Discrete Choice Modeling

Part 01
Introduction

A simple example (1/4)

	Journey to Work	
	Route 1	**Route 2**
Travel Cost	10 Yuan	20 Yuan
Travel Time	60 mins	30 mins
My Choice		✓

- Generally, the satisfaction of Route 2 is higher than the satisfaction of Route 1
- Specifically, 10-Yuan-saving cannot compensate for 30-mins-loss

A simple example (2/4)

Observation 1	Journey to Work	
	Route 1	Route 2
Travel Cost	10 Yuan	20 Yuan
Travel Time	60 mins	30 mins
My Choice		✓

Observation 2	Journey to Work	
	Route 1	Route 2
Travel Cost	15 Yuan	25 Yuan
Travel Time	45 mins	30 mins
My Choice	✓	

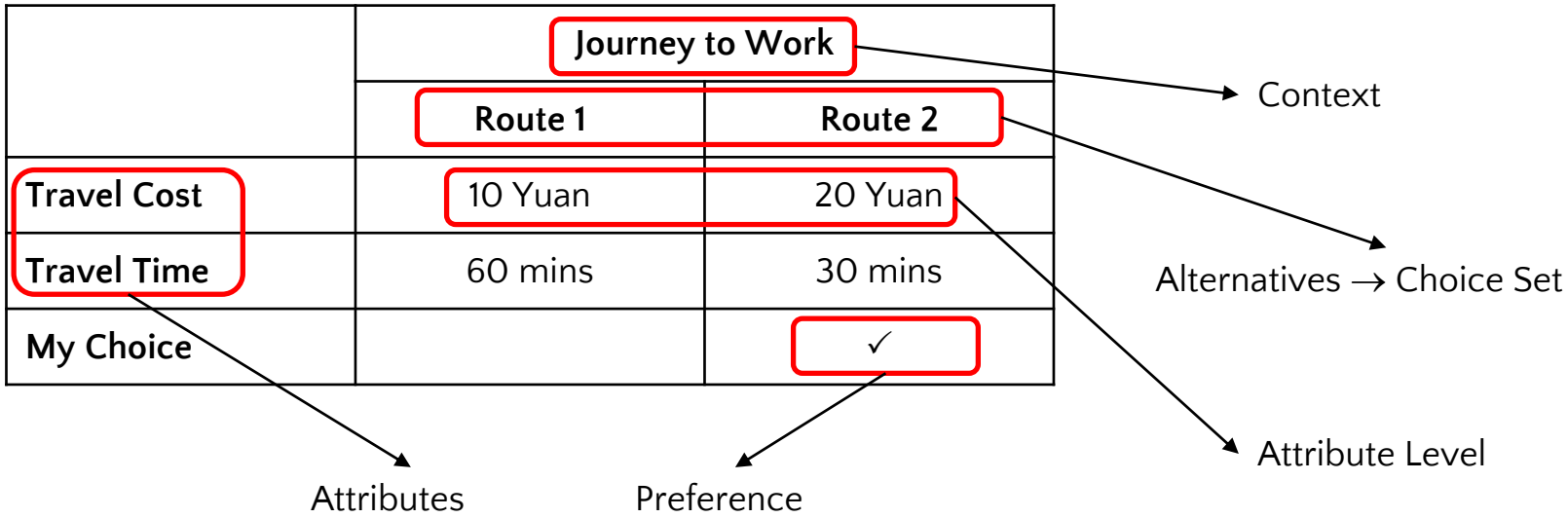
Observation 1:

- Rout 2 is 30 mins faster, 10 Yuan more expensive → choose Route 2
- Value of (saving) time > 10 Yuan per 30 mins (i.e. 20 Yuan per hour)

Observation 2:

- Rout 2 is 15 mins faster, 10 Yuan more expensive → choose Route 1
- Value of (saving) time < 10 Yuan per 15 mins (i.e. 40 Yuan per hour)

A simple example (3/4)



A simple example (4/4)

Individuals' choices are observed (in real life or experiments)

- Which travel mode to use for their commute
- Whether to buy a new car, where to live and work, etc.

From these choices, preferences and trade-offs are inferred

- Preference for car over train
- Trade-off between travel time and cost

Based on preferences and trade-offs, future choices are predicted

- Market share for a new travel mode
- Effects of road pricing, higher transit fares

As well as the benefits of policies

- How valuable is a reduction in total network travel time
- Is the transit service / transport policy worth funding

First term: utility (1/6)

	Journey to Work	
	Route 1	**Route 2**
Travel Cost	10 Yuan	20 Yuan
Flight Time	60 mins	30 mins
My Choice		✓

- Utility is a measure of preferences over some set of goods that satisfies human needs
- Utility represents satisfaction experienced by the consumer of a good

--- Wikipedia

- Utility cannot be measured directly
- For the above example, utility of Route 2 > utility of Route 1

First term: utility (2/6)

Cardinal utility:

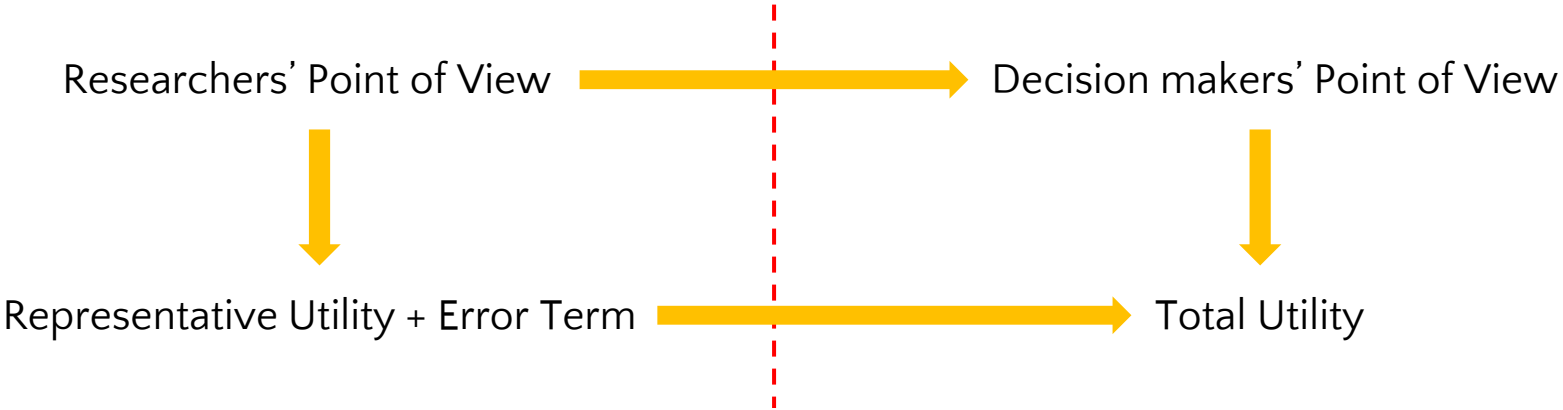
- Utility could be positive, negative or even zero
- Utility = 0 does not mean “nothing”, it actually means something
- Utility is relative, the absolute value means nothing
- Summation and subtraction are suitable, while multiplication and division are not

e.g. Car A (utility = 100), Car B (utility = 50), Car C (utility = 0)

- The utility of Car C is 0, that does not means Car C has no utility
- Car A is better than Car B by exactly the same amount by which Car B is better than Car C
- One cannot say Car A is twice better than Car B
- Let utility of Car A = 50, utility of Car B = 0, utility of Car C = -50, their relationship does not change

Only Difference of Utility Matters!

First term: utility (4/6)

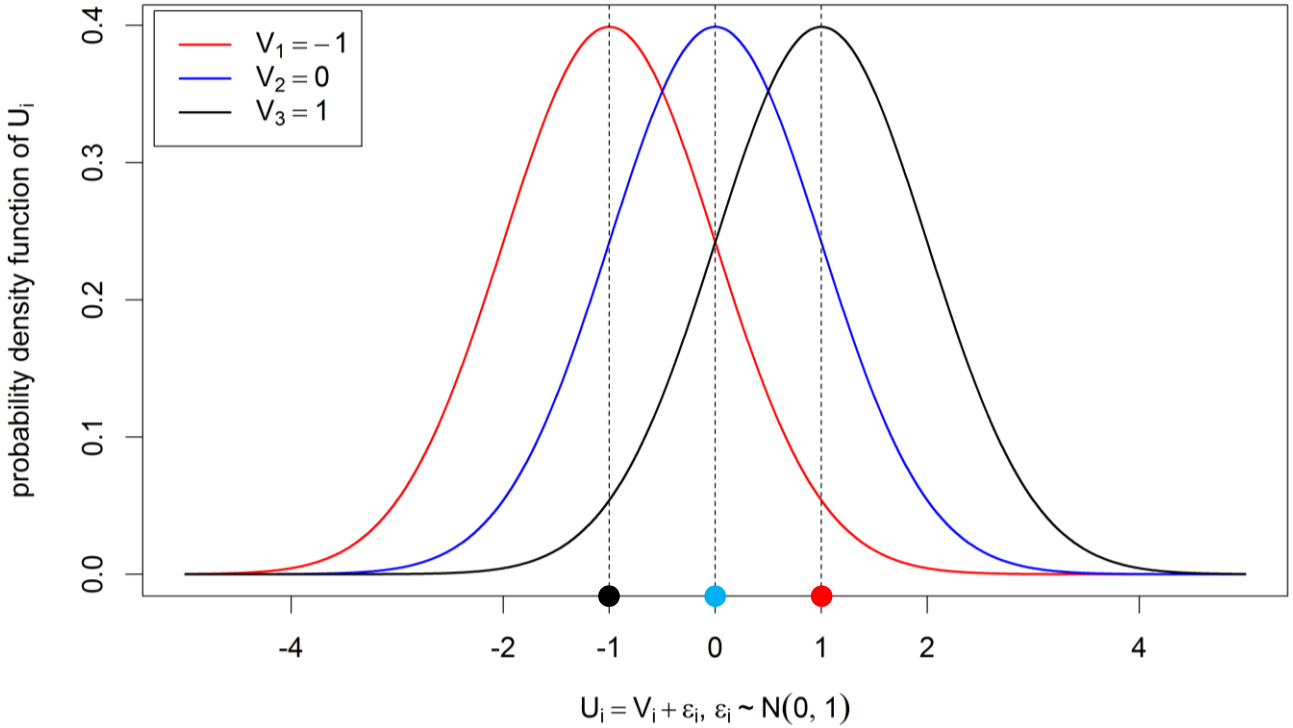


Random Utility Maximization (RUM) Theory

- Representative utility: everything that can be related to observed determinants
- Error term: everything else not considered, assumed to follow a certain distribution
- Alternative is chosen if its ***Total utility*** is highest
- Researchers can only predict choices up to a probability

$$\textit{Total Utility} = \textit{Representative Utility} + \textit{Error Term}$$

First term: utility (6/6)



Second term: preference (1/6)

Discrete choice of one alternative from a set of competing ones

Alternatives for journey to work	Commuter chooses
Bus	
Bicycle	
Drive own auto	✓
Carpool	
Walk	

- auto > bus, bicycle, carpool, walk

Second term: preference (2/6)

“Yes, I like this alternative” and “No, I don’t like this alternative”

Alternatives for journey to work	Commuter will consider
Bus	No
Bicycle	Yes
Drive own auto	Yes
Carpool	No
Walk	No

- auto > bus, carpool, walk
- bicycle > bus, carpool, walk

Second term: preference (3/6)

A complete ranking of alternatives from most to least preferred

Alternatives for journey to work	Ranking by likelihood use
Bus	4
Bicycle	2
Drive own auto	1
Carpool	3
Walk	5

- auto > bus, bicycle, carpool, walk
- bicycle > bus, carpool, walk
- carpool > bus, walk
- bus > walk

Second term: preference (4/6)

A complete ranking of alternatives from most to least preferred

- Based on decision makers' strongly reliable and valid cognitive abilities
- The middle part of a ranking is ambiguous and unreliable

"I like this alternative the best" and "I like that alternative the worst"

	Travel Time	Travel Cost	Most Liked	Least Liked
Route 1	30 mins	20 Yuan	✓	
Route 2	60 mins	10 Yuan		✓
Route 3	45 mins	15 Yuan		

Second term: preference (5/6)

Expressing degree of preference by rating alternatives on a scale

Alternatives for journey to work	rating by likelihood use (0 -10)
Bus	6
Bicycle	7
Drive own auto	10
Carpool	6
Walk	4

- auto > bus, bicycle, carpool, walk
- bicycle > bus, carpool, walk
- bus = carpool
- bus > walk
- carpool > walk

Second term: preference (6/6)

Expressing degree of preference by rating alternatives on a scale

- Based on decision makers' strongly reliable and valid cognitive abilities
- Hard to interpret the meaning of a difference between a rating of "4" and a rating "7"
- Respondents should be noticed with the best and the worst cases in advance

Third term: SP data (1/3)

Revealed preference (RP) data:

- Describe individuals' behavior or existing circumstances in the real world
  - how many cars do you have?
  - how did you come to school today?
  - how far from your home to your company?

Stated preference (SP) data:

- Describe individuals' potential behavior in a hypothetical context
  - if you are faced with two routes to wok, the only difference are their travel cost and time. Which one do you prefer?
- **We are talking about SP data only!**

Observation 1	Journey to Work	
	Route 1	Route 2
Travel Cost	10 Yuan	20 Yuan
Travel Time	60 mins	30 mins
My Choice		✓

Third term: SP data (2/3)

Revealed preference (RP) data:

- Depict the world as it is now
- Process inherent relationship between attributes
- Have only existing alternatives as observables
- Have high reliability and face validity
- Yield one observation per respondents at each observation point
- **RP data are to help us understand preferences within an existing market and technology structure**

Third term: SP data (3/3)

Stated preference (RP) data:

- Describe hypothetical or virtual decision contexts
- Control relationships between attributes, which permits mapping of utility functions with technologies different from existing ones
- Can include existing and/or proposed and/or generic (i.e., unlabeled) choice alternatives
- Seem to be reliable when respondents understand, are committed to and can respond to tasks
- Can yield multiple observations per respondents at each observation point
- **SP data provide insights into problems involving shifts in technological frontiers**

Level of measurement (1/5)

Measuring:

- Assigning numbers to empirical phenomena
- Numbers are easier to deal with in analysis than text

Level of measurement:

- How to interpret the numbers
- Determines which analysis approaches are allowed
- Nominal, Ordinal, Interval, and Ratio

Level of measurement (2/5)

Nominal:

- $1 \neq 2 \neq 3$
  - numbers just indicate the different categories
  - thus no ordering
  - numbers are fully interchangeable between categories
  - colors, gender, means of transport
- Example (mean of transport):
  "1=car; 2=bike; 3=train" is equivalent to "2=car; 3=bike; 1=train"
- Dichotomous:
  - variable of nominal level with only 2 categories
  - gender: 1=female, 2=male

Level of measurement (3/5)

Ordinal:

- 1 < 2 < 3 or 1 > 2 > 3
  - there is an order between categories
  - no equal differences between consecutive categories
  - hierarchical levels, level of education, rank order
  - when transforming, keep the order between categories the same
- Example (education level):
  - “1=bachelor degree; 2=master degree; 3=doctoral degree”
    is equivalent to “2=bachelor degree; 3=master degree; 7=doctoral degree”

Level of measurement (4/5)

Interval:

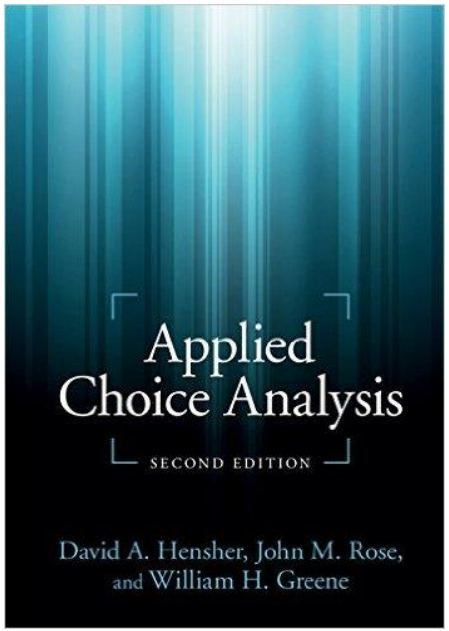
- 2–1 = 4–3; but 2 ≠ 2 * 1
  - order with equal differences between categories
  - no absolute zero value
  - preferences on rating scale, intelligence, °F, °C
  - when transforming, keep equal differences
  - any linear transformation wokrs (i.e., °F = 32 + 9/5 x °C)
- Example (temperature in Celsius or Fahrenheit):
  - right: difference 20°C–10°C = 2 x (15°C–10°C)
  - wrong: if temperature decreases from 20°C to 10°C, it does not become twice as cold (rate in °F: 68/50 = 1.36 ≠ 2)

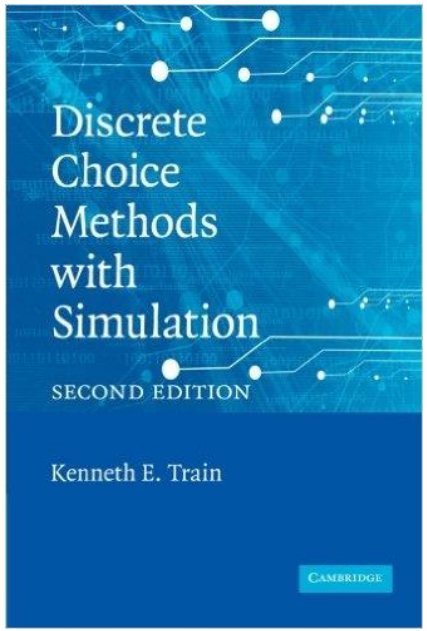
Level of measurement (5/5)

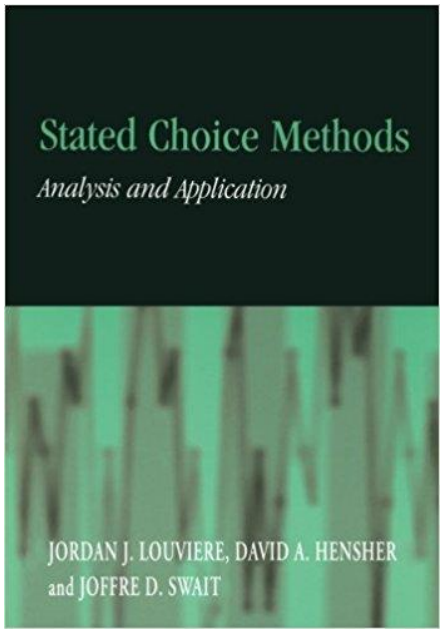
Ratio:

- 2 = 2*1
  - order & equal intervals & absolute zero value
  - weight, distance, age, temperature in Kelvin
  - when transforming, keep equal proportions
- Example:
  - 20 kilometers is twice as much as 10 kilometers
  - weight 40 kilos is twice as heavy as 20 kilos

Reference (1/1)







Thanks!

Any questions?

You can find me at

– pan.x@outlook.com