Lecture 7. Polytomous Data

Nan Ye

School of Mathematics and Physics
University of Queensland

Polytomous Response

Polytomous response: a response taking one of $K > 2$ fixed values (response categories).

Main types

- Ordinal scales: the categories are ordered.

    e.g. first, second, ...

- Interval scales: the categories are ordered with scores attached to categories.

    e.g. height groups

- Nominal scales: no structure at all.

    e.g. red, green, blue

This Lecture

- Modelling ordinal scales
- Modelling nominal scales

Models for Ordinal Scales

Reduction to binary problems

- Assume the categories are $1, 2, \ldots, K$.

- Model each cumulative distribution $p_j(\mathbf{x}) = P(Y \leq j \mid \mathbf{x})$ by a logistic regression model

$$p_j(\mathbf{x}) = logistic(\mathbf{x}^\top \beta_j).$$

- Equivalently,

$$logit(p_j(\mathbf{x})) = \mathbf{x}^\top \beta_j.$$

- This may not guarantee that $p_j(\mathbf{x}) \geq p_i(\mathbf{x})$ for $j \geq i$.

Proportional odds model

- If we further assume that $\boldsymbol{\beta}_1 = \ldots = \boldsymbol{\beta}_{K-1} = \boldsymbol{\beta}$, we get the proportional odds model

$$logit(p_j(\mathbf{x})) = \theta_j + \mathbf{x}^\top \boldsymbol{\beta}.$$

- The model need to satisfy

$$\theta_1 \leq \theta_2 \leq \ldots \leq \theta_{K-1}.$$

- If we move from $\mathbf{x}$ to $\mathbf{x}'$, we have

$$\frac{p_j(\mathbf{x}')/(1 - p_j(\mathbf{x}'))}{p_j(\mathbf{x})/(1 - p_j(\mathbf{x}))} = \exp(\boldsymbol{\beta}^\top(\mathbf{x}' - \mathbf{x})).$$

- That is, the odds changes by a factor of $\exp(\boldsymbol{\beta}^\top(\mathbf{x}' - \mathbf{x}))$ independent of the class.

- A unit increase in x_i changes the odds by a factor of $\exp(\beta_i)$.

Proportional hazards model

- We can use other link functions in proportional odds model.
- Proportional odds model uses cloglog instead of logit link

$$cloglog(p_j(\mathbf{x})) = \theta_j + \mathbf{x}^\top \boldsymbol{\beta}.$$

Example

Data

Number of pneumonia cases and exposure time to a certain bacteria.

exposure.time	normal	mild	severe
5.8	98	0	0
15.0	51	2	1
21.5	34	6	3
27.5	35	5	8
33.5	32	10	9
39.5	23	7	8
46.0	12	6	10
51.5	4	2	5

Fit a proportional odds model

```
> library(VGAM)
> # the pneumo dataset is part of the VGAM library
> fit.pom = vglm(cbind(normal, mild, severe) ~ log(exposure.time),
    data=pneumo, cumulative(parallel=T, link='logit'))
```

$1 =$ normal, $2 =$ mild, $3 =$ severe

Inspect the proportional odds model

```
> summary(fit.pom)
Pearson residuals:
                  Min      1Q Median     3Q    Max
logit(P[Y<=1]) -1.248 -0.07164 0.1441 0.3086 0.7714
logit(P[Y<=2]) -1.044 -0.18415 0.3093 0.3353 0.5048

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept):1      9.6761     1.3241   7.308 2.72e-13 ***
(Intercept):2     10.5817     1.3454   7.865 3.69e-15 ***
log(exposure.time) -2.5968     0.3811  -6.814 9.50e-12 ***
```

The fitted models are

$$logitP(Y \leq 1 \mid \mathbf{x}) = 9.6761 - 2.5968 \log(exposure.time)$$
$$logitP(Y \leq 2 \mid \mathbf{x}) = 10.5817 - 2.5968 \log(exposure.time)$$

```
Warning: Hauck-Donner effect detected in the following estimate(s):
'(Intercept):1'

Exponentiated coefficients:
log(exposure.time)
        0.07451115
```

- Hauck-Donner effect: Wald's test of significance is misleading.
  This often happens when the data is separable (for this data set,
  the log exposure time can be used to perfectly predict whether
  $Y \leq 1$).
- Increasing log exposure time by one unit changes all the odds by a
  factor of 0.07451115.

Fit a proportional hazards model

```
> fit.phm = vglm(cbind(normal, mild, severe) ~ log(exposure.time),
    data=pneumo, cumulative(parallel=T, link='cloglog'))
```

Inspect the proportional hazards model

```
> summary(fit.phm)
Pearson residuals:
                     Min      1Q   Median      3Q     Max
cloglog(P[Y<=1]) -0.6916 -0.4561 -0.04129  0.4381  0.5379
cloglog(P[Y<=2]) -1.0037 -0.3217 -0.03335  0.2006  0.7345

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept):1       4.3457     0.6299   6.900 5.22e-12 ***
(Intercept):2       4.8283     0.6417   7.524 5.31e-14 ***
log(exposure.time) -1.2407     0.1897  -6.540 6.15e-11 ***
```
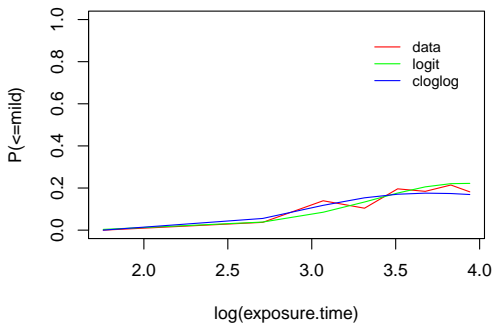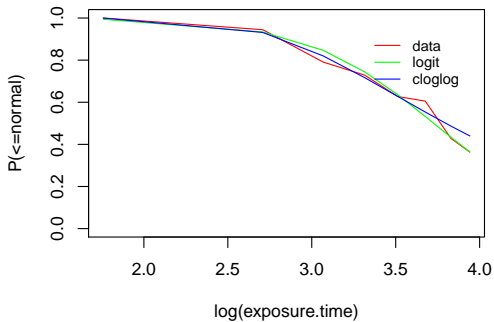
The fitted models are

$$cloglog P(Y \leq 1 \mid \mathbf{x}) = 4.3457 - 1.2407 \log(exposure.time)$$
$$cloglog P(Y \leq 2 \mid \mathbf{x}) = 4.8283 - 1.2407 \log(exposure.time)$$

Fitted probabilities and observed probabilities

Models for Nominal scales

Multi-class logistic regression

- Recall: in binary logistic regression,

$$\ln \frac{p(Y = 1 \mid \mathbf{x}, \boldsymbol{\beta})}{p(Y = 0 \mid \mathbf{x}, \boldsymbol{\beta})} = \boldsymbol{\beta}^\top \mathbf{x},$$

  That is, the log odds is linear.

- When the classes are $1, \ldots, K$, we want

$$\ln \frac{p(Y = i \mid \mathbf{x}, \boldsymbol{\beta})}{p(Y = 1 \mid \mathbf{x}, \boldsymbol{\beta})} = \boldsymbol{\beta}_i^\top \mathbf{x}.$$

- This implies

$$p(Y = i \mid \mathbf{x}, \boldsymbol{\beta}_{1:K}) = \frac{\exp(\boldsymbol{\beta}_i^\top \mathbf{x})}{\sum_j \exp(\boldsymbol{\beta}_j^\top \mathbf{x})},$$

  where $\boldsymbol{\beta}_1 = \mathbf{0}$, and $\boldsymbol{\beta}_{1:K}$ denotes $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K$.
- Also known as multinomial logistic regression.

Decision boundary

- Given $\mathbf{x}$, we predict its class as $\arg\max_y p(y \mid \mathbf{x}, \boldsymbol{\beta})$.
- The set of $\mathbf{x}$ in class y is the convex polytope described by the contraints

$$\boldsymbol{\beta}_y^\top \mathbf{x} \geq \boldsymbol{\beta}_1^\top \mathbf{x},$$
$$\cdots$$
$$\boldsymbol{\beta}_y^\top \mathbf{x} \geq \boldsymbol{\beta}_K^\top \mathbf{x}.$$

- The boundary between different classes must be linear.

Linearly separable data

- When the data is linearly separable, MLE diverges (it fails for simple data)!
- There are various ways to fix this problem (for example, through regularization, or using objective functions which search for hyperplanes which are optimal in some sense, like support vector machines).

Linearly separable data with K classes: there are K vectors $\beta_1, \ldots, \beta_K$ such that $\mathbf{x}$ is in class y iff $\beta_y^\top \mathbf{x} \geq \beta_i^\top \mathbf{x}$ for all i.

Multi-class logistic regression in R

```
> fit.mlr <- vglm(Species ~ ., multinomial, iris)
> # compute fitted probabilities
> predict(fit.mlr, type='response')
> # compute probabilities on new data
> predict(fit.mlr, newdata=iris, type='response')
```

What You Need to Know

- Modelling ordinal scales
    proportional odds model, proportional hazards model
- Modelling nominal scales
    multiclass logistic regression
- Working with polytomous response data in R