Week 10: Advanced Discrete Data

Multinomial Choice, Count Data

Max H. Farrell
The University of Chicago Booth School of Business
Discrete Responses

Today we will look at discrete responses.

In Week 4 we covered binary: $Y = 0$ or $1$.

Two extensions today:

- More categories: $Y = 0, 1, 2, 3, 4$
  - Unordered: buy product A, B, C, D, or nothing
  - Ordered: rate 1–5 stars
- Count: $Y = 0, 1, 2, 3, 4, \ldots$
  - How many products bought in a month?

The goal varies depending on the type of response.

At heart, we will keep using linear models.
First, more categories:

\[ Y = 0, 1, 2, \ldots, M. \]

These categories can be:

- Unordered: Buy product A, B, C, or nothing
- Ordinal: Low, medium, high risk
- Cardinal: Rate 1-5 stars

Predict choices/actions based on characteristics... then do classification into multiple categories.

Our main tool will be multinomial logistic regression, but there are others. For count data, Poisson regression may be better.
Discrete Choice

Units (firms, customers, etc) face a set of choices, labeled \( \{0, 1, 2, \ldots, M\} \). Or we want to classify units into one of these categories.

Choose based on utility (happiness/profit/etc)

- Utility unit \( i \) gets from option \( m \): \( U_{i,m} = X_{i,m} \beta_m + \varepsilon_{i,m} \).
- Chooses \( m^* \) if \( U_{i,m^*} \) is the biggest.
- So we want to model probabilities like \( P[U_{i,m^*} > U_{i,m}] \)

Gives rise to a set of log-odds, one for each category:

\[
\log \left( \frac{P[Y_i = m \mid X_i]}{P[Y_i = 0 \mid X_i]} \right) = \beta_{0,m} + \beta_{1,m} X_i.
\]

Just like regular (binary) logistic regression.

⇒ Interpretation is just as easy (hard?) as before.
Words of **warning**!

As usual, we’re skipping a lot of statistical details. But now, we’re skipping economic/substantive stuff too:

- ▶ What are we assuming about the agents behavior?
- ▶ What are the random shocks $\varepsilon_{i,j}$? Are the $X$ variables specific to the product (e.g. price) or the person (e.g. income)? Are the choices structured (e.g. ordered)?
- ▶ Should the intercepts and/or coefficients vary over the alternatives $(\beta_{0,m}, \beta_{1,m})$ or not?

Depending on the situation, other models may be appropriate. You’ll hear about ordinal, conditional, nested . . . . Even harder to interpret!

- ▶ Multinomial logit is a good all-purpose choice, often just as good as more specialized models.
Remember: The more complex the model, the more the assumptions matter!

It’s worth talking about the big one here: The **Independence of Irrelevant Alternatives**

- If $m^*$ is chosen from $\{0, 1, 2, \ldots, M\}$, then $m^*$ is chosen from any subset of $\{0, 1, 2, \ldots, M\}$.

- The log-odds of $m$ vs. $0$ doesn’t depend on other options.

- So if we change/remove options from the choice set, the odds of choosing $m$ vs. $0$ do not change.

Does that seem realistic?

- Does it matter? Maybe not for classification.

What to do?

- For now, we’ll ignore the problem, but you can test if IIA holds, then run a different model if need be.
Example: Choice of cracker brand

```r
> names(cracker <- read.csv("cracker_choice.csv"))
> table(cracker$choice)
kleebler  nabisco  private  sunshine
     226     1792    1035     239
> cracker$choice <- relevel(cracker$choice, "private")
```

Perfect data for relationship or inference questions:

- How does my price relate to choice?
- How do others’ prices relate to choice?
- Does an ad help? Anyone’s ad?

But the $X$’s are about the product, not the person:

- Can’t figure out what customer type buys a which brand.
- What would prediction/classification mean here?
Descriptives and Visualizations
Not as simple as linear regression (no lines to graph) but still an important step.

Look at the price of one option when others are chosen

```r
> colMeans(cracker[cracker$choice!='sunshine',2:5])
        pr.private  pr.keebler   pr.nabisco pr.sunshine
68.011253  107.74  96.44
> colMeans(cracker[cracker$choice=='sunshine',2:5])
        pr.private  pr.keebler   pr.nabisco pr.sunshine
68.9011340  110.22  86.26
```

![Sunshine Price](chart_sunshine.png)

![Private Price](chart_private.png)
The price of all the options when **Sunshine** is chosen:
On to regression analysis

Word of warning:

- There are a few packages out there, and they all do things a little differently
- Especially different choices for common coefficients

The `mlogit` package works well and is very flexible

... but has some startup cost

- Decide which coefficients should be common
- Get the data in the right shape
- Get the formula correct
Different slopes:  \[
\log \left( \frac{P[Y_i = m \mid X_i]}{P[Y_i = 0 \mid X_i]} \right) = \beta_0 + \beta_{1,m} X_i
\]

\texttt{fit1 <- mlogit(choice ~ 1 | pr.private + pr.keebler + pr.nabisco + pr.sunshine - 1, data=mlogit.data(cracker, choice="choice", shape="wide"), reflevel="private")}

\texttt{summary(fit1)}

Common slopes:  \[
\log \left( \frac{P[Y_i = m \mid X_i]}{P[Y_i = 0 \mid X_i]} \right) = \beta_{0,m} + \beta_1 X_i
\]

\texttt{fit2 <- mlogit(choice ~ pr, data=mlogit.data(cracker, choice="choice", shape="wide", varying=c(2:5)), reflevel="private")}

\texttt{summary(fit2)}
Interpretation

Exponentiate coefficients for interpretation

Single-slope model is easy:

```r
> exp(coef(fit2))

  keebler:(intercept)  nabisco:(intercept)  sunshine:(intercept)
  0.9922845           7.1344553           0.5718713
      pr
  0.9661083
```

- If the price of any name brand goes up by $0.05, then the odds of buying go down by a factor of $0.966^5 = 0.84$.
- Or, the odds of buying the store brand increase by $1/0.84, \approx 20\%$
Interpretation

Exponentiate coefficients for interpretation

Different slopes model is also interesting:

\[ \exp(\text{matrix(coef(fit1)), ...}) \]

<table>
<thead>
<tr>
<th></th>
<th>pr.private</th>
<th>pr.keebler</th>
<th>pr.nabisco</th>
<th>pr.sunshine</th>
</tr>
</thead>
<tbody>
<tr>
<td>keebler</td>
<td>1.02</td>
<td>0.957</td>
<td>1.012</td>
<td>1.007</td>
</tr>
<tr>
<td>nabisco</td>
<td>1.02</td>
<td>1.020</td>
<td>0.973</td>
<td>0.999</td>
</tr>
<tr>
<td>sunshine</td>
<td>1.01</td>
<td>1.023</td>
<td>1.007</td>
<td>0.942</td>
</tr>
</tbody>
</table>

➤ Why is \( b > 1 \) for pr.private always?
➤ Own price coefficient < 1?
➤ What if Keebler lowers their price by $0.05?
  ➔ people \( 1/0.957^5 \approx 25\% \) more likely to buy Keebler
  ➔ Nabisco sales go down by \( 1/1.020^5 \approx 10\% \)
  ... versus the store brand
Question: Do advertisements increase sales?

To tackle this, we have to decide if coefficients should be common or not:

```r
> fit1.ad <- mlogit(choice ~ 1 | pr.private + pr.keebler + pr.nabisco +
+ pr.sunshine + ad.private + ad.keebler + ad.nabisco + ad.sunshine +
+ - 1, data=mlogit.data(cracker, choice="choice",
+ shape="wide"), reflevel="private")
```

```r
> fit2.ad <- mlogit(choice ~ ad | pr.private + pr.keebler + pr.nabisco +
+ pr.sunshine - 1, data=mlogit.data(cracker, choice="choice",
+ shape="wide", varying=c(6:9)), reflevel="private")
```

```r
> fit3.ad <- mlogit(choice ~ pr + ad, data=mlogit.data(cracker, choice="choice"
+ shape="wide", varying=c(2:9)), reflevel="private")
```

You have to narrow the question, and make assumptions.

▶ Do customers respond to ads for brands the same way?
▶ Does Keebler running an ad increases Nabisco sales?
▶ Do ads in general raise awareness of crackers?
Question: Do advertisements increase sales?

Try Model 2: common slope on ad, different price effects

```r
> exp(coef(fit2.ad)[1])
  adTRUE
1.129938
```

```r
> exp(matrix(coef(fit2.ad)[2:13], nrow=3, ......
                     pr.private pr.keebler pr.nabisco pr.sunshine
keebler     1.02   0.957   1.011   1.008
nabisco     1.02   1.019   0.974   0.999
sunshine    1.01   1.022   1.006   0.944
```

Ads do increase\(^1\) sales of name brand crackers!

- Raise awareness/desire overall?
- But why do name brands benefit more than store brand?
- Even store brand ads contribute to this.

\(^{1}\) I meant “are associated with”
**Question:** Do advertisements increase sales?

**Try Model 1:**

```r
> exp(matrix(coef(fit1.ad)[1:12], ....
   pr.private pr.keebler pr.nabisco pr.sunshine
keeble 1.02 0.961 1.004 1.008
nabisco 1.02 1.020 0.971 0.998
sunshine 1.01 1.015 1.012 0.941
> exp(matrix(coef(fit1.ad)[13:24], ....
   ad.privateTRUE ad.keeblerTRUE ad.nabiscoTRUE ad.sunshineTRUE
keeble 1.07 2.51 0.892 1.11
nabisco 1.09 2.17 1.250 1.05
sunshine 1.08 1.29 2.094 1.22
```

▶ How much more likely are people to buy **Keebler** (vs. store brand) if they run an **ad**?

→ 2.5× as likely! (OR multiplied by 2.51.)

▶ **Keebler** ad makes people **2.17** as likely to buy **Nabisco**??

▶ Others are more or less sensitive.

▶ Price effects are similar to before, reassuring pattern
We can do all the usual things

- Predictions
  
  ```r
  > mlogit.probs <- fitted(fit1.ad, type="probabilities")
  > dim(mlogit.probs)
  [1] 3292 4
  ```

- Classification
  
  - May not make sense in some discrete choice applications, but useful in general
  
  - Cut-off rule problem:
    
    what if two categories have $P[Y = m | X] > c$?
  
  - One simple idea: highest predicted probability:
    
    ```r
    > colnames(mlogit.probs)[apply(mlogit.probs,1,which.max)]
    ```
Count Data

Poisson regression is a GLM designed for count data. How?

Remember the Poisson distribution:

- Integers only
- One parameter: \( \lambda > 0 \)
- \( \lambda = \text{Mean and Variance} \)

Extensions: Negative Binomial Model, Zero-Inflated Models
Example: Number of take-over bids (NUMBIDS) received by 126 US firms during 1978–85.

Covariates include

- **Continuous variables:**
  - BIDPREM, INSTHOLD, SIZE
- **Indicators:**
  - LEGLREST, FINREST, REGULATN, WHTKNGHT

Looks like Poisson!
Browsing the data

▶ More bids on average for the indicators being “Yes”:

<table>
<thead>
<tr>
<th>Indicator</th>
<th>No</th>
<th>1.46</th>
<th>1.69</th>
<th>1.70</th>
<th>1.18</th>
<th>1.67</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>2.11</td>
<td>1.96</td>
<td>2.08</td>
<td>2.12</td>
<td>1.91</td>
</tr>
</tbody>
</table>

▶ Pattern in the continuous variables?
Poisson regression results:

\[
\text{poisson.fit <- glm(NUMBIDS ~ BIDPREM+INSTHOLD+SIZE+LEGLREST + +REALREST+FINREST+ WHTKNGHT+REGULATN, family="poisson")}
\]

\[
\text{summary(poisson.fit)}
\]

Coefficients:

|                     | Estimate | Std. Error | z value | Pr(>|z|) |
|---------------------|----------|------------|---------|----------|
| (Intercept)         | 1.10489  | 0.53443    | 2.067   | 0.03870  *|
| BIDPREM             | -0.76982 | 0.37676    | -2.043  | 0.04103  *|
| INSTHOLD            | -0.13180 | 0.40027    | -0.329  | 0.74195  |
| SIZE                | 0.03428  | 0.01865    | 1.839   | 0.06598  |
| LEGLREST            | 0.25191  | 0.15005    | 1.679   | 0.09318  |
| REALREST            | -0.02024 | 0.17726    | -0.114  | 0.90909  |
| FINREST             | 0.09190  | 0.21688    | 0.424   | 0.67176  |
| WHTKNGHT            | 0.49821  | 0.15822    | 3.149   | 0.00164  **|
| REGULATN            | 0.01178  | 0.16297    | 0.072   | 0.94237  |

⇒ The expected log(NUMBIDS) increases by 0.498 \approx 1/2 if a White Knight is involved. What?
Just like logistic regression, exponentiating aids interpretation.

\[ \exp(\text{coef(poisson.fit)}) \]

(Intercept)  BIDPREM  INSTHOLD  SIZE  LEGLREST  REALREST  FINREST  WHTKNGHT  REGULATN
3.0188810  0.4630954  0.8765198  1.0348739  1.2864796  0.9799633  1.0962533  1.6457787  1.0118518

⇒ The incidence rate is 1.65 times higher if a White Knight is involved.

The incidence rate ratio (IRR, or risk ratio) is almost like the odds ratio from logistic regression: \( \beta_j > 0 \Leftrightarrow \text{IRR} > 1 \).

▶ The number of bids (the count we’re modeling) is expected to be 65% higher if a White Knight is involved.

▶ For a one-unit increase in the % held by institutions, number of bids falls by \( 1 - 0.88 = 12\% \).
The End

Whew! We made it!

Thanks for all your hard work. I know it’s been tough.

I hope everyone got something (a lot?) out of it.

▶ Good luck!
▶ Keep working on projects.
▶ Feedback please!
What we did

W1: Introduction, The simple linear regression (SLR) model
W2: Inference for SLR
W3: Multiple linear regression (MLR)
W4: Regression diagnostics and data transformations
W5: MLR: Causal inference, model building I
W6: Midterm
W7: MLR: model building II, data mining
W8: Generalized Linear Models, classification
W9: Time series models and autoregression
W10: Advanced topics in dependence (VAR, Clustering, Panels)
W11: Final exam and Project due
Course Feedback Hall of Fame

▶ “One weakness is that you assume your students know anything.”
▶ “Either a final or a project. Please. For the love of god.”
▶ “try to make class more interesting.”
▶ “less amount of work”
▶ “propensity to be too reductionist or too detailed”
▶ “shorten the sentences”