Multinomial Models

Motivation and Set-up:

Instead of binary outcomes, there are often some K arbitrary number of outcomes.

- (0) worst, (1) okay, (2) great

  notice the ordinal pairing

- (0) blue, (1) yellow, (2)green

  notice the nominal pairing

Suppose $\boldsymbol{Y} \sim$ Multinomial with $j = 1, ..., J+1$ categories; consider the following setup [1]:

$$\boldsymbol{Y} = \{Y_0, ..., Y_J\}^T \qquad \boldsymbol{\pi} = \{\pi_0, \pi_1, ..., \pi_J\}$$
$$\pi_j = P(Y = j) \qquad E[Y_j] = n\pi_j$$
$$V[Y_j] = n\pi_j(1-\pi_j) \qquad Cov(Y_j, Y_k) = -n\pi_j\pi_k$$

$$f_Y(y) = \frac{n!}{y_1!...y_J!} p_1^{y_1}...p_j^{y_J}$$

$$f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\pi}) = \exp\{n \log \pi_0 + \sum_{j=1}^{J} Y_j \log(\frac{\pi_j}{\pi_0}) + \log C(n; \boldsymbol{Y})\}$$

$$t(\boldsymbol{Y}) = \boldsymbol{Y} \qquad \boldsymbol{\theta} = \log(\frac{\pi_j}{\pi_0})$$

Note that:

$$\sum_{j=1}^{J} e^{\theta_j} = \frac{1}{\pi_0} \sum_{j=1}^{J} \pi_j = \frac{1}{\pi_0}(1-\pi_0)$$

where

$$\pi_0 = \frac{1}{1+e^{\theta_j}} \qquad \pi_j = \frac{e^{\theta_j}}{1+e^{\theta_j}}$$

and

$$b(\boldsymbol{\theta}) = -n \log(\pi_0)$$

Using the canonical link and setting $\theta_j = \log(\pi_j/\pi_0)$ The model becomes:

$$\log(\frac{\pi_{ij}}{\pi_{i0}}) = x_i^T \beta_j \qquad j = 1, ..., J$$

Notice that this is essentially fitting J logistic regression models, using the reference probability in each. Each of these models will have distinct intercepts and regression coefficients.

$$P(Y_i = j|\boldsymbol{x}_i) = \pi_{ij} = \frac{\exp\{\boldsymbol{x}_i^T \boldsymbol{\beta}_j\}}{1 + \sum_{j=1}^{J} \exp\{\boldsymbol{x}_i^T \boldsymbol{\beta}_j\}} \qquad P(Y_i = 0|\boldsymbol{x}_i) = \pi_{i0} = \frac{1}{1 + \sum_{j=1}^{J} \exp\{\boldsymbol{x}_i^T \boldsymbol{\beta}_j\}}$$

[1] see last page for full derivation of exponential family

Likelihood function and derivatives

say that $\boldsymbol{\beta} = [\beta_1^T, \beta_2^T, ..., \beta_J^T]$

$$
\begin{aligned}
L(\boldsymbol{\pi}) &= \prod_{i=1}^{n}\prod_{j=0}^{J} \pi_{ij}^{Y_ij} \\
\Rightarrow l(\boldsymbol{\beta}) &= \sum_{i=1}^{n}\Big[-Y_{i0}\log(1+\sum_{l=1}^{J} e^{\boldsymbol{x}_i^T\boldsymbol{\beta}_l}) + \sum_{j=1}^{J} Y_{ij}\log\Big(\frac{e^{\boldsymbol{x}_i^T\boldsymbol{\beta}_j}}{1+\sum_{l=1}^{J} e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}}\Big)\Big] \\
&= \sum_{i=1}^{n}\Big[-\log(1+\sum_{l=1}^{J} e^{\boldsymbol{x}_i\boldsymbol{\beta}_l}) + \sum_{j=1}^{J} Y_{ij}\boldsymbol{x}_i^T\boldsymbol{\beta}_j\Big]
\end{aligned}
$$

The score function then becomes

$$
U(\boldsymbol{\beta}) = \begin{bmatrix} U_1(\boldsymbol{\beta}) \\ \vdots \\ U_J(\boldsymbol{\beta}) \end{bmatrix}
$$

with jth sub-vector

$$
U_j(\boldsymbol{\beta}) = \sum_{i=1}^{n}(Y_{ij} - \pi_{ij})\boldsymbol{x}_i
$$

and information matrix $J(\boldsymbol{\beta}) \in \mathbb{R}^{(J(q+1))\times(J(q+1))}$

$$
J(\boldsymbol{\beta}) = \begin{bmatrix} J_{11}(\boldsymbol{\beta}) & ... & J_{1J}(\boldsymbol{\beta}) \\ \vdots & J_{22}(\boldsymbol{\beta}) & \vdots \\ ... & ... & J_{JJ}(\boldsymbol{\beta}) \end{bmatrix}
$$

with the (j,j) th block: $\sum_{i=1}^{n}\pi_{ij}(1-\pi_{ij})\boldsymbol{x}_i^T\boldsymbol{x}_i$ and $(j,k)th$ block: $-\sum_{i=1}^{n}\pi_{ij}\pi_{ik}\boldsymbol{x}_i^T\boldsymbol{x}_i$

Examples

Generalized Logit Model

	$Y_i = 0$	$Y_i = 1$	$Y_i = 2$	total
$F_i = 0$	24	14	22	60
$F_i = 1$	21	10	9	40

Model:

$$
\log(\frac{\pi_{ij}}{\pi_{i0}}) = \beta_{0j} + \beta_{1j}F_i \qquad\qquad j = 1, 2
$$

Since this model is saturated it can be estimated by the cell ratios, in addition to a fitted model.

$$\begin{aligned}
\log\big(\frac{P(Y=1|F=0)}{P(Y=0|F=0)}\big) &= \hat{\beta_{01}} = \log(\frac{14/60}{24/60}) \\
\log\big(\frac{P(Y=1|F=0)}{P(Y=0|F=0)}\big) &= \hat{\beta_{02}} = \log(\frac{22/60}{24/60}) \\
\log\big(\frac{P(Y=1|F=1)}{P(Y=0|F=1)}\big) &= \hat{\beta_{01}} + \hat{\beta_{11}} = \log(\frac{10/40}{21/40}) \\
\log\big(\frac{P(Y=2|F=1)}{P(Y=0|F=1)}\big) &= \hat{\beta_{02}} + \hat{\beta_{12}} = \log\big(\frac{9/40}{21/40}\big) \\
\Rightarrow \hat{\beta_{11}} &= \log(\frac{10/40}{21/40}) - \log(\frac{14/60}{24/60}) \\
\Rightarrow \hat{\beta_{12}} &= \log\big(\frac{9/40}{21/40}\big) - \log(\frac{22/60}{24/60})
\end{aligned}$$

It can be seen from the above that:

- β_{11} is the log odds ratio of $Y = 1$ for a male vs. female - given that $Y \neq 2$
- β_{22} is the log odds ratio of $Y = 2$ for a male vs. female given that $Y \neq 1$
- $\exp\{\hat{\beta}_{11}\} = 0.81$ is the effect of being a female on $P(Y = 1)$
- $\exp\{\hat{\beta}_{12}\} = 0.48$ is the effect of being a female on $P(Y = 2)$

Cumulative Logit Model

Generaliezd Logit models do not take advantage of the inherent ordering of the data in the interpretation. Even though they can still be used in this context, cumulative logit models offer the advantage of exploiting any ordering the multinomial data may have.

Model:

$$\log\left(\frac{P(Y_i \leq j)}{P(Y_i > j)}\right) = \log\left(\frac{\pi_0 + \pi_1 + ...\pi_j}{\pi_{j+1} + ... + \pi_J}\right) = \gamma_{0j} + \boldsymbol{x}_i^T \boldsymbol{\gamma}_j \quad j = 0, 1, ..., (J-1)$$

Properties:

- $J + 1$ response levels
- $\boldsymbol{x}_i : q \times 1$ covariate

  number of parameters $(q + 1) \times J$

Cumulative Logit Model Example:

to be added later

Proportional Odds Model

- Same essential set up as Cumulative Logit Model but...
- Assume $\boldsymbol{\gamma} = \gamma_0 = \gamma_1 = ... = \gamma_{J-1}$
- Proportionality assumption assumes a simpler model.
- new model

$$\log\left(\frac{P(Y_i \leq j)}{P(Y_i > j)}\right) = \gamma_{0j} + \boldsymbol{x}_i^T \boldsymbol{\gamma}$$

Properties

- Odds: $\text{odds}(Y_i \leq j|\boldsymbol{x}_1) = \frac{P(Y_i \leq j|\boldsymbol{x}_1)}{P(Y_i > j|\boldsymbol{x}_1)} = e^{\gamma_{0j}} e^{\boldsymbol{x}_1^T \boldsymbol{\gamma}}$
- Odds Ratio: $\frac{\text{odds}(Y_i \leq j|\boldsymbol{x}_1)}{\text{odds}(Y_i \leq j|\boldsymbol{x}_2)} = \exp\{(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T \boldsymbol{\gamma}\}$

  Notice this is independent of j
- Number of parameters: $J + q$
- Score test is used to test proportionality assumption

  compare proportional model to cumulative logit model

  reference distribution is χ^2_{J-1}

Prop Odds Model Example:

Consider the following data

Treatment	Sex	Progressive disease	No Change	Partial Remission	Complete Remission
Sequential	Male	28	45	29	26
	Female	4	12	5	2
Alternating	Male	41	44	20	20
	Female	12	7	3	1

Fitting a proportional odds model to estimate the probabilities for each response category taking treatment and sex effects into account.

$$\log\left(\frac{P(Y_i \leq j)}{P(Y_i > j)}\right) = \beta_{0j} + \beta_1 I(x_i = \text{female}) + \beta_2 I(x_i = \text{Alt TX})$$

with the following coding for disease states.

- $j = 0$: Progressive disease
- $j = 1$: No Change
- $J = 2$ Partial remission
- $J = 3$ Complete remission.

This model can be fit with the following code in SAS:

```
proc genmod data=data;
model resp = sex trt / dist=multinomial link=clogit aggregate;
run;
```

This results in the following estimates:

$$log(\frac{P(Y_i \leq j)}{P(Y_i > j)} = \beta_{0j} + 0.541 I(x_i = \text{female}) + 0.580 I(x_i = \text{Alt TX})$$

$$\beta_{00} = -1.318$$

$$\beta_{01} = 0.249$$

$$\beta_{02} = 1.300$$

The following probability estimates are derived by iteratively estimating each stratum (j) and then subtracting out a reference probability
e.g. for the case $j = 0, x_1 = 0, x_2 = 0$:

$$\log\left(\frac{P(Y<0|x_1=0,x_2=0)}{P(Y>0|x_1=0,x_2=0)}\right) = \log\left(\frac{P(Y=0|x_1=0,x_2=0)}{P(Y\neq 0|x_1=0,x_2=0)}\right) - 1.318$$
$$\Rightarrow \frac{P(Y=0|x_1=0,x_2=0)}{P(Y\neq 0|x_1=0,x_2=0)} = e^{-1.318}$$
$$\Rightarrow P(Y=0|x_1=0,x_2=0) = e^{-1.318} \underbrace{\frac{1}{1+e^{-1.318}}}_{P(Y\neq 0|x_1=0,x_2=0)} \approx .221$$

TX (X2)	Sex (X1)	Outcome Probability (Y)			
		Disease Progression (Y=0)	No Change (Y=1)	Partial Remission(Y=2)	Complete
Alternating (X2=1)	Male (X1=0)	0.324	0.373	0.171	0.132
	Female(X1=1)	0.451	0.346	0.121	0.081
Sequential (X2=0)	Male(X1=0)	0.211	0.351	0.224	0.214
	Female(X1=1)	0.315	0.373	0.175	0.137

Hypothesis Testing

- The Proportionality Assumption can be checked via assessment of the score test:

  $H_0: \theta = \theta_0, H_1: \theta \neq \theta_0$

  - $U(\theta_0)^T[I(\theta_0)]^{-1}U(\theta_0) \sim \chi^2_{9-5=4}$

  - note that the degrees of freedom for the χ^2 is the difference in parameter estimates between the full model and prop odds model

- The adequacy of model fit can be checked via the Deviance

  H_0 : model fit is adequate ; H_1 : model fit is inadequate

  TS = $5.5667 \sim \chi^2_7$

- Wald Test could be used to check the proportionality assumption as well, or the significance of one (or more) parameters

  $H_0: \beta_2 = 0$; $H_1: \beta_2 \neq 0$

  $TS = \frac{\hat{\beta}_j^2}{S.E.(\hat{\beta}_j)^2} \sim \chi^2_1$

Derivations

to be added later