Ordinal Graphical Models: A Tale of Two Approaches

Arun Sai Suggala [1]

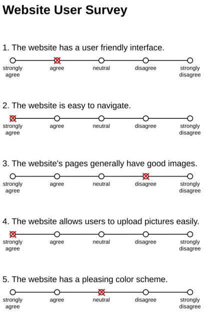
Joint work with Eunho Yang[2], Pradeep Ravikumar[1].

[1]Carnegie Mellon University, [2]Korea Advanced Institute of Science and Technology.

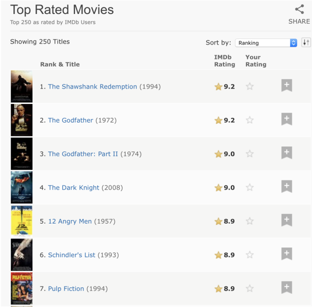
August 7, 2017

Multivariate Ordinal Data

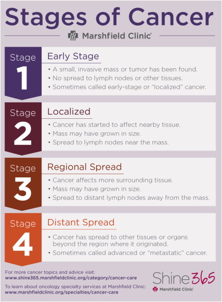
Many real world applications involve ordered categorical data also known as ordinal data.



surveys



movie ratings



medical data

Figure: Examples of ordinal data.

Figures from https://en.wikipedia.org/wiki/Likert_scale, http://www.imdb.com, http://shine365.marshfieldclinic.org/cancer-care/

Multivariate Ordinal Distributions

- Multivariate ordinal models, especially graphical models, help us understand dependencies between variables of interest.
- Existing models have one or more of the following problems:
  - estimators that do not scale well to high dimensions.
  - estimators with no strong statistical guarantees.
  - no graphical model representation.

Aim

Develop multivariate ordinal graphical model distributions

- and provide computationally tractable estimators, that come with strong statistical guarantees.

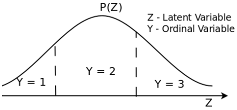
Univariate Ordinal Distributions

- A huge line of work exists on designing ordinal univariate conditional models
  - also called as ordinal regression models.
- Most of these fall into the following two categories:
  - Logit Models.
  - Latent Variable Models.

Logit Models

- Let $Y \in \{1, \dots M\}$ be a ordinal random variable and **X** be the vector of covariates.
- Logit models parametrize various Log-odds ratios of Y.
- Popular models in this class include:
  - cumulative logit model: $\log \left[\frac{\mathbb{P}(Y \leq k | \mathbf{X})}{\mathbb{P}(Y > k | \mathbf{X})} \right] = \theta_k + \mu(\mathbf{X})$.
  - continuation logit model: $\log \left[\frac{\mathbb{P}(Y = k | \mathbf{X})}{\mathbb{P}(Y > k | \mathbf{X})} \right] = \theta_k + \mu(\mathbf{X})$.
  - consecutive logit model: $\log \left[\frac{\mathbb{P}(Y = k | \mathbf{X})}{\mathbb{P}(Y = k+1 | \mathbf{X})} \right] = \theta_k + \mu(\mathbf{X})$.

- These models treat the ordinal variable as a discretized version of a continuous latent random variable.
- Let Z be the latent random variable. Then



$$Y = k \text{ iff } Z \in [\theta_{k-1}+\mu(\mathbf{X}), \theta_k+\mu(\mathbf{X}))$$

- Popular choices for Z include gaussian, logistic distributions.

Designing Multivariate Ordinal Distributions

- Leverage the univariate ordinal distributions to construct multivariate ordinal distributions.
- Two possible approaches, corresponding to each class of univariate distributions:
  - **Logit Models:** Specify conditional logit distributions that result in a consistent joint distribution.
  - **Latent Variable Models:** Assume that the ordinal random vector is generated through quantization of a continuous latent random vector.

Multivariate Models from Univariate Logit Models

- We use univariate Logit models to specify conditional distributions of each variable given the rest.
- then study when these lead to a consistent joint distribution via Hammersley-Clifford-type analyses.

- Let $\mathbf{Y} = (Y_1, \ldots, Y_p)$ be a ordinal random vector.
- Suppose, the conditional distributions either follow cumulative logit or continuation logit distributions:

$$\text{(cumulative logit)} \;\; \log\left[\frac{\mathbb{P}(Y_s \leq j | Y_{\setminus s})}{\mathbb{P}(Y_s > j | Y_{\setminus s})}\right] = \theta_{s;j} + \mu_s(Y_{\setminus s})$$

$$\text{(continuation logit)} \;\; \log\left[\frac{\mathbb{P}(Y_s = j | Y_{\setminus s})}{\mathbb{P}(Y_s > j | Y_{\setminus s})}\right] = \theta_{s;j} + \mu_s(Y_{\setminus s}).$$

Theorem

For any choice of functions $\{\mu_s(.)\}_{s\in[p]}$, the conditional distributions ***are not consistent*** *with any joint distribution over* $\mathbf{Y}$.

Consecutive Logit Model

Suppose, the conditional distributions have the following cumulative logit distribution:

$$\log \left[\frac{\mathbb{P}(Y_s = j | Y_{\setminus s})}{\mathbb{P}(Y_s = j+1 | Y_{\setminus s})} \right] = \theta_{s;j} + \mu_s(Y_{\setminus s}).$$

Theorem

The conditional distributions are consistent with a pairwise graphical model distribution w.r.t an undirected graph $G = (V, E)$, if and only if each $\mu_s(Y_{\setminus s})$ has the following form

$$\mu_s(Y_{\setminus s}) = \sum_{t \in N(s)} \theta_{st} \big(M - Y_t\big).$$

The corresponding joint distribution is given by

$$\mathbb{P}(\mathbf{Y}) \propto \exp \Bigg(\sum_{s \in V, j \in [M-1]} \theta_{s;j} \mathcal{I}[Y_s \leq j] + \sum_{(s,t) \in E} \theta_{st} \big(M - Y_s\big)\big(M - Y_t\big) \Bigg).$$

Estimation of Consecutive Logit Model

We solve a regularized node conditional log likelihood maximization problem at each node s:

$$\arg\min_{\theta_{s\cdot}} \; -\mathbb{E}_n \left[\log \mathbb{P}(Y_s | Y_{\setminus s})\right] + \lambda_n \sum_{t \neq s} |\theta_{st}|,$$

where $\mathbb{E}_n[f(Y)]$ is the sample mean of $f(Y)$.

Statistical Guarantees

Guarantees for estimators of exponential family graphical models [YRAL15, TPSR15] carry over to the consecutive ratio model.

Multivariate Models from Univariate Latent Variable Models

- The ordinal random vector is modeled as quantization of a continuous latent random vector.
- We study the probit model, where the multivariate latent random vector is multivariate Gaussian.

Probit Model

Let $\mathbf{Y} = (Y_1, \ldots, Y_p)$ and $\mathbf{Z} = (Z_1, \ldots, Z_p)$ be the ordinal and latent random vectors.

Probit Model

$$\text{(Latent Vector)} \quad \mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \Sigma), \quad \text{where } \operatorname{diag}(\Sigma) = 1,$$

$$\text{(Ordinal Vector)} \quad Y_j = k, \text{ iff } Z_j \in [\theta_{j,k-1}, \theta_{j,k}), \quad \forall j \in [p].$$

Multistage Estimation of Probit Model

Stage I - Estimation of Thresholds (θ)

To estimate the thresholds for Y_j, we maximize the univariate marginal log likelihood for Y_j.

Stage I - Estimation of Thresholds (θ)

To estimate the thresholds for Y_j, we maximize the univariate marginal log likelihood for Y_j.

Stage II - Raw Estimate of Correlation Matrix (Σ)

To estimate Σ_{jk}, we maximize the bivariate marginal log likelihood for (Y_j, Y_k).

Multistage Estimation of Probit Model

Stage I - Estimation of Thresholds (θ)

To estimate the thresholds for Y_j, we maximize the univariate marginal log likelihood for Y_j.

Stage II - Raw Estimate of Correlation Matrix (Σ)

To estimate Σ_{jk}, we maximize the bivariate marginal log likelihood for (Y_j, Y_k).

Stage III - Smoothed Estimate of Correlation Matrix (Σ)

We plug-in the estimate from Stage II into GLASSO estimator.

Theorem

Suppose the true correlation matrix satisfies incoherence condition and the bivariate likelihood functions satisfy certain regularity conditions. Then if GLASSO is run with $\lambda_n \asymp \sqrt{\log p'/n}$ *and n is lower bounded as* $n \gtrsim d^2 \log \max\{n, p\}$*, where d is the maximum node degree in the latent graph, then the inverse of estimate* $\widehat{\Sigma}$ *from Stage III satisfies the following bound with high probability*

$$\|\widehat{\Sigma}^{-1} - (\Sigma^*)^{-1}\|_\infty \lesssim \sqrt{\frac{\log p'}{n}}.$$

New Estimators

- Consec Model - Estimator for Consecutive Logit Model.
- ProbitDirect - New estimator for Probit Model.

Baselines

- Discrete Model - which ignores the ordering of categories.
- ProbitEM, ProbitEMApprox - EM based approaches for estimating Probit Model.
- Oracle which has access to latent variables in Probit Model.

Data generated from Probit Model

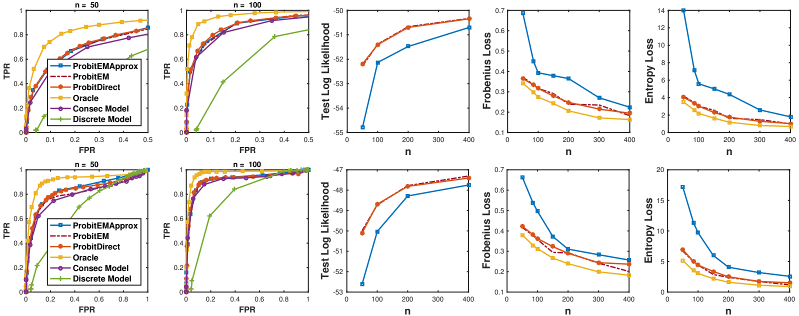


Figure: Comparison of various estimators when the data is generated from a probit model with chain graph structure, $p = 50$. Top and bottom rows correspond to chain and grid graphs respectively.

Frobenius Loss $= \frac{\|\Sigma^{*-1} - \widehat{\Sigma}^{-1}\|_F}{\|\Sigma^{*-1}\|_F}$. Entropy Loss $= \langle\langle \Sigma^*, \widehat{\Sigma}^{-1} \rangle\rangle - \log\det(\Sigma^* \widehat{\Sigma}^{-1}) - p$.

- ProtbitDirect is 1-2 orders of magnitude faster than ProbitEM.

Data generated from Consecutive Logit Model

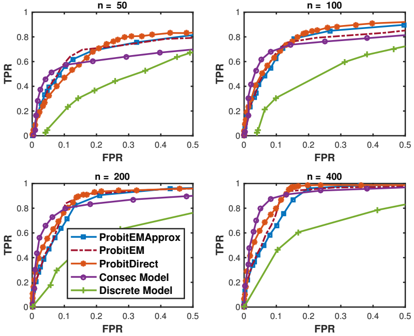


Figure: ROC plots for graph structure recovery. Data generated from a Consecutive Logit model with 2D grid structure (10 $\times$ 5 grid).

Health Information National Trends Survey study

- A survey conducted by National Cancer Institute (NCI) on Americal public.
- Collected information about tobacco product use and risk perceptions.

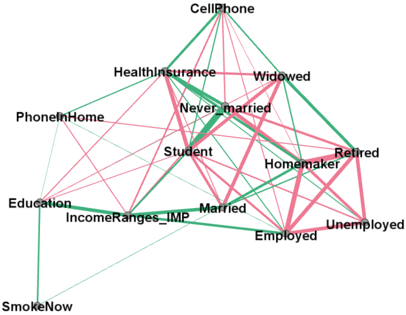


Figure: Associations of smoking behavior with socio-demographic indicators

Conclusion and Future Work

Conclusion

- Investigated two categories of multivariate ordinal distributions
- Proposed estimators for these that are both computationally tractable and come with strong statistical guarantees.

Questions?

References I

[TPSR15] W. Tansey, O. H. M. Padilla, A. S. Suggala, and P. Ravikumar, *Vector-space markov random fields via exponential families.*, ICML, 2015, pp. 684–692.

[YRAL15] E. Yang, P. Ravikumar, G. I. Allen, and Z. Liu, *Graphical models via univariate exponential family distributions*, Journal of Machine Learning Research **16** (2015), 3813–3847.