Models of Discrete Choice

Jonathan Wand

Polisci 350C
Stanford University

April 12, 2006

Likelihood for a dichotomous choice

Standard parameterizations for k-vector of covariates x_i,

1. Probit:
   1.1 $a(x, \beta) = x\beta$
   1.2 $F(x\beta) = \int_{-\infty}^{x\beta} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{z^2}{2}\} = \int_{-\infty}^{x\beta} \phi(z) = \Phi(x\beta)$
   1.3 $L_i = y \log(\Phi(x_i\beta)) + 1 - y \log(1 - \Phi(x_i\beta))$
2. Logit
   2.1 $a(x, \gamma) = x\gamma$
   2.2 $F(x\gamma) = \frac{1}{1+\exp\{-x\gamma\}} = \Lambda(x\gamma)$
   2.3 $L_i = y \log(\Lambda(x_i\gamma)) + 1 - y \log(1 - \Lambda(x_i\gamma))$

Particularly important concepts from last lecture

1. How to derive binary choice from models using normal and extreme value distributions
2. How to interpret the meaning of the coefficients: what is identified by reduced form parameters?
3. The (log)likelihood and score of a Bernoulli process
4. How to parametrize a logit and probit (log) likelihood, and derivation of score
5. How the logit and probit differ (i) in motivation (2) in tail behavior

Interpretation of parametric model

How should we interpret β?

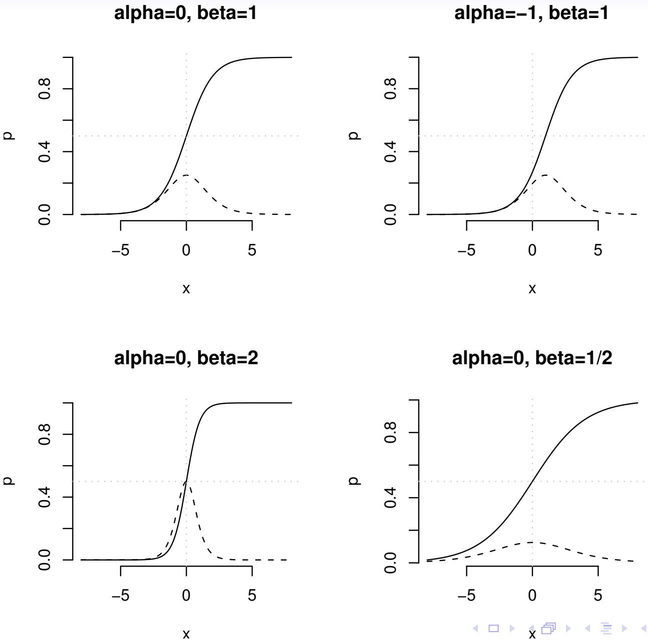
In OLS, straightforward interpretation

$$\frac{\partial y}{\partial x} = \frac{\partial x\beta}{\partial x} = \beta$$

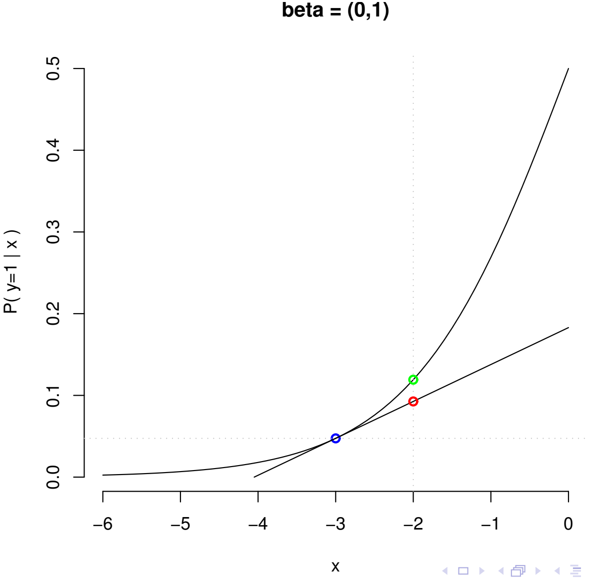
However,

$$\frac{\partial P(y=1|x)}{\partial x} = \frac{\partial F(x\beta)}{\partial x} = f(x\beta)\beta$$

is more complicated; effect of β will depend on x_i


alpha=0, beta=1
alpha=−1, beta=1
alpha=0, beta=2
alpha=0, beta=1/2
p
x
0.0
0.4
0.8
−5
0
5

Partial versus finite differences



Londregan, Bienen, and van de Walle (1995)

Londregan, Bienen; and van de Walle. 1995. Ethnicity and Leadership Succession in Africa. *ISQ*.

Question: What is the effect of leader's own ethnic population share on non-constitutional replacement of leader?

Shows larger share increase probability of non-constitutional replacement (but often replaced from within own ethnic group).

Key measure is (E)thnic (S)ize (D)ominance.

Thus we create a measure which adjusts the ethnic share of the leader's group for the degree of diffusion among the country's ethnic groups. We call this measure "ESD_1" (ethnic size dominance). This measure accounts for both size and dispersion of ethnic groups, and is derived from what is called a Herfindahl index. (See Herfindahl, 1950; Stigler, 1968; Hart, 1971.) Our ESD_1 measure for leader L is defined as follows:

$$ESD_1 = \frac{S_L}{\sqrt{S_1^2 + S_2^2 + \ldots + S_N^2}}$$

Case study: Londregan et al

TABLE 2. Descriptive Statistics

Variable	*Mean*	*Std. Dev.*	*Min.*	*Median*	*Maximum*	*% Within*
Annual growth rate	0.01	0.07	−0.40	0.01	0.44	92.27
ln(income)	6.76	0.56	5.40	6.79	8.40	10.28
Openness to trade	0.34	0.22	0.04	0.28	1.42	28.33
Leader's ethnic share	0.36	0.27	0.01	0.29	0.99	10.21
ESD_1	0.59	0.30	0.02	0.63	0.99	27.42
ESD_2	0.29	0.29	0.00	0.18	0.99	8.43
Political exit	0.08	0.27	0.00	0.00	1.00	91.94
Nonconstitutional exit	0.07	0.25	0.00	0.00	1.00	91.63
Nonconstitutional entrant	0.40	0.49	0.00	0.00	1.00	40.82
Inter-ethnic leadership transition	0.05	0.21	0.00	0.00	1.00	93.40

Case study: Londregan et al

Table 3. Sample Correlations

	Log of Lagged Income	*Openness to Trade*	*Ethnic Herf. Index*	*Leader's Ethnic Share*	*ESD_1*	*ESD_2*	*Political Exit*
Income growth rate	–0.06	–0.01	–0.06	0.10	0.06	0.11	–0.14
	(–1.64)[a]	(–0.19)	(–1.53)	(2.99)	(1.67)	(3.19)	(–3.87)
Log of lagged income		0.37	–0.20	0.26	0.22	0.27	0.00[b]
		(13.59)	(–5.21)	(8.52)	(7.10)	(8.99)	(–0.12)
Openness to trade			–0.02	0.11	0.06	0.14	–0.08
			(–0.44)	(3.34)	(1.82)	(4.31)	(–2.27)
Ethnic Herfindahl index				–0.72	–0.43	–0.70	0.02
				(–15.76)	(–10.44)	(–15.47)	(0.45)
Leader's ethnic share					0.89	0.99	–0.05
					(78.70)	(372.52)	(–1.37)
ESD_1						0.88	–0.03
						(71.43)	(–0.71)
ESD_2							–0.05
							(–1.33)

[a]T–ratios in parentheses.

[b]The estimated correlation between the log of lagged real per capita income and our political exit variable is –0.004, which is 0.00 to two decimal places.

Case study: Londregan et al

TABLE 5. Ethnicity and Nonconstitutional Succession

Dependent Variable: Nonconstitutional Exit				
Variable	*(1)*	*(2)*	*(3)*	*(4)*
Previous year's log of per capita income	−2.47	−2.53	−3.21	−3.73
	(1.15)[a]	(4.46)	(1.28)	(1.49)
Nonconstitutional ruler	2.95	2.41	3.10	2.88
	(1.13)	(1.37)	(1.24)	(1.29)
ln(leader's ethnic share)			5.46	
			(3.17)	
ln(leader's ethnic share)^2			1.70	
			(1.02)	
ln(ESD_1)				8.70
				(4.06)
ln(ESD_1)^2				5.37
				(2.69)
Log likelihood function[b]	−22.24	−6.61	−19.79	−18.02
Sample size	67	67	67	67

[a]Standard errors in parentheses.

[b]In column (2) this is the log of the conditional likelihood function corresponding to the conditional logit, and is not comparable with the logit likelihoods reported in columns (1), (3), and (4).

Case study: Londregan et al

The net impact of our ESD_1 variable is a weighted average of the coefficients of the log of ESD_1 and its square. More precisely, the estimated impact is given by:

$$\frac{p'(\underline{x}'\underline{\hat{\beta}})\left\{\hat{\beta}_1 + 2\hat{\beta}_2 \ln(ESD_1)\right\}}{ESD_1}$$

Estimates of this effect are shown in Figure 2. These estimates are calculated for a probability of nonconstitutional transition at the mean for the subsample of exit observations: 0.82. There we see that for low values of ESD_1 the estimated impact of an increase in ESD_1 on the probability a leadership transition is nonconstitutional is negative. However, these estimated negative effects are statistically insignificant, as indicated by the wide confidence bands, which encompass 0. At higher levels of ESD_1, the effect reverses: for values above 0.57, increases in ESD_1 significantly increase the probability that leadership transitions take place by nonconstitutional means. This critical value is just below the sample median of 0.63, so that, for just over half of our sample, the effects of increasing the relative size of the leader's ethnic group are directly counter to the prediction of the hypothesis

Case study: Londregan et al

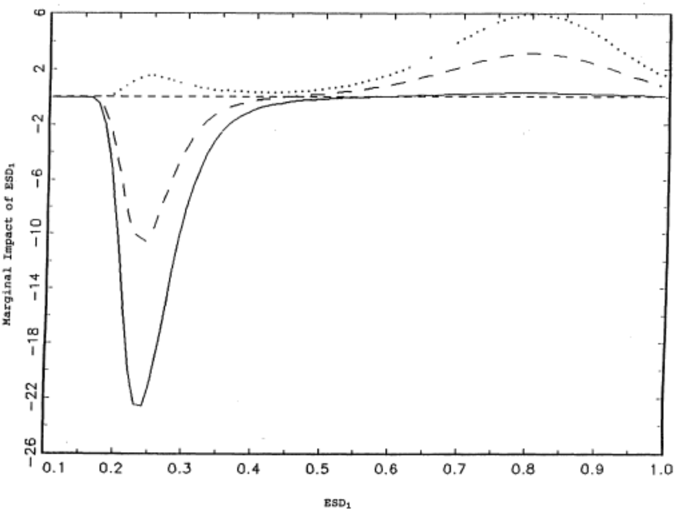


FIG. 2. ESD_1 vs. its marginal impact on the probability a leadership transition takes place nonconstitutionally.

Case study: Fearon and Laitin (2003)

Fearon, James D. and David D. Laitin. 2003. Ethnicity, nsurgency, and Civil War. *American Political Science Review*

Question: Is ethnic diversity associated with risk of civil war?

Case study: Fearon and Laitin

TABLE 1. Logit Analyses of Determinants of Civil War Onset, 1945–99

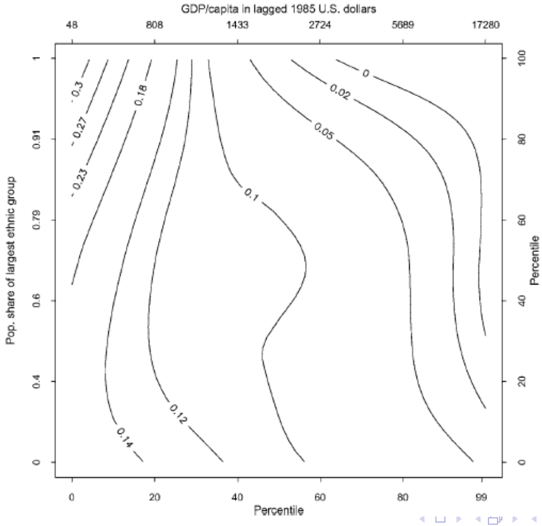
	Model				
	(1) Civil War	(2) "Ethnic" War	(3) Civil War	(4) Civil War (Plus Empires)	(5) Civil War (COW)
Prior war	−0.954**	−0.849*	−0.916**	−0.688**	−0.551
	(0.314)	(0.388)	(0.312)	(0.264)	(0.374)
Per capita income[a,b]	−0.344***	−0.379***	−0.318***	−0.305***	−0.309***
	(0.072)	(0.100)	(0.071)	(0.063)	(0.079)
log(population)[a,b]	0.263***	0.389***	0.272***	0.267***	0.223**
	(0.073)	(0.110)	(0.074)	(0.069)	(0.079)
log(% mountainous)	0.219**	0.120	0.199*	0.192*	0.418***
	(0.085)	(0.106)	(0.085)	(0.082)	(0.103)
Noncontiguous state	0.443	0.481	0.426	0.798**	−0.171
	(0.274)	(0.398)	(0.272)	(0.241)	(0.328)
Oil exporter	0.858**	0.809*	0.751**	0.548*	1.269***
	(0.279)	(0.352)	(0.278)	(0.262)	(0.297)
New state	1.709***	1.777***	1.658***	1.523***	1.147**
	(0.339)	(0.415)	(0.342)	(0.332)	(0.413)
Instability[a]	0.618**	0.385	0.513*	0.548*	0.584*
	(0.235)	(0.316)	(0.242)	(0.225)	(0.268)
Democracy[a,c]	0.021	0.013			
	(0.017)	(0.022)			
Ethnic fractionalization	0.166	0.146	0.164	0.490	−0.119
	(0.373)	(0.584)	(0.368)	(0.345)	(0.396)
Religious fractionalization	0.285	1.533*	0.326		1.176*
	(0.509)	(0.724)	(0.506)		(0.563)
Anocracy[a]			0.521*		0.597*
			(0.237)		(0.261)
Democracy[a,d]			0.127		0.219
			(0.304)		(0.354)
Constant	−6.731***	−8.450***	−7.019***	−6.801***	−7.503***
	(0.736)	(1.092)	(0.751)	(0.681)	(0.854)
N	6327	5186	6327	6360	5378

Note: The dependent variable is coded "1" for country years in which a civil war began and "0" in all others. Standard errors are in parentheses. Estimations performed using Stata 7.0. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

[a] Lagged one year.
[b] In 1000's.
[c] Polity IV; varies from −10 to 10.
[d] Dichotomous.

Case study: Fearon and Laitin

FIGURE 2. Probability of Civil War Onset per Five-Year Period



Interpretation of parametric model

Partial change

1. Case study (slope at a given point x_i^*)

$$\delta_1 = \frac{\partial P(y = 1 \mid x = x^*)}{\partial x} = F'(a(x^*, \beta))\frac{\partial a(x^*, \beta)}{\partial x}$$

   For Logit with $a(x, \beta)$, we have $\Lambda(x^*\beta)[1 - \Lambda(x^*\beta)]\beta$.

2. Average/expected value of case study derivative by integrating over values of x,

$$\delta_2 = \int g(x^*)F'(a(x^*, \beta))\frac{\partial a(x^*, \beta)}{\partial x} = \frac{1}{N}\sum \delta_1(x)$$

   where g() is the density of x.

Interpretation of parametric model

Discrete difference

Partition $x = (x_1, x_2)$, e.g., $x_1 \in \{0, 1\}$.

1. Case study, change between two sets of values, fixed x_2.

$$\begin{aligned}\delta_4(x_1) &= P(y = 1 \mid x = \{1, x_2\}) - P(y = 1 \mid x = \{0, x_2\}) \\ &= F[a(\{1, x^2\})] - F[a(\{0, x^2\})]\end{aligned}$$

2. Average, aggregate finite difference with respect to distribution of x_2

$$\begin{aligned}\delta_5 &= \int_{x_2} \{P(y = 1 \mid x = \{1, x_2\}) - P(y = 1 \mid x = \{0, x_2\})\} \\ &= \frac{1}{N} \sum \delta_4(x_1)\end{aligned}$$

Interpretation of parametric model

Comments

- Choose substantively interesting case studies
- Don't use average of a variable where average does not exist as observed value (e.g,. dichotomous explanatory variables)
- Asking "What happens on average?" and "What happens in the middle?" are not he same
- Partial differences may take you out of range of (0,1)
- Reflect on assumptions that you use when setting fixed other variables (case study) or integrating over values (expected values)—consider covariance between variable of interest and other variables.

Generalized Choice: multinomial

Consider the vote choice of Australians in recent Parliamentary elections:

Individuals at election time are faced with a number of choices, including

1. vote Australian Labor Party
2. vote Liberal Party
3. vote Australian Greens
4. vote National Party

NOTE: voting is mandatory (plausible to ignore abstention as choice).

Generalized Choice: multinomial

Let's just focus on choice between three parties for ease,

1. party 1 utility $u_1 = \mu_1 + \epsilon_1$
2. party 2 utility $u_2 = \mu_2 + \epsilon_2$
3. party 3 utility $u_3 = \mu_3 + \epsilon_3$

and let $y \in \{1, 2, 3\}$ be an index of the selected choice.

Generalized Choice: multinomial

So we have three decision rules,

1. vote party 1 , $y = 1$, if $u_1 > u_2$ and $u_1 > u_3$
2. vote party 2 , $y = 2$, if $u_2 > u_1$ and $u_2 > u_3$
3. vote party 3 , $y = 3$, if $u_3 > u_2$ and $u_3 > u_1$

... and three probabilities

1. prob of voting for party 1, $P(y = 1) = P(u_1 > u_2 \ \& \ u_1 > u_3)$
2. prob of voting for party 2, $P(y = 2) = P(u_2 > u_1 \ \& \ u_2 > u_3)$
3. prob of voting for party 3, $P(y = 3) = P(u_3 > u_2 \ \& \ u_3 > u_1)$

Generalized Choice: multinomial

$$
\begin{aligned}
P(y=1) &= \int_{-\infty}^{\infty} \lambda(\epsilon_1) \int_{-\infty}^{\mu_1+\epsilon_1-\mu_2} \lambda(\epsilon_2) \int_{-\infty}^{\mu_1+\epsilon_1-\mu_3} \lambda(\epsilon_3) \partial\epsilon_1 \partial\epsilon_2 \partial\epsilon_3 \\
&= \int_{-\infty}^{\infty} \lambda(\epsilon_1) \Lambda(\mu_1+\epsilon_1-\mu_2) \Lambda(\mu_1+\epsilon_1-\mu_3) \partial\epsilon_1 \\
&= \int_{-\infty}^{\infty} e^{-\epsilon_1} e^{-e^{-\epsilon_1}} e^{-e^{-(\mu_1+\epsilon_1-\mu_2)}} e^{-e^{-(\mu_1+\epsilon_1-\mu_3)}} \partial\epsilon_1 \\
&= \int_{-\infty}^{\infty} e^{-\epsilon_1} e^{-e^{-\epsilon_1}} e^{-e^{-\epsilon_1} e^{\mu_2-\mu_1}} e^{-e^{-\epsilon_1} e^{\mu_3-\mu_1}} \partial\epsilon_1
\end{aligned}
$$

Generalized Choice: multinomial

$$P(y=1) = \int_{-\infty}^{\infty} e^{-\epsilon_1} e^{-e^{-\epsilon_1}} e^{-e^{-\epsilon_1}(1+e^{\mu_2-\mu_1}+e^{\mu_3-\mu_1})} \partial\epsilon_1$$

$$= \frac{1+e^{\mu_2-\mu_1}+e^{\mu_3-\mu_1}}{1+e^{\mu_2-\mu_1}+e^{\mu_3-\mu_1}} \int_{-\infty}^{\infty} e^{-\epsilon_1} e^{-e^{-\epsilon_1}(1+e^{\mu_2-\mu_1}+e^{\mu_3-\mu_1})} \partial\epsilon_1$$

$$= \frac{1}{1+e^{\mu_2-\mu_1}+e^{\mu_3-\mu_1}}$$

$$P(y=2) = \frac{1}{1+e^{\mu_1-\mu_2}+e^{\mu_3-\mu_2}}$$

$$P(y=3) = \frac{1}{1+e^{\mu_1-\mu_3}+e^{\mu_2-\mu_3}}$$

Generalized Choice: multinomial

Since,

$$1 = P(y=1) + P(y=2) + P(y=3)$$

by rearrangement,

$$P(y=3) = 1 - P(y=1) - P(y=2)$$

which can also be seen by,

$$\frac{e^{\mu_3}}{e^{\mu_1} + e^{\mu_2} + e^{\mu_3}} = 1 - \frac{e^{\mu_1}}{e^{\mu_1} + e^{\mu_2} + e^{\mu_3}} - \frac{e^{\mu_2}}{e^{\mu_1} + e^{\mu_2} + e^{\mu_3}}$$

What can be identified? What if $\mu_i^* = \mu_i + \eta$ for all i?

Generalized Choice: multinomial

If $\mu_i^* = \mu_i + \eta$ for all i, same probability: MNL only identifies differences in utilities:

$$
\begin{aligned}
P(y=1) &= \frac{1}{1+e^{\mu_2-\mu_1}+e^{\mu_3-\mu_1}} = \frac{1}{1+e^{\mu_2^*-\mu_1^*}+e^{\mu_3^*-\mu_1^*}} \\
P(y=2) &= \frac{1}{1+e^{\mu_1-\mu_2}+e^{\mu_3-\mu_2}} = \frac{1}{1+e^{\mu_1^*-\mu_2^*}+e^{\mu_3^*-\mu_2^*}} \\
P(y=3) &= \frac{1}{1+e^{\mu_1-\mu_3}+e^{\mu_2-\mu_3}} = \frac{1}{1+e^{\mu_1^*-\mu_3^*}+e^{\mu_2^*-\mu_3^*}}
\end{aligned}
$$

Identification acheived by setting one utility to a constant.

Generalized Choice: multinomial

Handy to use zero as the constant, since $e^0 = 1$. Consider $\mu_j = x\beta_j$, then set $\beta_j = 0$ for a single category j.

$$
\begin{aligned}
P(y=1) &= \frac{1}{1+e^{x\beta_2}+e^{x\beta_3}} \\
P(y=2) &= \frac{e^{x\beta_1}}{1+e^{x\beta_2}+e^{x\beta_3}} \\
P(y=3) &= \frac{e^{x\beta_1}}{1+e^{x\beta_2}+e^{x\beta_3}} = 1 - P(y=1) - P(y=2)
\end{aligned}
$$

Axiomatic Foundations of Choice Models

Luce Lemma 3 (Independence from irrelevant alternatives):
For $x, y \in S$,

$$\frac{P(x,y)}{P(y,x)} = \frac{P_S(x)}{P_S(y)}$$

Proof:
By Axiom we have

$$P_S(x) = P(x,y)[P_S(x) + P_S(y)]$$

So

$$\begin{aligned} P_S(x) &= P(x,y)[P_S(x) + P_S(y)] \\ P_S(x) &= P(x,y)P_S(x) + P(x,y)P_S(y) \\ (1 - P(x,y))P_S(x) &= P(x,y)P_S(y) \\ P(y,x)P_S(x) &= P(x,y)P_S(y) \\ \frac{P(x,y)}{P(y,x)} &= \frac{P_S(x)}{P_S(y)} \end{aligned}$$

Axiomatic Foundations of Choice Models

Note that,

$$
\begin{aligned}
P(y=1) &= \frac{1}{1+e^{\mu_2-\mu_1}+e^{\mu_3-\mu_1}} = \frac{e^{\mu_1}}{e^{\mu_1}+e^{\mu_2}+e^{\mu_3}} \\
P(y=2) &= \frac{1}{1+e^{\mu_1-\mu_2}+e^{\mu_3-\mu_2}} = \frac{e^{\mu_2}}{e^{\mu_1}+e^{\mu_2}+e^{\mu_3}} \\
P(y=3) &= \frac{1}{1+e^{\mu_1-\mu_3}+e^{\mu_2-\mu_3}} = \frac{e^{\mu_3}}{e^{\mu_1}+e^{\mu_2}+e^{\mu_3}}
\end{aligned}
$$

So,

$$
\frac{P(y=1)}{P(y=2)} = \frac{e^{\mu_1}}{e^{\mu_2}}
$$

Axiomatic Foundations of Choice Models

And recall from logit,

$$\begin{aligned} P(y=1) &= \frac{1}{1+e^{\mu_2-\mu_1}} = \frac{e^{\mu_1}}{e^{\mu_1}+e^{\mu_2}} \\ P(y=2) &= \frac{1}{1+e^{\mu_1-\mu_2}} = \frac{e^{\mu_2}}{e^{\mu_2}+e^{\mu_2}} \end{aligned}$$

So,

$$\frac{P(y=1)}{P(y=2)} = \frac{e^{\mu_1}}{e^{\mu_2}}$$

Axiomatic Foundations of Choice Models

Putting this together, let $S \in \{1, 2, 3\}$, then

$$\frac{P(1,2)}{P(2,1)} = \frac{e^{\mu_1}}{e^{\mu_2}} = \frac{P_S(1)}{P_S(2)}$$

MNL conforms to Choice Axiom/IIA.

See Yellot (1977) and McFadden (1973) for details of these and related results.

Axiomatic Foundations of Choice Models

Luce Lemma 3 (Independence from irrelevant alternatives):

- Most famous implication: restriction that the relative probability of choosing two alternatives is invariant to the composition of the larger set of alternatives.
- Again, only ratio that is invariant not probabilities themselves
- Might also hear that log-odds of two choices are constant: $\log(P_S(x)) - \log(P_S(y)) = c$.
- Neither holds in general for models of choice (by design) nor is it plausible that it in general holds empirically.
- Benefit: can estimate parameters defining utility of choices even with only a subset. Not generally possible if IIA does not hold (e.g., correlation between utilities of choices)—then to estimate any choice must model all choices.

Generalized Choice: Absention due to indifference

Consider the vote choice in many parts of the US:
Individuals at election time are faced with three choices:

1. vote Democratic
2. vote Republican
3. or Abstain

Unlike other models of this fundamental choice process, Sanders (1999) builds on a spatial theory of voting which posits that abstention is result of indifference between parties.

Note: For easy introduction to this and more theory relating to spatial model choices, see Munger and Hinich, *Analytical Politics*.

Generalized Choice: Absention due to indifference

For some $T \geq 0$,

1. vote D if $U_D - U_R > T$
2. vote R if $U_R - U_D > T$
3. abstain if $-T < U_D - U_R < T$

Same general framework as for logit or multinomial logit,

$$
\begin{aligned}
P(D) &= P[(\mu_D + \epsilon_D) - (\mu_R + \epsilon_R) > T] \\
&= P[(\mu_D + \epsilon_D) - \mu_R - T > \epsilon_R] \\
&= \int_{-\infty}^{\infty} \lambda(\epsilon_D) \int^{\mu_D + \epsilon_D - \mu_R - T} \lambda(\epsilon_R) \partial\epsilon_R \partial\epsilon_D \\
&= \frac{1}{1 + \exp\{-(\mu_D - \mu_R)\}e^T}
\end{aligned}
$$

Similarly, we could calculate $P(R)$.

Generalized Choice: Absention due to indifference

Given $P(D)$ and $P(R)$, $P(A)$ what is left over,

$$
\begin{aligned}
P(D) &= \frac{1}{1+\exp\{-(\mu_D - \mu_R)\}e^T} \\
P(R) &= \frac{1}{1+\exp\{-(\mu_R - \mu_D)\}e^T} \\
P(A) &= 1 - P(D) - P(R)
\end{aligned}
$$

Additional extensions: Could also incorporate upper bound on absolute distance (for alienation).

Particularly important ideas from this lecture

1. How to interpret/present the effect of (reduced form) coefficients in a logit or probit model
2. Approach for deriving generalizations of choice models based on discriminal process
3. IIA concept and implications

Fearon, James D. and David D. Laitin. 2003. Ethnicity, Insurgency, and Civil War. *American Political Science Review*, 97(1):75–90.

Londregan, John, Henry Bienen, and Nicolas van de Walle. 1995. Ethnicity and Leadership Succession in Africa. *International Studies Quarterly*, 39:1–25.

McFadden, Daniel. 1973. Conditional logit analysis of qualitative choice behavior. In P Zarembka, editor, *Frontiers of Econometrics*, New York: Acadmic Press. pages 105–42.

Sanders, Mitchell S. 1999. Unified models of turnout and vote choice for two-candidate and three candidate elections. In Walter R. Mebane Jr., editor, *Political Analysis*, Ann Arbor: University of Michigan Press, volume 7. pages 89–116.

Yellot, John I. Jr. 1977. The Relationship between Luce's Axiom, Thurstone's Theory of Comparative Judgment, and the Double Exponential Distribution. *Journal of Mathematical Psychology*, 15:109–144.