Statistical Methods III: Spring 2013

Jonathan Wand

Stanford University

Choice models: generalize

Outline

Generalized Choice: multinomial, example

Consider the vote choice of Australians in recent Parliamentary elections:

Individuals at election time faced with 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, 3 choices

Let $y \in \{1, 2, 3\}$ index of choice.
where assume choice is between three parties,

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

Generalized Choice: multinomial, 3 choices

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, $P(y = 3) = P(u_3 > u_2 \ \& \ u_3 > u_1)$

Thurstone: discriminal process

With two choices, and ϵ iid Gumbel, then

$$\begin{aligned}
P(j,k) &= P(u_j > u_k) \\
&= P(\mu_j - \mu_k + \epsilon_j > \epsilon_k) \\
&= \int_{-\infty}^{\infty} \lambda(\epsilon_j) \int_{-\infty}^{\mu_j - \mu_k + \epsilon_j} \lambda(\epsilon_k) \\
&= \int_{-\infty}^{\infty} \lambda(\epsilon_j) \Lambda(\mu_j - \mu_k + \epsilon_j) \\
&= \int_{-\infty}^{\infty} \lambda(\epsilon_j) \Lambda(\mu_j - \mu_k + \epsilon_j) \\
&= \frac{1}{w} \int_{-\infty}^{\infty} -e^{-\epsilon_j} w \exp\{-e^{-\epsilon_j} w\} \\
&= \frac{1}{w} = \frac{1}{1 + \exp\{-(\mu_j - \mu_k)\}}
\end{aligned}$$

Generalized Choice: multinomial

So, how to calc prob of voting for party 1 with three choices?

$$
\begin{aligned}
P(y=1) &= P(u_1 > u_2 \text{ and } u_1 > u_3) \\
&= P(\mu_1 + \epsilon_1 > \mu_2 + \epsilon_2 \text{ and } \mu_1 + \epsilon_1 > \mu_3 + \epsilon_3) \\
&= P(\mu_1 + \epsilon_1 - \mu_2 > \epsilon_2 \text{ and } \mu_1 + \epsilon_1 - \mu_3 > \epsilon_3) \\
&=?P(\mu_1 + \epsilon_1 - \mu_2 > \epsilon_2)P(\mu_1 + \epsilon_1 - \mu_3 > \epsilon_3)
\end{aligned}
$$

no...

$$
\begin{aligned}
P(y=1) &= \int_{-\infty}^{\infty} \lambda(\epsilon_1) P(\mu_1 + \epsilon_1 - \mu_2 > \epsilon_2 \mid \epsilon_1) P(\mu_1 + \epsilon_1 - \mu_3 > \epsilon_3 \mid \epsilon_1) \\
&= \int_{-\infty}^{\infty} \lambda(\epsilon_1) \left[\int_{-\infty}^{\mu_1+\epsilon_1-\mu_2} \lambda(\epsilon_2)\partial\epsilon_2 \cdot \int_{-\infty}^{\mu_1+\epsilon_1-\mu_3} \lambda(\epsilon_3)\partial\epsilon_3 \right] \partial\epsilon_1
\end{aligned}
$$

Generalized Choice: multinomial

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

Generalized Choice: multinomial

$$
\begin{aligned}
P(y=1) &= \int_{-\infty}^{\infty} e^{-\epsilon_1} e^{-e^{-\epsilon_1}} e^{-e^{-\epsilon_1}(1+e^{\mu_2-\mu_1}+e^{\mu_3-\mu_1})} \partial\epsilon_1 \\
&= \frac{1+e^{\mu_2-\mu_1}+e^{\mu_3-\mu_1}}{1+e^{\mu_2-\mu_1}+e^{\mu_3-\mu_1}} \int_{-\infty}^{\infty} e^{-\epsilon_1} e^{-e^{-\epsilon_1}(1+e^{\mu_2-\mu_1}+e^{\mu_3-\mu_1})} \partial\epsilon_1 \\
&= \frac{1}{1+e^{\mu_2-\mu_1}+e^{\mu_3-\mu_1}} \\
P(y=2) &= \frac{1}{1+e^{\mu_1-\mu_2}+e^{\mu_3-\mu_2}} \\
P(y=3) &= \frac{1}{1+e^{\mu_1-\mu_3}+e^{\mu_2-\mu_3}}
\end{aligned}
$$

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?

Generalized Choice: multinomial

MNL only identifies differences in utilities (like logit...)

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

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.

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

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 (1998) 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).

Axiomatic Foundations of Choice Models

Assumptions

- D1 Let $R \subset S \subset T \subset U$.
- D2 Let $x, y, z \in T$, arbitrary elements of choice set.
- D3 Let $P(x, y)$ be the probability of choosing x instead of y, $0 < P(x, y) < 1$.
- D4 $P_S(R)$ is the probability of choosing R given choice from among alternatives in S.

Choice Axiom

- (i) $P_T(R) = P_S(R)P_T(S)$
- (ii) If $P(x, y) = 0$ for some $x, y \in T$, $P_T(S) = P_{T-\{x\}}(S - \{x\})$

Axiomatic Foundations of Choice Models

Axiom of Choice

- Defines relationship which defines how choices within subsets are related in the context of an individual making probabilistic choices.
- Can be rewritten as $P_T(R \mid S)P_T(S) = P_T(R)$
- Two core implications,
  Lemma 3: Independence of Irrelevant Alternatives (IIA)
  Theorem 3: Probability must satisfy a ratio scale

Axiomatic Foundations of Choice Models

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

Luce Lemma 3 (Independence from irrelevant alternatives):
What does this mean?

- relative probability of choosing two alternatives is invariant to the composition of the larger set of alternatives.
- Only ratio is invariant, not probabilities themselves
- Might also hear that log-odds of two choices are constant: $\log(P_S(x)) - \log(P_S(y)) = c$.

Axiomatic Foundations of Choice Models

Luce Lemma 3 (Independence from irrelevant alternatives):
Why so cool?

- 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.
- ** Neither holds in general for models of choice (by design) nor is it plausible that it in general holds empirically.

Axiomatic Foundations of Choice Models

Theorem 3: choice probability is ratio scale
$\exists v : T \to \Re_+$, unique up to multiplication by $k > 0$, such that

$$P_S(x) = \frac{v(x)}{\sum_{y \in S} v(y)} = \frac{1}{1 + \sum_{y \in S - \{x\}} v(y)/v(x)}$$

McFadden and Yellot have each pointed out the connection to logit models by setting $v(x) = e^x$.

E.g.,

$$P(x, y) = \frac{v(x)}{v(x) + v(y)} = \frac{e^{\mu_x}}{e^{\mu_x} + e^{\mu_y}} = \frac{1}{1 + e^{\mu_y}/e^{\mu_x}} = \frac{1}{1 + e^{-(\mu_x - \mu_y)}}$$

Yellot shows that discriminal process based on Type I discrete value distribution is uniquely equivalent to Choice Axiom.

Axiomatic Foundations of Choice Models

Let $S \in \{1, 2, 3\}$, and $P_S(j)$ be probability of choosing j from S,

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

So,

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

Axiomatic Foundations of Choice Models

Let $T \in \{1, 2\}$, and $P_T(j)$ be probability of choosing j from T,
Recal logit (special case of MNL),

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

So,

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

Axiomatic Foundations of Choice Models

Comparing probabilities of choosing 1 and 2 in logit and MNL,

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

MNL conforms to Choice Axiom/IIA.

See Yellot (1977) and McFadden (1973) for connections between Luce and EV Type 1.