Q2. a]

Summary

Purpose

The purpose of this paper is to investigate & develop model that can be used to analyze the data measured on an ordinal scale.

Scope: The models discussed in the paper has a single response measured on ordinal scale & there can be multiple explanatory factors on covariates.

Note: All the models in the paper share the property that the categories can be thought of as a contiguous ~~varia~~ intervals on some continuous scale.

Model 1: The Proportional Odd Model.

- Suppose, there are K-ordered categories of the response that have probabilites $\pi_1(x), \pi_2(x), \pi_3(x) \ldots \pi_K(x)$, where x is a covariate.

- Then according to the proportional odd model ~~where j is~~ odds (kj) ~~probability~~ of a response $y \leq j$, where j is the category ~~the category~~ is given by

$$k_j(x) = k_j \, e^{-(\beta^T x)} \qquad (1 \leq j \leq K)$$

- Suppose we have two covariates x_1, x_2, the ratio of corresponding odds =

$$\frac{k_j(x_1)}{k_j(x_2)} = e^{-\beta^T(x_1 - x_2)}$$

- From this, we can say that odds are independent in covariates of category & only depends on difference

- Let $\gamma_j(x) = \pi_1(x) + \pi_2(x) \cdots + \pi_j(x)$,

The odds of evevent $y \le j$ can be written as

$$= \frac{\gamma_j(x)}{1 - \gamma_j(x)}$$

which is identical to linear logistic model.

$$\log\left[\frac{\gamma_j(x)}{1 - \gamma_j(x)}\right] = \theta_j - \beta^T x.$$

$$= P(x \le j) = \theta_j - B^T x.$$

- Likelyhood of commulative transformation is given as,

Models

$$\left\{\left(\frac{\gamma_1}{\gamma_2}\right)^{R_1}\left(1 - \frac{\gamma_1}{\gamma_2}\right)^{R_2 - R_1} \cdot \left(\frac{\gamma_2}{\gamma_3}\right)^{R_2}\left(1 - \frac{\gamma_2}{\gamma_3}\right)^{R_3 - R} \cdots \left(\frac{\gamma_{K-1}}{\gamma_K}\right)^{R_{K-1}}\left(\cancel{K} - \frac{\gamma_{K-1}}{\gamma_K}\right)^{R_K - R_{K-1}}\right.$$

$$= \sum_{i=1}^{K-1} \left(\frac{\gamma_i}{\gamma_{i+1}}\right)^{R_i} \cdot \left(1 - \frac{\gamma_i}{\gamma_{i+1}}\right)^{R_{i+1} - R_i}$$

log-likelyhood =

$$= \sum_{i=1}^{K-1} R_i \log\left(\frac{\gamma_i}{\gamma_{i+1}}\right) + (R_{i+1} - R_i) \log\left(1 - \frac{\gamma_i}{\gamma_{i+1}}\right)$$

Model 3: Proportional Hazards Models

- Hazard function $h(t;x)$ is defined as instantaneous failure ~~property~~ probability at time t, provided it has survived upto time t.

It is given as
$$h(t,x) = h_0(t)\, e^{-\beta^T x}$$

where $h_0(t)$ is hazard function when $x=0$.

- We will use this function $s(t;x)$ to find the probability of surviving beyond time t given covariate x.

$$-\log[S(t,x)] = \Lambda_0(t)\, e^{-\beta^T x}.$$

where $\Lambda_0(t) = \int_0^t h_0(s)\, ds.$

- If we have two covariates x_1, x_2, then

$$\frac{\log(S(t,x_1))}{\log(S(t,x_2))} = e^{-\beta(x_1 - x_2)}$$

This shows that ~~the~~ hazard function also depends on difference of covariate values.

- For discrete data, the model can be written as

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

$$\Rightarrow \log(-\log(1-\gamma_j(x))) = \theta_j - \beta^T x.$$

where $1-\gamma_i(x)$ is probability of survival beyond category j.

- As we can see, the ~~prop~~ difference between complementary log-log is constant $(\beta^T(x_2 - x_1))$, the properties of Hazard models parallels the ~~log~~ proportional odds model

- Difference between likelyhood & odd ratio for Ordinal regression, multiclass classification & regression.

Likelyhood:

1) In ordinal regression, likelyhood gives the probability of a covariate being in some categories. ~~It~~

It uses cummulative probabilities for each category & uses link function like logit to find the probability of a covariate falling into some category.

2) Whereas, in multiclass classification where the classes are mutually exclusive, we find the probability of a covariate falling in one category ~~give~~ out of all classes.

3) In ~~non~~ regression, we find likelyhood with the help of some distribution like gaussian distribution, to find the probability of the dependent variable. This can be discrete or continuous values.

Odd ratios

1) In ordinal regression, the odd ratios are use to find whether the odds of observation will fall in lower category compare to odds of it falling in higher category

2) In regression & multiclass classification we do not have concept of odd ratios as the output is categorical.

Q2b] MLE of proportional odds model

we know that

$$P(K \le j) = \theta_j - \beta^T x.$$

We, have to find the value of θ_j & β^T.

suppose we have a matrix of features, (n_{ij}),

Let n_1, n_2 be the row totals,.

The cummulative sum rows are R_{ij}, i.e $n_i = R_{ij}$

$$\therefore R_1 = n_1, \quad R_2 = R_1 + n_2 \quad \ldots \quad R_K = \sum n_j = n.$$

Let B be a ~~matri~~ vectors $= (\theta_1, \theta_2 \ldots \theta_{K-1}, \beta_1 \ldots \beta_p)$.

Then,

The likelyhood can be written as.

$$L = \left[\left(\frac{\gamma_1}{\gamma_2}\right)^{R_1}\left(1-\frac{\gamma_1}{\gamma_2}\right)^{R_2-R_1}\right]\left[\left(\frac{\gamma_2}{\gamma_3}\right)^{R_2}\left(1-\frac{\gamma_2}{\gamma_3}\right)^{R_3-R_2}\right]\cdots$$

$$\left[\left(\frac{\gamma_{K-1}}{\gamma_K}\right)^{R_{K-1}}\left(1-\frac{\gamma_{K-1}}{\gamma_K}\right)^{R_K-R_{K-1}}\right]$$

$$= \prod_{i=1}^{K-1}\left(\frac{\gamma_i}{\gamma_{i+1}}\right)^{R_i}\left(1-\frac{\gamma_i}{\gamma_{i+1}}\right)^{R_{i+1}-R_i}$$

Taking log.

$$\Rightarrow \ell = \log\left(\prod_{i=1}^{K-1}\left(\frac{\gamma_i}{\gamma_{i+1}}\right)^{R_i}\left(1-\frac{\gamma_i}{\gamma_{i+1}}\right)^{R_{i+1}-R_i}\right)$$

$$= \sum_{i=1}^{K-1} R_i \log\left(\frac{\gamma_i}{\gamma_{i+1}}\right) + (R_{i+1}-R_i)\log\left(1-\frac{\gamma_i}{\gamma_{i+1}}\right)$$

we, can write $\log\left(\frac{\gamma_i}{\gamma_{i+1}}\right) = \frac{1}{1+e^{-B^T x}}$

($\because$ considering logistic)

$$= \sum_{i=1}^{K-1} R_i \log\left(\frac{1}{1+e^{-\beta^T x}}\right) + (R_{i+1} - R_i)\log\left(1 - \frac{1}{1+e^{-\beta^T x}}\right)$$

$$= \sum_{i=1}^{K-1} R_i \left[\log\left(\frac{1}{1+e^{-\beta^T x}}\right) - \log\left(\frac{e^{-\beta^T x}}{1+e^{-\beta^T x}}\right)\right] + R_{i+1} \log\left(\frac{e^{-\beta^T x}}{1+e^{-\beta^T x}}\right)$$

$$= \sum_{i=1}^{K-1} R_i \log(e^{\beta^T x}) + R_{i+1} \log\left(\frac{e^{-\beta^T x}}{1+e^{-\beta^T x}}\right)$$

$$l(\beta) = \sum_{i=1}^{K-1} R_i \beta^T x + R_{i+1} \log\left(\frac{1}{1+e^{\beta^T x}}\right)$$

As this is trancendental equation, we will use Newton-Raphson method to calculate β value.

$$\therefore \beta^{n+1} = \beta^n - \frac{\frac{\partial}{\partial \beta} l(\beta)}{\frac{\partial^2 l(\beta)}{\partial \beta^2}}$$

$$\frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^{K-1} R_i x_i - (R_{i+1}) \frac{e^{\beta^T x}}{1+e^{\beta^T x}} \cdot x_i$$

$$= \sum_{i=1}^{K-1} R_i x_i - \frac{R_{i+1}}{1+e^{-\beta^T x}} x_i$$

$$\frac{\partial^2 l(\beta)}{\partial^2 \beta} = -\sum_{i=1}^{K-1} \frac{\partial}{\partial \beta}\left(\frac{R_{i+1}}{1+e^{-\beta^T x}} x_i\right)$$

$$= -\sum_{i=1}^{K-1} R_{i+1} \left[\frac{e^{-\beta^T x_i}}{1+e^{-\beta^T x_i}}\right]\left[\frac{1}{1+e^{-\beta^T x_i}}\right] x_i^T x_i$$

we can also write this as

$$= -\sum_{i=1}^{K-1} R_{i+1} P(x_i)(1-P(x_i)) x_i^T x_i$$

$$\therefore \beta^{n+1} = \beta^n + \frac{\sum_{i=1}^{K-1} \left[R_i - \frac{R_{i+1}}{1+e^{-\beta^T x}}\right] x_i}{\sum_{i=1}^{K-1} R_{i+1} x_i^T \left[\frac{e^{-\beta^T x_i}}{1+e^{-\beta^T x_i}}\right]\left[\frac{1}{1+e^{-\beta^T x_i}}\right] x_i}$$