

Ordinal categorical outcomes?

Bayesian Modeling for Socio-Environmental Data

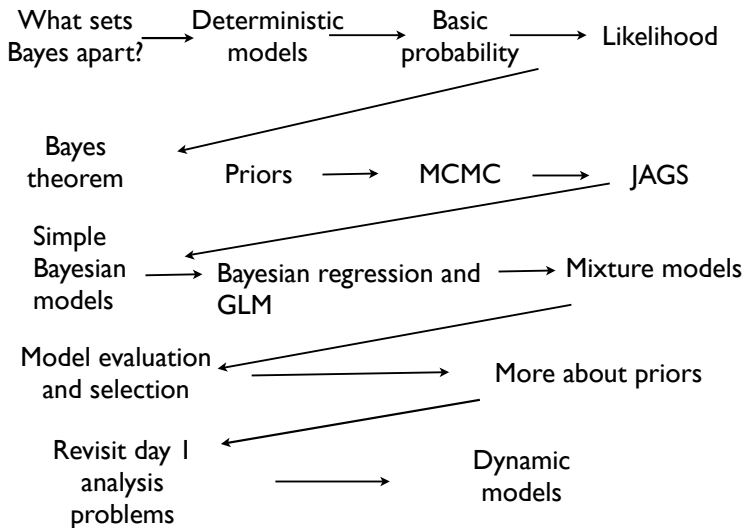
Mary B. Collins

August 28, 2015

Please fill out the course survey

Click here

The screenshot shows a SharePoint site for the "SESYNC Bayesian Modeling Short Course". The browser address bar displays the URL: https://collab.seSYNC.org/groups/bayescourse/_layouts/15/start.aspx#/SitePages/Home.aspx. The site has a green header with "SharePoint", "Newsfeed", and "OneDrive" links. A left sidebar contains navigation links: Home, Notebook, Documents, Wiki, Email List, Discussion Board, Research Support, and Site Contents. The main content area is titled "Newsfeed" and includes a "Start a conversation" input field. Below this, there are three posts. The first post is from Mary Collins, dated "Yesterday at 10:25 AM", and contains the text "Please take our course evaluation!" followed by a green link: <https://www.surveymonkey.com/r/CQ6MRYP>. A black arrow points from a green box labeled "Click here" to this link. The second post is from Chris Che-Castaldo, dated "Wednesday at 12:37 PM", and mentions an updated HTML file and 5 PDF papers. The third post is from William Admin. On the right side of the newsfeed, there is a "Documents" section with a "new document or drag files here" button and a list of course materials including "Day 4 - Priors and MCMC", "Day 10 - Bayesian Models IV", "Day 9 - Bayesian Models III", "Day 8 - Bayesian Models II", "Day 7 - Bayesian Models I (Regression)", "Day 6 - JAGS", "Day 3 - Likelihood and Bayes Derived", "Day 2 - Intro to Probability", and "Day 1 - What Sets Bayes Apart".



Considering dependent variables that are ordinal:

- Likert scales (very dissatisfied to very satisfied)
- Size categories (small, medium, large)
- Achievement groupings (some high school to graduate degree)
- etc.

Ordinal-level quantities are sometimes called *polychotomous*.

Besides Hobbs and Hooten, 2015...

...I also rely heavily on

- 1 Gelman, Andrew & Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- 2 Jackman, Simon. 2009. *Bayesian Analysis for the Social Sciences*. John Wiley & Sons, Ltd.

Ordinary regression analysis is inappropriate.

- Regression is an incoherent model for discrete variables (makes continuous predictions that are impossible for the data).
- We can extend the GLM for binary data to model ordinal data with predictors.
- “Cumulative link GLM” (McKelvey & Zavoina, 1975; Aitchison & Silvey 1957). Logistic regression can be extended to multiple categories (ordered or unordered).
- Caveat for 7+ ordinal categories: treat as continuous.

The ordered multinomial logit model

Consider a categorical outcome y that can take the values $1, 2, \dots, K$

$$\Pr(y > 1) = \text{logit}^{-1}(X\beta) \quad (1)$$

$$\Pr(y > 2) = \text{logit}^{-1}(X\beta - c_2) \quad (2)$$

$$\Pr(y > 3) = \text{logit}^{-1}(X\beta - c_3) \quad (3)$$

$$\dots \quad (4)$$

$$\Pr(y > K - 1) = \text{logit}^{-1}(X\beta - c_{K-1}) \quad (5)$$

In this model c are thresholds or *cutpoints*

Probabilities of individual outcomes:

$$\Pr(y = k) = \Pr(y > k - 1) - \Pr(y > k) \quad (6)$$

$$= \text{logit}^{-1}(X\beta - c_{k-1}) - \text{logit}^{-1}(X\beta - c_k) \quad (7)$$



Fracking and risk perception example.

Data background: 2500 US respondents took a ~30 web-based environmental risk perception survey. Due to survey length, not all respondents were asked all questions.

Dependent variable: 4-point ordinal acceptability of fracking.

Independent variables: political ideology (liberal – conservative), 3-class political party identification, environmental values index (NEP battery), income, gender, race (white vs. non-white).

DAG

