September 25-27, 2015
The 14th Summer course for Behavior Modeling in Transportation Networks
@The University of Tokyo

Advanced behavior models

Recent development of discrete choice models

Makoto Chikaraishi[1], Hideki Yaginuma[2]

[1] Hiroshima University

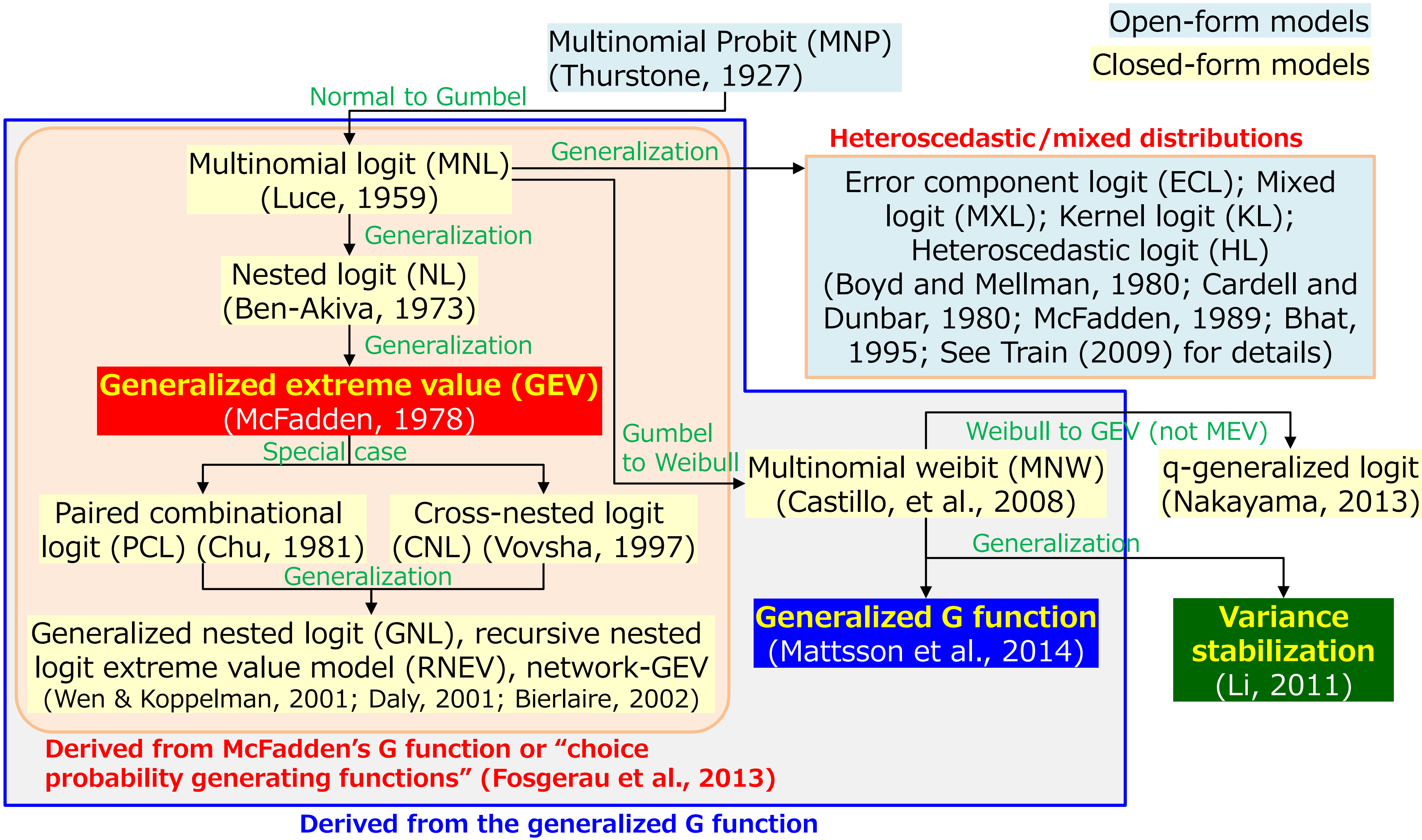
[2] The University of Tokyo

Contents

Introduction

- This lecture introduces advanced discrete choice models, including
  - advanced closed-form models, and
  - advanced open-form models
- Understanding such advanced models are important not only for utilizing advanced models, but also for understanding the limitations of conventional models
  - Advanced models are often costly (computational cost, etc.), but need to be understood even when the conventional models are just applied

Genealogy of discrete choice models
[based on Hato (2002)]



Models without specifying error distributions

Closed-form discrete choice models

G FUNCTION & SOME EXAMPLES

(McFadden, 1978)

McFadden's G function

The properties that the G function must exhibit

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

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

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

④ The cross partial derivatives of G satisfy:

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

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

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

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

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

Derivation of G function

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

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

multivariate extreme value (MEV) distribution (**NOT** GEV)

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

$$P_{i1} = \int_{\epsilon=-\infty}^{+\infty} F_1\left(\epsilon, V_{i1} - V_{i2} + \epsilon, \dots, V_{i1} - V_{iJ} + \epsilon\right) d\epsilon$$

$$= \int_{\epsilon=-\infty}^{+\infty} \begin{bmatrix} e^{-\epsilon} G_1(e^{-\epsilon}, e^{-\epsilon - V_{i1} + V_{i2}}, \dots, e^{-\epsilon - V_{i1} + V_{iJ}}) \\ \times \exp\{-G(e^{-\epsilon}, e^{-\epsilon - V_{i1} + V_{i2}}, \dots, e^{-\epsilon_{i1} - V_{i1} + V_{iJ}})\} \end{bmatrix} d\epsilon$$

$$= \int_{\epsilon=-\infty}^{+\infty} \begin{bmatrix} e^{-\epsilon} G_1(e^{V_{i1}}, e^{V_{i2}}, \dots, e^{V_{iJ}}) \\ \times \exp\{-e^{-\epsilon} e^{-V_{i1}} G(e^{V_{i1}}, e^{V_{i2}}, \dots, e^{V_{iJ}})\} \end{bmatrix} d\epsilon$$

Uses the linear homogeneity

$$= \frac{e^{V_{i1}} G_1(e^{V_{i1}}, e^{V_{i2}}, \dots, e^{V_{iJ}})}{G(e^{V_{i1}}, e^{V_{i2}}, \dots, e^{V_{iJ}})}$$

Some examples

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

* $y_{ij} \coloneqq \exp(V_{ij})$

Strengths and limitations

- **Strengths**
  - A closed-form discrete choice model **without assuming specific error distributions**
  - This allow us to derive a number of **behaviorally understandable** models
    - Nested logit, Cross-nested logit, Paired combinational logit, etc.
- **Limitations**
  - Only for **additive utility**, i.e., $u_{ij} = V_{ij} + \epsilon_{ij}$
    - V_{ij} and ϵ_{ij} can be dependent each other
  - Only for ~~**GEV**~~ **MEV family**
    - Some other distributions can be useful in some context

VARIANCE STABILIZATION & SOME EXAMPLES

Variance stabilization

Two fundamental ideas:

1. A stable class of distributions w.r.t. the minimum operation

Suppose the random disutility X_{ij} from the following CDF:

$$F_{ij}(x) = \Pr\{X_{ij} < x\} = 1 - [1 - F(x)]^{\alpha_{ij}}$$

Unspecified base distribution function

The minimum random disutility X_{ij} under the assumption of independence can be written as:

$$\Pr\{\min_{j \in C_i} X_{ij} < x\} = 1 - \Pi_{j \in C_i} \Pr\{1 - F_{ij}(x)\} = 1 - [1 - F(x)]^{\alpha_{i0}}$$

$\alpha_{i0} = \Sigma_{j \in C_i} \alpha_{ij}$

2. Variance-stabilizing transformations

Consider the transformation of $F_{ij}(x)$ to the Gumbel distribution:

$$F_{ij}(x) = \Pr\{X_{ij} < x\} = 1 - [1 - F(x)]^{\alpha_{ij}}$$

A transformation function $h(x)$ which stabilize the variance can be defined as:

$$h(x) = \theta^{-1} \log\{-\log[1 - F(x)]\}$$

The transformed random variable $Z_{ij} = h(X_{ij})$ follows:

$$G(z; \theta, \alpha_{ij}) = 1 - \exp[-\alpha_{ij} \exp(\theta z)] \quad \text{[Gumbel]}$$

(Li, 2011)

Derivation of choice probability

$Z_{ij} = h(X_{ij})$ where $h(\cdot)$ is a monotonically increasing transformation

$$P_{ij} = \Pr\{X_{ij} \le \min_{j'(\neq j)} X_{ij'}\} = \Pr\{Z_{ij} \le \min_{j'(\neq j)} Z_{ij'}\}$$

$$= \int_{z\in\Omega_i} Q_{i1}(z) \cdots Q_{ij-1}(z) f_{ij}(z) Q_{ij+1}(z) \cdots Q_{iJ}(z) dz$$

where,

$Q_{ij}(z) = 1 - F_{ij}(z) = \exp\left[-\alpha_{ij} \exp \theta z\right]$, and

$f_{ij}(z) = \theta\alpha_{ij} \exp\left[-\alpha_{ij} \exp(\theta z)\right] \exp(\theta z)$

$$P_{ij} = \theta\alpha_{ij} \int_{z\in\Omega_i} \exp[-\alpha_{i0} \exp(\theta z)] \exp(\theta z)\, dz$$

$$= \frac{\alpha_{ij}}{\alpha_{i0}} = \frac{\alpha_{ij}}{\sum_{j'\in C_i} \alpha_{ij'}} = \frac{H(V_{ij})}{\sum_{j'\in C_i} H(V_{ij'})}$$

How to specify α_{ij}?

Since $h(X_{ij})$ follows the Gumbel where the CDF is $1 - \exp[-\alpha_{ij} \exp(\theta x)]$, $E\left[h(X_{ij})\right] = -\{\log(\alpha_{ij}) + \gamma\}/\theta$. Thus, $\alpha_{ij} = \exp\{-\gamma - \theta E\left[h(X_{ij})\right]\}$

Some examples

- The models with the distributions of: Exponential, Parato, Type II generalized logistic, Gompertz, Rayleigh, Weibull, and Gumbel (some types of distributions need approximations)

Table 1
Special cases of the distribution family (1).

	Underlying distribution $F_{in}(t)$	Base distribution $F(t)$	Expectation V_{in}	Variance σ_{in}^2
Exponential	$1-\exp\{-\alpha_{in}t\}$	$1-\exp\{-t\}$	α_{in}^{-1}	α_{in}^{-2}
Pareto	$1-t^{-\alpha_{in}}(t \geqslant 1)$	$1-t^{-1}$	$\alpha_{in}/(\alpha_{in}-1)$	$\alpha_{in}/[(\alpha_{in}-1)^2(\alpha_{in}-2)]$
Type II generalized logistic	$1-[1+\exp(t)]^{-\alpha_{in}}$	$1-1/[1+\exp(t)]$	$\psi(1)-\psi(\alpha_{in})$	$\psi'(1)-\psi'(\alpha_{in})$
Gompertz	$1-\exp\{-\alpha_{in}[\exp(\theta t)-1]\}$	$1-\exp\{-[\exp(\theta t)-1]\}$		
Rayleigh	$1-\exp\{-\alpha_{in}t^2/2\}$	$1-\exp\{-t^2/2\}$	$[\pi/(2\alpha_{in})]^{1/2}$	$(4-\pi)/(2\alpha_{in})$
Weibull	$1-\exp\{-\alpha_{in}t^{\theta}\}$	$1-\exp\{-t^{\theta}\}$	$\alpha_{in}^{-1/\theta}\Gamma(1+1/\theta)$	$\alpha_{in}^{-2/\theta}\{\Gamma(1+\frac{2}{\theta})-[\Gamma(1+\frac{1}{\theta})]^2\}$
Gumbel	$1-\exp\{-\alpha_{in}\exp(\theta t)\}$	$1-\exp\{-\exp(\theta t)\}$	$-\{\log(\alpha_{in})+\gamma\}/\theta$	$\pi^2/(6\theta^2)$

Table 2
The variance-stabilizing transformations, mean functions, and sensitivity functions for some distributions in family (1).

	Variance-stabilizing transformation $h(t)$	Mean function $H(t)$	Sensitivity function $S(t)$
Exponential	$\theta^{-1}\log(t)$	t^{-1}	$-\log(t)$
Pareto	$\theta^{-1}\log\{\log(t)\}$	$t/(t-1)$	$\log(t)-\log(t-1)$
Type II generalized logistic	$\theta^{-1}\log\{\log[1+\exp(t)]\}$	$\psi^{-1}(\psi(1)-\psi(t))$	$\log\{\psi^{-1}(\psi(1)-\psi(t)\}$
Gompertz	$\theta^{-1}\log\{\exp(\theta t)-1\}$		
Rayleigh	$\theta^{-1}\log(t^2)$	$\pi/(2t^2)$	$-2\log(t)$
Weibull	$\log(t)$	$\{\Gamma(1+1/\theta)/t\}^{\theta}$	$-\theta\log(t)$
Gumbel	t	$\exp(-\gamma-\theta t)$	$-\theta t$

(Li, 2011)

Further generalization

"Scale parameter is absorbed into $H(\cdot)$ so it is not identifiable. Hence, extending the multinomial logit model by allowing an unspecified functional form $H(\cdot)$ can address both the issue of non-linearity in the mean function and the issue of variance stabilization" (p. 465)

Since $H\left(V_{ij}\right)$ $[= \alpha_{ij}]$ should be non-negative, it is natural to assume:

$$\frac{H\left(V_{ij}\right)}{\sum_{j' \in C_i} H\left(V_{ij'}\right)} = \frac{\exp\left\{S\left(\boldsymbol{\beta}\mathbf{x_{ij}}\right)\right\}}{\sum_{j' \in C_i} \exp\left\{S\left(\boldsymbol{\beta}\mathbf{x_{ij}}\right)\right\}}$$

where $S(\cdot)$ is a sensitivity function

Semi-parametric approach (such as P-splines approach) can be used as an approximation of any base distribution F

Distribution/linearity: an example

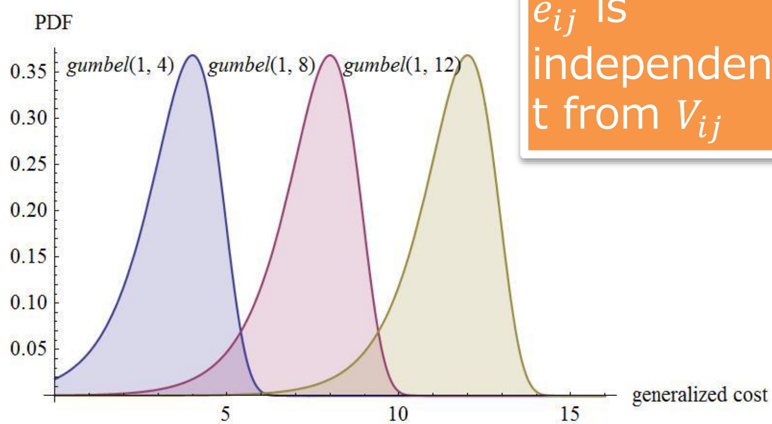
(1) Differences in distribution assumption

$$u_{ij} = g\left(V_{ij}, \varepsilon_{ij}\right)$$

u_{ij}: Random utility
V_{ij}: Systematic utility (linear in parameters)
ε_{ij}: Error term

Gumbel distribution

e_{ij} is independent from V_{ij}

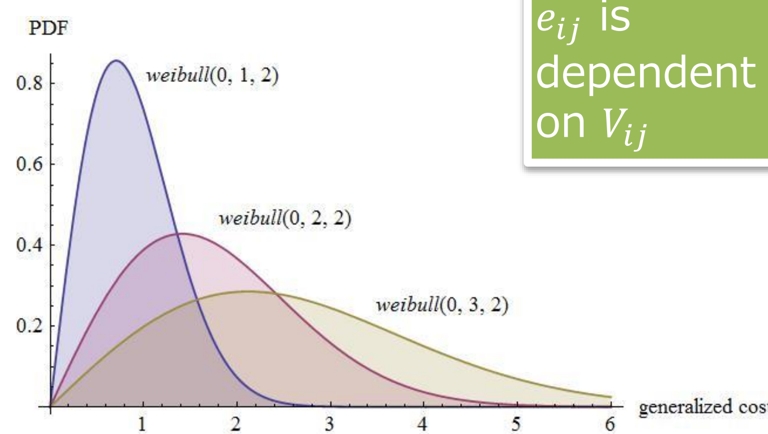


Logit model

$$p_{ij} = \frac{\exp\left(-\frac{1}{\theta}V_{ij}\right)}{\sum_k \exp\left(-\frac{1}{\theta}V_{ik}\right)}$$

Weibull distribution

e_{ij} is dependent on V_{ij}



Weibit (or multiplicative) model

$$p_{ij} = \frac{V_{ij}^{-\frac{1}{\theta}}}{\sum_k V_{ik}^{-\frac{1}{\theta}}}$$

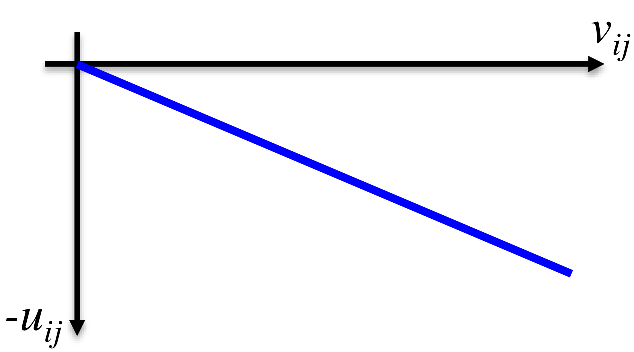
Distribution/linearity: an example

(2) Difference in systematic utility

$$u_{ij} = f\left(V_{ij}\right) + \varepsilon_{ij}$$

ε_{ij}: Gumbel distribution

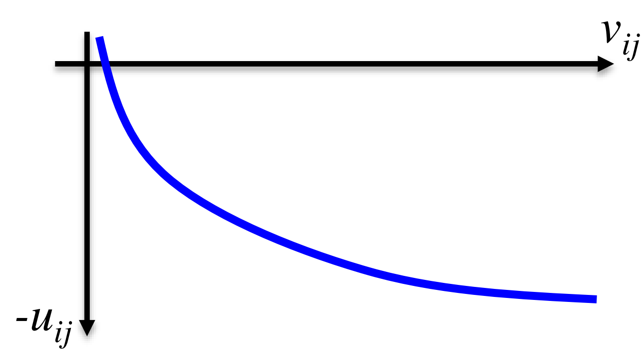
Linear systematic utility



Logit model

$$p_{ij} = \frac{\exp\left(-\frac{1}{\theta}V_{ij}\right)}{\sum_k \exp\left(-\frac{1}{\theta}V_{ik}\right)}$$

Logarithm systematic utility

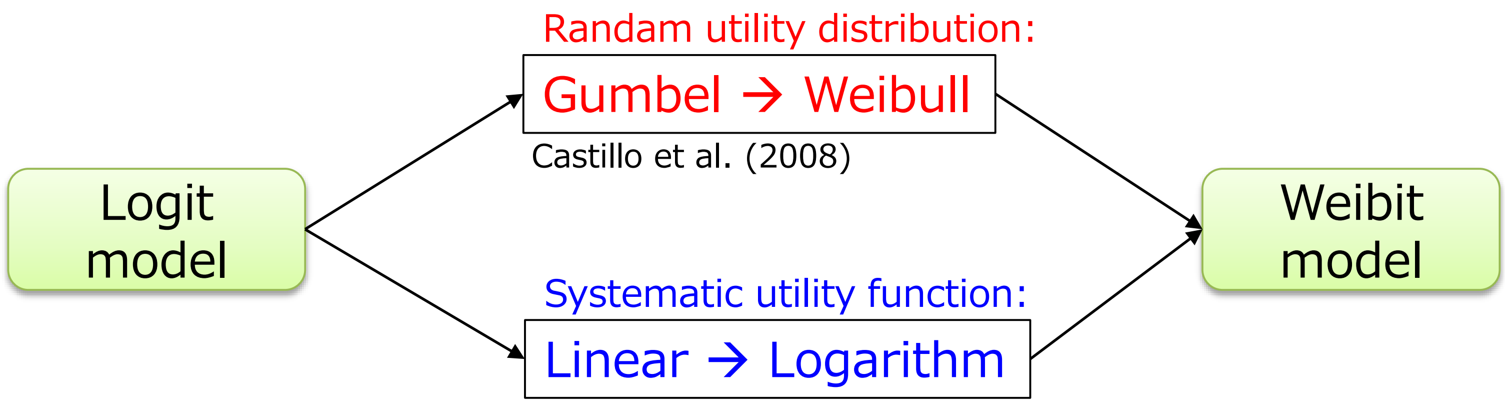


Weibit (or multiplicative) model

$$p_{ij} = \frac{V_{ij}^{-\frac{1}{\theta}}}{\sum_k V_{ik}^{-\frac{1}{\theta}}}$$

Distribution/linearity: an example

$$u_{ij} = g\left(f\left(V_{ij}\right), \varepsilon_{ij}\right)$$



(See Castillo et al. (2008) for elegant explanations)

Strengths and limitations

- **Strengths**
  - Not limited to the MEV distribution. **A larger class of distributions** can be assumed in the development of closed-form choice models
  - A **semi-parametric** discrete choice model can approximate any base distribution F
- **Limitations**
  - Only **under the assumption of independence**
    - Unobserved terms need to be independent across alternatives
  - **Behavioral foundations** of some types of distributions has not been well established
    - Increase the difficulty to use the models in practice

GENERALIZED G FUNCTION & SOME EXAMPLES

(Mattsson et al., 2014)

Generalized G (A) function

The properties that the A function must exhibit

① $A(y_{i1}, y_{i2}, \dots, y_{iJ_i}) \geq 0$

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

③ $\lim_{y_{ij} \to -\infty} A(y_{i1}, y_{i2}, \dots, y_{iJ_i}) = \infty$

④The cross partial derivatives of A satisfy:

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

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

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

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

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

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

(Mattsson et al., 2014)

Derivation of A function

Suppose $u_{ij} = f(w_{ij}, x_{ij})$, where $(x_{i1}, \dots, x_{iJ})$ is distributed F defined as:

$$F(x_{i1}, \dots, x_{iJ}) = \exp\{-A(-w_{i1}\ln[\Psi(x_{i1})], \dots, -w_{iJ}\ln[\Psi(x_{iJ})])\}$$

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

$$P_{i1} = \int_{x\in\Omega_i} F_1(x, x, \dots, x)dx$$

$$= \int_{x\in\Omega_i} \left[\begin{array}{c} e^{-A(-w_{i1}\ln[\Psi(x_{i1})], \dots, -w_{iJ}\ln[\Psi(x_{iJ})])} \times \\ A_1(-w_{i1}\ln[\Psi(x_{i1})], \dots, -w_{iJ}\ln[\Psi(x_{iJ})]) \cdot w_{i1} \cdot \frac{\psi(x)}{\Psi(x)} \end{array} \right] dx$$

$$= w_{i1} \cdot \frac{A_1(w)}{A(w)} \int_{x\in\Omega_i} \underbrace{A(w)[\Psi(x)]^{A(w)-1}\psi(x)}_{\text{=density function of } F} dx$$

Uses the linear homogeneity

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

Assuming the statistical independence,

$P_{i1} = \frac{w_{i1}}{\Sigma_{j\in C_i} w_{ij}}$ which is equivalent to Li's (2011) model

(Mattsson et al., 2014)

Some examples [1/2]

	G function	Choice probability
Under the assumption of independence		
Logit (Gumbel)	A: summation, $w_{ij} = e^{\beta V_{ij}}$, $\Psi(x_{ij}) \sim Gumbel(\beta, 0)$	$P_{ij} = \frac{\exp(V_{ij})}{\sum_{j'=1}^{J} \exp(V_{ij'})}$
Weibit-type (Frechet)	A: summation, $w_{ij} = {V_{ij}}^{\beta}$, $\Psi(x_{ij}) \sim Frechet(\beta, 1)$	$P_{ij} = \frac{V_{ij}^{\beta}}{\sum_{j'=1}^{J} V_{ij'}^{\beta}}$
Weibit (Weibull)	A: summation, $w_{ij} = {V_{ij}}^{-\beta}$, $\Psi(x_{ij}) \sim Weibull(\beta, 1)$	$P_{ij} = \frac{V_{ij}^{-\beta}}{\sum_{j'=1}^{J} V_{ij'}^{-\beta}}$
Under the statistical dependence		
Nested logit, Paired combinational logit, Cross-nested logit, etc. (Same as the models derived from G function), **AND** some other models (see the next page)		

(Mattsson et al., 2014)

Some examples [2/2]

An example of A function under the statistical dependence

Let $m \leqslant n$ and suppose that $X = (X_1, \ldots, X_n)$ has a c.d.f. $F \in \mathcal{G}^n$ for some seed function $\Psi \in \mathcal{F}$, positive weights $w = (w_1, \ldots, w_n)$, and aggregation function A of the form

$$A(y) = \left(\sum_{i=1}^{m} y_i^{\rho}\right)^{1/\rho} + \left(\sum_{i=m+1}^{n} y_i^{\tau}\right)^{1/\tau} \quad \forall y \in \mathbb{R}_+^n \tag{12}$$

for some $\rho,\ \tau \geqslant 1$. This is still an aggregation function that satisfies the alternating-signs condition, and F is a c.d.f. by Lemma 2. When both $\rho,\ \tau > 1$, there is statistical dependence within the subset $I_1 = \{1, \ldots, m\}$ of the first m random variables, as well as within the remaining set $I_2 = \{m+1, \ldots, n\}$ of random variables. Rewrite $A(y_1, \ldots, y_n) = A_1(y_1, \ldots, y_m) + A_2(y_{m+1}, \ldots, y_n)$. This set-up arises naturally in travel demand, location choice, industrial organization and international trade (in which case I_1 might be a travel mode, a geographical area, a category of goods, an industry or a country; and likewise for I_2). We then have, for each $k \in I_1$:

$$\Pr[k \in \arg\max_{i \in I} X_i] = \frac{A_1(w_1, \ldots, w_m)}{A_1(w_1, \ldots, w_m) + A_2(w_{m+1}, \ldots, w_n)} \cdot \frac{w_k^{\rho}}{\sum_{i=1}^{m} w_i^{\rho}}, \tag{13}$$

At this moment, the behavioral foundations have not been well established

Strengths and limitations

- **Strengths**
  - Extend McFadden's G function
    - **From MEV to GEV (but not fully GEV)**
  - The model can deal with **the statistical dependence** among alternatives
    - G-function-based GEV models are the special cases
- **Limitations**
  - **Behavioral foundations** of some types of distributions has not been well established
    - Increase the difficulty to use the models in practice

Summary of closed-form models

- The new types of closed-form models can still be developed
- The biggest remaining problem may be the lack of behavioral foundation
  - **The task of behavioral modelers**