Statistical Methods III: Spring 2013

Jonathan Wand

Stanford University

Choice models: generalize II

Outline

1. Philosophy
2. Models of choice: review
3. Axioms of choice and IIA

A bit of philosophy

1. How to learn
   - see one
   - do one
   - teach one
2. Arguments by authority
   - are contemptible as an intellectual stance
   - offer only reference for demonstration
   - of course, there is usually a layer of deeper understanding that is assumed
3. what is our goal? ability to
   - evaluate properties of statistical methods
     what is it we learn from the use of a model? why?
   - better map existing methods to application, and vice versa
     evaluate suitability of application
     choose better /proper methods
   - extend and develop statistical

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

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

Definitions

Definition (Likelihood)

$\mathcal{L}(\theta \mid y)$ is a *known* density evaluated as function parameter values θ given data y.

NOTE:

- Likelihood of n observations

$$\mathcal{L}(\theta \mid y) = \mathcal{L}(\theta \mid y_1, y_2, \ldots, y_n) = f(y_1, y_2, \ldots, y_n \mid \theta)$$

- if iid observations

$$\mathcal{L}(\theta \mid y_1, y_2, \ldots, y_n) = f(y_1 \mid \theta) f(y_2 \mid \theta) \cdots f(y_n \mid \theta)$$

- almost always easier to deal with natural log

$$L(\theta \mid y_1, y_2, \ldots y_n) = \sum_n \mathscr{ln} f(y_i \mid \theta)$$

Definitions

Definition (MLE)

Maximum Likelihood Estimate (MLE) $\hat{\theta}$

Type I: $L(\hat{\theta} \mid y) \geq L(\theta \mid y) \quad \text{for all} \quad \theta$

Type II: $S(\hat{\theta} \mid y) = 0$

Definition (Score)

$$S(\theta \mid y) = \frac{\partial}{\partial \theta} L(\theta \mid y_1, y_2, \ldots y_n)$$

Likelihood for a dichotomous choice

Assume each y_i is drawn independently.

What does independence imply for the joint probability of events?

Joint likelihood of n observations of choices,

$$\mathcal{L} = \prod_{i=1}^{n} \mathcal{L}_i = \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}$$

And, log-likelihood,

$$L = \sum_{i=1}^{n} L_i = \sum_{i=1}^{n} [y_i \log(p) + (1-y_i)\log(1-p)]$$

Likelihood for a dichotomous choice

- Homogeneous / same p for all draws

$$L = \sum_{i=1}^{n} L_i = \sum_{i=1}^{n} [y_i \log(p) + (1 - y_i) \log(1 - p)]$$

- allow p_i to differ by draws (i.e., by person)

$$L_i = y_i \log(p_i) + (1 - y_i) \log(1 - p_i)$$

How many parameters would this require for n people?

- reduce n unknowns to the k unkowns, with γ being k-vector

$$L_i = y_i \log(\Lambda(x_i\gamma)) + (1 - y_i) \log(1 - \Lambda(x_i\gamma))$$

Likelihood for a dichotomous choice

Standard parameterizations for k-vector of covariates x_i,

1. Probit:
   1. $a(x, \beta) = x\beta$
   2. $F(x\beta) = \int_{-\infty}^{x\beta} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{z^2}{2}\} = \int_{-\infty}^{x\beta} \phi(z) = \Phi(x\beta)$
   3. $L_i = y_i \log(\Phi(x_i\beta)) + (1 - y_i) \log(1 - \Phi(x_i\beta))$
2. Logit
   1. $a(x, \gamma) = x\gamma$
   2. $F(x\gamma) = \frac{1}{1+\exp\{-x\gamma\}} = \Lambda(x\gamma)$
   3. $L_i = y_i \log(\Lambda(x_i\gamma)) + (1 - y_i) \log(1 - \Lambda(x_i\gamma))$

Questions:

- what does mnl imply about relationship between choices
- how to interpret model? $dP(y|x)/dx$? $P(y|x=1) - P(y|x=0)$?
- what is in μ? endogeneity of choices?

Spatial model

Consider three parties in unidimensional model

- where should parties locate?
  depends on whether there is threat of entry or not? stackleberg equilibrium
  most stat models treat location as given
- what party should a party vote for?
- how should voter make choice?

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.