DATA 606: Statistical Methods in Data Science

—— Multinomial regression

Wenjun Jiang

Department of Mathematics & Statistics
The University of Calgary

Lecture 11

Nominal response Y

Baseline-category logit model

- Nominal response Y has $J > 2$ levels:

		Y			
		1	2	$\cdots$	J
X	x_1 x_2 $\vdots$ x_I				

- Given data (x_i, y_i), we are interested in

$$\pi_1(x_i) = \mathbf{P}(Y = 1|x_i), \quad \pi_2(x_i) = \mathbf{P}(Y = 2|x_i), \ \ldots, \ \pi_J(x_i) = \mathbf{P}(Y = J|x_i),$$

where $\pi_1(x_i) + \pi_2(x_i) + \cdots + \pi_J(x_i) = 1$.

Baseline-category logit model

- We would like to model the relationship between $\{\pi_1(x), \pi_2(x), \ldots, \pi_J(x)\}$ and x.
- We first pick up a reference probability, e.g. $\pi_J(x)$, then model $\pi_j(x)/\pi_J(x)$ as

$$\log(\frac{\pi_1(x)}{\pi_J(x)}) = \alpha_1 + \beta_1 x,$$
$$\log(\frac{\pi_2(x)}{\pi_J(x)}) = \alpha_2 + \beta_2 x,$$
$$\vdots$$
$$\log(\frac{\pi_{J-1}(x)}{\pi_J(x)}) = \alpha_{J-1} + \beta_{J-1} x.$$

Baseline-category logit model

- Given the model, we can compare any 2 categories:

$$\log(\frac{\pi_1(x)}{\pi_2(x)}) = (\alpha_1 - \alpha_2) + (\beta_1 - \beta_2)x.$$

Baseline-category logit model

- Given the model, we can compare any 2 categories:

$$\log\left(\frac{\pi_1(x)}{\pi_2(x)}\right) = (\alpha_1 - \alpha_2) + (\beta_1 - \beta_2)x.$$

- We can also figure out $\pi_j(x)$ explicitly for any j with any x:

$$\begin{aligned}
\pi_1(x) &= \pi_J(x)e^{\alpha_1+\beta_1 x}, \\
\pi_2(x) &= \pi_J(x)e^{\alpha_2+\beta_2 x}, \\
&\cdots \\
\pi_{J-1}(x) &= \pi_J(x)e^{\alpha_{J-1}+\beta_{J-1} x}.
\end{aligned}$$

Baseline-category logit model

- Given the model, we can compare any 2 categories:

$$\log(\frac{\pi_1(x)}{\pi_2(x)}) = (\alpha_1 - \alpha_2) + (\beta_1 - \beta_2)x.$$

- We can also figure out $\pi_j(x)$ explicitly for any j with any x:

$$\begin{aligned}
\pi_1(x) &= \pi_J(x)e^{\alpha_1+\beta_1 x}, \\
\pi_2(x) &= \pi_J(x)e^{\alpha_2+\beta_2 x}, \\
&\dots \\
\pi_{J-1}(x) &= \pi_J(x)e^{\alpha_{J-1}+\beta_{J-1} x}.
\end{aligned}$$

- Since $\pi_1(x) + \pi_2(x) + \cdots + \pi_J(x) = 1$, we have

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

An example (job satisfaction)

Job satisfaction

		Job satisfaction			
		Very dissatisfied	Little dissatisfied	Moderately satisfied	Very satisfied
Income (1000s)	< 5	2	4	13	3
	$5 - 15$	2	6	22	4
	$15 - 25$	0	1	15	8
	> 25	0	3	13	8

An example

The prediction equations are

$$\log(\frac{\hat{\pi}_1}{\hat{\pi}_4}) = 0.564 - 0.199x,$$
$$\log(\frac{\hat{\pi}_2}{\hat{\pi}_4}) = 0.645 - 0.071x,$$
$$\log(\frac{\hat{\pi}_3}{\hat{\pi}_4}) = 1.819 - 0.047x.$$

The explicit prediction function for $Y = \text{VS}$ is given by

$$\hat{\pi}_4(x) = \frac{1}{1 + \exp(0.564 - 0.199x) + \exp(0.645 - 0.071x) + \exp(1.819 - 0.047x)}.$$

Goodness-of-fit

- **How to check the goodness-of-fit?**

Goodness-of-fit

- **How to check the goodness-of-fit?** Using Pearson Chi-square statistic.

Goodness-of-fit

- **How to check the goodness-of-fit?** Using Pearson Chi-square statistic.

- How to determine the degree of freedom?

Goodness-of-fit

- **How to check the goodness-of-fit?** Using Pearson Chi-square statistic.
- How to determine the degree of freedom? df$=(J-1)\times(I-1-\dim(X))$.

Another example

Alligator's food choice: fish (F), invertebrates (I), others (O).

Table 6.1. Alligator Size (Meters) and Primary Food Choice,[a] for 59 Florida Alligators

1.24 I	1.30 I	1.30 I	1.32 F	1.32 F	1.40 F	1.42 I	1.42 F
1.45 I	1.45 O	1.47 I	1.47 F	1.50 I	1.52 I	1.55 I	1.60 I
1.63 I	1.65 O	1.65 I	1.65 F	1.65 F	1.68 F	1.70 I	1.73 O
1.78 I	1.78 I	1.78 O	1.80 I	1.80 F	1.85 F	1.88 I	1.93 I
1.98 I	2.03 F	2.03 F	2.16 F	2.26 F	2.31 F	2.31 F	2.36 F
2.36 F	2.39 F	2.41 F	2.44 F	2.46 F	2.56 O	2.67 F	2.72 I
2.79 F	2.84 F	3.25 O	3.28 O	3.33 F	3.56 F	3.58 F	3.66 F
3.68 O	3.71 F	3.89 F					

[a] F = Fish, I = Invertebrates, O = Other.

Source: Thanks to M. F. Delany and Clint T. Moore for these data.

We are interested in how alligator's size (length) affects their food choice.

Alligator's food choice

The predicted probability function

$$\hat{\pi}_F = \frac{\exp(1.618 - 0.110x)}{1 + \exp(1.168 - 0.110x) + \exp(5.697 - 2.465x)},$$
$$\hat{\pi}_I = \frac{\exp(5.697 - 2.465x)}{1 + \exp(1.168 - 0.110x) + \exp(5.697 - 2.465x)},$$
$$\hat{\pi}_O = \frac{1}{1 + \exp(1.168 - 0.110x) + \exp(5.697 - 2.465x)}.$$

Alligator's food choice

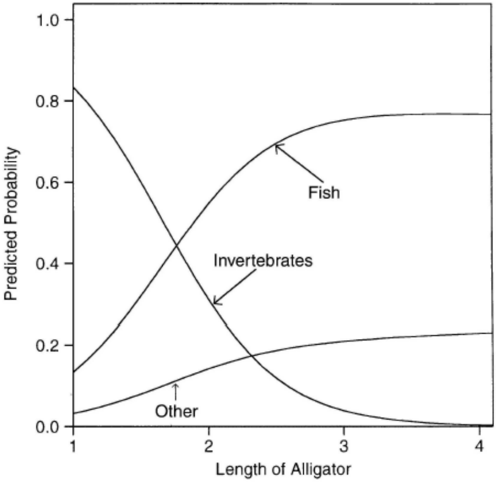

Figure 6.1. Estimated probabilities for primary food choice.

Third example

Belief in afterlife

Table 6.4. Belief in Afterlife by Gender and Race

		Belief in Afterlife		
Race	Gender	Yes	Undecided	No
White	Female	371	49	74
	Male	250	45	71
Black	Female	64	9	15
	Male	25	5	13

Source: General Social Survey.

Third example

Belief in afterlife

Table 6.6. Estimated Probabilities for Belief in Afterlife

		Belief in Afterlife		
Race	Gender	Yes	Undecided	No
White	Female	0.76	0.10	0.15
	Male	0.68	0.12	0.20
Black	Female	0.71	0.10	0.19
	Male	0.62	0.12	0.26

Cumulative logit model for ordinal response

- Ordinal response Y has $J > 2$ levels (e.g., $1 < 2 < \cdots < J$)

Y			
1	2	$\cdots$	J
$\pi_1(x)$	$\pi_2(x)$	$\cdots$	$\pi_J(x)$

Cumulative logit model for ordinal response

- Ordinal response Y has $J > 2$ levels (e.g., $1 < 2 < \cdots < J$)

Y			
1	2	$\cdots$	J
$\pi_1(x)$	$\pi_2(x)$	$\cdots$	$\pi_J(x)$

- We could use baseline-category logit model, but we would like to take the ordinal scale into account for a better prediction.

Cumulative logit model for ordinal response

- Ordinal response Y has $J > 2$ levels (e.g., $1 < 2 < \cdots < J$)

Y			
1	2	$\cdots$	J
$\pi_1(x)$	$\pi_2(x)$	$\cdots$	$\pi_J(x)$

- We could use baseline-category logit model, but we would like to take the ordinal scale into account for a better prediction.

- One way is to model the *cumulative probabilities:*

$$\tau_j(x) = \mathbf{P}(Y \leq j \mid x) = \pi_1(x) + \pi_2(x) + \cdots + \pi_j(x), \; j = 1, 2, \ldots, J - 1.$$

Cumulative logit model

- Consider the following logistic model for $\tau_j(x)$

$$\log(\frac{\tau_j(x)}{1-\tau_j(x)}) = \alpha_j + \beta x,\ j = 1, 2, \ldots, J-1.$$

Cumulative logit model

- Consider the following logistic model for $\tau_j(x)$

$$\log(\frac{\tau_j(x)}{1-\tau_j(x)}) = \alpha_j + \beta x,\ j = 1, 2, \ldots, J-1.$$

- Since $\tau_1(x) < \tau_2(x) < \cdots < \tau_{J-1}(x)$, we have

$$\alpha_1 < \alpha_2 < \cdots < \alpha_{J-1}.$$

Cumulative logit model

- Consider the following logistic model for $\tau_j(x)$

$$\log(\frac{\tau_j(x)}{1-\tau_j(x)}) = \alpha_j + \beta x,\ j = 1, 2, \ldots, J-1.$$

- Since $\tau_1(x) < \tau_2(x) < \cdots < \tau_{J-1}(x)$, we have

$$\alpha_1 < \alpha_2 < \cdots < \alpha_{J-1}.$$

- The expression of $\tau_j(x)$ is

$$\tau_j(x) = \frac{e^{\alpha_j+\beta x}}{1+e^{\alpha_j+\beta x}}.$$

  Therefore, we have

$$\pi_1(x) = \tau_1(x), \quad \pi_2(x) = \tau_2(x) - \tau_1(x), \ldots, \pi_J(x) 1 - \tau_{J-1}(x).$$

An example

Political ideology and party affiliation

Table 6.7. Political Ideology by Gender and Political Party

		Political Ideology				
Gender	Political Party	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

Source: General Social Survey.

An example

Political ideology and party affiliation

Table 6.7. Political Ideology by Gender and Political Party

		Political Ideology				
Gender	Political Party	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

Source: General Social Survey.

Our goal is to predict the probability of each cell as per cumulative logit model, where

$$\text{logit}(\tau_j(x, z)) = \alpha_j + \beta_1 x + \beta_2 z + \beta_3 x \times z$$

where x is indicator of democrat/republican and z is indicator of male/female.

An example (cont.)

With fitted model, we can estimate 4 cumulative probabilities:
Female Democrats: $x = 1, z = 0 : \tau_j's = 0.174, 0.365, 0.762, 0.881$
$\Rightarrow$ cell probs: $\pi_j's : 0.174, 0.190, 0.397, 0.119, 0.119$

Female Republicans: $x = 0, z = 0 : \tau_j's = 0.090, 0.212, 0.601, 0.776$
$\Rightarrow$ cell probs: $\pi_j's : 0.090, 0.122, 0.388, 0.176, 0.234$

Male Democrats: $x = 1, z = 1 : \tau_j's = 0.196, 0.398, 0.787, 0.895$
$\Rightarrow$ cell probs: $\pi_j's : 0.196, 0.202, 0.389, 0.108, 0.105$

Male Republicans: $x = 0, z = 1 : \tau_j's = 0.065, 0.157, 0.510, 0.707$
$\Rightarrow$ cell probs: $\pi_j's : 0.065, 0.093, 0.353, 0.196, 0.293$

Example with continuous/categorical x

- Mental impairment example: 40 subjects

Subj	Mental	SES	Life	Subj	Mental	SES	Life
1	Well	1	1	21	Mild	1	9
2	Well	1	9	22	Mild	0	3
3	Well	1	4	23	Mild	1	3
4	Well	1	3	24	Mild	1	1
5	Well	0	2	25	Moderate	0	0
6	Well	1	0	26	Moderate	1	4
7	Well	0	1	27	Moderate	0	3
8	Well	1	3	28	Moderate	0	9
9	Well	1	3	29	Moderate	1	6
10	Well	1	7	30	Moderate	0	4
11	Well	0	1	31	Moderate	0	3
12	Well	0	2	32	Impaired	1	8
13	Mild	1	5	33	Impaired	1	2
14	Mild	0	6	34	Impaired	1	7
15	Mild	1	3	35	Impaired	0	5
16	Mild	0	1	36	Impaired	0	4
17	Mild	1	8	37	Impaired	0	4
18	Mild	1	2	38	Impaired	1	8
19	Mild	0	5	39	Impaired	0	8
20	Mild	1	5	40	Impaired	0	9

Example with continuous/categorical x

- Y: mental impairment, has 4 levels

Y			
1	2	3	4
Well	Mild	Moderate	Impaired

- x_1: life event index.
- x_2: social-economic status (ses).

We would like to model

$$\log \frac{\mathbf{P}(Y \leq j)}{1 - \mathbf{P}(Y \leq j)} = \alpha_j + \beta_1 x_1 + \beta_2 x_2, \quad j = 1, 2, 3.$$

Invariance to choice of response categories

Suppose we group the middle 2 categories and form a new table

$\tilde{Y}$		
1 Well	2 Mild or Moderate	3 Impaired

Invariance to choice of response categories

We have the following observations

$\widehat{\alpha}_1 = -0.0468(SE = 0.642)$, compared to -0.282 $(SE = 0.623)$ from the original model.
$\widehat{\alpha}_3 = 2.482(SE = 0.783)$, compared to 2.210 $(SE = 0.717)$ from the original model.
$\widehat{\beta}_1 = -0.355(SE = 0.129)$, compared to -0.319 $(SE = 0.119)$ from the original model.
$\widehat{\beta}_2 = 0.933(SE = 0.640)$, compared to 1.111 $(SE = 0.614)$ from the original model.

Continuation-ratio logit model

- Ordinal response Y has $J > 2$ levels

Y			
1	2	$\cdots$	J
$\pi_1(x)$	$\pi_2(x)$	$\cdots$	$\pi_J(x)$

- We may consider the following model (in R, reverse=FALSE)

$$\log\left(\frac{\pi_1(x)}{\pi_2(x) + \cdots + \pi_J(x)}\right) = \alpha_1 + \beta x,$$

$$\log\left(\frac{\pi_2(x)}{\pi_3(x) + \cdots + \pi_J(x)}\right) = \alpha_2 + \beta x,$$

$$\cdots$$

$$\log\left(\frac{\pi_{J-1}(x)}{\pi_J(x)}\right) = \alpha_{J-1} + \beta x.$$

Continuation-ratio logit model

- Or we may have (in R, reverse=TRUE)

$$
\begin{aligned}
&\log\left(\frac{\pi_2(x)}{\pi_1(x)}\right) = \alpha_1 + \beta x, \\
&\log\left(\frac{\pi_3(x)}{\pi_1(x) + \pi_2(x)}\right) = \alpha_2 + \beta x, \\
&\ldots \\
&\log\left(\frac{\pi_J(x)}{\pi_1(x) + \cdots + \pi_{J-1}(x)}\right) = \alpha_{J-1} + \beta x.
\end{aligned}
$$