

REGRESSION MODELS FOR ORDINAL DATA: A CONCISE SUMMARY

ABHAY SHANKAR K: CS21BTECH11001 AND KARTHEEK SRIRAM TAMMANA:
CS21BTECH11028

1. PAPER SUMMARY

(I) Introduction

Wall of text, will parse and write later.

(II) The Proportional Odds model

- The model

The probabilities of the k ordered categories of the response variable Y are given by $\{\pi_1, \dots, \pi_k\}$, as a function of the covariant vector \mathbf{x} .

Given the following quantities,

- j : The paper maps each ordinal class to a contiguous interval of the real line, and j represents the class of the response variable Y .
- κ_j : The cumulative odds of the response variable Y being less than or equal to j .
- β : The vector of regression coefficients.
- x : The vector of covariates.

the cumulative odds are given by

$$\kappa_j \propto \exp(-\beta^T \mathbf{x})$$

with proportionality constant denoted κ_j .

Naturally, for a given j and the corresponding $\kappa = \kappa_j$ it follows that

$$\frac{\kappa(\mathbf{x}_1)}{\kappa(\mathbf{x}_2)} = \exp(\beta^T (\mathbf{x}_2 - \mathbf{x}_1))$$

Taking the logarithm and expressing in terms of cumulative probabilities, we have

$$\lambda_j = \log\left(\frac{\gamma_j}{1 - \gamma_j}\right) = \theta_j - \beta^T \mathbf{x}$$

where $\gamma_j = \sum_{i=1}^j \pi_i$ and $\theta_j = \log \kappa_j$. As a special case, in a two group problem, **(II)** reduces to

$$\lambda_{ij} = \theta_j + (-1)^i \frac{1}{2} \Delta \quad \forall j \in [k], i \in \{1, 2\}$$

where $\Delta = \beta^T (\mathbf{x}_2 - \mathbf{x}_1)$. {Copilot says this, not sure. MF has not declared Delta.}

- Generalised Empirical Logit transform

Consider the data matrix with cumulative row sums R_{ij} and $n_i = R_{ik}$. Furthermore, $\Sigma_i R_{ij} = R_{.j}$ and $\Sigma_i n_{ij} = n_{.j}$. We have the j th sample logit $\tilde{\lambda}_{ij} = \log \frac{R_{ij}+0.5}{n_i-R_{ij}+0.5}$. The extra 0.5 is due to reduce bias and avoid zeros.

The paper references the results of another paper and asserts that the expectation of the sample logit is $\lambda_{ij} + \mathcal{O}(n_i^{-2})$. Let

This implies that for any normalised weight vector \mathbf{w} , we have

$$\mathcal{E} [Z_i = \Sigma_j w_j \tilde{\lambda}_{ij}] = (-1)^i + \Sigma w_j \theta_j + \mathcal{O}(n_i^{-2})$$

which yields $\mathcal{E} [Z_2 - Z_1] = \Delta + \mathcal{O}(n_1^{-2}, n_2^{-2})$. For a similar estimator, the paper references another paper and states the weights minimizing the variance of $Z_2 - Z_1$ when $\Delta = 0$.

$$w_j \propto \gamma_j (1 - \gamma_j) (\pi_j + \pi_{j+1})$$

where $\gamma_j = \frac{\kappa_j}{1+\kappa_j}$ from **(II)** and **(II)**. Using these weights, the asymptotic variance of $\tilde{\Delta} = Z_2 - Z_1$ is

$$\mathbf{Var} [\tilde{\Delta}] = \left\{ \frac{n_1 n_2}{n} \sum_{j=1}^{k-1} \gamma_j (1 - \gamma_j) (\pi_j + \pi_{j+1}) \right\}^{-1} + \mathcal{O}(\Delta^2)$$

The quantity Z_i with weights given by

$$w_j \propto R_{.j} (1 - R_{.j}) (n_{.j} + n_{.j+1})$$

is called the generalised empirical logistic transform for the i 'th group.

(III) The Proportional Hazards model

The hazard function is defined as the probability of failure at time t , conditional on survival up to time t . The proportional hazards model assumes that the hazard function is of the form

$$\lambda(t) = \lambda_0(t) \exp(-\beta^T \mathbf{x})$$

where \mathbf{x} is the vector of covariates and β is a vector of unknown parameters.

Thus, the survivor function takes on the form

$$-\log \{S(t; \mathbf{x})\} = \Lambda_0(t) \exp(-\beta^T \mathbf{x})$$

where $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$

Analogous to the proportional odds model, we have

$$\frac{\log S(t; \mathbf{x}_1)}{\log S(t; \mathbf{x}_2)} = \exp(\boldsymbol{\beta}^T (\mathbf{x}_2 - \mathbf{x}_1))$$

.

For discrete data, **(III)** becomes

$$-\log \{1 - \gamma_j(\mathbf{x})\} = \exp(\theta_j - \boldsymbol{\beta}^T \mathbf{x})$$

Taking logarithm, we obtain the complementary log-log transform.

$$\log [-\log \{1 - \gamma_j(\mathbf{x})\}] = \theta_j - \boldsymbol{\beta}^T \mathbf{x}$$

(IV) Properties of related linear models

Both the above models have the same general form, namely

$$\text{link} \{\gamma_j(\mathbf{x})\} = \theta_j - \boldsymbol{\beta}^T \mathbf{x}$$

The paper then suggests a few alternative link functions, such as the inverse Gaussian or the inverse Cauchy transform.

The paper goes on to discuss invariances of the models, and proposes invariance under reversal of the ordering of categories as an appropriate property.

(V) Elaboration

The rest of the paper contains a more thorough analysis of the two above models.

2. PARAMETER ESTIMATION

What even?

3. CODE

Refer to `wine.ipynb`.