Advanced behavior models

Recent development of discrete choice models

Makoto Chikaraishi
Hiroshima University
Why advanced models are needed? A case of route choice

(a) Route overlap

(b) Route length

(c) Route enumeration
Closed-form and open-form

• **Closed-form expression**
  – A mathematical expression that can be evaluated in a finite number of operations

\[
P_{ij} = \frac{e^{\beta x_{ij}/\lambda_k} \left( \sum_{j \in B_k} e^{\beta x_{ij}/\lambda_l} \right)^{\lambda_k-1}}{\sum_{l=1}^{\lambda_k} \left( \sum_{j \in B_k} e^{\beta x_{ij}/\lambda_l} \right)^{\lambda_l}}
\]

• **Open-form expression**

\[
P_{ij} = \int_{\beta_i \in D_{\beta_i}} \frac{\exp(\beta_i x_{ij})}{\sum_{j' = 1}^{J} \exp(\beta_i x_{ij})} f(\beta_i) d\beta_i
\]
Pros and cons

• **Closed-form expression**
  – Pros
    • Easy to use in practice
    • Can be embedded into a larger modeling system as a subcomponent
  – Cons
    • Not flexible enough in some cases

• **Open-form expression**
  – Pros
    • Very flexible and any kind of closed-form models can be approximately modeled
  – Cons
    • Behavioral understanding of the model is sometimes difficult
1. McFadden’s G function (McFadden, 1978)
   - Route overlap

2. Generalized G function (Mattsson et al., 2014)
   - Route overlap and route length

3. Recursive logit (Fosgerau et al., 2013)
   - Route enumeration
Discrete choice models
[based on Hato (2002)]

Multinomial logit (MNL)
(Luce, 1959)

Nested logit (NL)
(Ben-Akiva, 1973)

Generalized extreme value (GEV)
(McFadden, 1978)

Generalized nested logit (GNL), recursive nested logit extreme value model (RNEV), network-GEV
(Wen & Koppelman, 2001; Daly, 2001; Bierlaire, 2002)

Paired combinational logit (PCL) (Chu, 1981)
Cross-nested logit (CNL) (Vovsha, 1997)

Generalization

Normal to Gumbel

Generalization

Multinomial Probit (MNP)
(Thurstone, 1927)

Generalization

Heteroscedastic/mixed distributions

Error component logit (ECL); Mixed logit (MXL); Kernel logit (KL);
Heteroscedastic logit (HL)
(Boyd and Mellman, 1980; Cardell and Dunbar, 1980; McFadden, 1989; Bhat, 1995; See Train (2009) for details)

Normal to Gumbel

Generalization

Weibull to GEV (not MEV)

Weibull to GEV (not MEV)

Generalization

Multinomial weibit (MNW)
(Castillo, et al., 2008)

q-generalized logit
(Nakayama, 2013, Nakayama and Chikaraishi, 2015)

Generalized G function
(Mattsson et al., 2014)

Variance stabilization
(Li, 2011)

Derived from McFadden’s G function or “choice probability generating functions” (Fosgerau et al., 2013)

Derived from the generalized G function
McFadden’s G function

The properties that the $G$ function must exhibit

1. $G(y_{i1}, y_{i2}, ..., y_{iJ_i}) \geq 0$

2. $G$ is homogeneous of degree $m$: $G(\alpha y_{i1}, ..., \alpha y_{iJ_i}) = \alpha^m G(y_{i1}, ..., y_{iJ_i})$

3. $\lim_{y_{ij} \to \infty} G(y_{i1}, y_{i2}, ..., y_{iJ_i}) = \infty$ for any $j$

4. The cross partial derivatives of $G$ satisfy:

$$(-1)^{k-1} \frac{\partial^k G(y_{i1}, y_{i2}, ..., y_{iJ_i})}{\partial y_{i1} \partial y_{i2} \cdots \partial y_{ik}} \geq 0$$

When all conditions are satisfied, the choice probability can be defined as:

$$P_{ij} = \frac{e^{V_{ij}} \cdot G_j(e^{V_{i1}}, e^{V_{i2}}, ..., e^{V_{iJ_i}})}{G(e^{V_{i1}}, e^{V_{i2}}, ..., e^{V_{iJ_i}})}$$

(where, $G_j = \partial G / \partial y_{ij}$)

Assumption: $F(\epsilon_{i1}, ..., \epsilon_{ij}) = \exp\{-G(e^{-\epsilon_{i1}}, ..., e^{-\epsilon_{ij}})\}$

$\star u_{ij} = V_{ij} + \epsilon_{ij}$
Derivation of choice probability

Suppose $u_{ij} = V_{ij} + \epsilon_{ij}$, where $(\epsilon_{i1}, ..., \epsilon_{ij})$ is distributed $F$ defined as:

$$F(\epsilon_{i1}, ..., \epsilon_{ij}) = \exp\{-G(e^{-\epsilon_{i1}}, ..., e^{-\epsilon_{ij}})\}$$

Then, the probability of the first alternative $P_{i1}$ satisfies:

$$P_{i1} = \int_{\epsilon=-\infty}^{+\infty} F_{1}(\epsilon, V_{i1} - V_{i2} + \epsilon, ..., V_{i1} - V_{iJ} + \epsilon) d\epsilon$$

$$= \int_{\epsilon=-\infty}^{+\infty} \left[ e^{-\epsilon} G_{1}(e^{-\epsilon}, e^{-\epsilon} - V_{i1} + V_{i2}, ..., e^{-\epsilon} - V_{i1} + V_{iJ}) \times \exp\{-G(e^{-\epsilon}, e^{-\epsilon} - V_{i1} + V_{i2}, ..., e^{-\epsilon} - V_{i1} + V_{iJ})\} \right] d\epsilon$$

$$= \int_{\epsilon=-\infty}^{+\infty} \left[ e^{-\epsilon} G_{1}(e^{V_{i1}}, e^{V_{i2}}, ..., e^{V_{ij}}) \times \exp\{-e^{-\epsilon} e^{-V_{i1}} G(e^{V_{i1}}, e^{V_{i2}}, ..., e^{V_{ij}})\} \right] d\epsilon$$

$$= \frac{e^{V_{i1}} G_{1}(e^{V_{i1}}, e^{V_{i2}}, ..., e^{V_{ij}})}{G(e^{V_{i1}}, e^{V_{i2}}, ..., e^{V_{ij}})}$$

multivariate extreme value (MEV) distribution (NOT GEV)

Uses the linear homogeneity
## Some examples

<table>
<thead>
<tr>
<th></th>
<th>G function</th>
<th>Choice probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logit</td>
<td>( G = \sum_{j=1}^{J} y_{ij} )</td>
<td>( P_{ij} = \frac{\exp(V_{ij})}{\sum_{j=1}^{J} \exp(V_{ij})} )</td>
</tr>
<tr>
<td>Nested logit</td>
<td>( G = \sum_{l=1}^{K} \left( \sum_{j \in B_l} y_{ij}^{1/\lambda_l} \right)^{\lambda_l} )</td>
<td>( P_{ij} = \frac{e^{V_{ij}/\lambda_k} \left( \sum_{j \in B_k} e^{V_{ij}/\lambda_l} \right)^{\lambda_k-1}} {\sum_{l=1}^{K} \left( \sum_{j \in B_k} e^{V_{ij}/\lambda_l} \right)^{\lambda_l}} )</td>
</tr>
<tr>
<td>Paired combinational logit</td>
<td>( G = \sum_{k=1}^{J-1} \sum_{l=k+1}^{J} \left( y_{ik}^{1/\lambda_{kl}} + y_{il}^{1/\lambda_{kl}} \right)^{\lambda_{kl}} )</td>
<td>( P_{ij} = \frac{\sum_{m \neq j} e^{V_{ij}/\lambda_{jm}} \left( \frac{V_{ij}}{e^{\lambda_{jm}}} + \frac{V_{im}}{e^{\lambda_{jm}}} \right)^{\lambda_{jm}-1}} {\sum_{k=1}^{J-1} \sum_{l=k+1}^{J} \left( \frac{V_{ik}}{e^{\lambda_{kl}}} + \frac{V_{il}}{e^{\lambda_{kl}}} \right)^{\lambda_{kl}}} )</td>
</tr>
<tr>
<td>Generalized nested logit</td>
<td>( G = \sum_{k=1}^{K} \left( \sum_{j \in B_k} (\alpha_{jk} y_{ij})^{1/\lambda_k} \right)^{\lambda_k} )</td>
<td>( P_{ij} = \frac{\sum_{k} (\alpha_{jk} e^{V_{ij}})^{1/\lambda_k} \left( \sum_{m \in B_k} (\alpha_{mk} e^{V_{im}})^{1/\lambda_k} \right)^{\lambda_k-1}} {\sum_{l=1}^{K} \left( \sum_{m \in B_k} (\alpha_{ml} e^{V_{im}})^{1/\lambda_l} \right)^{\lambda_l}} )</td>
</tr>
</tbody>
</table>

\* \( y_{ij} := \exp(V_{ij}) \)
Illustration

Path 1 = \{Link 1\} [travel time: 20]
Path 2 = \{Link 2, Link 3\} [travel time: 20]
Path 3 = \{Link 2, Link 4\} [travel time: 20]

Nested logit

\[ P_1 = \frac{\exp(\beta x)}{\exp(\beta x) + \exp\left(\frac{1}{\rho} \Lambda\right)} \]

\[ P_2 = P_3 = \frac{1}{2} \cdot \frac{\exp\left(\frac{1}{\rho} \Lambda\right)}{\exp(\beta x) + \exp\left(\frac{1}{\rho} \Lambda\right)} \]

※ \( \Lambda = \ln(\exp(\rho \beta x) + \exp(\rho \beta x)) \)

\( \beta \) is fixed as \(-0.2\)
Generalized G (A) function

The properties that the A function must exhibit

① $A(y_{i1}, y_{i2}, ..., y_{ij}) \geq 0$

② $A$ is homogeneous of degree one: $A(\alpha y_{i1}, ..., \alpha y_{ij}) = \alpha A(y_{i1}, ..., y_{ij})$

③ $\lim_{y_{ij} \to \infty} A(y_{i1}, y_{i2}, ..., y_{ij}) = \infty$ for any $j$

④ The cross partial derivatives of $A$ satisfy:

$$(-1)^{k-1} \frac{\partial^k A(y_{i1}, y_{i2}, ..., y_{ij})}{\partial y_{i1} \partial y_{i2} \cdots \partial y_{ik}} \geq 0$$

When all conditions are satisfied, the choice probability can be defined as:

$$P_{ij} = \frac{w_{ij} \cdot A_j(w_{i1}, w_{i2}, ..., w_{ij})}{A(w_{i1}, w_{i2}, ..., w_{ij})}$$

(where, $A_j = \partial A / \partial w_{ij}$)

Assumption: $F(x_{i1}, ..., x_{ij}) = \exp\{-A(-w_{i1}\ln[\Psi(x_{i1})], ..., -w_{ij}\ln[\Psi(x_{ij})])\}$

When $w_j = e^{V_{ij}}$ and $\Psi(x_j) \sim i.i.d. Gumbel$, $A$ function becomes McFadden's $G$ function
Derivation of choice probability

Note that \( \Pr[\max_{j \in J} X_{ij} \leq x] = F(x, x, ..., x) \), where \( F \) is defined as:

\[
F(x_{i1}, ..., x_{ij}) = \exp\{-A(-w_{i1} \ln[\Psi(x_{i1})], ..., -w_{ij} \ln[\Psi(x_{ij})])\}
\]

Then, the probability of the first alternative \( P_{i1} \) satisfies:

\[
P_{i1} = \int_{x \in \Omega_{i}} F_1(x, x, ..., x) dx
\]

\[
= \int_{x \in \Omega_{i}} \left[ e^{-A(-w_{i1} \ln[\Psi(x)], ..., -w_{ij} \ln[\Psi(x)])} \times \right.
\]

\[
\left. A_1(-w_{i1} \ln[\Psi(x)], ..., -w_{ij} \ln[\Psi(x)]) \cdot w_{i1} \cdot \frac{\psi(x)}{\psi(x)} \right] dx
\]

\[
= w_{i1} \cdot \frac{A_1(w)}{A(w)} \int_{x \in \Omega_{i}} A(w)[\Psi(x)]^{A(w)-1}\psi(x) dx
\]

\[
= w_{i1} \cdot \frac{A_1(w)}{A(w)}
\]

Uses the linear homogeneity
## Some examples

<table>
<thead>
<tr>
<th>G function</th>
<th>Choice probability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Under the assumption of independence</strong> <em>(Mattsson et al., 2014)</em></td>
<td></td>
</tr>
<tr>
<td>Logit (Gumbel)</td>
<td>$A: \text{summation, } w_{ij} = e^{\beta V_{ij}}, \psi(x_{ij}) \sim \text{Gumbel}(\beta, 0)$</td>
</tr>
<tr>
<td></td>
<td>$P_{ij} = \frac{\exp(\beta V_{ij})}{\sum_{j=1}^{J} \exp(\beta V_{ij})}$</td>
</tr>
<tr>
<td>Weibit-type (Frechet)</td>
<td>$A: \text{summation, } w_{ij} = V_{ij}^\beta, \psi(x_{ij}) \sim \text{Frechet}(\beta, 1)$</td>
</tr>
<tr>
<td></td>
<td>$P_{ij} = \frac{V_{ij}^\beta}{\sum_{j=1}^{J} V_{ij}^\beta}$</td>
</tr>
<tr>
<td>Weibit (Weibull)</td>
<td>$A: \text{summation, } w_{ij} = V_{ij}^{-\beta}, \psi(x_{ij}) \sim \text{Weibull}(\beta, 1)$</td>
</tr>
<tr>
<td></td>
<td>$P_{ij} = \frac{V_{ij}^{-\beta}}{\sum_{j=1}^{J} V_{ij}^{-\beta}}$</td>
</tr>
<tr>
<td><strong>Under the statistical dependence</strong> <em>(Chikaraishi and Nakayama, 2016)</em></td>
<td></td>
</tr>
<tr>
<td>Nested logit</td>
<td>$A = \sum_{l=1}^{K} \left( \sum_{j \in B_l} w_{ij}^{1/\lambda_l} \right)^{\lambda_l}, \text{ } w_{ij} = e^{\beta (a_{il} + b_{ij})}, \psi(x_{ij}) \sim \text{Gumbel}(\beta, 0)$</td>
</tr>
<tr>
<td></td>
<td>$P_{ij} = \frac{\exp[\beta b_{ij} / \lambda_l]}{\sum_{j \in J_l} \exp[\beta b_{ij} / \lambda_l]} \cdot \frac{\exp[\beta a_{il} + \lambda_l \tilde{b}<em>{0il}]}{\sum</em>{l'=1}^{L} \exp[\beta a_{il'} + \lambda_{l'} \tilde{b}<em>{0il'}]} \cdot \tilde{b}</em>{0il} = \ln \sum_{j \in J_l} \exp(\beta b_{ij} / \lambda_l)$</td>
</tr>
<tr>
<td>Nested weibit</td>
<td>$A=\sum_{l=1}^{K} \left( \sum_{j \in B_l} w_{ij}^{1/\lambda_l} \right)^{\lambda_l}, \text{ } w_{ij} = (a_{il} b_{ij})^{-\beta}, \psi(x_{ij}) \sim \text{Weibull}(\beta, 1)$</td>
</tr>
<tr>
<td></td>
<td>$P_{ij} = \frac{\beta^{\lambda_l}}{b_{ij}^{\lambda_l}} \cdot \frac{\exp[-\beta b_{ij} / \lambda_l]}{\sum_{j \in J_l} \exp[-\beta b_{ij} / \lambda_l]} \cdot \frac{\exp[-\beta (a_{il} - \beta (\tilde{b}<em>{0il})^{\lambda_l})]}{\sum</em>{l'=1}^{L} \exp[-\beta (a_{il'} - \beta (\tilde{b}<em>{0il'})^{\lambda_l})]} \cdot \tilde{b}</em>{0il} = \sum_{j \in J_l} b_{ij}^{-\beta / \lambda_l}$</td>
</tr>
</tbody>
</table>
Illustration

(b) Route length

Logit

Weibit

Choice probability of Path 1

Choice probability of Path 2

Choice probability of Path 3
Illustration

Path 1 = \{\text{Link 1}\} [travel time: \(x\)]
Path 2 = \{\text{Link 2, Link 3}\} [travel time: \(x + 5\)]
Path 3 = \{\text{Link 2, Link 4}\} [travel time: \(x + 10\)]

Nested logit

\[
\rho = \frac{1}{\lambda} = 1.0 \text{ (logit)}
\]

\[
\rho = \frac{1}{\lambda} = 1.5
\]

\[
\rho = \frac{1}{\lambda} = 2.0
\]

Nested weibit

\[
\rho = \frac{1}{\lambda} = 1.0 \text{ (logit)}
\]

\[
\rho = \frac{1}{\lambda} = 1.5
\]

\[
\rho = \frac{1}{\lambda} = 2.0
\]
Recursive logit  Fasgerau et al. (2013)

The recursive logit model corresponds to a dynamic discrete choice model where the path choice problem is formulated as a sequence of link choices (same as Akamatsu (1996))

\[ u(a|k) = v(a|k) + V(a) + \mu \varepsilon(a) \]

where
\[ V(k) = E \left[ \max_{a \in A(k)} (v(a|k) + V(a) + \mu \varepsilon(a)) \right] \]

Instantaneous cost  i.i.d. error terms (Gumbel)

The expected maximum utility to the destination
Recursive logit

\[ u(a|k) = v(a|k) + V(a) + \mu \varepsilon(a) \]

where \[ V(k) = E[ \max_{a \in A(k)} (v(a|k) + V(a) + \mu \varepsilon(a))] \]

Link choice probability:

\[ P(a|k) = \frac{\frac{1}{e^{\mu(v(a|k)+V(a)}}}{\sum_{a' \in A(k)} \frac{1}{e^{\mu(v(a'|k)+V(a'))}}} \]

Route choice probability:

\[ P(\sigma) = \prod_{i=0}^{I-1} P(k_{i+1}|k_i) = \prod_{i=0}^{I-1} e^{v(k_{i+1}|k_i)+V(k_{i+1})-V(k_i)} = e^{-V(k_0)} \prod_{i=0}^{I-1} e^{v(k_{i+1}|k_i)} \]

Log-likelihood:

\[ LL(\beta) = \ln \prod_{n=1}^{N} P(\sigma_n) = \frac{1}{\mu} \sum_{n=1}^{N} \left( \sum_{i=0}^{I_{n-1}} v(k_{i+1}|k_i) - V(k_0) \right) \]

Can be analytically obtained
Illustration

Incidence matrix $L$

\[
\begin{pmatrix}
0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\]
Vector of the expected maximum utility $\mathbf{Z} (\mu = 1)$

$$e^V(0) = \sum_{a \in A} L_{0a} \cdot e^{v(a|0) + V(a)}$$

$$e^V(1) = \sum_{a \in A} L_{1a} \cdot e^{v(a|1) + V(a)}$$

$$e^V(2) = \sum_{a \in A} L_{2a} \cdot e^{v(a|2) + V(a)}$$

$$e^V(3) = \sum_{a \in A} L_{3a} \cdot e^{v(a|3) + V(a)}$$

$$e^V(4) = \sum_{a \in A} L_{4a} \cdot e^{v(a|4) + V(a)}$$

$$e^V(d) = 1$$

Matrix defining instantaneous utility $\mathbf{M} (\mu = 1)$

$$\begin{pmatrix}
0 & e^{v(1|0)} & e^{v(2|0)} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & e^0 \\
0 & 0 & 0 & e^{v(3|2)} & e^{v(4|2)} & 0 \\
0 & 0 & 0 & 0 & 0 & e^0 \\
0 & 0 & 0 & 0 & 0 & e^0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
\end{pmatrix}$$

Elements of $\mathbf{z}$ and $\mathbf{M}$:

$$z_k = e^V(k) = \begin{cases} 
\sum_{a \in A} L_{ka} \cdot e^{v(a|k) + V(a)} & \forall k \in A \\
1 & k = d
\end{cases}$$

$$M_{ka} = \begin{cases} 
L_{ka} \cdot e^{v(a|k)} & a \in A(k) \\
0 & k = d
\end{cases}$$

$$\mathbf{z} = \mathbf{Mz} + \mathbf{b} \implies \mathbf{z} = (\mathbf{I} - \mathbf{M})^{-1}\mathbf{b} \quad \mathbf{b}' = (0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1)$$

$V(k)$ can be analytically obtained
Generalization of recursive logit

Recursive logit (Fosgerau et al., 2013)

\[ u(a|k) = v(a|k) + V(a) + \mu \varepsilon(a) \]

where

\[ V(k) = E \left[ \max_{a \in A(k)} (v(a|k) + V(a) + \mu \varepsilon(a)) \right] \]

Nested recursive logit (Mai et al., 2015)

\[ u(a|k) = v(a|k) + V(a) + \mu_k \varepsilon(a) \]

where

\[ V(k) = E \left[ \max_{a \in A(k)} (v(a|k) + V(a) + \mu_k \varepsilon(a)) \right] \]

Generalized recursive logit (Mai, 2016)

\[ u(a|k) = v(a|k) + V(a) + \mu \varepsilon(a) \]

where

\[ V(k) = E \left[ \max_{a \in A(k)} \left( v(a|k) + V(a) + \varepsilon(a|k) - \frac{\gamma}{\mu_k} \right) \right] \]

Following the MEV distribution (expressed through G function)

Generalization leads to difficulties in model estimation (as usual)
Highly recommended!

- Kenneth E. Train
- Discrete Choice Methods with Simulation
- Cambridge University Press
- Second edition, 2009

- [https://eml.berkeley.edu/books/choice2.html](https://eml.berkeley.edu/books/choice2.html)

Chapter 1. Introduction
Chapter 2. Properties of Discrete Choice Models
Chapter 3. Logit
Chapter 4. GEV
Chapter 5. Probit
Chapter 6. Mixed Logit
Chapter 7. Variations on a Theme
Chapter 8. Numerical Maximization
Chapter 9. Drawing from Densities
Chapter 10. Simulation-Assisted Estimation
Chapter 11. Individual-Level Parameters
Chapter 12. Bayesian Procedures
Chapter 13. Endogeneity
Chapter 14. EM Algorithms
References


References

References