Global Initiative of Academic Networks (GIAN)

UNDERSTANDING THE INCREASING RELEVANCE OF CHOICE MODELS FOR ADVANCING TRANSPORTATION MODELLING IN SMART CITIES

MAY 22 -26, NAGPUR, INDIA

Module 3

Naveen Eluru, *University of Central Florida*

COURSE MODULES

Introduction	• Introduction to Smart City Technologies, their impact on Transportation
Stated Preference Module	• Background on Data Collection Approaches • Stated Preference Design and application
Traditional Discrete Choice Models	• Binary logit, multinomial logit, ordered logit, and count models
Advanced Discrete Choice Models	• Nested logit, mixed logit, maximum simulated likelihood estimation, regret minimization, discrete continuous models
Transportation Planning	• Current state of the art and recent advances

IN THIS MODULE

I will introduce choice modeling approaches for data analysis including binary logit, multinomial logit, ordered logit and count models

MODELS AND THEIR ROLE

- What is a model?
  - A mechanism to represent something of interest
    - Car engine
    - Human blood circulation
    - Transportation system
- Abstract and/or Physical in nature
  - A computer model of the house
  - A physical model of a house (in reduced scale)
- The problem of interest determines the class of models we are interested in
- In this class we focus our attention on abstract models – more specifically mathematical models

MATHEMATICAL MODELS

- Mathematical models represent the system of interest employing the underlying scientific theory
- What do we want to do with these models
  - Gain an understanding of the system
  - Replicate the performance of a system
  - Develop a common platform that can be employed by different interested parties
- The objective of model development is two fold
  - Understand the system under consideration
  - Predict how the systems will evolve (sometimes influence the change)
- Are the models perfect?
  - If the model involves completing an integral probably we can develop a perfect one
  - If the model is to interpret your thoughts (probably so that you can make sense of life!) … then we have a long way to go
  - Depending on what we are interested in the "perfection" of model varies

HOW TO DEVELOP MODELS

- We need to identify what we are modelling
  - Spread of epidemics, religion, language, internet etc.
  - Patterns of time spent on Facebook (for advertisers!)
  - Travel destination choice [Orlando or Las Vegas]
  - Travel mode choice
- Study the system and understand the underlying behavior
  - So if you want to deal with fever, it is important to understand what it does to a human system
  - Examining mode choice would require us to understand travel patterns of individuals
- In this class, we will examine models developed through empirical data collection
  - We see how the system behaves based on our recording of current behavior
  - So data becomes a very important component

DETERMINISTIC VERSUS PROBABILISTIC

- Models are either deterministic or probabilistic
  - Deterministic models are Laws such as gravity, Newton's Law – valid always
  - Probabilistic models work differently
- A deterministic model developed can be invalidated if there is one contradiction
  - When examining deterministic models such as any of the laws, we can refute them by just one contradiction
- In probabilistic models, to incorporate all possibilities a stochastic element is introduced, so as opposed to making exact statements, we are providing a probabilistic description
- In this class we restrict ourselves to probabilistic models that involve data
- The downside to this is to say a model is bad becomes a challenge – something we will learn in this class

8

CHOICE MODELS

MOTIVATION

- Choice modeling and econometric models provide an important analytical tool to study
  - Travel Behavior (travel mode, response to congestion pricing)
  - Land-Use Transportation Interactions (residential location, residential tenure, effect of built environment)
  - Activity Time-use Modeling (activities pursued, time invested, location of activity, travel mode)
  - Transportation Safety (driver injury severity, non-motorists' injury severity, safety concerns among specific demographic segments)
  - Non-motorized Travel (walking/bicycling to work, bicycle route choice)
  - Physical Heath (physical activity participation among children and adults)

CHOICE THEORY

- We are interested in understanding behavior of a large number of individuals or firms
- The overall behavior we observe is the manifestation of individual decisions
- For example, we measure traffic flow on the roadways as 1000 vehicles per hour.
- This is the results of many individuals deciding to travel on that road to perform something of interest to them
  - What we see is just the result of their choices
- Hence, to gain an understanding of different processes it is necessary to examine choices made by individual components (or Decision makers)

AGGREGATE VERSUS DISAGGREGATE

- Aggregate versus Disaggregate - Huge debate point
- Consider a simple example
  - We have two households in a zone: HH1 is low income and HH2 is high income. HH1 made 4 trips and HH2 made 6 trips
  - So at the aggregate level all you see is 10 trips being made by 2 HH, so we can create a simple model that gives trips as 5*(# of HHs)
  - Now, if we looked at the disaggregate level, we notice that the low income HH made 4 trips and the high income HH, so considering this information I can create a model that predicts trips as: 4(# of HHs)+2*(# of high income HHs)
  - Now lets examine trips for a zone with 2 high income households
    - The aggregate model will predict 10 trips (because we don't look at the type of the household)
    - The disaggregate model will predict (4*2+2*2) = 12 trips
- The disaggregate model is more likely to be true because we have focussed on the type of the HH or to put it simply we examined the decision makers choice rather than lumping all DMs as homogenous quantities

CHOICE THEORY

- A choice can be viewed as an outcome of a sequential decision-making process
  - Definition of the problem
  - Alternative generation
  - Evaluation of attributes of the alternative
  - Choice
  - Implementation of the choice
- Consider how you travel to the university
  - What is the best way to get to the university
  - Car, bus, metro, walk and bike
  - Attributes: time, cost and comfort
    - Travel time (Car 7 minutes, bus 15, metro 12, walk 20, bike 10)
    - Travel cost (car 3$, bus 1$, metro 2$, walk 0, bike 1$),
    - Comfort (Car Comfortable, Bus Uncomfortable, Metro Comfortable, walk Uncomfortable, Bike comfortable)
  - Choice: walk
  - Implementation: walk to work

CHOICE THEORY

- A choice is a collection of processes that define the following elements
  - Decision maker (in the example YOU)
  - Alternatives
  - Attributes of alternatives
  - Decision rule
- Please remember, not every choice is made so elaborately
  - For example, you don't decide how to get to work everyday; you made your decision once and stick with it as a habit
  - Individuals can follow habits, follow convention, or imitate someone else
  - We can represent such behavior in a well-developed model
    - For example, a person who walks to work regularly, we can generate only one alternative for that person i.e., to mimic his choice process we must realize how s/he generates the alternatives and if we could do that (which is a big IF mind you) we can even model such behavior)

ELEMENTS OF CHOICE PROCESS

- Decision Maker
  - Individuals or groups based on the choice of interest
    - Examining travel mode choice DM – individual
    - Vehicle ownership DM – household
    - Vacation Decisions DM - household
    - Land-use models DM – Travel Analysis Zone etc.
  - DMs might have varying tastes
  - DMs might face different choices
    - For example, for a person without a car, driving is not an alternative
- Alternatives
  - Any choice is made from a non-empty set of alternatives
  - Universal choice set: all the alternative offered by the environment to the population
  - Feasible choice set: alternatives feasible for a DM (if I have a car then driving to school is feasible)
  - Evoked choice set: alternatives that are actually considered by the individual at the time of decision making

ELEMENTS OF CHOICE PROCESS

- Alternative attributes
  - Alternatives are characterized by attributes from the point of view of the DM
  - Involves both certain and uncertain values
    - For example, when we model travel mode choice, we assume a travel time for all modes, but the travel time value is affected by congestion (it is hard to predict the extent of this effect)
- Decision rule
  - The internal mechanism used by the DM to process the information and arrive at the unique choice
  - Different rules
    - Dominance
    - Satisfaction
    - Lexicographic
    - Utility

ELEMENTS OF CHOICE PROCESS - DECISION RULE

- Dominance
  - Under this rule, for one alternative at least one attribute is better and for all other attributes it is no worse
  - No controversy over this process
  - It is rarely the case in reality – probably helpful in eliminating inferior choices
  - You can make it better by defining what is "better" through a pre-determined threshold
- Satisfaction
  - A level of aspiration based on decision makers expectation is generated to develop a level that serves a satisfaction criterion
  - For example, in terms of travel time, I can set a limit of 50 minutes, so any alternative that fails this rule will be ignored
  - Again not necessary that you will end up with one option
  - Typically employed to eliminate inferior alternatives

ELEMENTS OF CHOICE PROCESS - DECISION RULE

- Lexicographic
  - Rank all attributes by level of importance
  - The DM picks the alternative that performs best on the top rated attribute
  - For example if travel time is the most important attribute, the alternative with lowest travel time is chosen
  - If there is a tie for the most important attribute for some alternatives, the next important alternative is chosen
  - You can consider a combination of lexicographic and satisfaction based rules!

ELEMENTS OF CHOICE PROCESS - DECISION RULE

- Utility
  - In this process, we try to generate a single scalar measure for each alternative through a function of the attributes
  - So for travel mode, you have a scalar for car, bus, walk etc. which is a function of time, cost and comfort – scalar is referred to as utility
  - The alternative that provides the highest value of utility is chosen!
  - The approach accommodates the compensatory effects
    - i.e. we try to identify trade-offs across the different attributes
  - In this rule, it is possible to choose an alternative that has higher cost, provided it somehow provides better comfort and time reduction.. Thus it compensates across attributes by capturing such trade-offs
  - In other rules we don't interact across attributes

DISCRETE CHOICE THEORY

- In this approach we have to come up with a way to compute utility .. Typically an additive form of utility is employed
- For mode choice example for alternative i
- $Ui = b_0 + b_1 time + b_2 cost + b_3 comfort$ where parameters express the tastes of the commuter
- The idea is that the alternative that provides with the highest U is chosen!
- Utility is a cardinal value. i.e. we cannot say anything about it; a utility of 10 or 1000 does not provide any information
- We can only compare across the alternatives and choose the highest utility
- Also, an interesting property referred to as transitive property holds i.e. if $U_A > U_B$ and $U_B > U_C$ we assume $U_A > U_C$
- This might not "truly" hold in some choice settings based on the individual or DM

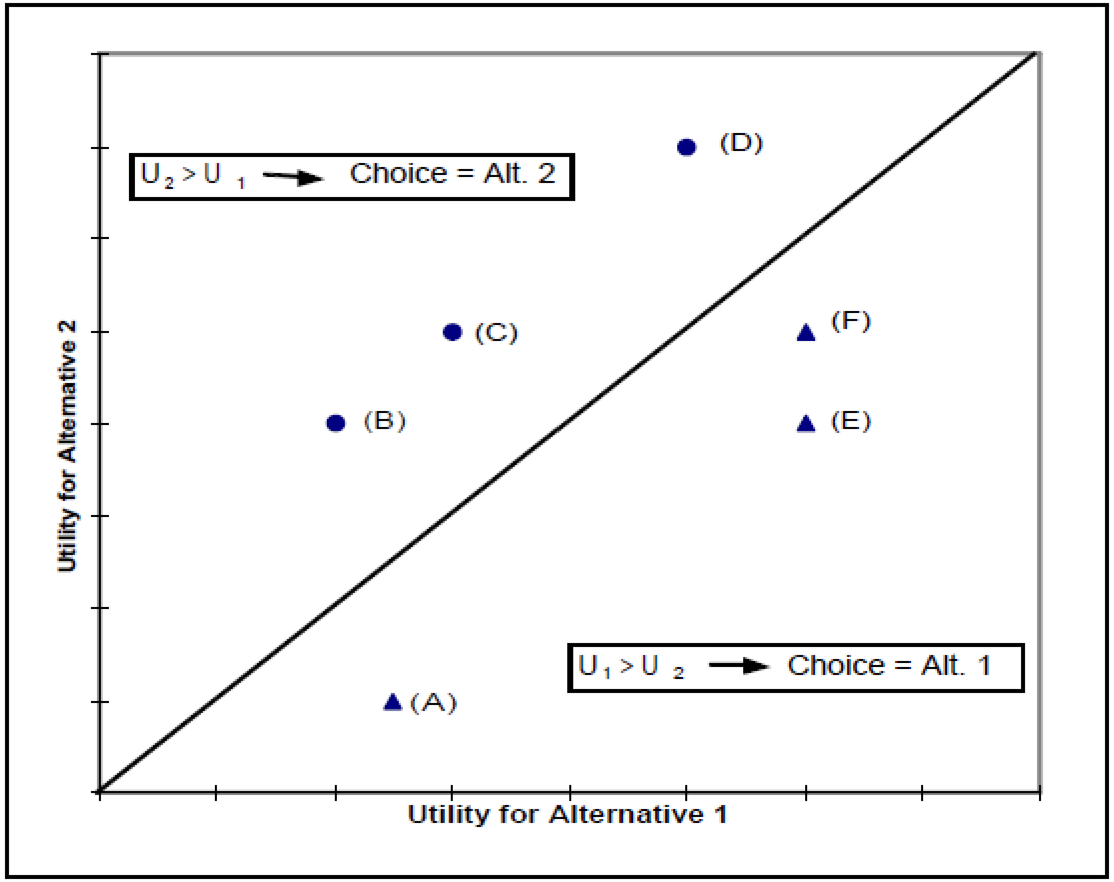
PROBABILISTIC CHOICE THEORY

- Utility theory directly cannot be applied in practice because people are not like machines i.e. we cannot predict how people act
- Sometimes it is observed that people do not choose the alternative with highest utility and sometimes the transitive property is violated -> so researchers started accounting for this weird (according to researchers) behavior through a error term
- Addressing this error in modeling choice processes gave rise to two schools of thought
  - Psychology
  - Economics

PSYCHOLOGICAL SCHOOL OF THOUGHT

- In psychology experiments are conducted in a controlled environment
- Hence in these experiments if the DM makes different choice under exactly identical utility measurements, the error is supposed to occurring because of the inherent probabilistic nature of the choice process
- So what psychologists claim is that they can exactly measure the choice process and the error is induced because the choice process is itself probabilistic
- So the error is because of this (not because of accuracy in utility computation)
- This results in a Constant utility Approach

ILLUSTRATION OF DETERMINISTIC CHOICE FOR 2 ALTERNATIVES



ECONOMIC SCHOOL OF THOUGHT

- In this school of thought, the researchers believed that the DM knows what s/he is exactly doing. However, because we cannot collect all the data that was employed in the choice process, the analyst misses some components that affect the choice and hence we have an error
- In this the error component refers to the "missing information"
- Economics experiments are rarely controlled and hence this is a natural assumption for economists
- This results in a Random Utility Approach (RUM)
- We will examine the RUM approach for remainder of the course

RANDOM UTILITY APPROACH (RUM)

- In the random utility approach, we assume that an individual always chooses the alternative with highest utility
- The utilities are unknown to the analyst with certainty; hence we treat these utilities as random variables
- From this perspective, for a DM "n" probability of choosing alternative i is equal to probability that utility of alternative i is greater than or equal to the utilities of all other alternatives in the choice set
  - $P(i|C_n) = Pr[U_{in} \geq U_{jn}, \text{ all } j \in C_n]$
- We derive choice probabilities by assuming a joint probability distributions for the set of random utilities $\{U_{in}, i \in C_n\}$

RANDOM UTILITY APPROACH (RUM)

- The basis for this distributional assumption is about different underlying sources of randomness
  - Unobserved attributes
    - For example, Data is not available on life styles
  - Unobserved taste variations
    - People have preferences.. Some people love driving (so they opt to drive)
  - Measurement errors and imperfect information
    - Income reporting is typically under-reported
  - Proxy variables
    - Some variables are not directly measured, but some proxies are measured

RANDOM UTILITY APPROACH (RUM)

- Random utility of an alternative is partitioned into two components: (1) observed utility (systematic) and (2) unobserved utility
- $U_{in} = V_{in} + \varepsilon_{in}$
- Hence we can write $P(i|C_n) = Pr[U_{in} \geq U_{jn}, \text{ all } j \in C_n]$ as
  - $P(i|C_n) = Pr[V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}, \text{ all } j \in C_n]$
  - To derive a probabilistic model we need to make assumptions on the error structures
  - ε_{jn} has a zero mean (random disturbance that is not observable across the data)

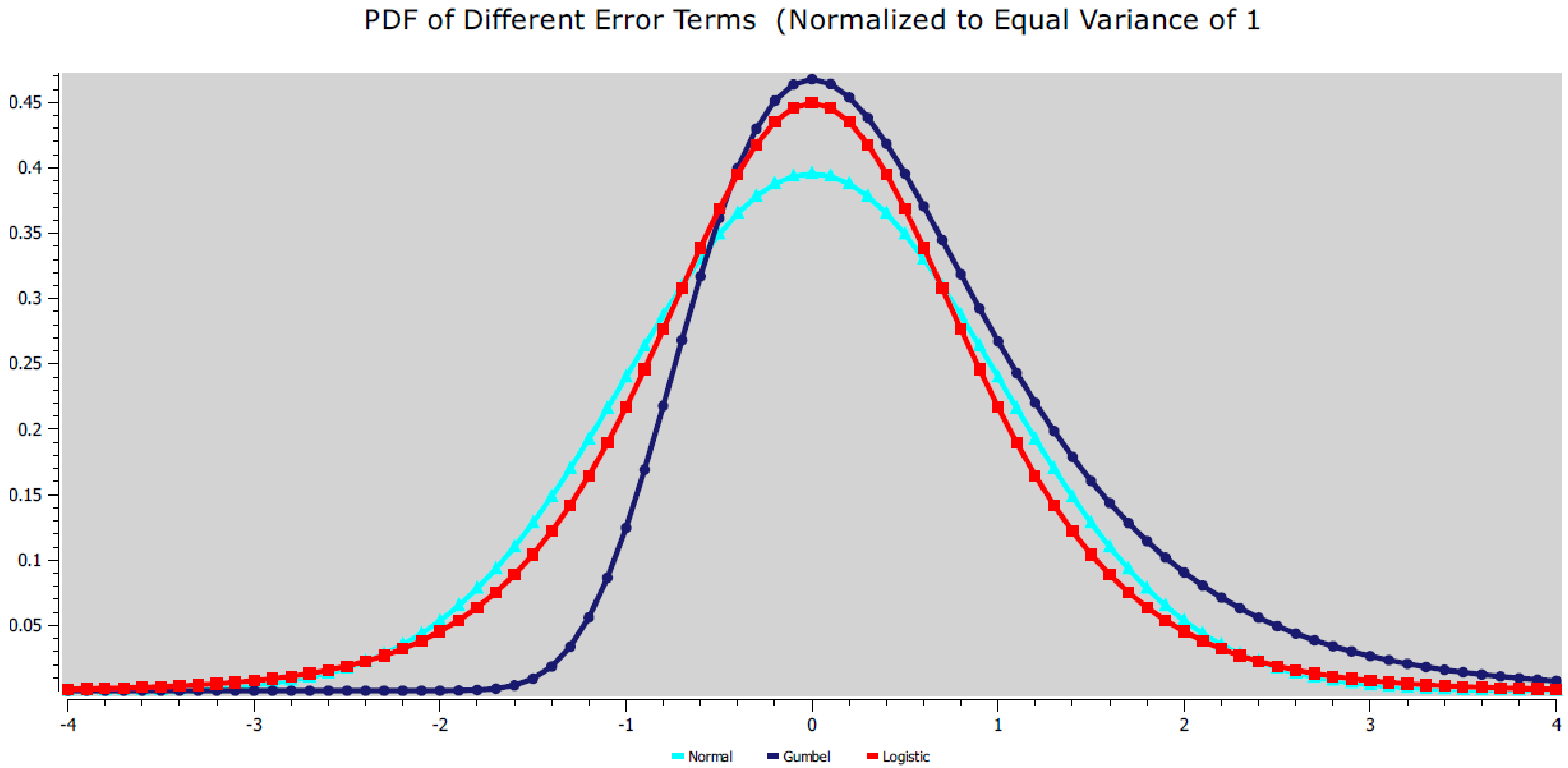
ERROR STRUCTURES

- Now the error structures we assume makes a huge difference to the model structure and form (and its implications)
- Lets review some properties of distributions we will use in this course
  - Normal
  - Gumbel
  - Logistic

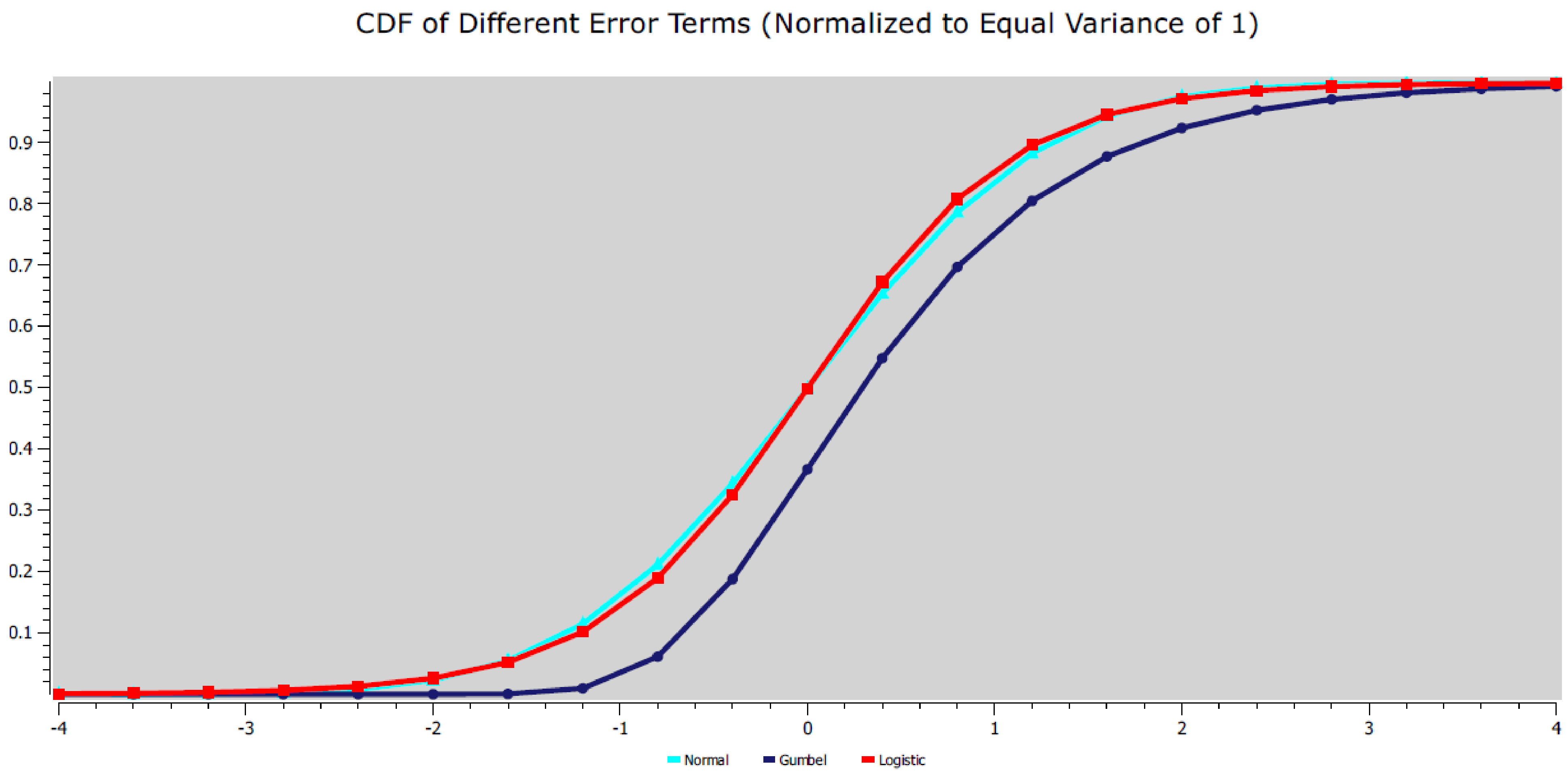
ERROR STRUCTURES

- Normal
  - PDF: $\frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}}$; CDF $= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{x^2}{2}} dx$
- Gumbel (Extreme value or Type I)
  - PDF $= \frac{e^{-u/\theta} e^{-e^{-u/\theta}}}{\theta}$; CDF $= e^{-e^{-u/\theta}}$
  - G(0, θ) where 0 is the mode and var $= {(\pi^2\theta^2)}/{6}$;
- Logistic
  - PDF $= \frac{e^{-u/\theta}}{(1+e^{-u/\theta})^2}$; CDF $= \frac{1}{(1+e^{-u/\theta})}$
  - L(0, θ) where 0 is the mean and var $= {(\pi^2\theta^2)}/{3}$

PLOTS - PDF


PDF of Different Error Terms (Normalized to Equal Variance of 1
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
-4
-3
-2
-1
0
1
2
3
4
Normal
Gumbel
Logistic

PLOTS - CDF



CDF of Different Error Terms (Normalized to Equal Variance of 1)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
-4
-3
-2
-1
0
1
2
3
4
Normal
Gumbel
Logistic

BINARY CHOICE MODELS

BINARY CHOICE MODELS

- Now lets start examining the case where we have two discrete alternatives in the choice set
- The reasons we are examining binary choice are:
  - Simplicity of binary choices allows us to develop a range of practical models
  - Many conceptual properties can be illustrated; Solutions from this can be applied to more complicated situations
- Individual n, alternatives i and j
  - Probability of *i* is: $P_n(i) = Pr(U_{in} \geq U_{jn})$
  - Probability of *j* is : $P_n(j) = 1 - P_n(i)$
- $U_{in} = V_{in} + \varepsilon_{in}$; $U_{jn} = V_{jn} + \varepsilon_{jn}$
- $P_n(i) = Pr(U_{in} \geq U_{jn})$
- $P_n(i) = Pr(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn})$
- $P_n(i) = Pr(\varepsilon_{jn} - \varepsilon_{in} \leq V_{in} - V_{jn})$

SOME PROPERTIES OF UTILITIES

- From the above expression we see that the probability is a function of the difference of utilities
  - So magnitude of the utilities do not matter, lets say I add 10 units to V_{in} and V_{jn}, it does not affect the probability, only differences matter
  - Similarly if we multiply the utilities also it does not make any difference to the eventual choice
- Effect of addition
- $U_{in} = V_{in} + \varepsilon_{in}$; $U_{jn} = V_{jn} + \varepsilon_{jn}$ ->Add both utilities with K
- $P_n(i) = Pr(K+U_{in} \geq K+U_{jn})$
- $P_n(i) = Pr(\varepsilon_{jn} - \varepsilon_{in}) \leq (V_{in} - V_{jn})$

SOME PROPERTIES OF UTILITIES

- Effect of scale
  - $U_{in} = V_{in} + \varepsilon_{in}$; $U_{jn} = V_{jn} + \varepsilon_{jn}$ ->Multiply both utilities with μ
  - $P_n(i) = Pr(\mu U_{in} \geq \mu U_{jn})$
  - $P_n(i) = Pr(\mu V_{in} + \mu\varepsilon_{in} \geq \mu V_{jn} + \mu\varepsilon_{jn})$
  - $P_n(i) = Pr(\mu(\varepsilon_{jn} - \varepsilon_{in}) \leq \mu(V_{in} - V_{jn}))$

- This is the reason why we can set the variance to any suitable value of choice

BINARY CHOICE MODELS

- Unobserved component
  - We have $P_n(i) = Pr(\varepsilon_{jn} - \varepsilon_{in} \leq V_{in} - V_{jn})$
- Now the equation above looks like a typical cdf term of a random variable
- If ε_{in} and ε_{jn} are random variables $\varepsilon_{jn} - \varepsilon_{in}$ will also be a random variable
- Lets assume that $\varepsilon_{in}, \varepsilon_{jn}$ are normally distributed; in this case $\varepsilon_{jn} - \varepsilon_{in}$ will also be normally distributed with mean given by mean(ε_{jn})+mean(ε_{in}) and variance given by var(ε_{jn})+var(ε_{in})

BINARY CHOICE MODELS

- To make this simple, we would ideally want the variance of the resulting error term (ε_{jn} - ε_{in}) to be 1. So, we can choose ε_{jn} and ε_{in} to have var of ½
- So we start with ε_{jn} , ε_{in} N ~ (0, ½)
- This distributional assumption results in a binary probit model N ~ (0, 1)
- $P_n(i) = \Phi(V_{in} - V_{jn})$; where Φ is the standard normal distribution
- This is referred to as the **BINARY PROBIT MODEL**

BINARY LOGIT

- If ε_{in} and ε_{jn} are random variable ε_{jn} - ε_{in} will also be a random variable
- Now instead of normal assumption, let us assume that ε_{in} and ε_{jn} are Gumbel distributed
- This assumption yields the logistic model
- $P_n(i) = CDF\ Logit(V_{in} - V_{jn})$
- $= \frac{1}{(1+e^{-u/\theta})}$ for our case (θ=1)
- $\frac{1}{(1+e^{-(V_{in} - Vjn)})} = \frac{e^{V_{in}}}{(e^{V_{in}} + e^{V_{jn}})} = \frac{e^{\beta' X_{in}}}{(e^{\beta' X_{in}} + e^{\beta' X_{jn}})}$
- This is the binary logit model

LOGIT VS PROBIT

- Binary logit expression
  - Numerator: exp(alternative utility)
  - Denominator: sum of exp(alternative utility)
- Advantage compared to binary probit?
  - Clear formula for alternative probability

BINARY CHOICE MODELS: PROBIT VS LOGIT

- How do the probit and logit compare?
- Probit variance is 1 and logit it is $^{\pi^2}/_3$
  - We started with std. gumbel, var = $^{\pi^2}/_6$; var of logistic = $^{\pi^2}/_6 + ^{\pi^2}/_6 = ^{\pi^2}/_3$
- Coefficients ratio for logit and probit is sqrt($^{\pi^2}/_3$) $= \frac{\pi}{\sqrt{3}}$ because that is the ratio of the standard deviation of error terms
  - It will hold approximately ($\frac{\pi}{\sqrt{3}} \approx 1.8$)

BINARY CHOICE MODELS

- Systematic component
- Let Drive and Transit be the modes available for individual n
- $V_D = \beta_D + \beta_{D,inc} * Inc_n$
- $V_T = \beta_T + \beta_{T,inc} * Inc_n$
- Since we said that we do not really care about magnitude we can manipulate both equations by the same amount; lets reduce V_D and V_T by $\beta_D + \beta_{D,inc} * Inc_n$
- Now $V_D = 0$; $V_T = (\beta_T - \beta_D) + (\beta_{T,inc} - \beta_{D,inc}) * Inc_n$
- Replace $(\beta_T - \beta_D)$ with (β_T) and $(\beta_{T,inc} - \beta_{D,inc})$ with $(\beta_{T,inc})$ because we cannot estimate 2 parameters as all that matters is difference

EXAMPLE

- Lets see how this works for an example: Drive vs Transit
- Binary probit and logit
- Attributes
  - Alternative specific constant (ASC), In-vehicle travel time(IVTT); Out-of-vehicle travel time (OVTT); Cost (cents); Income (in 000s)

	ASC	IVTT	OVTT	Cost	Inc	Chosen
D	0	12	7	1.5	0	0
Tr	1	10	8	0.5	30	1

EXAMPLE

	ASC	IVTT	OVTT	Cost	Inc	Chosen
D	0	12	7	150	0	0
Tr	1	10	8	50	30	1

Coefficients

	ASC	IVTT	OVTT	Cost	Inc
Probit	-1.3	-0.04	-0.06	-0.004	-0.007
Logit	-2.3	-0.072	-0.11	-0.007	-0.013

	Utility (Dr)	Utility (TR)	Probability	
			Dr	TR
Probit	-1.5	-2.59	0.86	0.14
Logit	-2.7	-4.67	0.88	0.12

MODEL ESTIMATION

- So far we have discussed how to compute the probabilities
- Now we will start examining how do we estimate the parameter values
  - How would you go about estimating the model for a dataset
  - In the dataset for each individual we have information on the choice made.
- For the travel mode choice, we will be provided with information on whether D or T are chosen
  - In linear regression we decided the parameters should be values that reduce the square of the difference between "dependent variable" and "predicted value of dependent variable"
- How will this be different for discrete choice case
- The dependent variable here is a choice between multiple alternatives
  - In the binary case between 2 alternatives
- Any ideas?

MODEL ESTIMATION

- The objective of the parameters should be such that we correctly predict the "choice"
- Consider the following example

	ASC	IVTT	OVTT	Cost	Inc	Chosen
D	0	12	7	150	0	0
Tr	1	10	8	50	30	1

- We want coefficients of different variables such that the probability of Tr is 1 and probability of D is 0. This is not possible
- So, we want to penalize deviation from 1 for the chosen alternative
- A possible approach Min (1-predicted prob for chosen alternative)^2

MODEL ESTIMATION

- Intuitive and easy; however people really did not like using a continuous based error approach to a discrete problem
- So a max. likelihood approach was suggested
- What you do here is try to maximize the predicted probability of the chosen alternative
- Max (Predicted prob for chosen alternative)
- Now we do this for all individuals in the dataset
- Likelihood function is $\prod_{n=1}^{N}$ (Predicted prob for chosen alternative)

MAXIMUM LIKELIHOOD APPROACH

- Lets say we have n individuals
- For every individual we have P_{Dn}, P_{Tn}
- Also, define δ_{Dn}, δ_{Tn} such that $\delta_{Dn} = 1$ if D is chosen by individual n and 0 otherwise, same for δ_{Tn}.
- Now our objective is to estimate parameters such that we maximize the chance to predict the chosen alternatives
- For example, lets say ind. 1 chose T, then we want our probability for T (P_{Tn}) as close to 1 as possible.

MAXIMUM LIKELIHOOD APPROACH

- For this purpose, we define what is called a likelihood function; For individual n this is how it will look like
  - $L_n(\beta_1, \beta_2, \ldots \beta_K) = {P_{Dn}}^{\delta_{Dn}} \, {P_{Tn}}^{\delta_{Tn}}$
- The function is defined such that the only contribution to the function comes from the chosen alternative (because one of the δs is 0)
- Now to get estimates for the entire dataset set
  - $L(\beta_1, \beta_2, \ldots \beta_K) = \prod_{n=1}^{N} {P_{Dn}}^{\delta_{Dn}} \, {P_{Tn}}^{\delta_{Tn}}$
- For example with 3 individuals in the data with first two choosing D and last one choosing T
  - $L(\beta_1, \beta_2, \ldots \beta_K) = P_{D1} * P_{D2} * P_{T3}$
- Now we want to maximize this function to obtain our parameters
- For the sake of convenience we take the log of the above function
  - $\mathcal{L}(\beta_1, \beta_2, \ldots \beta_K) = \sum_{n=1}^{N} (\delta_{Dn} ln P_{Dn} + \delta_{Tn} ln P_{Tn})$

MAXIMUM LIKELIHOOD APPROACH

- Now we maximize the likelihood function to estimate the β vector. This approach is referred to as the maximum likelihood approach
- So we formulate the problem as
  - Max $L(\beta_1, \beta_2, \dots \beta_K) = \sum_{n=1}^{N}(\delta_{Dn} lnP_{Dn} + \delta_{Tn} lnP_{Tn})$
- We can solve for β vector by differentiating the above function w.r.t each β_k
- $L(\beta_1, \beta_2, \dots \beta_K) = \sum_{n=1}^{N}(\delta_{Dn} lnP_{Dn} + \delta_{Tn} lnP_{Tn})$
- $\frac{\partial L}{\partial \beta_k} = \sum_{n=1}^{N}\left(\delta_{Dn}\frac{1}{P_{Dn}}\frac{\partial P_{Dn}}{\partial \beta_k} + \delta_{Tn}\frac{1}{P_{Tn}}\frac{\partial P_{Tn}}{\partial \beta_k}\right) = 0$;
  - for k =1, 2, 3…K

MAXIMUM LIKELIHOOD APPROACH

- $\sum_{n=1}^{N}\left(\delta_{\mathrm{Dn}}\frac{1}{P_{Dn}}\frac{\partial P_{Dn}}{\partial\beta_{\mathrm{k}}}+\delta_{\mathrm{Tn}}\frac{1}{P_{Tn}}\frac{\partial P_{Tn}}{\partial\beta_{\mathrm{k}}}\right)=0$
- In the linear regression case the derivative allowed us to get a formula for β
- In the discrete choice case can we get the formula for β ?
  - No
- So we will have to solve for the solution with this formula
- We can show that LL function is globally concave i.e. single optimal solution
- Lets investigate the expression further for binary logit

MAXIMUM LIKELIHOOD APPROACH

- $\sum_{n=1}^{N}\left(\delta_{\mathrm{Dn}}\frac{1}{P_{Dn}}\frac{\partial P_{Dn}}{\partial\beta_{\mathrm{k}}}+\delta_{\mathrm{Tn}}\frac{1}{P_{Tn}}\frac{\partial P_{Tn}}{\partial\beta_{\mathrm{k}}}\right)=0$ - Eqn (A)
- $P_{Dn}=\frac{e^{\beta' X_{Dn}}}{\left(e^{\beta' X_{Dn}}+e^{\beta' X_{Tn}}\right)}=\frac{e^{\beta' X_{Dn}}}{Q}$ where $Q=e^{\beta' X_{Dn}}+e^{\beta' X_{Tn}}$
- $\frac{\partial P_{Dn}}{\partial\beta_{\mathrm{k}}}=\frac{Q*e^{\beta' X_{Dn}}*X_{Dn}-e^{\beta' X_{Dn}}*\left(e^{\beta' X_{Dn}}*X_{Dn}+e^{\beta' X_{Tn}}*X_{Tn}\right)}{Q^2}$
- $=P_{Dn}*x_{Dn}-P_{Dn}(P_{Dn}*x_{Dn}+P_{\mathrm{Tn}}*x_{Tn})$
- Similarly, $\frac{\partial P_{\mathrm{Tn}}}{\partial\beta_{\mathrm{k}}}=P_{\mathrm{Tn}}*x_{Tn}-P_{\mathrm{Tn}}(P_{Dn}*x_{Dn}+P_{\mathrm{Tn}}*x_{Tn})$
- Substitute these in Equation (A)
- $(\delta_{\mathrm{Dn}}*[x_{Dn}-(P_{Dn}*x_{Dn}+P_{\mathrm{Tn}}*x_{Tn})]+\delta_{\mathrm{Tn}}*[x_{Tn}-(P_{Dn}*x_{Dn}+P_{\mathrm{Tn}}*x_{Tn})])$
- $=(\delta_{\mathrm{Dn}}*x_{Dn}+\delta_{\mathrm{Tn}}*x_{Tn}-(\delta_{\mathrm{Dn}}+\delta_{\mathrm{Tn}})(P_{Dn}*x_{Dn}+P_{\mathrm{Tn}}*x_{Tn}))$
- $=(\delta_{\mathrm{Dn}}*x_{Dn}+\delta_{\mathrm{Tn}}*x_{Tn}-(P_{Dn}*x_{Dn}+P_{\mathrm{Tn}}*x_{Tn}))$

MAXIMUM LIKELIHOOD APPROACH

- $\sum_{n=1}^{N}\left(\delta_{\mathrm{Dn}}\frac{1}{P_{Dn}}\frac{\partial P_{Dn}}{\partial\beta_{\mathrm{k}}}+\delta_{\mathrm{Tn}}\frac{1}{P_{Tn}}\frac{\partial P_{Tn}}{\partial\beta_{\mathrm{k}}}\right)=0$
- $\sum_{n=1}^{N}((\delta_{\mathrm{Dn}}-P_{Dn})\,x_{Dn}+(\delta_{\mathrm{Tn}}-P_{\mathrm{T}n})x_{Tn})=0$
- Lets say we have only one ASC for Tr in the model
- $x_{\mathrm{Tn}}=1$ and $x_{\mathrm{Dn}}=0$
- $\frac{\partial\mathrm{L}}{\partial\beta_{\mathrm{ASCT}}}=\sum_{n=1}^{N}(\delta_{\mathrm{Tn}}-P_{\mathrm{T}n})=0$
- $\sum_{n=1}^{N}\delta_{\mathrm{Tn}}=\sum_{n=1}^{N}P_{\mathrm{Tn}}$
- Divide both sides by N
- $(\sum_{n=1}^{N}\delta_{\mathrm{Tn}})/\mathrm{N}=(\sum_{n=1}^{N}P_{\mathrm{Tn}})/N$
- Sample share for Transit $\mathrm{S}_{\mathrm{Tn}}=(\sum_{n=1}^{N}P_{\mathrm{Tn}})/N$
- Hence when ASC for Tr is the only variable, sample share is same as predicted share

REMARKS

- This is the equivalent of the sample mean information in linear regression
- The worst prediction you can do is provide the sample share from the population
- So if 20 out of 100 people use transit
- The easiest estimate of probability is 0.2 for transit mode
- Even worse than this is the equal share model. If there are two modes, we can always guess a 0.5 value for each mode
- Why do we care about these?
  - They are the yardsticks with which we measure
- You just need to apply the estimation method by employing the appropriate probability computation
  - Logit, probit or any other distributions
- ML approach is generic to all models

SOLVE FOR THE ESTIMATE

- $\sum_{n=1}^{N}\left(\delta_{\mathrm{Dn}}\frac{1}{P_{Dn}}\frac{\partial P_{Dn}}{\partial \beta_{\mathrm{k}}}+\delta_{\mathrm{Tn}}\frac{1}{P_{Tn}}\frac{\partial P_{Tn}}{\partial \beta_{\mathrm{k}}}\right)=0$
- We need to find β_k that satisfy this condition
- Unfortunately we can't do it easily
- So what we do is we set $\beta_k = 0$ (for example)
- Then evaluate $\nabla L, \nabla^2 L$
- We use the Newton-Raphson method for this
- In this approach
- $\beta_{kn} = \beta_{k(n-1)} - [\nabla^2 L(\beta_{k(n-1)})]^{-1}[\nabla L(\beta_{k(n-1)})]$
- We stop when difference between β_k from n and n-1 iterations is small
- An illustration
- There is huge research on doing this better
- Other methods BFGS, BHHH, DFP etc. (A course in Non-linear Optimization)

GOODNESS OF FIT MEASURES

- Benchmarks
  - Equal share model
    - L(0) =N * ln(1/K) where K is no. of alternatives
  - Market share model – Constants only model
    - $L(C) = \sum_{i=1}^{K}\{N_i * \ln(Ni/N)\}$
  - Perfect Model – for perfect model what is the value of L
    - 0
- Measure 1
  - $\rho_0{}^2 = 1 - \frac{L(\beta)}{L(0)}$
- Measure 2
  - $\rho_c{}^2 = 1 - \frac{L(\beta)}{L(C)}$
  - Adjusted $\overline{\rho_c{}^2} = 1 - \frac{L(\beta) - (\text{no. of parameters excluding constant})}{L(C)}$
- The comparison with the constants model is the most appropriate comparison

GOODNESS OF FIT MEASURES

- Other measures
  - % right measure $= \frac{100}{N} \sum_{i=1}^{N} \{y_n\}$ where y_n is 1 if the predicted probability for the chosen share is the highest; 0 otherwise
  - Avg. probability of correct prediction $= \frac{1}{N} \sum_{n=1}^{N} (\delta_{Dn} P_{Dn} + \delta_{Tn} P_{Tn})$
- The measure used most often however is the Log-likelihood $(\sum_{n=1}^{N} (\delta_{Dn} lnP_{Dn} + \delta_{Tn} lnP_{Tn})$
  - The log-likelihood penalizes error substantially
  - When the probability is close to 1 the penalty is small, while we go away from 1 to say 0.2 the penalty is very high
    - ln(0.9) = -0.105; ln(0.2) = -1.60, ln(0.1) = -2.3 and ln(0.0001) = -9.2

TESTING BETWEEN MODELS

- Remember parameter significance we use the same t-stats as the linear regression
- Now to test between models (equivalent to F-test)
  - We use the Log-likelihood ratio test
  - The statistic is $2(L_{UR} – L_{R})$; follows a chi-square distribution with degrees of freedom given by no. of restrictions
    - Null hypothesis: restricted model is same as unrestricted
    - Alternate hypothesis: UR is better than R model
- This is a test for models that are nested within each other (i.e. we can impose some restrictions on the UR model to get the R model)

MARGINAL EFFECTS

- Now we estimated a model, we have the estimates of travel time and travel cost; we need to examine how changing travel time for transit mode affects probability of choosing transit and drive modes
- This process is called marginal effect measurement
- Definition: Change in probability due to change in independent variable
- If we measure the impact of change in transit independent variable on transit probability – it is referred to as self-marginal effect
- If we measure the impact of change in transit independent variable on drive probability – it is referred to as cross-marginal effect
- Can we compute them?

MARGINAL EFFECTS

- $P_{Dn} = \frac{e^{\beta' X_{Dn}}}{\left(e^{\beta' X_{Dn}} + e^{\beta' X_{Tn}}\right)} = \frac{e^{\beta' X_{Dn}}}{Q}$ $P_{\text{Tn}} = \frac{e^{\beta' X_{Tn}}}{Q}$;
- where $Q = e^{\beta' X_{Dn}} + e^{\beta' X_{Tn}}$
- Let k[th] variable for transit be altered
- Self: $\frac{\partial P_{Tn}}{\partial x_{T k}} = \frac{Q * e^{\beta' X_{Tn}} * \beta_{T,k} - e^{\beta' X_{Tn}} * \left(e^{\beta' X_{Dn}} * 0 + e^{\beta' X_{Tn}} * \beta_{T,k}\right)}{Q^2}$
- $= \beta_{T,k} [P_{\text{Tn}}] - \beta_{T,k} [P_{\text{Tn}}]^2$
- $= \beta_{T,k} [P_{\text{Tn}}] [1 - P_{\text{Tn}}] = \beta_{T,k} [P_{\text{Tn}}] [P_{\text{Dn}}]$
- Cross: $\frac{\partial P_{\text{Dn}}}{\partial x_{T k}} = -\beta_{T,k} [P_{\text{Tn}}] [P_{\text{Dn}}]$
- The relationship from marginal effects is informative, however, we still don't know what is the percentage change in probability for a delta change in x

ELASTICITY EFFECTS

- Slightly different definition from marginal effects
- Self elasticity : $\frac{\partial P_{Tn}}{P_{Tn}} \Big/ \frac{\partial x_{T,k}}{x_{T,k}} = \frac{\partial P_{Tn}}{\partial x_{T,k}} * \frac{x_{T,k}}{P_{Tn}}$
- $= \beta_{T,k}[P_{Tn}]\,[P_{Dn}] * \frac{x_{T,k}}{P_{Tn}}$
- $= \beta_{T,k}[P_{Dn}]\, x_{T,k}$
- Cross-elasticity: $\frac{\partial P_{Dn}}{P_{Dn}} \Big/ \frac{\partial x_{T,k}}{x_{T,k}} = -\beta_{T,k}[P_{Tn}]\, x_{T,k}$
  - Please remember these effects exist only for variables that have values in both equations such as travel time and travel cost (attributes that change for alternatives)
- But when we have a variable like income, it can exist in only one alternative i.e. the other alternative is base – individual level attributes
  - These attributes have only self-elasticity effect

REMARKS

- The effects we measured so far are the changes at the individual level
- Now we want to examine the impact on the total dataset i.e. what happens to the overall share
- This involves just adding the probability change across the population

SPECIFICATION

BINARY CHOICE MODELS

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

Total utility for person *q* by choosing mode *m*

Deterministic or Observed component of utility for person *q* by choosing mode *m*

Is a function of characteristics of the traveler q and the mode m

Random or Unobserved component of utility for person *q* by choosing mode *m*

Has a probability distribution function associated with it

For Example:

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 TC_C$$

$$V_{qT} = \beta_1 TT_T + \beta_2 TC_T + \beta_3 NT_T$$

$$TT_m = \text{Travel time by mode } m$$

$$TC_m = \text{Travel cost by mode } m$$

$$NT_T = \text{Number of transit transfers}$$

$$\beta s = \text{model parameters to be estimated from data}$$

BINARY CHOICE MODELS

Probability that person *q* chooses car = $\text{Prob}_q(\text{car})$ = $\dfrac{\exp(V_{qC})}{\exp(V_{qC}) + \exp(V_{qT})} = \dfrac{1}{1 + \exp\left[-\left(V_{qC} - V_{qT}\right)\right]}$

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 TC_C$$

$$V_{qT} = \beta_1 TT_T + \beta_2 TC_T + \beta_3 NT_T$$

$$\left(V_{qC} - V_{qT}\right) = \beta_0 + \beta_1\left(TT_C - TT_T\right) + \beta_2\left(TC_C - TC_T\right) - \beta_3 NT_T$$

As travel time by car increases relative to transit (i.e., the difference in travel time between car and transit increases),

The utility of car decreases relative to transit (i.e, the difference in utility between car and transit decreases) AND

The probability of choosing car decreases

=> We would expect a negative coefficient on the travel time variable β_1

BINARY CHOICE MODELS

Probability that person *q* chooses car = $\text{Prob}_q(\text{car})$ = $\dfrac{\exp(V_{qC})}{\exp(V_{qC}) + \exp(V_{qT})} = \dfrac{1}{1 + \exp\left[-\left(V_{qC} - V_{qT}\right)\right]}$

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 TC_C$$

$$V_{qT} = \beta_1 TT_T + \beta_2 TC_T + \beta_3 NT_T$$

$$\left(V_{qC} - V_{qT}\right) = \beta_0 + \beta_1\left(TT_C - TT_T\right) + \beta_2\left(TC_C - TC_T\right) - \beta_3 NT_T$$

As number of transit transfers increases,

The utility of car increases relative to transit (i.e, the difference in utility between car and transit increases) AND

The probability of choosing car increases [or the probability of choosing transit decreases]

=> We would expect a negative coefficient on the number of transfers variable β_3

BINARY CHOICE MODELS

Probability that person q chooses car = $\text{Prob}_q(\text{car})$ = $\dfrac{\exp(V_{qC})}{\exp(V_{qC}) + \exp(V_{qT})} = \dfrac{1}{1+\exp[-(V_{qC} - V_{qT})]}$

$$(V_{qC} - V_{qT}) = \beta_0 + \beta_1(TT_C - TT_T) + \beta_2(TC_C - TC_T) - \beta_3 NT_T$$

Lets say, for a traveler q, $(TT_C = TT_T)$, $(TC_C = TC_T)$, and $(NT_T = 0)$

Hence, for this person, $(V_{qC} - V_{qT}) = \beta_0$

If $\beta_0 > 0$ $\quad (V_{qC} > V_{qT})$ and $\text{Prob}_q(\text{car}) > \text{Prob}_q(\text{transit})$

If $\beta_0 < 0$ $\quad (V_{qC} < V_{qT})$ and $\text{Prob}_q(\text{car}) < \text{Prob}_q(\text{transit})$

If $\beta_0 = 0$ $\quad (V_{qC} = V_{qT})$ and $\text{Prob}_q(\text{car}) = \text{Prob}_q(\text{transit})$

Constant term captures a generic preference for a mode

BINARY CHOICE MODELS

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 TC_C$$
$$V_{qT} = \beta_1 TT_T + \beta_2 TC_T + \beta_3 NT_T$$

If TT_m decreases by 1 unit the utility of mode *m* increases by β_1

If TC_m increases by 1 unit the utility of mode *m* decreases by β_2

If TC_m increases by (β_1 / β_2) units

the utility of mode *m* decreases by $(\beta_1 / \beta_2) * \beta_2 = \beta_1$

If TT_m decreases by 1 unit and TC_m increases by (β_1 / β_2) units

Net change in the utility of mode *m* is $+\beta_1 - [(\beta_1 / \beta_2) * \beta_2] = 0$

BINARY CHOICE MODELS

If TT_m decreases by 1 unit and TC_m increases by (β_1 / β_2) units

Net change in the utility of mode *m* is $+\beta_1 - [(\beta_1 / \beta_2) * \beta_2] = 0$

The traveler is willing to accept an increase in travel cost of (β_1 / β_2) if it will decrease his/her travel time by 1 unit

The traveler is willing to pay (β_1 / β_2) to save 1 unit of travel time

The traveler is willing to incur 1 more unit of travel time to save (β_1 / β_2) in costs

Money value of 1 unit of travel time (VOTT) = (β_1 / β_2)

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 TC_C$$
$$V_{qT} = \beta_1 TT_T + \beta_2 TC_T + \beta_3 NT_T$$

Ratio of the coefficients on the attributes reflect the marginal rate of substitution

BINARY CHOICE MODELS

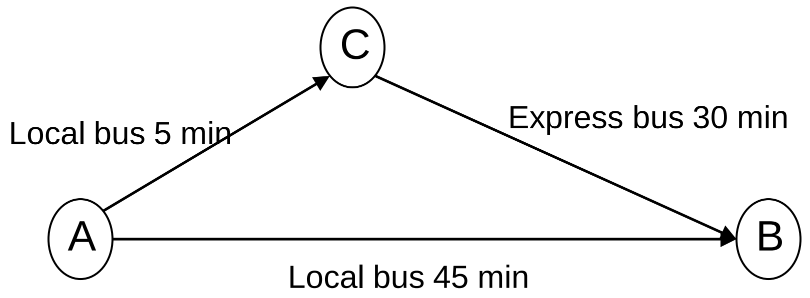
$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 TC_C$$
$$V_{qT} = \beta_1 TT_T + \beta_2 TC_T + \beta_3 NT_T$$

Time value of a transfer = (β_1 / β_3)

Amount of additional travel time a person is willing to incur to reduce one transfer

The reduction in the travel time that will make a person accept one more transfer

BINARY CHOICE MODELS



A-B: 45 mins. + 0 Transfers

A-C-B   35 mins. + 1 Transfer

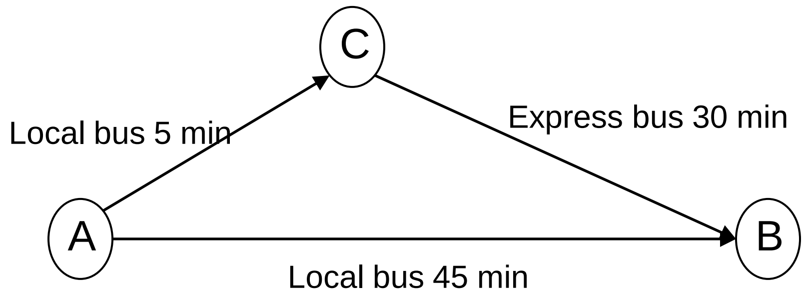
If the time value of a transfer (β_1 / β_3) is 12 min/transfer

The person is willing to accept 12 more minutes of travel time to save 1 transfer

By choosing A-B over A-C-B, the person incurs only 10 more minutes of travel time, but saves one transfer

The person prefers A-B

BINARY CHOICE MODELS



A-B: 45 mins. + 0 Transfers

A-C-B 35 mins. + 1 Transfer

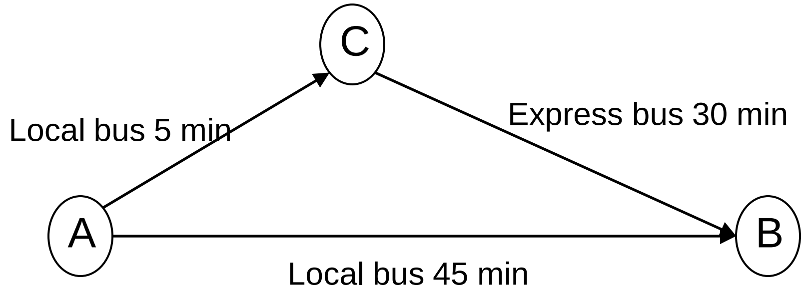
If the time value of a transfer (β_1 / β_3) is 8 min/transfer

The travel time should reduce by 8 minutes for this person to accept one more transfer

By choosing A-C-B over A-B, the person has 10 fewer minutes of travel time, but saves one transfer

The person prefers A-C-B

BINARY CHOICE MODELS



A-B: 45 mins. + 0 Transfers

A-C-B 35 mins. + 1 Transfer

If the time value of a transfer (β_1 / β_3) is 10 min/transfer

The two options are equally attractive to this person

BINARY CHOICE MODELS

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 TC_C$$
$$V_{qT} = \beta_1 TT_T + \beta_2 TC_T + \beta_3 NT_T$$

Money value of a transfer = (β_3 / β_2)

Amount of additional cost a person is willing to incur to reduce one transfer

The reduction in the cost that will make a person accept one more transfer

BINARY CHOICE MODELS

Lets now include the characteristics of the traveler in the utility equations

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 Male_q + \beta_3 Income_q$$

$$V_{qT} = \beta_1 TT_T$$

$$TT_m = \text{Travel time by mode } m$$

$$Male_q = 1 \text{ if traveler } q \text{ is male; } 0 \text{ if female}$$

$$Income_q = \text{Income of traveler } q$$

$$\beta s = \text{model parameters to be estimated from data}$$

NOTE:

The characteristics of the traveler enters the utility expression of only one of the two modes

Probability that person q chooses car = $\text{Prob}_q(\text{car})$ =

$$\frac{\exp(V_{qC})}{\exp(V_{qC}) + \exp(V_{qT})} = \frac{1}{1 + \exp\left[-\left(V_{qC} - V_{qT}\right)\right]}$$

BINARY CHOICE MODELS

Probability that person q chooses car = $\text{Prob}_q(\text{car})$ = $\dfrac{\exp(V_{qC})}{\exp(V_{qC})+\exp(V_{qT})} = \dfrac{1}{1+\exp\left[-\left(V_{qC}-V_{qT}\right)\right]}$

$$\left(V_{qC}-V_{qT}\right)=\beta_0+\beta_1\left(TT_C-TT_T\right)+\beta_2 Male_q+\beta_3 Income_q$$

Consider two travelers:

- Both have the same gender
- Both have the same travel time by car
- Both have the same travel time by transit
- One has higher income than other

If β_3 is positive, $\left(V_{qC}-V_{qT}\right)_{HI} > \left(V_{qC}-V_{qT}\right)_{LI}$ ⇒ $\text{Prob}_{HI}(\text{car}) > \text{Prob}_{LI}(\text{car})$

The higher income person is more likely to choose car than an identical lower income person

BINARY CHOICE MODELS

Probability that person *q* chooses car = $\text{Prob}_q(\text{car})$ = $\dfrac{\exp(V_{qC})}{\exp(V_{qC}) + \exp(V_{qT})} = \dfrac{1}{1 + \exp\left[-\left(V_{qC} - V_{qT}\right)\right]}$

$$\left(V_{qC} - V_{qT}\right) = \beta_0 + \beta_1\left(TT_C - TT_T\right) + \beta_2 Male_q + \beta_3 Income_q$$

Consider two travelers:

Both have the same income

Both have the same travel time by car

Both have the same travel time by transit

Differ only in gender

If β_2 is positive, $\left(V_{qC} - V_{qT}\right)_{MALE} > \left(V_{qC} - V_{qT}\right)_{FEMALE}$

$$\text{Prob}_{MALE}(\text{car}) > \text{Prob}_{FEMALE}(\text{car})$$

Men are more likely to choose car compared to identical women

BINARY CHOICE MODELS

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 Male_q + \beta_3 Income_q$$
$$V_{qT} = \beta_1 TT_T$$

Probability that person *q* chooses car = $Prob_q$(car) = $\dfrac{\exp(V_{qC})}{\exp(V_{qC}) + \exp(V_{qT})} = \dfrac{1}{1 + \exp\left[-\left(V_{qC} - V_{qT}\right)\right]}$

Probability of choosing a mode depends on the **difference** in the utility between the two modes

By introducing the traveler characteristics in the utility expression of any one mode, we allow for the utility difference to vary across travelers.

It is adequate to introduce the traveler characteristics in the utility expression of any one of the two alternatives in binary choice models

BINARY CHOICE MODELS

Further enhancements to the utility specifications

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 TC_C$$

$$V_{qT} = \beta_1 TT_T + \beta_2 TC_T + \beta_3 Income_q$$

$$TT_m = \text{Travel time by mode } m$$

$$TC_m = \text{Travel cost by mode } m$$

$$Income_q = \text{Income of traveler } q$$

$$\beta s = \text{model parameters to be estimated from data}$$

If β_3 is negative, $(V_{qC} - V_{qT})_{HI} > (V_{qC} - V_{qT})_{LI}$ ⇒ $\text{Prob}_{HI}(\text{car}) > \text{Prob}_{LI}(\text{car})$

The higher income person is more likely to choose car than an identical lower income person

BINARY CHOICE MODELS

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 TC_C$$
$$V_{qT} = \beta_1 TT_T + \beta_2 TC_T + \beta_3 Income_q$$

Ratio of the coefficients on the attributes reflect the marginal rate of substitution

Irrespective of the income levels of the person,
Money value of 1 unit of travel time (VOTT) = (β_1 / β_2)

However, one may expect a person's value of travel time to depend on his/her income

Alternately, a unit increase in cost may affect a low income person much more than a high income person

This specification does not accommodate differential sensitivity to cost between high and low income persons

BINARY CHOICE MODELS

Accommodating differential sensitivity to cost based on income:

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 \frac{TC_C}{Income_q}$$

$$V_{qT} = \beta_1 TT_T + \beta_2 \frac{TC_T}{Income_q} + \beta_3 Income_q$$

Implication of this specification:

A unit increase in travel cost of a mode decreases the utility of that mode to a person with income = $Income_q$ by a amount = $\left(\frac{\beta_2}{Income_q}\right)$

A unit increase in travel cost affects a low income person more than a high income person

BINARY CHOICE MODELS

Accommodating differential sensitivity to cost based on income:

Value of travel time for a person with income = $Income_q$ =

$$\frac{\beta_1}{\left(\frac{\beta_2}{Income_q}\right)} = \frac{\beta_1}{\beta_2}\left(Income_q\right)$$

A higher income person has a higher value of time

A higher income person is willing to pay more to save 1 unit of travel time compared to a lower income person

MODE CHOICE

MODE CHOICE MODEL SPECIFICATION

- Mode choice model
- Typical representation
- $\mathrm{U_{TR}} = \beta_{TR} + \beta_t TT_{TR} + \beta_c TC_{TR} + \cdots$
- $\mathrm{U_{DA}} = 0 \quad + \beta_t TT_{DA} + \beta_c TC_{DA} + \cdots$
- Now we can split travel time IVTT and OVTT
- In that case
- $\mathrm{U_{TR}} = \beta_{TR} + \beta_{ivtt} IVTT_{TR} + \beta_{ovtt} OVTT_{TR} + \beta_c TC_{TR} + \cdots$
- $\mathrm{U_{DA}} = 0 \quad + \beta_{ivtt} IVTT_{DA} + \beta_{ovtt} OVTT_{D} + \beta_c TC_{DA} + \cdots$
- The reason being the impact of out of vehicle time is expected to be larger on mode choice

MODE CHOICE MODEL SPECIFICATION

- Now, it is possible that impact of OVTT reduces with overall travel distance i.e. people travelling 2kms are likely to feel more burdened by waiting time than people travelling for 10 kms
- So OVTT/Distance is commonly used
- So we can add OVTT/distance variable to OVTT variable in the above specification
- The travel time and travel cost are alternative attributes
- Now, there are individual level attributes that affect alternative utilities
  - However, you can only estimate the alternative specific effect

MONEY VALUE OF TIME

- One of the most important objectives of the model is to understand the willingness to pay measure for mode choice
- We can evaluate the value of time placed by individuals in the mode choice
  - i.e. how many $ people are willing to pay to reduce travel time by 1 minute
- $U_{TR} = \beta_t TT_{TR} + \beta_c TC_{TR} + \cdots$
- What are the units of utility?
  - No units
- TT_{TR} – minutes, TC_{TR} - $
- => β_t – 1/minutes and β_c – 1/$

MONEY VALUE OF TIME

- Now lets try to generate a measure which has the same units as TC_{TR}
- So, $\frac{\beta_t}{\beta_c} TT_{TR} + TCTR$ => $\frac{\beta_t}{\beta_c}$ is the money value of time (check the units - $/minute)
- Now if we were using IVTT and OVTT, money value of time can be estimated separately for IVTT and OVTT
- If we are using OVTT/distance we will need to account for the change in dimensions appropriately
  - We consider an average distance measure and use that to generate the money value of ovtt

REFERENCES

- Ben-Akiva, M. E. and S. R. Lerman (1985). Discrete choice analysis: theory and application to travel demand, The MIT Press.
- Ortuzar, J. D. and L. G. Willumsen (2011). Modelling transport, Wiley. 4th ed.