, \Sigma_2)$

$\mathcal{H}_2 = 1$: $Y = 2$; $\mathcal{H}_2 = 0$: Generate $\mathcal{H}_3 \sim Bern(\Delta_3), \Delta_3 = \varphi(\mathbf{x}^T\boldsymbol{\beta}_3), \boldsymbol{\beta}_3 \sim N(\boldsymbol{\mu}_3, \Sigma_3)$

$\mathcal{H}_3 = 1$: $Y = 3$; $\mathcal{H}_3 = 0$: $\cdots$

It can also be viewed as a process that determines the stick-breaking weights.



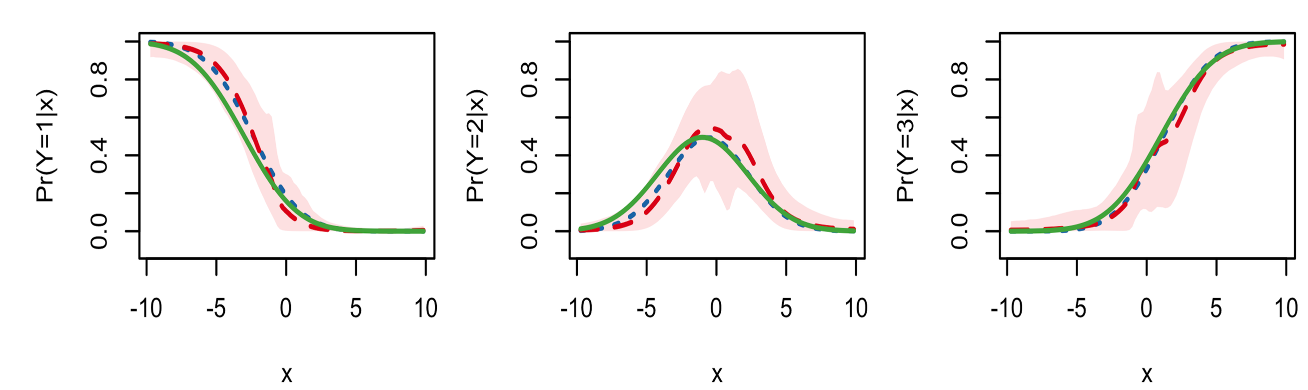
The symmetric structure in the weights and the atoms of the induced mixture model leads to the following benefits:

1. **Flexible**: allow general forms for ordinal response distribution and ordinal regression relationship.
2. **Easy**: clear prior specification strategy, efficient posterior inference algorithm, and easy to work with expressions for quantities of interest.

Simulation Study

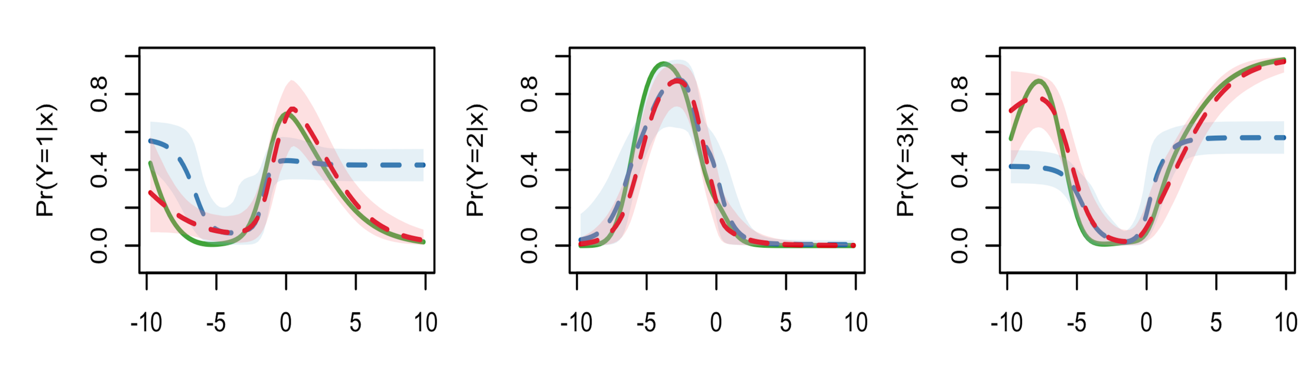
First experiment

In the first experiment, we check on potential overfitting of the model. The true probability responses curves, given in green solid lines, have standard shape. The model estimates capture the truth, even though it is substantially more complex than the data generating mechanism.



Second experiment

In the second experiment, we seek to highlight the model's capacity in capturing nonstandard shapes of probability response curves (green solid lines). The red dashed lines and shadow area provide the point and interval estimate of our proposed model. The blue dashed lines and shadow area are the estimate from a competing method. Contrasting the performance highlights the practical utility of our models in effective estimation of non-standard probability response curves.



Real Data Examples

Credit ratings of U.S. firms

The dataset contains Standard and Poors (S&P) credit ratings for 921 U.S. firms. For each firm, a credit rating on a five-point ordinal scale is available, along with five characteristics. We are interested in estimating the first-order marginal probability curves and the second-order marginal probability surfaces.

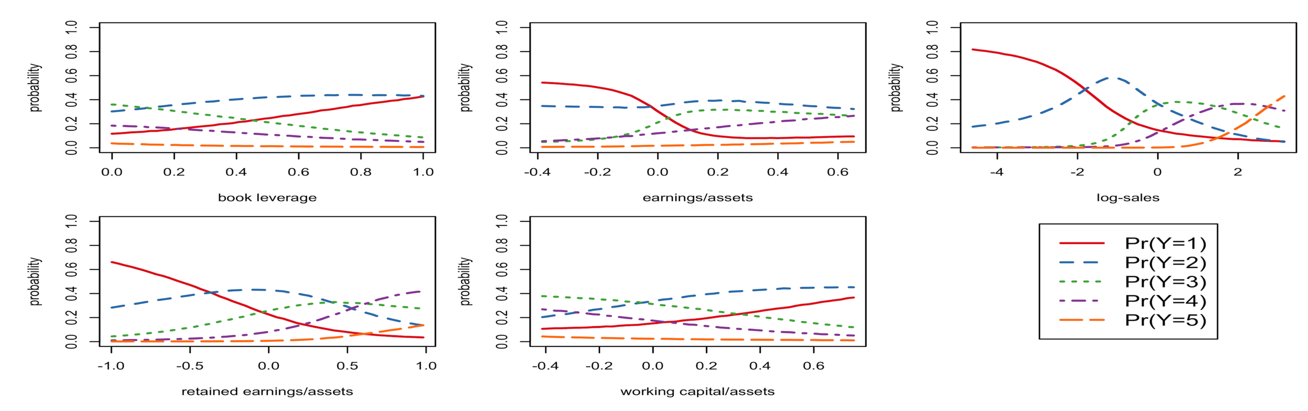


Figure 3. First-order marginal probability curves.

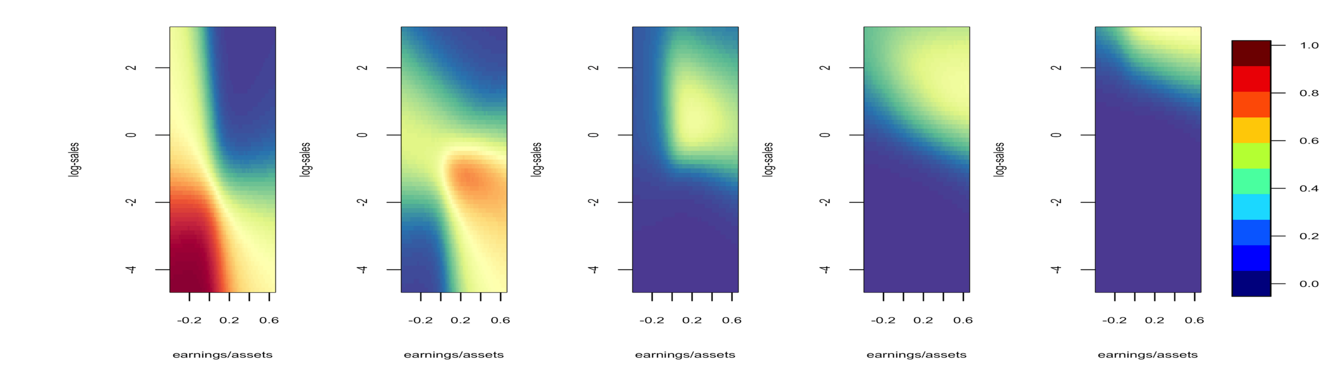


Figure 4. Second-order marginal probability surfaces. The corresponding credit rating decreases from left to right.

Developmental toxicity study (DYME)

Segment II designs data of an organic solvent, diethylene glycol dimethyl ether (DYME). At dose level $x_i, i = 1, \cdots, N$, n_i pregnant laboratory animals (dams) exposed to the toxin at level x_i. For the jth dam at dose x_i, the experiment record (i) m_{ij}: number of implants; (ii) R_{ij}: number of resorptions and prenatal deaths; (iii) y_{ij}: number of live pups with a malformation. Researchers are interested in three dose-response relationships: (i) probability of embryolethality: $Pr(R^* = 1|G_{\mathbf{x}})$; (ii) probability of malformation: $Pr(Y^* = 1|R^* = 0, G_{\mathbf{x}})$ (iii) combined risk function $Pr(R^* = 1 \text{ or } Y^* = 1|G_{\mathbf{x}})$.

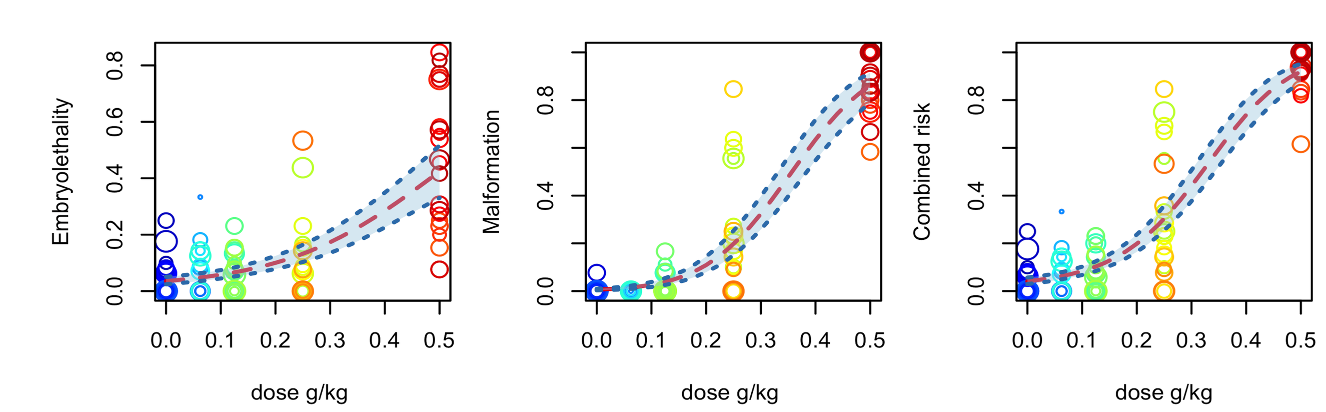


Figure 5. Posterior point and 95% interval estimate of three dose-response curves. The circles are the original data.

Conclusion and Future Work

We proposed a modeling framework for ordinal regression that achieves a good balance between model flexibility and implementation difficulty. It has at least the following two possible directions for future extensions.

1. **Scalability of the algorithm**: for large amount of data with moderate to large number of possible categories.
2. **Extension to longitudinal studies**: for individuals with ordinal responses that measured repeatedly over time.

We are interested in potential collaboration opportunities. Please contact me if you have suitable applications/datasets.