Slide 1

Reference: Agresti, Sections 6.2-6.3

Cumulative link models

- Suppose the response categories, $1, 2, \ldots, c$ are ordered.

- Let $\pi_{ij} = p(y_i = j)$ then the corresponding cumulative probability is

$$P(y_i \leq j) = \pi_{i1} + \pi_{i2} + \ldots + \pi_{ij}, \qquad j = 1, \ldots, c.$$

- The class of models that we consider are of the form

$$g(P(y_i \leq j)) = \alpha_j + \boldsymbol{x_i'}\boldsymbol{\beta}.$$

- α_j represents the baseline value of the transformed cumulative probability for category j.

- β_k represents the increase in *all* transformed cumulative probabilities for a one-unit increase in x_{ik}.

Slide 2

Proportional odds cumulative logit model

- A direct extension of the usual logistic regression model applies the logit transformation to the cumulative probabilities,

$$\text{logit}(P(y_i \leq j)) = \log \frac{P(y_i \leq j)}{1 - P(y_i \leq j)} = \alpha_j + \boldsymbol{x_i'}\boldsymbol{\beta}.$$

- The cumulative logits are defined

$$\log\left(\frac{\pi_{i1}}{\pi_{i2} + \pi_{i3} + \ldots + \pi_{ic}}\right)$$
$$\log\left(\frac{\pi_{i1} + \pi_{i2}}{\pi_{i3} + \pi_{i4} + \ldots + \pi_{ic}}\right)$$
$$\vdots$$
$$\log\left(\frac{\pi_{i1} + \pi_{i2} + \ldots + \pi_{i,r-1}}{\pi_{ic}}\right)$$

- Each cumulative logit uses all c response categories.

Slide 3

- The log-**cumulative odds ratio** is proportional to the distance between $\boldsymbol{x}_i = \boldsymbol{u}$ and $\boldsymbol{x}_i = \boldsymbol{v}$

$$\log \frac{p(y_i \leq j | \boldsymbol{x}_i = \boldsymbol{u})/p(y_i > j | \boldsymbol{x}_i = \boldsymbol{u})}{p(y_i \leq j | \boldsymbol{x}_i = \boldsymbol{v})/p(y_i > j | \boldsymbol{x}_i = \boldsymbol{v})} = (\boldsymbol{u} - \boldsymbol{v})' \boldsymbol{\beta}.$$

- For each j, adjusting for the other covariates in the model, for a 1-unit increase in x_{ik}, the odds that $y_i \leq j$ is multiplied by $\exp(\beta_k)$.

- The model constrains the $c - 1$ response curves to have the same shape. The same proportionality constant applies to each logit. Hence the name **proportional odds model**.

Slide 4

- The odds of $y_i \leq j$ are

$$\frac{P(y_i \leq j)}{1 - P(y_i \leq j)} = \exp(\alpha_j + \boldsymbol{x}_i' \boldsymbol{\beta}).$$

- $\exp(\alpha_j)$ may be interpreted as the **baseline** odds of a response in category j or below when $\boldsymbol{x} = \boldsymbol{0}$.

- Note that α_j increases as j increases, since for $j < j'$, $P(y_i \leq j) < P(y_i \leq j')$.

- β_k is the change in log-odds of falling into or below any category for a one unit increase in x_k, holding all other covariates fixed.

Slide 5

- The likelihood function viewed as a function of $(\{\alpha_j\}, \boldsymbol{\beta})$ is

$$\prod_{i=1}^{N}\prod_{j=1}^{c} \pi_{ij}^{y_{ij}} \;=\; \prod_{i=1}^{N}\prod_{j=1}^{c} \left(P(y_i \leq j|\boldsymbol{x}_i) - P(y_i \leq j-1|\boldsymbol{x}_i)\right)^{y_{ij}}$$

$$= \prod_{i=1}^{N}\prod_{j=1}^{c} \left(\frac{\exp(\alpha_j + \boldsymbol{x_i'\beta})}{1 + \exp(\alpha_j + \boldsymbol{x_i'\beta})} - \frac{\exp(\alpha_{j-1} + \boldsymbol{x_i'\beta})}{1 + \exp(\alpha_{j-1} + \boldsymbol{x_i'\beta})} \right)^{y_{ij}}.$$

- The log-likelihood is

$$\log \mathcal{L} = \sum_{i=1}^{N}\sum_{j=1}^{c} y_{ij} \log \left[\frac{\exp(\alpha_j + \boldsymbol{x_i'\beta})}{1 + \exp(\alpha_j + \boldsymbol{x_i'\beta})} - \frac{\exp(\alpha_{j-1} + \boldsymbol{x_i'\beta})}{1 + \exp(\alpha_{j-1} + \boldsymbol{x_i'\beta})} \right]$$

Slide 6

Latent variable motivation

- The model can be interpreted in terms of a *latent variable*.

- Suppose the observed response y_i results from grouping an underlying continuous variable y_i^* using cut-points $\alpha_1 < \alpha_2 < \ldots < \alpha_{c-1}$, so that

$$y_i = j \;\; \text{if } \alpha_{j-1} < y_i^* \leq \alpha_j$$

$$\text{then,} \quad P(y_i \leq j) = P(y_i^* \leq \alpha_j).$$

Slide 7

- Suppose that the underlying continuous variable follows a linear model of the form

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i,$$

where the error term ε_i has c.d.f. $F(\varepsilon_i)$, then

$$P(y_i^* \leq \alpha_j) = P(\varepsilon_i \leq \alpha_j - \boldsymbol{x}_i'\boldsymbol{\beta}) = F(\alpha_j - \boldsymbol{x}_i'\boldsymbol{\beta})$$

and

$$F^{-1}(P(y_i \leq j)) = \alpha_j - \boldsymbol{x}_i'\boldsymbol{\beta}^*$$

Slide 8

- Note that the latent variable formulation results in linear predictor $\alpha_j - \boldsymbol{x_i'}\boldsymbol{\beta}$ rather than $\alpha_j + \boldsymbol{x_i'}\boldsymbol{\beta}$.

- When using software check which parametrization is used, $\alpha_j + \boldsymbol{x_i'}\boldsymbol{\beta}$ or $\alpha_j + \boldsymbol{x_i'}\boldsymbol{\beta}$, and interpret β_k accordingly.

- Note that we have the same parameters $\boldsymbol{\beta}$ for the effects on Y regardless of how the cutpoints $\{\alpha_j\}$ are chosen.

- Two investigators who use different response categories in studying an association should reach similar conclusions.

Slide 9

- If $\varepsilon \sim$ logistic, then F^{-1} is the logit link.

- Normality for ε implies a probit link for the cumulative probabilites.

$$\boldsymbol{\Phi}^{-1}(P(y_i \leq j) = \alpha_j + \boldsymbol{x}_i'\boldsymbol{\beta}$$

- In general, the fit of a logit and probit model are quite similar.

- Another choice is the complementary log-log link, which leads to the model

$$\log(-\log(1 - P(y_i \leq j))) = \alpha_j + \boldsymbol{x}_i'\boldsymbol{\beta}.$$

Slide 10

Example: Cheese-tasting experiment

Data from McCullagh and Nelder (1989), Table 5.1. Subjects were randomly assigned to taste one of four different cheeses. Response categories are 1=strong dislike to 9=excellent taste.

Cheese	Response category								
	1	2	3	4	5	6	7	8	9
A	0	0	1	7	8	8	19	8	1
B	6	9	12	11	7	6	1	0	0
C	1	1	6	8	23	7	5	1	0
D	0	0	0	1	3	7	14	16	11

Slide 11

Note that the function `polr(.)` in the R package `MASS` uses the following parametrization for the cumulative logit model

$$\log \frac{P(y_i \leq j)}{1 - P(y_i \leq j)} = \alpha_j - \boldsymbol{x_i'\beta}$$

That is,

$$\log \frac{P(y_i \leq 1)}{P(y_i > 1)} = \alpha_1 - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i}$$

$$\log \frac{P(y_i \leq 2)}{P(y_i > 2)} = \alpha_2 - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i}$$

$$\vdots$$

$$\log \frac{P(y_i \leq 8)}{P(y_i > 8)} = \alpha_8 - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i}$$

Slide 12

Using cheese A as the reference, the three dummy variables are

$$x_{1i} = \begin{cases} 1 & \text{if subject } i \text{ tasted cheese B} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{2i} = \begin{cases} 1 & \text{if subject } i \text{ tasted cheese C} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{3i} = \begin{cases} 1 & \text{if subject } i \text{ tasted cheese D} \\ 0 & \text{otherwise} \end{cases}$$

A positive β_k indicates a greater liking for cheese x_k compared to the reference.

Slide 13

```
library(MASS)
polr.fit = polr(as.factor(response) ~ type, weights=count, data=cheese)


> summary(polr.fit)


Coefficients:
        Value Std. Error t value
typeB -3.352     0.4287  -7.819
typeC -1.710     0.3715  -4.603
typeD  1.613     0.3805   4.238


Intercepts:
    Value    Std. Error t value
1|2  -5.4674   0.5236   -10.4413
2|3  -4.4122   0.4278   -10.3148
3|4  -3.3126   0.3700    -8.9522
```

Slide 14

```
4|5  -2.2440    0.3267    -6.8680
5|6  -0.9078    0.2833    -3.2037
6|7   0.0443    0.2646     0.1673
7|8   1.5459    0.3017     5.1244
8|9   3.1058    0.4057     7.6547


Residual Deviance: 711.3479
AIC: 733.3479
```

Slide 15

- $\hat{\alpha}_1 = -5.47$ is the estimated log-odds of falling into category 1 (strong dislike) versus all other categories for cheese A.

- $\hat{\alpha}_8 = 3.11$ is the estimated log-odds of responses in $1, \ldots, 8$ versus 9 for cheese A.

- $\hat{\beta}_1 = -3.352$ and has p-value $= 9.41 \times 10^{-15}$ (2*pt(-7.819, 1653) where $df = 208 \times 8 - 11 = 1653$. This indicates that cheese B does not taste as good as cheese A.

- The results imply that D is the preferred cheese, followed by A, then C, and B comes last.

Slide 16

We can also use the R package VGAM to fit the proportional odds model, but we need to first put the data in an ungrouped format.

```
cheese.ungrp = as.data.frame(lapply(cheese, function(x,p) rep(x,p), cheese$count))
vglm.fit = vglm(response ~ type, family=cumulative(parallel=TRUE), data=cheese.ungrp)

> summary(vglm.fit)
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept):1 -5.46737    0.52022 -10.510  < 2e-16 ***
(Intercept):2 -4.41218    0.42472 -10.389  < 2e-16 ***
(Intercept):3 -3.31262    0.36967  -8.961  < 2e-16 ***
(Intercept):4 -2.24401    0.32617  -6.880 5.99e-12 ***
(Intercept):5 -0.90776    0.27477  -3.304 0.000954 ***
(Intercept):6  0.04426    0.25984   0.170 0.864758
(Intercept):7  1.54592    0.30419   5.082 3.73e-07 ***
(Intercept):8  3.10577    0.40443   7.679 1.60e-14 ***
```

```
typeB              3.35184     0.42351    7.914 2.48e-15 ***
typeC              1.70989     0.37305    4.583 4.57e-06 ***
typeD             -1.61279     0.37778   -4.269 1.96e-05 ***
---


Residual deviance: 711.3479 on 1653 degrees of freedom
Log-likelihood: -355.674 on 1653 degrees of freedom
```

Slide 17

- Note that the function `vglm(.)` with `family = cumulative`
  uses the parametrization

$$\log \frac{P(y_i \le j)}{1 - P(y_i \le j)} = \alpha_j + \boldsymbol{x_i'\beta}$$

  so, a positive β_k indicates a greater dislike for cheese x_k
  compared to the reference.

- The estimates of α_j are the same as those using `polr(.)`, but
  the signs for the estimates of β_k are different.

We can test the proportional odds assumption

H_0 : same slope for all cumulative logits   vs. H_1 : different slopes

with a likelihood ratio test comparing the model with varying
slopes for the cumulative logits and the model above

```
noprop.fit = vglm(response ~ type, family=cumulative, data=cheese.ungrp)
---
```

Slide 18

```
Residual deviance: 691.0397 on 1632 degrees of freedom
Log-likelihood: -345.5199 on 1632 degrees of freedom
---


> 1-pchisq(-2*(logLik(vglm.fit)-logLik(noprop.fit)),
                df=df.residual(vglm.fit)-df.residual((noprop.fit)))
[1] 0.5018223
```

We fail to reject the null and can assume a proportional odds
structure.