Chapter 6 Multicategory Logit Models

Response Y has $J > 2$ categories.

Extensions of logistic regression for nominal and ordinal Y assume a multinomial distribution for Y.

Review of Multinomial Distribution

If n trials are performed:

- in each trial there are $J > 2$ possible outcomes (categories)
- $\pi_j = \mathrm{P}(\text{category } j)$, for each trial, $\sum_{j=1}^{J} \pi_j = 1$
- trials are **independent**
- Y_j = number of trials fall in category j out of n trials

then the joint distribution of $(Y_1, Y_2, \dots, Y_J)$ is said to have a **multinomial distribution**, with n trials and category probabilities $(\pi_1, \pi_2, \dots, \pi_J)$, denoted as

$$(Y_1, Y_2, \dots, Y_J) \sim \text{Multinom}(n; \pi_1, \pi_2, \dots, \pi_J),$$

with probability function

$$\mathrm{P}(Y_1 = y_1, Y_2 = y_2, \dots, Y_J = y_J) = \frac{n!}{y_1!\, y_2! \cdots y_J!} \pi_1^{y_1} \pi_2^{y_2} \cdots \pi_J^{y_J}$$

where $0 \le y_j \le n$ for all j and $\sum_j y_j = n$.

Odds for Multi-Category Response Variable

For a binary response variable, there is only one kind of odds that we may consider

$$\frac{\pi}{1-\pi}.$$

For a multi-category response variable with $J > 2$ categories and category probabilities $(\pi_1, \pi_2, \dots, \pi_J)$, we may consider various kinds of odds, though some of them are more interpretable than others.

- odds between two categories: π_i/π_j.
- odds between a group of categories vs another group of categories, e.g.,

  $$\frac{\pi_1 + \pi_3}{\pi_2 + \pi_4 + \pi_5}.$$

  Note the two groups of categories should be non-overlapping.

Odds for Multi-Category Response Variable (Cont'd)

E.g., if Y = source of meat (in a broad sense) with 5 categories

beef, pork, chicken, turkey, fish

We may consider the odds of

- beef vs. chicken: $\pi_{\text{beef}}/\pi_{\text{chicken}}$
- red meat vs. white meat:

  $$\frac{\pi_{\text{beef}} + \pi_{\text{pork}}}{\pi_{\text{chicken}} + \pi_{\text{turkey}} + \pi_{\text{fish}}}$$

- red meat vs. poultry:

  $$\frac{\pi_{\text{beef}} + \pi_{\text{pork}}}{\pi_{\text{chicken}} + \pi_{\text{turkey}}}$$

Odds for Ordinal Variables

If Y is ordinal with ordered categories:

$$1 < 2 < \ldots < J$$

we may consider the odds of $Y \le j$

$$\frac{P(Y \le j)}{P(Y > j)} = \frac{\pi_1 + \pi_2 + \cdots + \pi_j}{\pi_{j+1} + \cdots + \pi_J}$$

e.g., $Y =$ political ideology, with 5 levels

very liberal < slightly liberal < moderate
< slightly conservative < very conservative

we may consider the odds

$$\frac{P(\text{very or slightly liberal''})}{P(\text{moderate or conservative})} = \frac{\pi_{\text{vlib}} + \pi_{\text{slib}}}{\pi_{\text{mod}} + \pi_{\text{scon}} + \pi_{\text{vcon}}}$$

Odds Ratios for XY When Y is Multi-Category

For any sensible odds between two (groups of) categories of Y can be compared across two levels of X.
E.g., if $Y =$ source of meat, we may consider

OR between Y (beef vs. chicken) and $X = 1$ or 2

$$= \frac{P(Y = \text{beef}|X = 1)/P(Y = \text{chicken}|X = 1)}{P(Y = \text{beef}|X = 2)/P(Y = \text{chicken}|X = 2)}$$

OR between Y (red meat vs. poultry) and $X = 1$ or 2

$$= \frac{\text{odds of red meat vs. poultry when } X = 1}{\text{odds of red meat vs. poultry when } X = 2}$$

$$= \frac{P(Y = \text{beef or pork}|X = 1)/P(Y = \text{chicken or turkey}|X = 1)}{P(Y = \text{beef or pork}|X = 2)/P(Y = \text{chicken or turkey}|X = 2)}$$

- Again, ORs can be estimated from both prospective and retrospective studies.
- Usually we need more than 1 OR to describe XY associations completely.

6.1 Baseline-Category Logit Models for Nominal Responses

6.1 Baseline-Category Logit Models for Nominal Responses

Let $\pi_j = Pr(Y = j)$, $j = 1, 2, \ldots, J$.

Baseline-category logits are

$$\log\left(\frac{\pi_j}{\pi_J}\right), \quad j = 1, 2, \ldots, J-1.$$

Baseline-category logit model has form

$$\log\left(\frac{\pi_j}{\pi_J}\right) = \alpha_j + \beta_j x, \quad j = 1, 2, \ldots, J-1.$$

or equivalently,

$$\pi_j = \pi_J \exp(\alpha_j + \beta_j x) \quad j = 1, 2, \ldots, J-1.$$

- Separate set of parameters (α_j, β_j) for each logit.
- Equation for π_J is not needed since $\log(\pi_J/\pi_J) = 0$

Choice of Baseline-Category Is Arbitrary

Equation for other pair of categories, say, categories a and b can then be determined as

$$\begin{aligned}\log\left(\frac{\pi_a}{\pi_b}\right) = \log\left(\frac{\pi_a/\pi_J}{\pi_b/\pi_J}\right) = \log\left(\frac{\pi_a}{\pi_J}\right) - \log\left(\frac{\pi_b}{\pi_J}\right) \\ = (\alpha_a + \beta_a x) - (\alpha_b + \beta_b x) \\ = (\alpha_a - \alpha_b) + (\beta_a - \beta_b)x\end{aligned}$$

Any of the categories can be chosen to be the baseline

- The model will fit equally well, achieving the same likelihood and producing the same fitted values.
- The coefficients α_j, β_j's will change, but their differences

  $$\alpha_a - \alpha_b \quad \text{and} \quad \beta_a - \beta_b$$

  between any two categories a and b will stay the same.

- The probabilities for the categories can be determined from that $\sum_{j=1}^{J} \pi_j = 1$ to be

$$\begin{aligned}\pi_j &= \frac{e^{\alpha_j + \beta_j x}}{1 + \sum_{k=1}^{J-1} e^{\alpha_k + \beta_k x}}, \qquad \text{for } j = 1, 2, \ldots, J-1 \\ \pi_J &= \frac{1}{1 + \sum_{k=1}^{J-1} e^{\alpha_k + \beta_k x}} \qquad \text{(baseline)}\end{aligned}$$

- Interpretation of coefficients: e^{β_j} is the multiplicative effect of a 1-unit increase in x on the odds of response j instead of response J.
- Could also use this model with ordinal response variables, but this would ignore information about ordering.

Example (Job Satisfaction and Income)

Data from General Social Survey (1991)

Income	Job Satisfaction			
	Dissat	Little	Moderate	Very
0-5K	2	4	13	3
5K-15K	2	6	22	4
15K-25K	0	1	15	8
>25K	0	3	13	8

- Response: Y = Job Satisfaction
- Explanatory Variable: X = Income.

Using x = income scores (3K, 10K, 20K, 35K), we fit the model

$$\log\left(\frac{\pi_j}{\pi_J}\right) = \alpha_j + \beta_j x, \quad j = 1, 2, 3.$$

for $J = 4$ job satisfaction categories.

Parameter Estimates

ML estimates for coefficients (α_j, β_j) in logit model can be found via R function `vglm` in package `VGAM` w/ multinomial family.

You will have to install the `VGAM` library first, by the following command. You only need to install ONCE!

```
> install.packages("VGAM")   # JUST RUN THIS ONCE!
```

Once installed, you need to load `VGAM` at every R session before it can be used.

```
> library(VGAM)
```

Now we can type in the data and fit the baseline category logit model.

```
> Income = c(3,10,20,35)
> Diss = c(2,2,0,0)
> Little = c(4,6,1,3)
> Mod = c(13,22,15,13)
> Very = c(3,4,8,8)
> jobsat.fit1 = vglm(cbind(Diss,Little,Mod,Very) ~ Income,
                     family=multinomial)
```

```
> coef(jobsat.fit1)
(Intercept):1 (Intercept):2 (Intercept):3
   0.42980117    0.45627479    1.70392894
     Income:1      Income:2      Income:3
  -0.18536791   -0.05441184   -0.03738509
```

The fitted model is

$$\log\left(\frac{\widehat{\pi}_1}{\widehat{\pi}_4}\right) = \widehat{\alpha}_1 + \widehat{\beta}_1 x = 0.430 - 0.185x \qquad \text{(Dissat. v.s. Very Sat.)}$$

$$\log\left(\frac{\widehat{\pi}_2}{\widehat{\pi}_4}\right) = \widehat{\alpha}_2 + \widehat{\beta}_2 x = 0.456 - 0.054x \qquad \text{(Little v.s. Very Sat.)}$$

$$\log\left(\frac{\widehat{\pi}_3}{\widehat{\pi}_4}\right) = \widehat{\alpha}_3 + \widehat{\beta}_3 x = 1.704 - 0.037x \qquad \text{(Moderate v.s. Very Sat.)}$$

- As $\widehat{\beta}_j < 0$ for $j = 1, 2, 3$, for each logit, estimated odds of being in less satisfied category (instead of very satisfied) decrease as $x =$ income increases.
- Estimated odd of being "dissatisfied" instead of "very satisfied" multiplied by $e^{-0.185} = 0.83$ for each 1K increase in income.

$$\widehat{\pi}_1 = \frac{e^{0.430-0.185x}}{1 + e^{0.430-0.185x} + e^{0.456-0.054x} + e^{1.704-0.037x}}$$

$$\widehat{\pi}_2 = \frac{e^{0.456-0.054x}}{1 + e^{0.430-0.185x} + e^{0.456-0.054x} + e^{1.704-0.037x}}$$

$$\widehat{\pi}_3 = \frac{e^{1.704-0.037x}}{1 + e^{0.430-0.185x} + e^{0.456-0.054x} + e^{1.704-0.037x}}$$

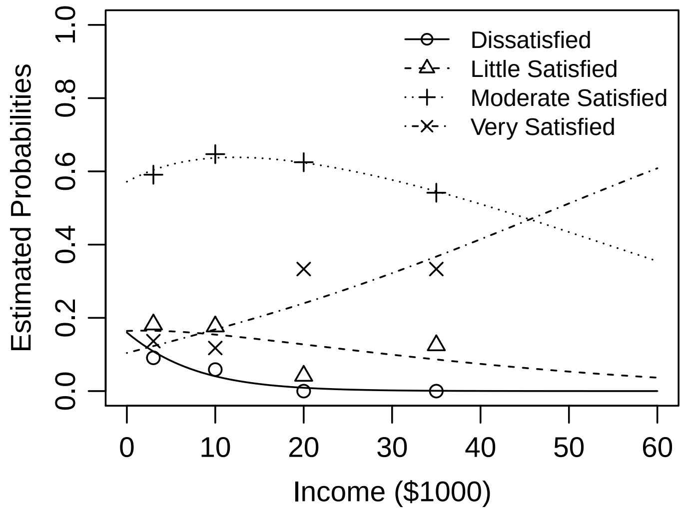
$$\widehat{\pi}_4 = \frac{1}{1 + e^{0.430-0.185x} + e^{0.456-0.054x} + e^{1.704-0.037x}}$$

E.g., at $x = 20$ (K), estimated prob. of being "very satisfied" is

$$\widehat{\pi}_4 = \frac{1}{1 + e^{0.430-0.185(20)} + e^{0.456-0.054(20)} + e^{1.704-0.037(20)}} \approx 0.240$$

Similarly, one can compute $\widehat{\pi}_1 \approx 0.009$, $\widehat{\pi}_2 \approx 0.127$, $\widehat{\pi}_3 \approx 0.623$, and observe $\widehat{\pi}_1 + \widehat{\pi}_2 + \widehat{\pi}_3 + \widehat{\pi}_4 = 1$.

Plot of sample proportions and estimated probabilities of Job Satisfaction as a function of Income



Observe that though π_j/π_J is a monotone function of x, π_j may NOT be **monotone** in x.

Deviance and Goodness of Fit

For grouped multinomial response data,

	conditions of trial (explanatory variables)				number of trials	multinomial counts			
Condition 1	x_{11}	x_{12}	$\dots$	x_{1p}	n_1	y_{11}	y_{12}	$\dots$	y_{1J}
Condition 2	x_{21}	x_{22}	$\dots$	x_{2p}	n_2	y_{21}	y_{22}	$\dots$	y_{2J}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
Condition N	x_{N1}	x_{N2}	$\dots$	x_{Np}	n_N	y_{N1}	y_{N2}	$\dots$	y_{NJ}

(Residual) Deviance for a Model M is defined as

$$\text{Deviance} = -2(L_M - L_S) = 2\sum_{ij} y_{ij} \log\left(\frac{y_{ij}}{n_i \widehat{\pi}_j(\mathbf{x}_i)}\right)$$

$$= 2\sum_{ij} (\text{observed}) \log\left(\frac{\text{observed}}{\text{fitted}}\right)$$

where $\widehat{\pi}_j(\mathbf{x}_i) =$ estimated prob. based on Model M

$L_M =$ max. log-likelihood for Model M

$L_S =$ max. log-likelihood for the saturated model

DF of Deviance

df for deviance of Model M is

$$N(J-1)-(\#\text{ of parameters in the model}).$$

If the model has p explanatory variables,

$$\log\left(\frac{\pi_j}{\pi_J}\right)=\alpha_j+\beta_{1j}x_1+\cdots+\beta_{pj}x_p,\quad j=1,2,\ldots,J-1.$$

there are $p+1$ coefficients per equation, so $(J-1)(p+1)$ coefficients in total.

df for deviance $=N(J-1)-(J-1)(p+1)=(J-1)(N-p-1)$.

```
> deviance(jobsat.fit1)
[1] 4.657999

> df.residual(jobsat.fit1)
[1] 6
```

Goodness of Fit

If the estimated expected counts $n_i\widehat{\pi}_j(\mathbf{x}_i)$ are large enough, the deviance has a large sample chi-squared distribution with df = df of deviance.

We can use deviance to conduct Goodness of Fit test

H$_0$: Model M is correct (fits the data as well as the saturated model)

H$_A$: Saturated model is correct

When H$_0$ is rejected, it means that Model M doesn't fit as well as the saturated model.

For the Job Satisfaction and Income example, the P-value for the Goodness of fit is 58.8%, no evidence of lack of fit. However, this result is not reliable because most of the cell counts are small.

```
> deviance(jobsat.fit1)
[1] 4.657999
> df.residual(jobsat.fit1)
[1] 6
> pchisq(4.657999, df=6, lower.tail=F)
[1] 0.5883635
```

Wald CIs and Wald Tests for Coefficients

- Wald CI for β_j is $\widehat{\beta}_j\pm z_{\alpha/2}\text{SE}(\widehat{\beta}_j)$.
- Wald test of H$_0$: $\beta_j=0$ uses $z=\frac{\widehat{\beta}_j}{\text{SE}(\widehat{\beta}_j)}$ or $z^2\sim\chi^2_1$

Example (Job Satisfaction)

```
> summary(jobsat.fit1)
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept):1  0.42980    0.94481   0.455 0.649176
(Intercept):2  0.45627    0.62090   0.735 0.462423
(Intercept):3  1.70393    0.48108   3.542 0.000397 ***
Income:1      -0.18537    0.10251  -1.808 0.070568 .
Income:2      -0.05441    0.03112  -1.748 0.080380 .
Income:3      -0.03739    0.02088  -1.790 0.073401 .
```

- 95% for β_1: $-0.185\pm1.96\times0.1025\approx(-0.386,0.016)$
- 95% for e^{β_1}: $(e^{-0.386},e^{0.016})\approx(0.680,1.016)$

  Interpretation: Estimated odd of being "dissatisfied" instead of "very satisfied" multiplied by 0.680 to 1.016 for each 1K increase in income w/ 95% confidence.

Likelihood Ratio Tests

Example (Job Satisfaction)

Overall test of income effect

$$\text{H}_0:\beta_1=\beta_2=\beta_3=0$$

is equivalent of the comparison of the two models

$$\begin{aligned}\text{H}_0:&\ \log(\pi_j/\pi_4)=\alpha_j,\quad j=1,2,3\\ \text{H}_1:&\ \log(\pi_j/\pi_4)=\alpha_j+\beta_jx,\quad j=1,2,3.\end{aligned}$$

$$\begin{aligned}\text{LRT}&=-2(L_0-L_1)=-2(-21.358-(-16.954))=8.808\\ &=\text{diff in deviances}=13.467-4.658=8.809\\ Df&=\text{diff. in number of parameters}=6-3=3\\ &=\text{diff. in residual df}=9-6=3\end{aligned}$$

P-value $=Pr(\chi^2_3>8.809)\approx0.03194.$

```
> lrtest(jobsat.fit1)
Likelihood ratio test

Model 1: cbind(Diss, Little, Mod, Very) ~ Income
Model 2: cbind(Diss, Little, Mod, Very) ~ 1
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   6 -16.954
2   9 -21.358  3 8.8093    0.03194 *

> jobsat.fit2 = vglm(cbind(Diss,Little,Mod,Very) ~ 1,
                     family=multinomial)
> deviance(jobsat.fit2)
[1] 13.4673
> df.residual(jobsat.fit2)
[1] 9
> deviance(jobsat.fit1)
[1] 4.657999
> df.residual(jobsat.fit1)
[1] 6
```

Note that H_0 implies job satisfaction is independent of income. We got some evidence (P-value $= 0.032$) of dependence between job satisfaction and income.

Note we get a different conclusion if we conduct Pearson's Chi-square test of independence:

$$X^2 = 11.5, \quad df = (4-1)(4-1) = 9 \quad , P\text{-value} = 0.2415$$

```
> jobsat = matrix(c(2,2,0,0,4,6,1,3,13,22,15,13,3,4,8,8), nrow=4)
> chisq.test(jobsat)

        Pearson's Chi-squared test

data:  jobsat
X-squared = 11.524, df = 9, p-value = 0.2415

Warning message:
In chisq.test(jobsat) : Chi-squared approximation may be incorrect
```

LR test of independence gives similar conclusion ($G^2 = 13.47$, $df = 9$, P-value $= 0.1426$)

Why Logit models give different conclusion from Pearson's test of independence?

6.2 Cumulative Logit Models for Ordinal Responses

6.2 Cumulative Logit Models for Ordinal Responses

Suppose the response Y is multinomial with **ordered** categories

$$\{1,\ 2,\ \ldots\ , J\}.$$

Let $\pi_i = \mathrm{P}(Y = i)$.

The **cumulative probabilities** are

$$\mathrm{P}(Y \le j) = \pi_1 + \cdots + \pi_j, \quad j = 1, 2, ..., J.$$

- Note $\mathrm{P}(Y \le 1) \le \mathrm{P}(Y \le 2) \le \ldots \le \mathrm{P}(Y \le J) = 1$
- If Y is not ordinal, it's nonsense to say "$Y \le j$."

The **cumulative logits** are

$$\begin{aligned}\text{logit}[\mathrm{P}(Y \le j)] &= \log\left(\frac{\mathrm{P}(Y \le j)}{1 - \mathrm{P}(Y \le j)}\right) = \log\left(\frac{\mathrm{P}(Y \le j)}{\mathrm{P}(Y > j)}\right)\\ &= \log\left(\frac{\pi_1 + \cdots + \pi_j}{\pi_{j+1} + \cdots + \pi_J}\right), \quad j = 1, ..., J-1.\end{aligned}$$

Cumulative Logit Models

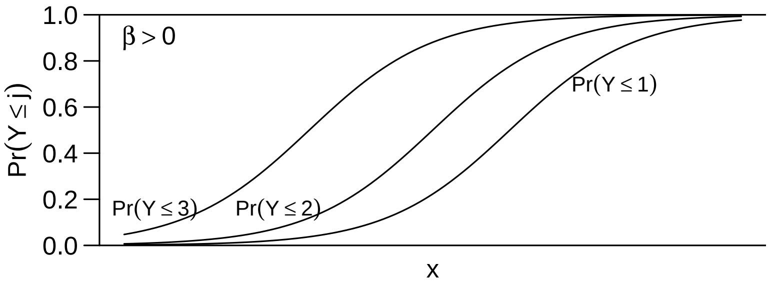
$$\text{logit}[\mathrm{P}(Y \le j|x)] = \alpha_j + \beta x, \quad j = 1, \ldots, J-1.$$

- separate intercept α_j for each cumulative logit
- same slope β for all cumulative logits

$\Rightarrow$ Curves of $\mathrm{P}(Y \le j|x)$ are "parallel", never cross each other.

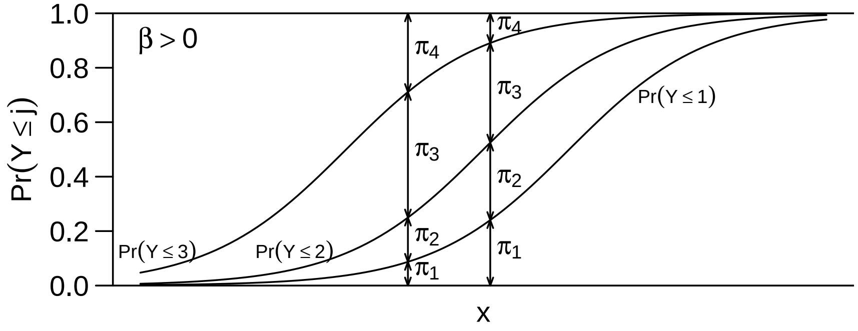
As long as $\alpha_1 \le \alpha_2 \le \ldots \le \alpha_{J-1}$, we can ensure that

$$\mathrm{P}(Y \le j|x) = \frac{e^{\alpha_j+\beta x}}{1+e^{\alpha_j+\beta x}} \le \frac{e^{\alpha_{j+1}+\beta x}}{1+e^{\alpha_{j+1}+\beta x}} = \mathrm{P}(Y \le j+1|x)$$



Cumulative Logit Models

$$\mathrm{P}(Y \le j|x) = \frac{e^{\alpha_j+\beta x}}{1+e^{\alpha_j+\beta x}}, \quad j = 1, \ldots, J-1.$$



$$\pi_j(x) = \mathrm{P}(Y = j|x) = \mathrm{P}(Y \le j|x) - \mathrm{P}(Y \le j-1|x)$$
$$= \frac{e^{\alpha_j+\beta x}}{1+e^{\alpha_j+\beta x}} - \frac{e^{\alpha_{j-1}+\beta x}}{1+e^{\alpha_{j-1}+\beta x}}$$

If $\beta > 0$, as $x \uparrow$, Y is more likely at the **lower** categories ($Y \le j$).
If $\beta < 0$, as $x \uparrow$, Y is more likely at the **higher** categories ($Y > j$).

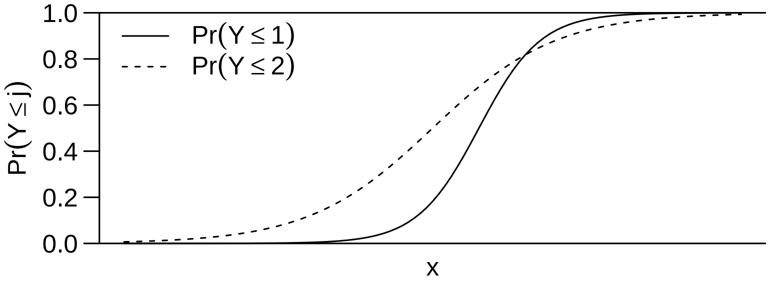
"Non-Parallel" Cumulative Logit Models

$$\text{logit}[\mathrm{P}(Y \le j|x)] = \alpha_j + \beta_j x, \quad j = 1, \ldots, J-1.$$

- separate intercept α_j for each cumulative logit
- **separate slope** β_j for each cumulative logit

However, $\mathrm{P}(Y \le j)$ curves in "non-parallel" cumulative logit models may cross each other and hence may not maintain that

$$\mathrm{P}(Y \le j|x) \le \mathrm{P}(Y \le j+1|x) \quad \text{for all } x$$



Latent Variable Interpretation for Cumulative Logit Models

- Suppose there is a unobserved continuous response Y^* under the observed ordinal response Y, such that we only observe

  $$Y = j \text{ if } \alpha_{j-1} < Y^* \le \alpha_j, \quad \text{for } j = 1, 2, \ldots, T$$

  where $-\infty = \alpha_0 < \alpha_1 < \alpha_2 < \ldots < \alpha_J = \infty$.
- Y^* has a linear relationship w/ explanatory x

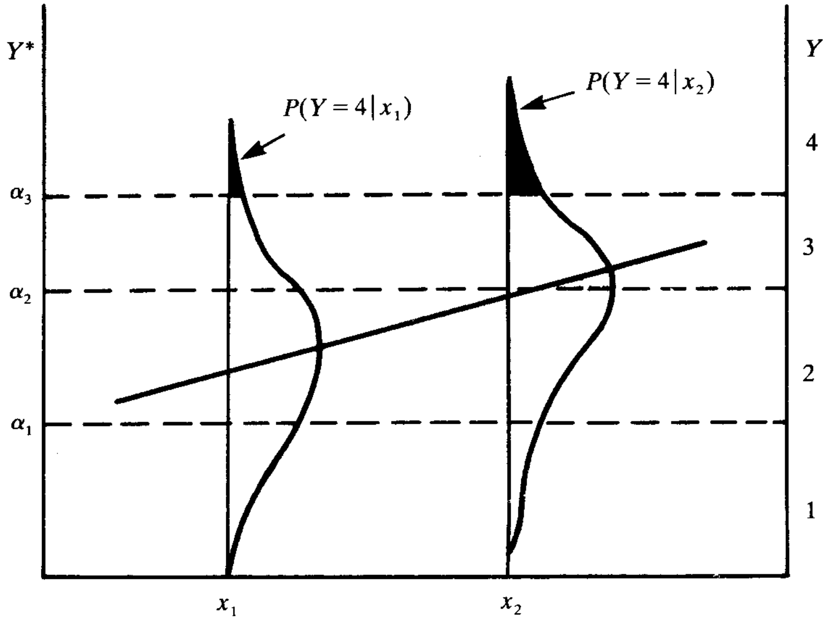
  $$Y^* = -\beta x + \varepsilon$$

  and the error term ε has a logistic distribution with cumulative distribution function

  $$\mathrm{P}(\varepsilon \le u) = \frac{e^u}{1+e^u}$$
- Then

  $$\mathrm{P}(Y \le j) = \mathrm{P}(Y^* \le \alpha_j) = \mathrm{P}(-\beta x + \varepsilon \le \alpha_j)$$
  $$= \mathrm{P}(\varepsilon \le \alpha_j + \beta x) = \frac{e^{\alpha_j+\beta x}}{1+e^{\alpha_j+\beta x}}$$

Latent Variable Interpretation for Cumulative Logit Models



Properties of Cumulative Logit Models

$$\text{odds}(Y \le j|x) = \frac{\mathrm{P}(Y \le j|x)}{\mathrm{P}(Y > j|x)} = e^{\alpha_j + \beta x}, \quad j = 1, \ldots, J-1.$$

- e^{β} = multiplicative effect of 1-unit increase in x on odds that $(Y \le j)$ (instead of $(Y > j)$).

  $$\frac{\text{odds}(Y \le j|x_2)}{\text{odds}(Y \le j|x_1)} = e^{\beta(x_2 - x_1)}$$

  So cumulative logit models are also called **proportional odds models**.
- ML estimates for coefficients (α_j, β) can be found via R function `vglm` in package `VGAM` w/ cumulative family.

Example (Income and Job Satisfaction from 1991 GSS)

Income	Job Satisfaction			
	Very Dissatisfied	Little Dissatisfied	Moderately Satisfied	Very Satisfied
0-5K	2	4	13	3
5K-15K	2	6	22	4
15K-25K	0	1	15	8
>25K	0	3	13	8

Using x = income scores (3K, 10K, 20K, 35K), we fit the model

$$\text{logit}[\mathrm{P}(Y \le j|x)] = \alpha_j + \beta x, \quad j = 1, 2, 3.$$

```
> jobsat.cl1 = vglm(cbind(Diss,Little,Mod,Very) ~ Income,
                   family=cumulative(parallel=TRUE))
> coef(jobsat.cl1)
(Intercept):1 (Intercept):2 (Intercept):3        Income
  -2.58287349   -0.89697897    2.07506012   -0.04485911
```

Fitted model:

$$\text{logit}[\widehat{\mathrm{P}}(Y \le j|x)] = \begin{cases} -2.583 - 0.045x, & \text{for } j = 1 \text{ (Dissat)} \\ -0.897 - 0.045x, & \text{for } j = 2 \text{ (Dissat or little)} \\ 2.075 - 0.045x, & \text{for } j = 3 \text{ (Dissat or little or mod)} \end{cases}$$

Estimated odds of satisfaction below any given level is multiplied by

$$e^{10\widehat{\beta}} = e^{(10)(-0.045)} = 0.64$$

for each 10K increase in income.

Remark

If **reverse ordering** of response, β **changes sign** but has same SE. With "very sat." < "moderately sat" < "little dissatisfied" < "dissatisfied":

```
> jobsat.cl1r = vglm(cbind(Very,Mod,Little,Diss) ~ Income,
                    family=cumulative(parallel=TRUE))
> coef(jobsat.cl1r)
(Intercept):1 (Intercept):2 (Intercept):3        Income
  -2.07506012    0.89697897    2.58287349    0.04485911
```

$\widehat{\beta} = 0.045$, estimated odds of satisfaction above any given level is multiplied by

$$e^{10\widehat{\beta}} = 1.566 = 1/0.64.$$

for each 10K increase in income

Wald Tests and Wald CIs for Parameters

Wald test of H_0: $\beta = 0$ (job satisfaction indep. of income):

$$z = \frac{\widehat{\beta} - 0}{\mathrm{SE}(\widehat{\beta})} = \frac{-0.0449}{0.0175} = -2.56, \quad (z^2 = 6.57, df = 1)$$

P-value $= 0.0105$

$$\begin{aligned} \underline{\text{95\% CI for } \beta} : \ \widehat{\beta} \pm 1.96\mathrm{SE}(\widehat{\beta}) &= -0.0449 \pm 1.96 \times 0.0175 \\ &= (-0.079, -0.011) \end{aligned}$$

$$\underline{\text{95\% CI for } e^{\beta}} : \ (e^{-0.079}, e^{-0.011}) = (0.924, 0.990)$$

```
> summary(jobsat.cl1)
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept):1 -2.58287    0.55842  -4.625 3.74e-06 ***
(Intercept):2 -0.89698    0.35499  -2.527   0.0115 *
(Intercept):3  2.07506    0.41582   4.990 6.03e-07 ***
Income        -0.04486    0.01750  -2.563   0.0104 *
```

LR Test for Parameters

LR test of H_0: $\beta = 0$ (job satisfaction indep. of income):

LR statistic $= -2(L_0 - L_1) = -2((-21.358) - (-18.000)) = 6.718$

P-value $= 0.0095$

```
> lrtest(jobsat.cl1)
Likelihood ratio test

Model 1: cbind(Diss, Little, Mod, Very) ~ Income
Model 2: cbind(Diss, Little, Mod, Very) ~ 1
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   8 -18.000
2   9 -21.358  1 6.7179   0.009545 **
```

Remark

For the Income and Job Satisfaction data, we obtained stronger evidence of association if we use a cumulative logits model treating Y (Job Satisfaction) as ordinal than obtained if we treat:

- Y as nominal (baseline category logit model) and X as ordinal:

  $$\log(\pi_j/\pi_4) = \alpha_j + \beta_j x.$$

  Recall P-value $= 0.032$ for LR test.
- X, Y both as nominal: Pearson test of independence had
  $X^2 = 11.5$, $df = 9$, P-value $= 0.24$
  $G^2 = 13.47$, $df = 9$, P-value $= 0.14$

Deviance and Goodness of Fit

Deviance can be used to test Goodness of Fit in the same way.

For cumulative logit model for Job Satisfaction data

$$\text{Deviance} = 6.7494, \quad df = 8, \quad P\text{-value } = 0.56$$

The Model fits data well.

```
> summary(jobsat.cl1)

Call:
vglm(formula = cbind(Diss, Little, Mod, Very) ~ Income,
               family = cumulative(parallel = TRUE))

Residual deviance: 6.7494 on 8 degrees of freedom

> pchisq(deviance(jobsat.cl1),df=8,lower.tail=F)
[1] 0.5638951
```

Remark. Generally, Goodness of fit test is appropriate if most of the fitted counts are ≥ 5. which is not the case for for the Job Satisfaction data. The P-value might not be reliable.

Example (Political Ideology and Party Affiliation)

Gender	Political Party	Political Ideology				
		very liberal	slightly liberal	moderate	slightly conservative	very conservative
Female	Democratic	44	47	118	23	32
	Republican	18	28	86	39	48
Male	Democratic	36	34	53	18	23
	Republican	12	18	62	45	51

Y = political ideology (1 = very liberal, $\ldots, 5$ = very conservative)

G = gender ($1 = M, 0 = F$)

P = political party (1 = Republican, 0 = Democratic)

Cumulative Logit Model:

$$\text{logit}[\mathrm{P}(Y \le j)] = \alpha_j + \beta_G G + \beta_P P, \quad j = 1, 2, 3, 4.$$

```
> Gender = c("F","F","M","M")
> Party = c("Dem","Rep","Dem","Rep")
> VLib = c(44,18,36,12)
> SLib = c(47,28,34,18)
> Mod  = c(118,86,53,62)
> SCon = c(23,39,18,45)
> VCon = c(32,48,23,51)

> ideo.cl1 =
   vglm(cbind(VLib,SLib,Mod,SCon,VCon) ~ Gender + Party,
   family=cumulative(parallel=TRUE))
> coef(ideo.cl1)
(Intercept):1 (Intercept):2 (Intercept):3 (Intercept):4
   -1.4517674    -0.4583355     1.2549929     2.0890430
      GenderM      PartyRep
   -0.1168560    -0.9636181
```

Fitted Model:

$$\text{logit}[\widehat{\mathrm{P}}(Y \le j)] = \widehat{\alpha}_j - 0.117G - 0.964P, \qquad j = 1, 2, 3, 4.$$

```
> summary(ideo.cl1)
              Estimate Std. Error z value Pr(>|z|)
(Intercept):1  -1.4518     0.1228 -11.818  < 2e-16 ***
(Intercept):2  -0.4583     0.1058  -4.333 1.47e-05 ***
(Intercept):3   1.2550     0.1145  10.956  < 2e-16 ***
(Intercept):4   2.0890     0.1292  16.174  < 2e-16 ***
GenderM        -0.1169     0.1268  -0.921    0.357
PartyRep       -0.9636     0.1294  -7.449 9.39e-14 ***
```

- Controlling for gender, estimated odds that a Republican is in liberal direction ($Y \le j$) rather than conservative ($Y > j$) are

  $$e^{\widehat{\beta}_P} = e^{-0.964} = 0.38$$

  times estimated odds for a Democrat.
  Same for all $j = 1, 2, 3, 4$.
- 95% CI for e^{β_P} is

  $$e^{\widehat{\beta}_P \pm 1.96\,\mathrm{SE}(\widehat{\beta}_P)} = e^{-0.964 \pm (1.96)(0.129)} = (0.30, 0.49)$$

- Based on Wald test, Party effect is significant (controlling for Gender) but Gender is not significant (controlling for Party).

LR Tests

LR test of H_0: $\beta_G = 0$ (no Gender effect, given Party):

```
> lrtest(ideo.cl1,1)     # LR test for Gender effect
Likelihood ratio test

Model 1: cbind(VLib, SLib, Mod, SCon, VCon) ~ Gender + Party
Model 2: cbind(VLib, SLib, Mod, SCon, VCon) ~ Party
  #Df  LogLik Df  Chisq Pr(>Chisq)
1  10 -47.415
2  11 -47.836  1 0.8427     0.3586
```

LR test of H_0: $\beta_P = 0$ (no Party effect, given Gender):

```
> lrtest(ideo.cl1,2)     # LR test for Party effect
Likelihood ratio test

Model 1: cbind(VLib, SLib, Mod, SCon, VCon) ~ Gender + Party
Model 2: cbind(VLib, SLib, Mod, SCon, VCon) ~ Gender
  #Df  LogLik Df  Chisq Pr(>Chisq)
1  10 -47.415
2  11 -75.838  1 56.847  4.711e-14 ***
```

LR tests give the same conclusion as Wald tests.

Interaction?

Model w/ Gender×Party interaction:

$$\text{logit}[\mathrm{P}(Y \le j)] = \alpha_j + \beta_G G + \beta_P P + \beta_{GP} G * P, \quad j = 1,2,3,4.$$

For H_0: $\beta_{GP} = 0$, LR statistic $= 3.99$, df $= 1$, P-value $= 0.046$

⇒ Evidence of effect of Party depends on Gender (and vice versa)

```
> ideo.cl2 =
    vglm(cbind(VLib,SLib,Mod,SCon,VCon) ~ Gender * Party ,
    family=cumulative(parallel=TRUE))
> lrtest(ideo.cl2,ideo.cl1)
Likelihood ratio test

Model 1: cbind(VLib, SLib, Mod, SCon, VCon) ~ Gender * Party
Model 2: cbind(VLib, SLib, Mod, SCon, VCon) ~ Gender + Party
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   9 -45.419
2  10 -47.415  1 3.9922    0.04571 *
```

```
> coef(ideo.cl2)
   (Intercept):1    (Intercept):2    (Intercept):3    (Intercept):4
      -1.5520853       -0.5549908        1.1646526        2.0012144
         GenderM         PartyRep GenderM:PartyRep
       0.1430828       -0.7562072       -0.5091332
```

Fitted model w/ Gender×Party interaction:

$$\text{logit}[\widehat{\mathrm{P}}(Y \le j)] = \widehat{\alpha}_j + 0.143G - 0.756P - 0.509G*P, \quad j = 1,2,3,4.$$

Estimated odds ratio for Party effect is

$$\begin{cases} e^{-0.756} = 0.47 & \text{for females } (G = 0) \\ e^{-0.756-0.51} = e^{-1.266} = 0.28 & \text{for males } (G = 1) \end{cases}$$

Diff between Dem. and Rep. are bigger among males than among females.

Observed percentages based on data:

Gender		Political Ideology				
	Party	Very Liberal	Slightly Liberal	Moderate	Slightly Conserve.	Very Conserve.
Female	Dem.	17%	18%	45%	9%	12%
	Rep.	8%	13%	39%	18%	22%
Male	Dem.	22%	21%	32%	11%	14%
	Rep.	6%	10%	33%	24%	27%

Fitted model w/ Gender×Party interaction:

$$\text{logit}[\widehat{\mathrm{P}}(Y \le j)] = \widehat{\alpha}_j + 0.143G - 0.756P - 0.509G*P, \quad j = 1,2,3,4.$$

Estimated odds ratio for Gender effect is

$$\begin{cases} e^{0.143} = 1.15 & \text{for Dems } (P = 0) \\ e^{0.143-0.51} = e^{-0.336} = 0.69 & \text{for Reps } (P = 1) \end{cases}$$

Among Dems, males tend to be more liberal than females.
Among Reps, males tend to be more conservative than females.

Goodness of Fit

$$\text{Deviance} = 11.063, \quad df = 9 \text{ (why?)}, \quad P\text{-value} = 0.44$$

The cumulative logits model w/ interaction fits data well.

```
> summary(ideo.cl2)

Call:
vglm(formula = cbind(VLib, SLib, Mod, SCon, VCon) ~
     Gender * Party, family = cumulative(parallel = TRUE))

Residual deviance: 11.0634 on 9 degrees of freedom
> # or
> deviance(ideo.cl2)
[1] 11.06338
> df.residual(ideo.cl2)
[1] 9
> 1-pchisq(11.0634 , df= 11)
[1] 0.4379672
```

Reversing Order of Responses

Reversing order of response categories **changes signs** of estimates (odds ratio ⟶ 1/odds ratio).

```
> ideo.cl2r =
     vglm(cbind(VCon,SCon,Mod,SLib,VLib) ~ Gender * Party ,
     family=cumulative(parallel=TRUE))
> coef(ideo.cl2r)
   (Intercept):1    (Intercept):2    (Intercept):3    (Intercept):4
      -2.0012144       -1.1646526        0.5549908        1.5520853
         GenderM         PartyRep GenderM:PartyRep
      -0.1430828        0.7562072        0.5091332
> coef(ideo.cl2)
   (Intercept):1    (Intercept):2    (Intercept):3    (Intercept):4
      -1.5520853       -0.5549908        1.1646526        2.0012144
         GenderM         PartyRep GenderM:PartyRep
       0.1430828       -0.7562072       -0.5091332
```

Collapsing Ordinal Responses to Binary

A loss of efficiency occurs collapsing ordinal responses to binary (and using ordinary logistic regression) in the sense of getting larger SEs.

```
> ideo.bin1 =
  glm(cbind(VLib+SLib, Mod+SCon+VCon) ~ Gender*Party, family=binomial)
> summary(ideo.bin1)
Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)       -0.6424     0.1295  -4.961 7.01e-07 ***
GenderM            0.3476     0.2042   1.702  0.08866 .
PartyRep          -0.6822     0.2104  -3.242  0.00119 **
GenderM:PartyRep  -0.6844     0.3300  -2.074  0.03807 *

> ideo.bin2 =
  glm(cbind(VLib+SLib+Mod, SCon+VCon) ~ Gender*Party, family=binomial)
> summary(ideo.cl.bin)
Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)        1.3350     0.1515   8.809  < 2e-16 ***
GenderM           -0.2364     0.2356  -1.004    0.316
PartyRep          -0.9181     0.2050  -4.478 7.54e-06 ***
GenderM:PartyRep  -0.2231     0.3096  -0.721    0.471
```