Multicategory Logit Models

Ordinal Responses

Demetris Athienitis

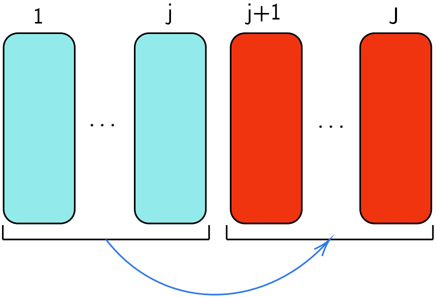
Motivation

Try to utilize the inherent information in ordinal responses to provide more accurate predictions.

Ordinal responses are often quantitative responses that have been simplified. E.g. a beverage can be small, medium or large. Underlying is a quantitative scale such as ml or oz. Sometimes it is harder to unearth the quantitative scale, e.g. happiness scale: very happy, happy, indifferent, sad, very sad.

Proportional Odds Model

$$
\begin{aligned}
\operatorname{logit}\left[P(Y \le j)\right] &= \log\left(\frac{P(Y \le j)}{1 - P(Y \le j)}\right) \\
&= \log\left(\frac{P(Y \le j)}{P(Y > j)}\right) \\
&= \alpha_j + \beta x, \quad j = 1, \dots, J-1
\end{aligned}
$$



Proportional Odds Model

$$P(Y \le j) = \frac{e^{\alpha_j + \beta x}}{1 + e^{\alpha_j + \beta x}}, \quad j = 1, 2, \ldots J - 1$$

- Separate intercept α_j for each cumulative logit
- Same (slope) coefficient β for each cumulative logit
- The term e^{β} = multiplicative effect of 1-unit increase in x on odds that $(Y \le j)$ instead of $(Y > j)$

$$\begin{aligned} \frac{\text{odds}(Y \le j | x_2)}{\text{odds}(Y \le j | x_1)} &= \frac{e^{\alpha_j + \beta x_2}}{e^{\alpha_j + \beta x_1}} \\ &= e^{\beta(x_2 - x_1)} \\ &= e^{\beta}, \quad \text{when } x_2 = x_1 + 1 \end{aligned}$$

Example (Job Satisfaction)

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

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

Example (continued)

```
> fit.clogit1=vglm(cbind(VD,LD,MS,VS)~income,
+ family=cumulative(parallel=TRUE),data=dat)
> summary(fit.clogit1)
Coefficients:
               Estimate Std. Error z value
(Intercept):1 -2.473156   0.568376 -4.3513
(Intercept):2 -0.781728   0.373724 -2.0917
(Intercept):3  2.211091   0.445123  4.9674
income        -0.056347   0.020871 -2.6998

Residual deviance: 5.9527 on 8 degrees of freedom
Log-likelihood: -17.60121 on 8 degrees of freedom
```

$$\text{logit}\left[\hat{P}(Y \le j|x)\right] = \hat{\alpha}_j - 0.056x \quad j = 1, 2, 3.$$

Example (continued)

- Odds of response at low end of job satisfaction scale decrease as x increases, i.e. $\exp(-0.056) = 0.95$
- Estimated odds of job satisfaction below any given level (instead of above it) multiply by 0.95 for a 1-unit increase in x (1-unit=$1000)
- For a $10,000 increase in income, i.e. 10 units, the estimated odds multiply by $\exp(10(-0.056)) = 0.57$
- If we were to reverse the order of the responses, then $\hat{\beta} = +0.056$
- Odds ratio is the same between *same* two categories of x irrespective of cutoff region for response categories
- Odds ratio is the same between categories $x = 10$ and $x = 20$, and $x = 20$ and $x = 30$ due to the same increment in x

Example (continued)

A goodness of fit test concludes that the model is a good fit

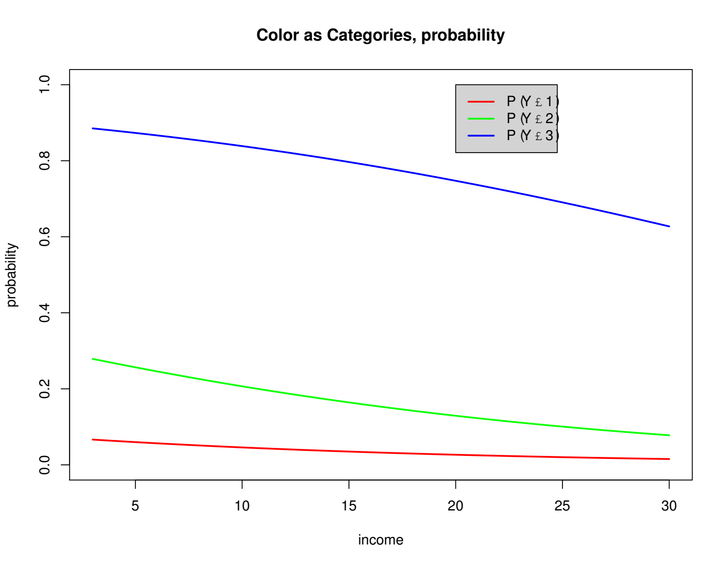
```
> 1-pchisq(deviance(fit.clogit1),df.residual(fit.clogit1))
[1] 0.6525305
```

Test of H_0 : job satisfaction independent of income, i.e. $\beta = 0$ yields

- A Wald z-stat of -2.6998 (or χ^2 of 7.17) and a p-value of 0.007.
- A LR statistic of $13.4673 - 5.9527 = 7.5146$ on 1 df and a p-value of 0.006. The null deviance was computed using

```
> vglm(cbind(VD,LD,MS,VS)~1,
+ family=cumulative(parallel=TRUE),data=dat)
```

Example (continued)



Example (continued)

A model with different β_j for $j = 1, 2, 3$ although more "flexible" does not significantly differ from the parallel lines model. To test $H_0 : \beta_1 = \beta_2 = \beta_3$ via L.R.T.

```
> fit.clogit2=vglm(cbind(VD,LD,MS,VS)~income,
+ family=cumulative(parallel=FALSE),data=dat)
> summary(fit.clogit2)
.
Residual deviance: 4.37717 on 6 degrees of freedom

> 1-pchisq(5.9527-4.37717,2)
[1] 0.4548603
```

and conclude that at we should be using one common β.

Example (Political Ideology)

An example with the following data yields

```
> ideow
  Gender      Party VLib SLib Mod SCon VCon
1 Female   Democrat   44   47 118   23   32
2 Female Republican   18   28  86   39   48
3   Male   Democrat   36   34  53   18   23
4   Male Republican   12   18  62   45   51
```

Example (continued)

```
> ideo.cl1=vglm(cbind(VLib,SLib,Mod,SCon,VCon)~Gender+Party,
+        family=cumulative(parallel=TRUE), data=ideow)
> summary(ideo.cl1)
Coefficients:
                Estimate Std. Error   z value
(Intercept):1   -1.45177    0.12284 -11.81819
(Intercept):2   -0.45834    0.10577  -4.33337
(Intercept):3    1.25499    0.11455  10.95598
(Intercept):4    2.08904    0.12916  16.17374
GenderMale      -0.11686    0.12681  -0.92147
PartyRepublican -0.96362    0.12936  -7.44917

Residual deviance: 15.05557 on 10 degrees of freedom
Log-likelihood: -47.41497 on 10 degrees of freedom
```

Example (continued)

- GoF with $G^2 = 15.056$ and 10 df with p-value of 0.13
- Testing for gender effect (controlling for party) we have a Wald statistic -0.921 indicating a lack of evidence
- Testing for party effect (controlling for gender)
  - Wald: $z = -7.449$
  - LR: $71.902 - 15.056 = 56.846$ with df $= 1$. (Deviance of 71.902 was obtained by fitting model with only gender effect)

Controlling for gender, estimated odds that a Republican's response ($x_2 = 0$ to $x_2 = 1$) is in liberal direction ($Y \leq j$) rather than conservative ($Y > j$) are $\exp(-0.964) = 0.38$ times estimated odds for a Democrat. The 95% CI for the odds ratio is (but best to use `confint`)

$$\exp(-0.964 \pm 1.96(0.129)) \rightarrow (0.30, 0.49)$$

Example (continued)

May be an interaction between gender and party affiliation.

```
> ideo.cl2=vglm(cbind(VLib,SLib,Mod,SCon,VCon)~Gender*Party,
+               family=cumulative(parallel=TRUE), data=ideow)
> summary(ideo.cl2)
Coefficients:
                            Estimate Std. Error   z value
(Intercept):1               -1.55209    0.13353 -11.62339
(Intercept):2               -0.55499    0.11703  -4.74225
(Intercept):3                1.16465    0.12337   9.44006
(Intercept):4                2.00121    0.13682  14.62633
GenderMale                   0.14308    0.17936   0.79772
PartyRepublican             -0.75621    0.16691  -4.53062
GenderMale:PartyRepublican -0.50913    0.25408  -2.00381

Residual deviance: 11.06338 on 9 degrees of freedom
```

Example (continued)

Interaction term appears significant.

- Wald: $z = -2.004$ with p-value=0.04507
- LR: $15.056 - 11.063 = 3.993$ with df=1 and p-value=0.0457

The goodness of fit test with $G^2 = 11.063$ residual deviance and df=9 wields a p-value of 0.2714153, a big improvement from 0.13 for the additive model. This is because the interaction takes into account the relationship between gender and party affiliation and how they affect political ideology.

Example (continued)

- Estimated odds ratio for party effect (x_2), (allowing gender to differ)

$$\exp(b_2) = \exp(-0.756) = 0.47 \quad \text{when } x_1 = 0 \text{ (F)}$$
$$\exp(b_2 + b_3) = \exp(-0.756 - 0.509) = 0.28 \quad \text{when } x_1 = 1 \text{ (M)}$$

- Estimated odds that a female Republican's response is in liberal direction rather than conservative are 0.47 times estimated odds for a female Democrat.
- Estimated odds that a male Republican's response is in liberal direction rather than conservative are 0.28 times estimated odds for a male Democrat.

Example (continued)

- Estimated odds ratio for gender effect (x_1)

$$\exp(b_1) = \exp(0.143) = 1.15 \quad \text{when } x_2 = 0 \text{ (Dem)}$$
$$\exp(b_1 + b_3) = \exp(0.143 - 0.509) = 0.69 \quad \text{when } x_2 = 1 \text{ (Rep)}$$

- Estimated odds that a male Democrat's response is in liberal direction rather than conservative are 1.15 times estimated odds for a female Democrat.
- Estimated odds that a male Republican's response is in liberal direction rather than conservative are 0.69 times estimated odds for a female Republican.

Example (continued)

$$\hat{P}(Y \leq j) = \frac{\exp(\hat{\alpha}_j + 0.143x_1 - 0.756x_2 - 0.509x_1x_2)}{1 + \exp(\hat{\alpha}_j + 0.143x_1 - 0.756x_2 - 0.509x_1x_2)}$$

- $\hat{P}(Y = 1) = \hat{P}(Y \leq 1)$. For $j = 1$ (very liberal) the probability for a male republican ($\hat{\alpha}_1 = -1.55, x_1 = 1, x_2 = 1$):

$$\hat{P}(Y = 1) = \frac{e^{-2.67}}{1 + e^{2.67}} = 0.065$$

- Similarly, $\hat{P}(Y = 2) = \hat{P}(Y \leq 2) - \hat{P}(Y \leq 1)$, etc. Note $\hat{P}(Y = 5) = \hat{P}(Y \leq 5) - \hat{P}(Y \leq 4) = 1 - \hat{P}(Y \leq 4)$.

We learned

- Utilized inherent information in ordinal responses to provide more accurate predictions
- Kept systematic component "parallel" so that there is no violation of cumulative probabilities.