# Discrete Choice Analysis I

**Moshe Ben-Akiva**

1.201 / 11.545 / ESD.210
Transportation Systems Analysis: Demand & Economics

Fall 2008

# Outline of 2 Lectures on Discrete Choice

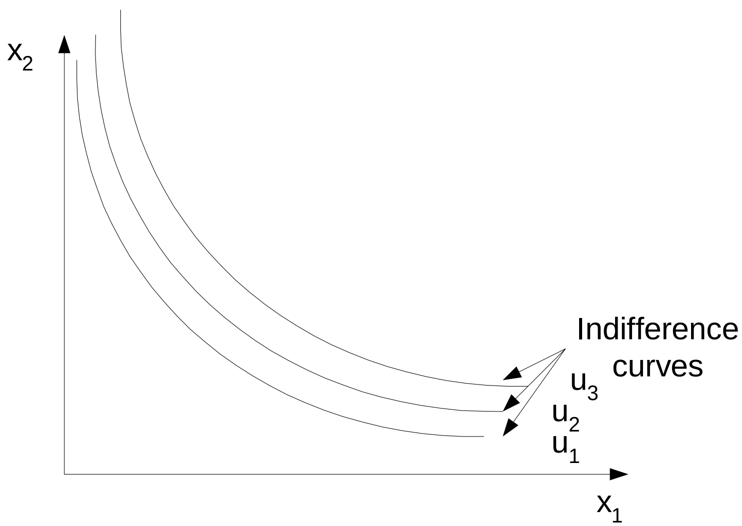
- Introduction
- A Simple Example
- The Random Utility Model
- Specification and Estimation
- Forecasting
- IIA Property
- Nested Logit

# Outline of this Lecture

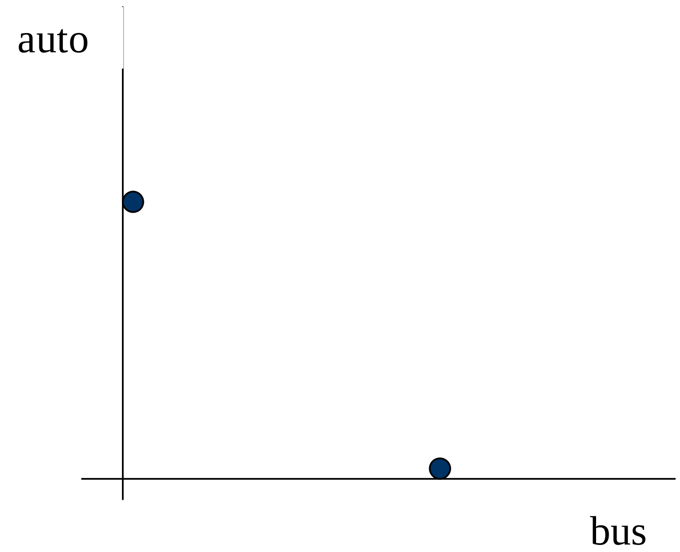
- Introduction
- A simple example – route choice
- The Random Utility Model
  - Systematic utility
  - Random components
- Derivation of the Probit and Logit models
  - Binary Probit
  - Binary Logit
  - Multinomial Logit

# Continuous vs. Discrete Goods

Continuous Goods




Discrete Goods



# Discrete Choice Framework

- Decision-Maker
  - Individual (person/household)
  - Socio-economic characteristics (e.g. Age, gender,income, vehicle ownership)
- Alternatives
  - Decision-maker *n* selects one and only one alternative from a choice set $C_n=\{1,2,\ldots,i,\ldots,J_n\}$ with $J_n$ alternatives
- Attributes of alternatives (e.g.Travel time, cost)
- Decision Rule
  - Dominance, satisfaction, utility etc.

# Choice: Travel Mode to Work

- Decision maker: an individual worker
- Choice: whether to drive to work or take the bus to work
- Goods: bus, auto
- Utility function: $U(X) = U(\text{bus, auto})$
- Consumption bundles: {1,0} (person takes bus)
  {0,1} (person drives)

# Consumer Choice

- Consumers maximize utility
  - Choose the alternative that has the maximum utility (and falls within the income constraint)

    If $U$(bus) > $U$(auto) → choose bus
    If $U$(bus) < $U$(auto) → choose auto

$U$(bus)=?

$U$(auto)=?

# Constructing the Utility Function

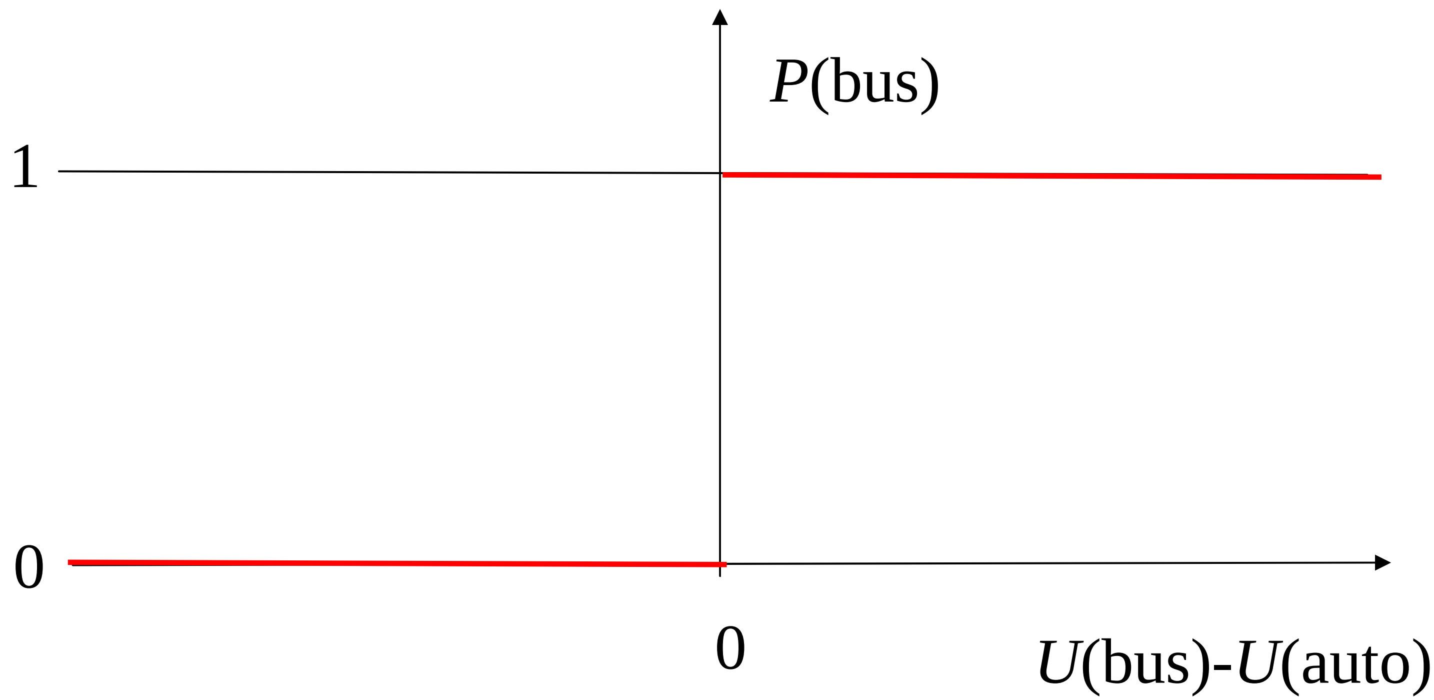
- $U$(bus) = $U$(walk time, in-vehicle time, fare, …)

  $U$(auto) = $U$(travel time, parking cost, …)
- Assume linear (in the parameters)

  $U(\text{bus}) = \beta_1 \times (\text{walk time}) + \beta_2 \times (\text{in-vehicle time}) + \ldots$
- Parameters represent tastes, which may vary over people.
  - Include socio-economic characteristics (e.g., age, gender, income)
  - $U(\text{bus}) = \beta_1 \times (\text{walk time}) + \beta_2 \times (\text{in-vehicle time}) + \beta_3 \times (\text{cost/income}) + \ldots$

# Deterministic Binary Choice

- If $U$(bus) - $U$(auto) > 0 , Probability(bus) = 1
  If $U$(bus) - $U$(auto) < 0 , Probability(bus) = 0



# Probabilistic Choice

- Random utility model

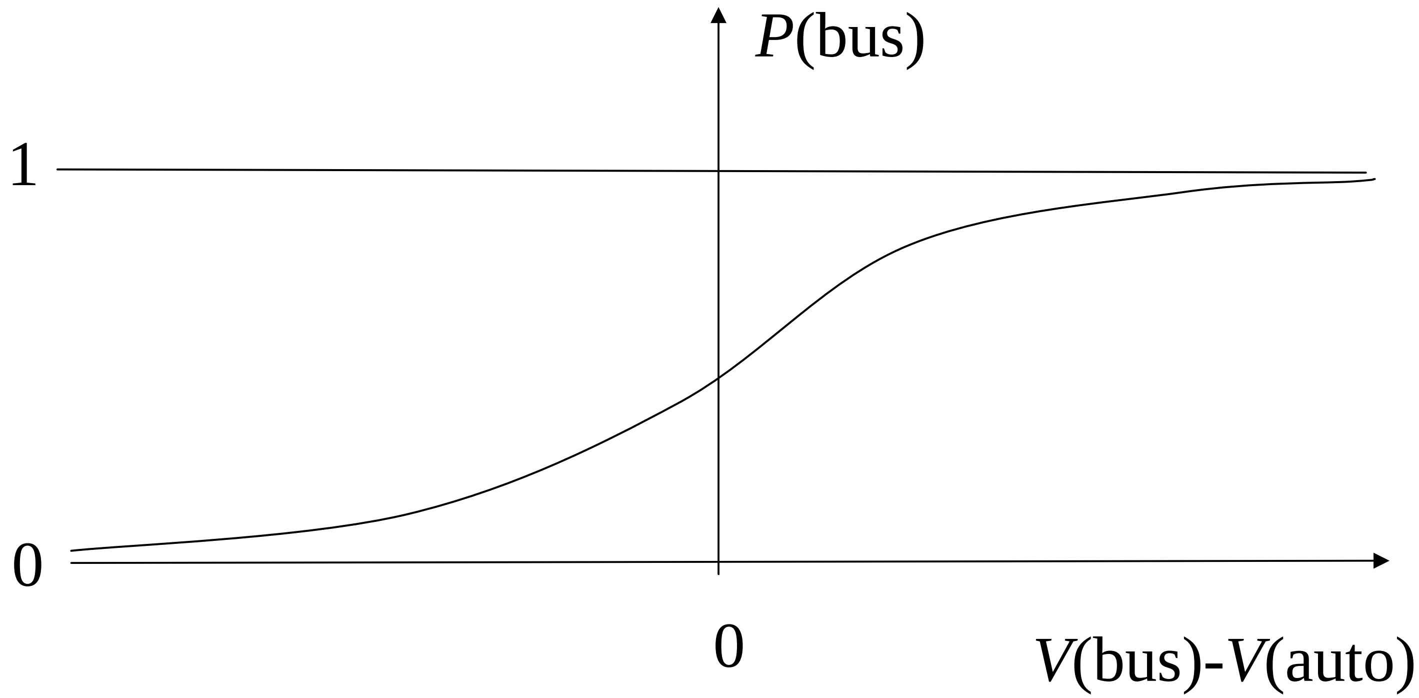
  $U_i$ = $V$(attributes of $i$; parameters) + $epsilon_i$

- What is in the epsilon?

  Analysts' imperfect knowledge:
  - Unobserved attributes
  - Unobserved taste variations
  - Measurement errors
  - Use of proxy variables

- $U$(bus) = $\beta_1 \times$(walk time) + $\beta_2 \times$(in-vehicle time + $\beta_3 \times$(cost/income) + … + *epsilon_bus*

# Probabilistic Binary Choice



# A Simple Example: Route Choice

- Sample size: $N$ = 600
- Alternatives: Tolled, Free
- Income: Low, Medium, High

| **Route choice** | **Income** | | | |
|---|---|---|---|---|
| | **Low ($k$=1)** | **Medium ($k$=2)** | **High ($k$=3)** | |
| **Tolled ($i$=1)** | 10 | 100 | 90 | 200 |
| **Free ($i$=2)** | 140 | 200 | 60 | 400 |
| | 150 | 300 | 150 | 600 |

# A Simple Example: Route Choice

Probabilities

- (Marginal) probability of choosing toll road $P(i = 1)$

$$\hat{P}(i=1) = 200/600 = 1/3$$

- (Joint) probability of choosing toll road and having medium income: $P(i=1, k=2)$

$$\hat{P}(i=1, k=2) = 100/600 = 1/6$$

$$\sum_{i=1}^{2}\sum_{k=1}^{3} P(i,k) = 1$$

# Conditional Probability *P(i|k)*

$$P(i,k) = P(i) \cdot P(k \mid i)$$
$$= P(k) \cdot P(i \mid k)$$

Independence

$$P(i \mid k) = P(i)$$
$$P(k \mid i) = P(k)$$

$$P(i) = \sum_k P(i,k)$$
$$P(k) = \sum_i P(i,k)$$
$$P(k \mid i) = \frac{P(i,k)}{P(i)}, \qquad P(i) \neq 0$$
$$P(i \mid k) = \frac{P(i,k)}{P(k)}, \qquad P(k) \neq 0$$

# Model : $P(i|k)$

- Behavioral Model~
  Probability (Route Choice|Income) = $P(i|k)$
- Unknown parameters

$$P(i=1|k=1)=\pi_1$$

$$P(i=1|k=2)=\pi_2$$

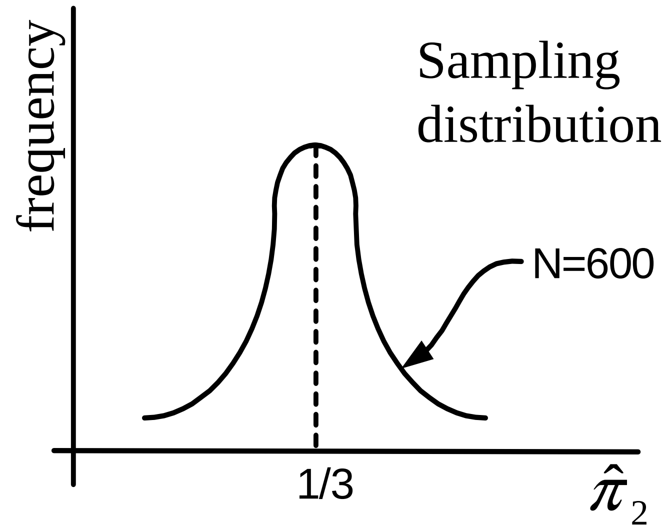
$$P(i=1|k=3)=\pi_3$$

# Example: Model Estimation

- Estimation

$$\hat{\pi}_1 = \frac{1}{15}, \quad \hat{\pi}_2 = \frac{1}{3}, \quad \hat{\pi}_3 = \frac{3}{5}$$

$$= 0.067 \quad = 0.333 \quad = 0.6$$




$$s_1 = \sqrt{\frac{\hat{\pi}_1 \cdot (1-\hat{\pi}_1)}{N_1}} = \sqrt{\frac{1/15 \cdot (1-1/15)}{150}} = 0.020$$

$$s_2 = \sqrt{\frac{\hat{\pi}_2 \cdot (1-\hat{\pi}_2)}{N_2}} = \sqrt{\frac{1/3 \cdot (1-1/3)}{300}} = 0.027$$

$$s_3 = \sqrt{\frac{\hat{\pi}_3 \cdot (1-\hat{\pi}_3)}{N_3}} = \sqrt{\frac{3/5 \cdot (1-3/5)}{150}} = 0.040$$

Standard errors

# Example: Forecasting

- Toll Road share under existing income distribution: 33%
- New income distribution

| Route h i | Income | | | | |
|---|---|---|---|---|---|
| | Low ($k$=1) | Medium ($k$=2) | High ($k$=3) | | |
| Tolled ($i$=1) | 1/15*45=3 | 1/3*300=100 | 3/5*255=153 | 256 | 43% |
| F ($i$ 2) | 42 | 200 | 102 | 344 | 57% |
| **New income distribution** | **45** | **300** | **255** | **600** | |
| *Existing income di ib i* | *150* | *300* | *150* | *600* | |

- Toll road share: 33%→43%

# The Random Utility Model

- Decision rule: Utility maximization
  - Decision maker *n* selects the alternative *i* with the highest utility $U_{in}$ among $J_n$ alternatives in the choice set $C_n$.

$$U_{in} = V_{in} + \varepsilon_{in}$$

$V_{in}$ =Systematic utility : function of observable variables

$\varepsilon_{in}$ =Random utility

# The Random Utility Model

- Choice probability:

$$P(i|C_n) = P(U_{in} \geq U_{jn},\ \forall j \in C_n)$$
$$= P(U_{in} - U_{jn} \geq 0,\ \forall j \in C_n)$$
$$= P(U_{in} = max_j\ U_{jn}, \forall j \in C_n)$$

- For binary choice:

$$P_n(1) = P(U_{1n} \geq U_{2n})$$
$$= P(U_{1n} - U_{2n} \geq 0)$$

# The Random Utility Model

| Routes | Attributes | | Utility (utils) |
|---|---|---|---|
| | Travel time ($t$) | Travel cost ($c$) | |
| Tolled ($i$=1) | $t_1$ | $c_1$ | $U_1$ |
| Free ($i$=2) | $t_2$ | $c_2$ | $U_2$ |

$$U_1 = -\beta_1 t_1 - \beta_2 c_1 + \varepsilon_1$$

$$U_2 = -\beta_1 t_2 - \beta_2 c_2 + \varepsilon_2$$

$$\beta_1, \beta_2 > 0$$

# The Random Utility Model

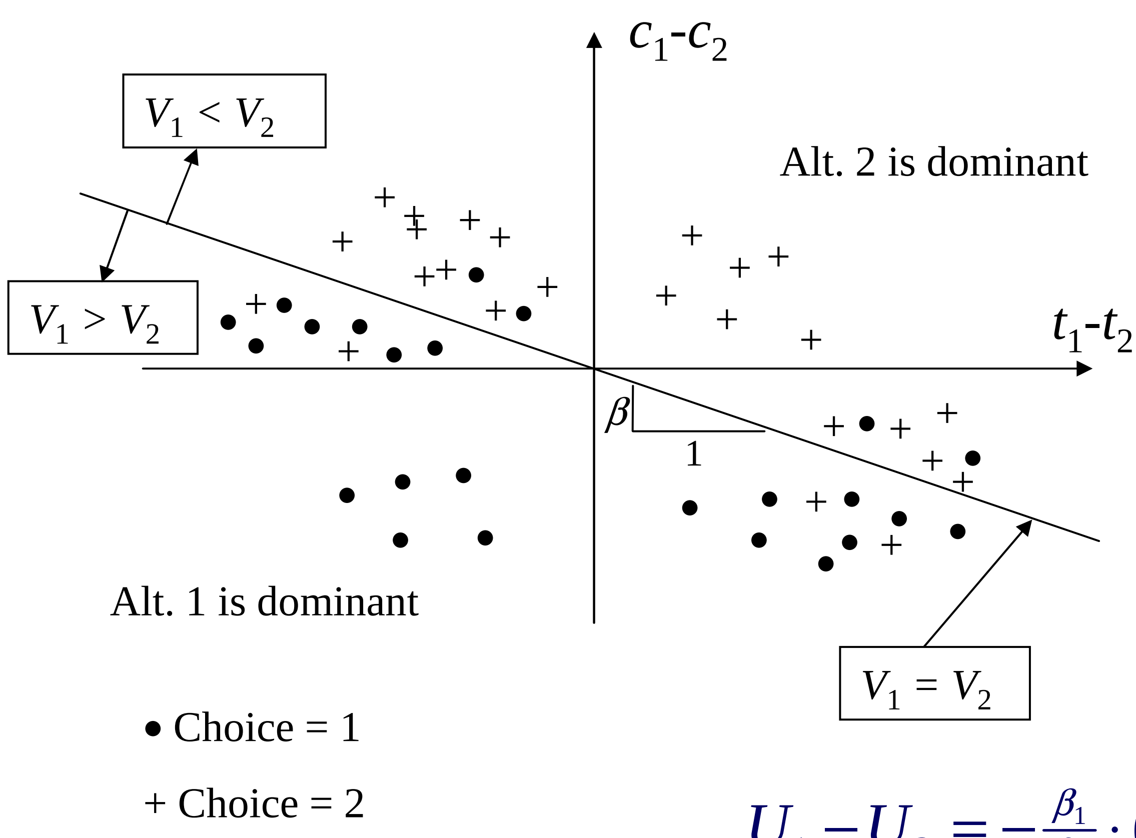
- Ordinal utility
  - Decisions are based on utility differences
  - Unique up to order preserving transformation

$$U_1 = (-\beta_1 t_1 - \beta_2 c_1 + \varepsilon_1 + K)\lambda$$

$$U_2 = (-\beta_1 t_2 - \beta_2 c_2 + \varepsilon_2 + K)\lambda$$

$$\beta_1, \beta_2, \lambda > 0$$

# The Random Utility Model




$$U_1 = -\frac{\beta_1}{\beta_2} \cdot t_1 - c_1 + \varepsilon_1$$

$$U_2 = -\frac{\beta_1}{\beta_2} \cdot t_2 - c_2 + \varepsilon_2$$

$$\beta = \frac{\beta_1}{\beta_2} = \text{"value of time"}$$

$$U_1 - U_2 = -\frac{\beta_1}{\beta_2} \cdot (t_1 - t_2) - (c_1 - c_2) + (\varepsilon_1 - \varepsilon_2)$$

# The Systematic Utility

- Attributes: describing the alternative
  - Generic vs. Specific
    - Examples: travel time, travel cost, frequency
  - Quantitative vs. Qualitative
    - Examples: comfort, reliability, level of service
  - Perception
  - Data availability
- Characteristics: describing the decision-maker
  - Socio-economic variables
    - Examples: income,gender,education

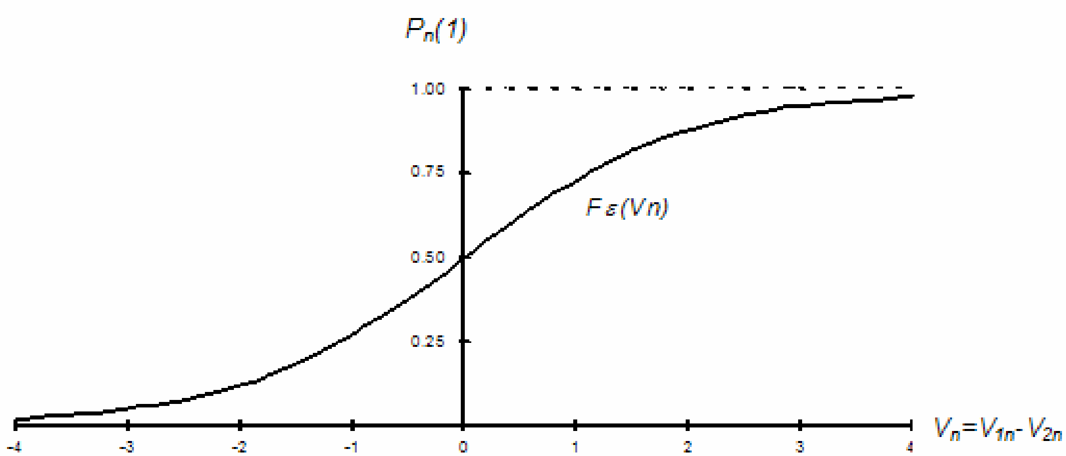
# Random Terms

- Capture imperfectness of information
- Distribution of *epsilons*
- Variance/covariance structure
  - Correlation between alternatives
  - Multidimensional decision
    - Example: Mode and departure time choice
- Typical models
  - Logit model (i.i.d. “Extreme Value” error terms, a.k.a Gumbel)
  - Probit model (normal error terms)

# Binary Choice

- Choice set $C_n = \{1,2\} \ \forall n$

$$P_n(1) = P(1|C_n) = P(U_{1n} \geq U_{2n})$$
$$= P(V_{1n} + \varepsilon_{1n} \geq V_{2n} + \varepsilon_{2n})$$
$$= P(V_{1n} - V_{2n} \geq \varepsilon_{2n} - \varepsilon_{1n})$$
$$= P(V_{1n} - V_{2n} \geq \varepsilon_n) = P(V_n \geq \varepsilon_n) = F_\varepsilon(V_n)$$



# Binary Probit

- "Probit" name comes from ***Prob**ability Un**it***

$\varepsilon_{1n} \sim N(0,\sigma_1^2)$

$\varepsilon_{2n} \sim N(0,\sigma_2^2)$

$\varepsilon_n \sim N(0,\sigma^2)$ *where* $\sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}$

$$f(\varepsilon) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\varepsilon}{\sigma}\right)^2}$$

$$P_n(1) = F_\varepsilon(V_n) = \int_{-\infty}^{V_n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\varepsilon}{\sigma}\right)^2} d\varepsilon = \Phi\left(\frac{V_n}{\sigma}\right)$$

where $\Phi(z)$ is the standardized cumulative normal distribution

# Binary Probit Normalization

- Relationship between Utility scale $\mu^*$ and Scale Parameter $\sigma$:

$$Var(\mu^* \varepsilon_n) = 1$$

*iff*

$$\mu^{*2} \operatorname{var}(\varepsilon_n) = 1$$

$$\Rightarrow \mu^* = \frac{1}{\sqrt{\operatorname{Var}(\varepsilon_n)}} = \frac{1}{\sigma}$$

- Usual normalization: $\sigma = 1$, implying $\mu^* = 1$

# Binary Logit Model

- "Logit" name comes from ***Log***istic Probability Un***it***

$\varepsilon_{1n}$ ~ *ExtremeValue (0,$\mu$)*

$\varepsilon_{2n}$ ~ *ExtremeValue (0,$\mu$)*

$\varepsilon_n$ ~ *Logistic (0,$\mu$)*

$$F_\varepsilon(\varepsilon_{1n}) = \exp\left[-e^{-\mu\varepsilon_{1n}}\right]$$

$$F_\varepsilon(\varepsilon_{2n}) = \exp\left[-e^{-\mu\varepsilon_{2n}}\right]$$

$$F_\varepsilon(\varepsilon_n) = \frac{1}{1+e^{-\mu\varepsilon_n}}$$

$$P_n(1) = F_\varepsilon(V_n) = \frac{1}{1+e^{-\mu V_n}}$$

# Why Logit?

- Probit does not have a closed form – the choice probability is an integral.
- The logistic distribution is used because:
  - It approximates a normal distribution quite well.
  - It is analytically convenient
  - Gumbel can also be “justified” as an extreme value distribution
- Logit does have “fatter” tails than a normal distribution.

# Logit Model Normalization

- Relationship between Utility Scale $\mu^*$ and Scale Parameter $\mu$

$$Var(\mu^* \varepsilon_n) = 1 \text{ iff}$$

$$\mu^* = \frac{1}{\sqrt{\mathrm{Var}(\varepsilon_n)}}$$

$$\text{where } Var(\varepsilon_n) = Var(\varepsilon_{2n} - \varepsilon_{1n}) = 2\pi^2/6\mu^2$$

# Logit Model Normalization

- Usual normalization: $\mu$ =1, implying $\mu^* = \frac{\sqrt{3}}{\pi}$

- Utility scale different from probit
  - Need to multiply probit coefficients by $\frac{\pi}{\sqrt{3}}$ to be comparable to logit coefficients

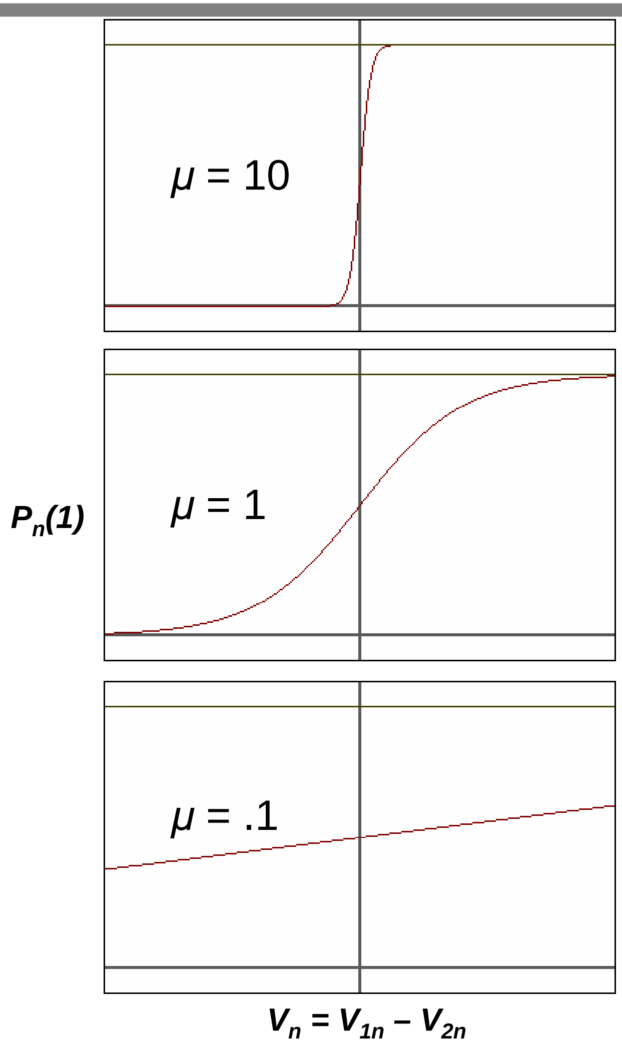
# Limiting Cases

- Recall: $P_n(1) = P(V_n \geqslant \varepsilon_n)$
  $= F_\varepsilon(V_{1n} - V_{2n})$

- With logit, $F_\varepsilon(V_n) = \dfrac{1}{1+e^{-\mu V_n}} = \dfrac{e^{\mu V_{1n}}}{e^{\mu V_{1n}} + e^{\mu V_{2n}}}$

- What happens as $\mu \rightarrow \infty$ ?

- What happens as $\mu \rightarrow 0$ ?



# Re-formulation

- $P_n(i) = P(U_{in} \geqslant U_{jn})$

$$= \frac{1}{1 + e^{-\mu (V_{in} - V_{jn})}}$$

$$= \frac{e^{\mu V_{in}}}{e^{\mu V_{in}} + e^{\mu V_{jn}}}$$

- If $V_{\text{in}}$ *and* $V_{\text{jn}}$ are linear in their parameters:

$$P_n(i) = \frac{e^{\mu \beta' x_{in}}}{e^{\mu \beta' x_{in}} + e^{\mu \beta' x_{jn}}}$$

# Multiple Choice

- Choice set $C_n$: $J_n$ alternatives, $J_n \geq 2$

$$P(i \mid C_n) = P[V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}, \forall j \in C_n]$$
$$= P\left[(V_{in} + \varepsilon_{in}) = \max_{j \in C_n} (V_{jn} + \varepsilon_{jn})\right]$$
$$= P\left[\varepsilon_{jn} - \varepsilon_{in} \leq V_{in} - V_{jn}, \forall j \in C_n\right]$$

# Multiple Choice

- Multinomial Logit Model
  - $\varepsilon_{jn}$ independent and identically distributed (i.i.d.)
  - $\varepsilon_{jn} \sim ExtremeValue(0,\mu) \quad \forall j$

$$F(\varepsilon) = \exp\left[-e^{-\mu\varepsilon}\right], \ \mu > 0$$

$$f(\varepsilon) = \mu e^{-\mu\varepsilon} \exp\left[-e^{-\mu\varepsilon}\right]$$

  - Variance: $\pi^2/6\mu^2$

$$\mathrm{P}(i \mid C_n) = \frac{e^{\mu V_{in}}}{\sum_{j \in C_n} e^{\mu V_{jn}}}$$

# Multiple Choice – An Example

- Choice Set $C_n = \{auto, bus, walk\} \; \forall n$

$$P(auto \mid C_n) = \frac{e^{\mu V_{auto,n}}}{e^{\mu V_{auto,n}} + e^{\mu V_{bus,n}} + e^{\mu V_{walk,n}}}$$

# Next Lecture

- Model specification and estimation
- Aggregation and forecasting
- Independence from Irrelevant Alternatives (IIA) property – Motivation for Nested Logit
- Nested Logit – specification, estimation and an example

1.201J / 11.545J / ESD.210J Transportation Systems Analysis: Demand and Economics
Fall 2008