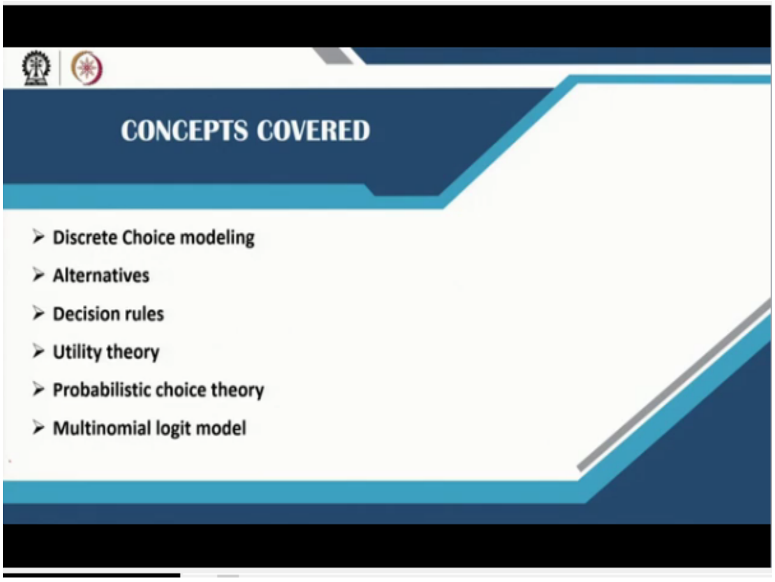
Urban Landuse and Transportation Planning
Prof. Debapratim Pandit
Department of Architecture and Regional Planning
Indian Institute of Technology-Kharagpur

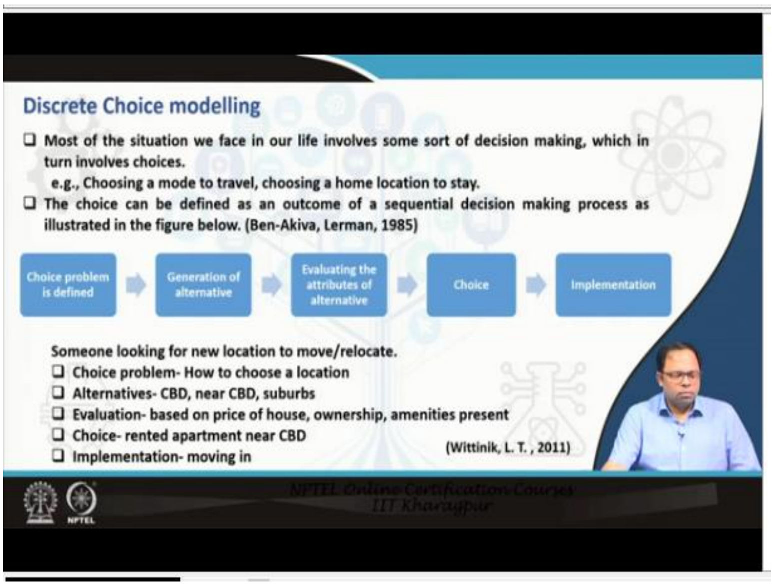
Lecture-26
Discrete Choice Theory

(Refer Slide Time: 00:34)



The different concepts covered in lecture 26 are discrete choice modelling, elements of decision making process like choice alternatives and decision rules. It also covers utility theory, probabilistic choice theory, and the multinomial logit model.

(Refer Slide Time: 00:51)

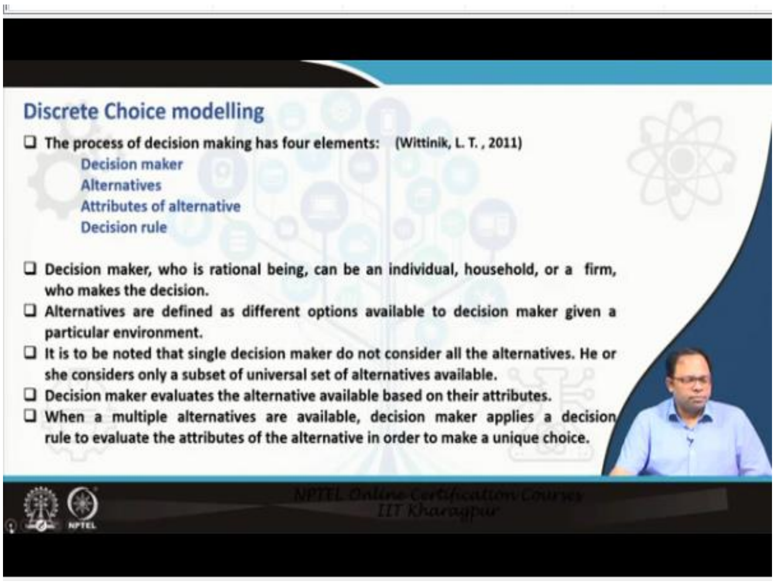


Discrete Choice Modelling

Most of the situations an individual faces in life involves some sort of decision making, which involves choosing something out of a set of alternatives for example choosing a mode of travel, choosing a location to stay and so on. All these different choices play a role in the urban land use transportation process.

A choice can be defined as an outcome of a sequential decision making process. This sequential process can be divided into five steps. Firstly, the choice problem is defined. For example, how to choose a location. Second step is the generation of alternatives for example the alternatives can be the location near CBD, in the CBD or the suburbs. Thirdly, the evaluation of alternatives. All the available locations are evaluated based on different criteria like price of house, ownership, regulations, amenities etc. and after that the decision maker arrives at a particular choice using a particular evaluation technique. For example, a decision maker chooses to stay in a rental apartment near CBD. Final step is the implementation of that choice, which is the moving in process to that particular house. So, this is the basic choice making process that every individual needs to go through.

(Refer Slide Time: 02:35)

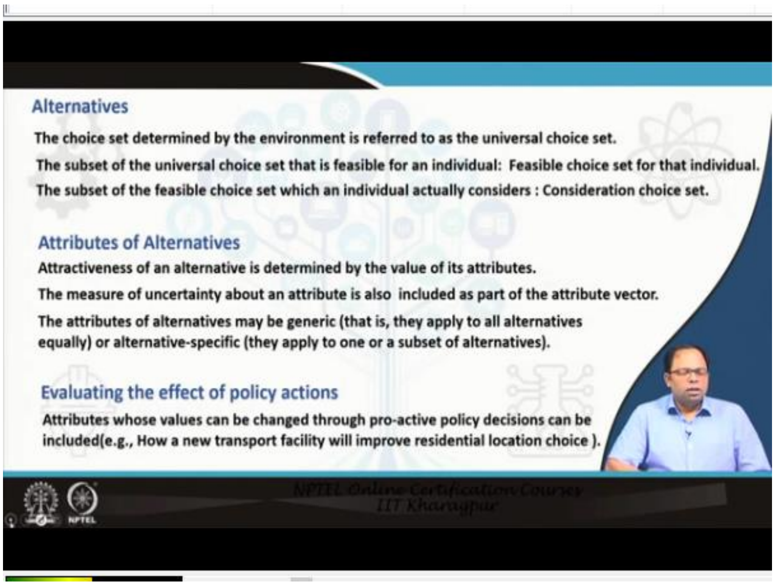


The process of decision making has got four elements i.e. the decision maker, the alternatives, the attribute of alternatives, and the final decision rule.

The decision maker is supposed to be a rational being, he can be an individual, a household, a firm, or an entity who takes a decision. The alternatives are defined as different options available to the decision maker in a particular context or a particular environment. It is important to note that, a single decision maker does not consider all the alternatives. Because an individual is not aware of all the alternatives available to him/her. For example, when a decision maker is choosing a house, he does not consider all the houses in an urban area, instead he considers certain houses or a subset of all the housing available in that particular area. So, a subset of the universal set of available alternatives is considered.

Now, the decision maker evaluates the available alternatives based on their respective attributes or characteristics. Finally, the decision maker applies the decision rule based on which the attributes of the alternatives are evaluated and the final alternative is selected among multiple available alternatives. So this is the basic process of discrete choice modelling.

(Refer Slide Time: 04:33)



As defined earlier, alternatives are the different options, or choices available to an individual for a specific choice scenario. But a set of these alternatives is defined as the choice set. If the choice set is determined by the environment, it is referred as the universal choice set. It includes all the alternatives. For example, all the houses in an urban area constitutes a universal choice set.

The subset of this universal choice set that is feasible for an individual is termed as feasible choice set. For example, among all the houses available in an urban area only some of the housing fits in the decision maker's budget. So, these alternatives constitute feasible choice set that a decision maker can look into.

Even though there may be a lot of options, which an individual can afford, for example, it is within his budget. But he still limits his choices only to certain choices, like which are more close to his work location. The subset of the feasible choice set which an individual actually considers is called the consideration choice set.

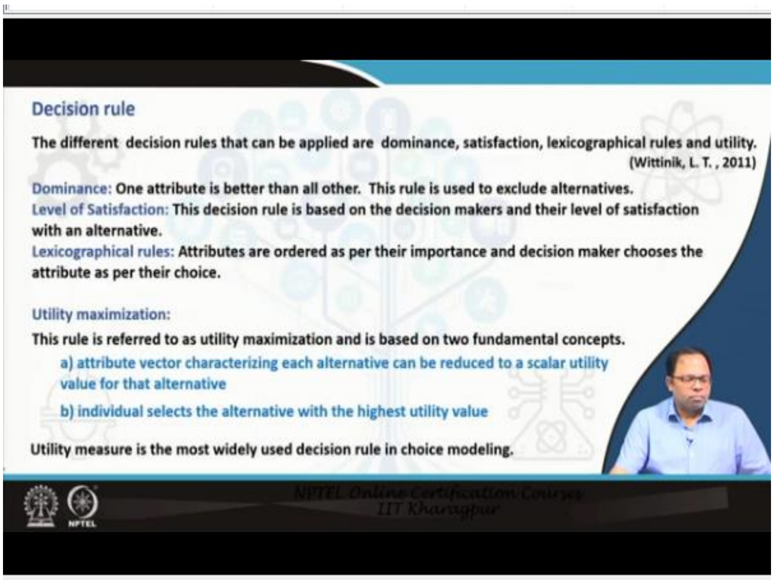
In the modelling process, the different levels of choice set play an important role. With millions of choices or alternatives, for example, housing options in a big urban area, it is difficult to model such a choice. Large choice set either leads to inaccurate results or an unstable model. So, a modeller emulates the choice behaviour of decision maker by considering different levels of choice set for better predictions.

Next element is attributes of alternatives. The alternatives among which a decision maker makes a choice are characterised by different properties (or attributes). The attractiveness of an alternative is determined by the value of its attributes. For example, housing unit can be characterised by the prestige of the location, cost, number of rooms, structural quality etc. Suppose for a single member household, housing unit with low rent/cost is of priority, then the attractiveness of alternatives with lower cost is more.

Next, the measure of uncertainty about an attribute is also included as part of the attribute vector, i.e., it is not only about the things that are known but there are unknown things about that particular attribute which are also considered. In addition, attributes of alternatives may be generic (that means they apply to all the alternatives equally), and alternative specific (that means they only apply to a subset of alternatives). For example, travel time in a mode choice problem. Consider three modes of travel i.e. car, bus and two-wheeler. Travel time is involved in each mode of travel. But in case of a bus, travel time constitutes travel time inside the bus and access time to the bus stop as well. Therefore, access time to bus stop is an alternative specific attribute.

In addition to all this, variables which are related to policies are also considered. Values of these attributes help in determining effects of proactive policy measures. For example, how a new transport facility will improve residential choices can be considered while doing this kind of discrete choice modelling.

(Refer Slide Time: 08:15)

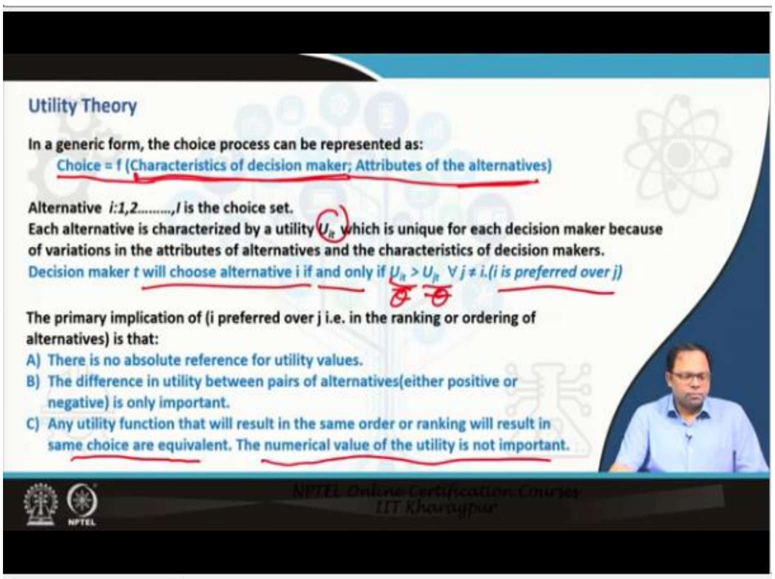


The next element in the decision making process is the decision rule. The decision rule is generally a function used to evaluate different alternatives and make a unique choice. There are several ways a decision maker can make a decision. For example, decision rules are dominance, satisfaction, lexicographical rules, and utility based.

In dominance, one alternative is better than another alternative when at least one attribute is better than all other. Generally, this rule is used to exclude poor alternatives from the choice set. Satisfaction based decision rule is based on the decision maker and their level of satisfaction with a particular alternative. So, the decision maker may be satisfied with one of the alternatives, or he can have different satisfaction with different orderings based on that he choose the alternative he wishes. In lexicographical rules, attributes are ordered as per their importance and decision maker chooses the attribute as per their choice, that means ranks are given to the different alternatives and then the decision maker decides on any one of those alternatives.

The most popular rule that is being used in choice modelling is utility maximization rule. It is based on two fundamental concepts. First, attribute vector characterizing each alternative can be reduced to a scalar utility value for that particular alternative which means that the different characteristics or variables that define a particular alternative can be brought into a scalar quantity or a unique value. Second, an individual selects the alternative which has got the highest utility value. So, this utility measure is the most widely used decision rule considered in the choice modelling process.

(Refer Slide Time: 10:16)



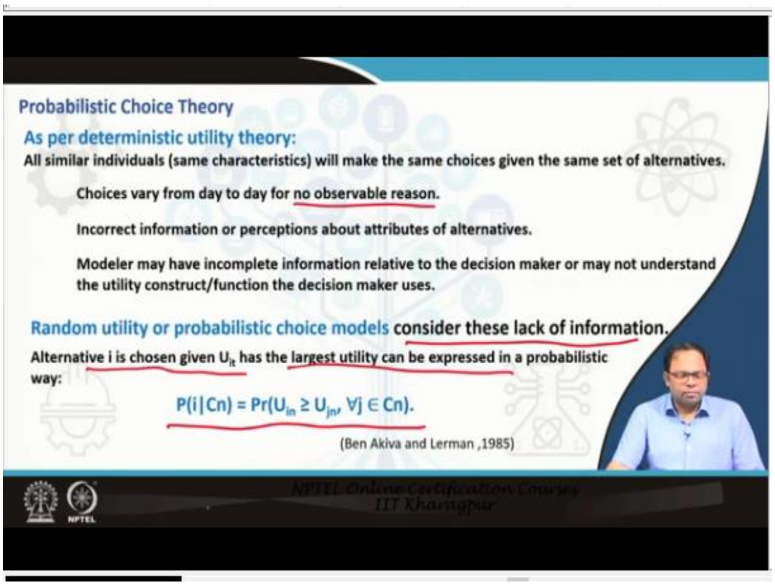
Utility Theory

In utility theory, it is stated that, the choice process can be represented as a function of characteristics of both decision makers and attributes of the alternatives. A utility value is created for each alternative which is a scalar quantity and depends on the different characteristics of the alternative and the decision maker. So, both play a role in the final utility of a particular alternative or choice.

Let us consider a choice set with alternatives 'i' where i=1, 2…j. Each alternative 'i' is characterised by a utility value U_{it} which is unique for each decision maker 't', because of variations in the attributes of alternatives and the characteristics of the decision maker. A decision maker 't' will choose alternative 'i' if and only if the utility of alternative 'i' is greater than the utility of alternative 'j' (or $U_{it} > U_{jt}$). This means that for decision maker 't', alternative 'i' is preferred over alternative 'j'.

It is important to note that only the difference between the utility values of the alternatives is considered. So, the primary implications of 'i' preferred over 'j' basically means that there is no absolute reference for utility values i.e. the actual values of the utility does not play any role. Moreover, the difference in utility between pairs of alternatives (Uit - Ujt), either positive or negative, is important. Also, any utility function that will result in the same order or ranking will result in the same choices are equivalent. So, if the utility value is divided or multiplied by another value θ the result remains the same. So, the numerical value of the utilities is not important.

(Refer Slide Time: 12:47)



Probabilistic choice theory

The deterministic utility theory states that, all similar individuals with the same characteristics will make the same choices given the same set of alternatives. So, whenever an individual or a group of individuals with same characteristics are presented with choices having same attributes, then based on utility theory they will choose the same alternative every time since the utility of one choice is higher than another.

In reality, it is not the case, because choices vary from day to day for no observable reasons. For example, 90% of the time an individual choose to take his vehicle to office, but ten days later he may choose to take a taxi. The reason maybe he does not feel like driving that particular day. Even though the decision maker in question knows the reason but the modeller or someone else does not have any idea of why the decision maker have changed the choice. So, the modeller will have no reason to understand this.

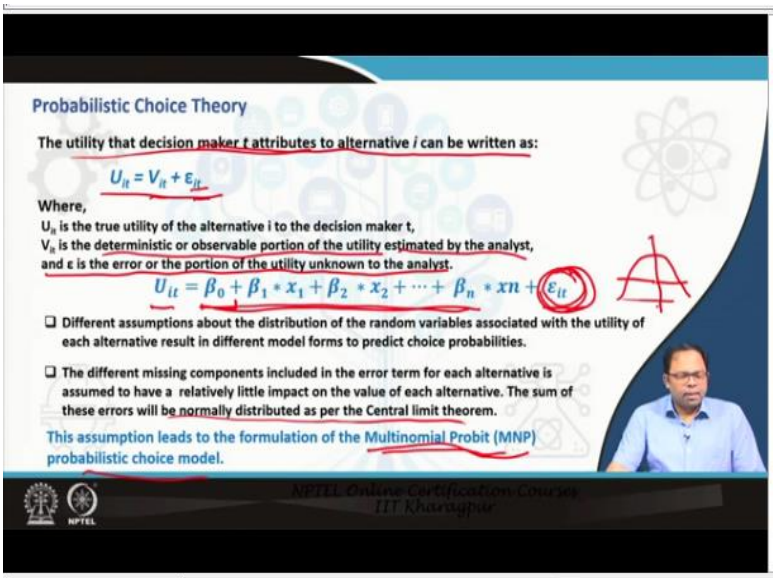
Also, in choice scenarios, people have incorrect information or perceptions about attributes of alternatives. So, either people do not know the exact information about every attribute of alternative or they build some perceptions i.e. they believe the value to be something but it is different.

Therefore, the modeller (the person who is developing these models) may have incomplete information relative to the decision maker or may not understand the utility construct/function that the decision maker actually uses to decide. So, these issues plague the modelling process. Random utility or probabilistic choice model considers these issues i.e.

lack of information on the part of the user or the part of the modeller, and choice variability due to no observable reasons.

In probabilistic choice model, the choice of alternative ‘i' given U_{in} has the largest utility, could be expressed in a probabilistic way as shown in the slide where P is the probability of choosing alternative ‘i’, C_n is the choice set specific to an individual ‘n’, and j is one of the other alternatives. An individual ‘n’ will choose alternative ‘i’, if the utility of alternative ‘i’ (U_{in}) is greater than the utility of alternative ‘j’ (U_{jn}).

(Refer Slide Time: 16:23)



The utility (U_{it}) that decision maker ‘t’ attributes to alternative ‘i’ can be written as:

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

where, U_{it} is the true utility of the alternative ‘i’ to the decision maker ‘t’. V_{it} is the deterministic or observable portion of the utility that is known to modeller or could be estimated by the modeller or the analyst. ε_{it} is the error term or the portion of the utility unknown to the analyst. So, U_{it} can be disintegrated into a deterministic part and the error term.

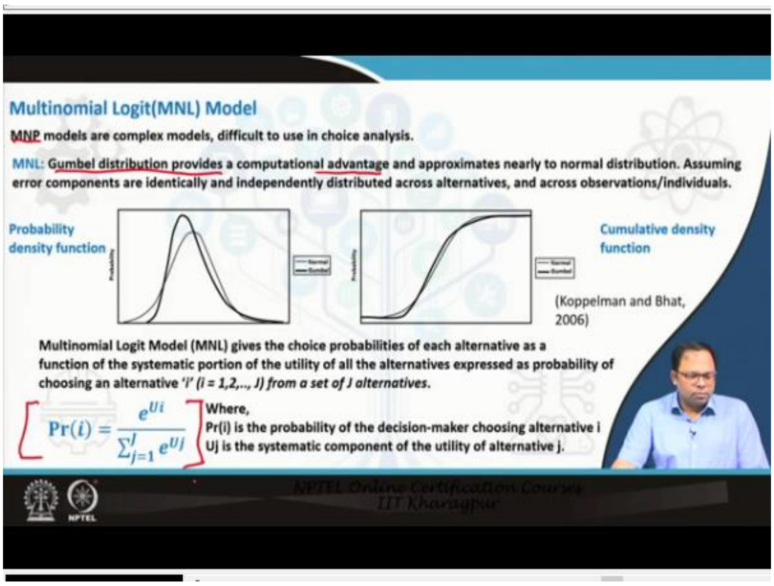
The deterministic part or V_{it} could be expressed as follows:

$$V_{it} = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \dots\dots\dots. + \beta_n * x_n$$

Where β’s are the coefficients or parameters and ‘x’ are the different characteristics of an alternative ‘i’. β_0 is the constant and takes care of the characteristics which are not included in the equation.

ε_{it} is the error term or the random component of utility. It is important to note that error terms are unmeasurable. So, different assumptions about the distribution of the random variables associated with the utility of each alternative, are considered and this result in different model forms to predict choice probabilities. If the modeller assumes that the different missing components included in the error term for each alternative is assumed to have a relatively little impact on the value of each alternative, and the sum of these errors will be normally distributed as per the central limit theorem. This assumption leads to the formulation of multinomial probit probabilistic choice model.

(Refer Slide Time: 19:21)



Multinomial logistic model

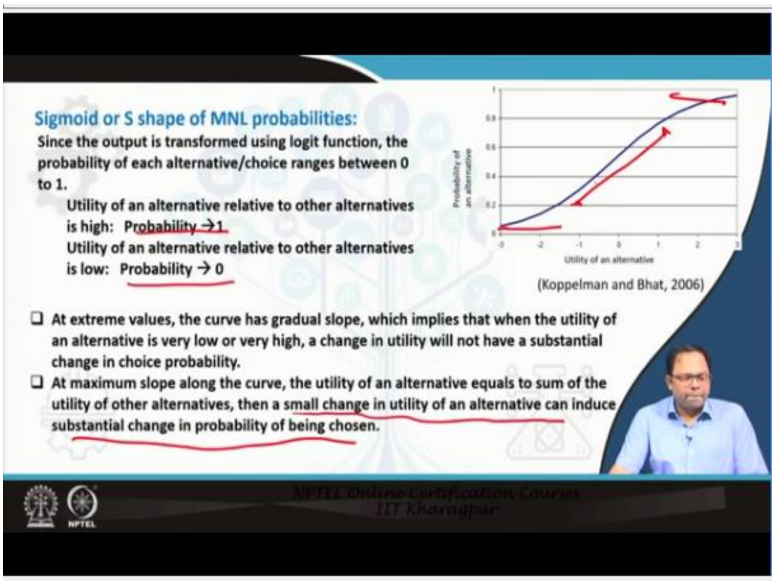
The multinomial probit model or MNP is very complex and difficult to use in choice analysis. Also, it is very difficult to determine the beta parameters. So, Gumbel distribution is selected instead of normal distribution. This distribution provides computational advantages and approximates nearly to the normal distribution as illustrated in the above figures where the thick line represents Gumbel distribution.

In addition to Gumbel distribution, error components are assumed to be identically and independently distributed across alternatives, and across observations and individuals. These three assumptions lead to formulation of multinomial logistic model. Multinomial logit model gives the choice probabilities of each alternative as a function of the systemic portion of the utility of all the alternatives expressed as probability of choosing alternative 'i' (i= 1 to J) from a set of J alternatives. The mathematical structure can be written as follows:

$$\Pr(i) = \frac{e^{U_i}}{\sum_{j=1}^{J} e^{U_j}}$$

Where Pr (i) is the probability of the decision maker choosing alternative 'i' and U_j is the systematic component of the utility of alternative j. So, this is the functional form of the multinomial logit model and this is possible after assuming Gumbel distribution for the error term instead of a normal distribution, which would have led to the multinomial probit model.

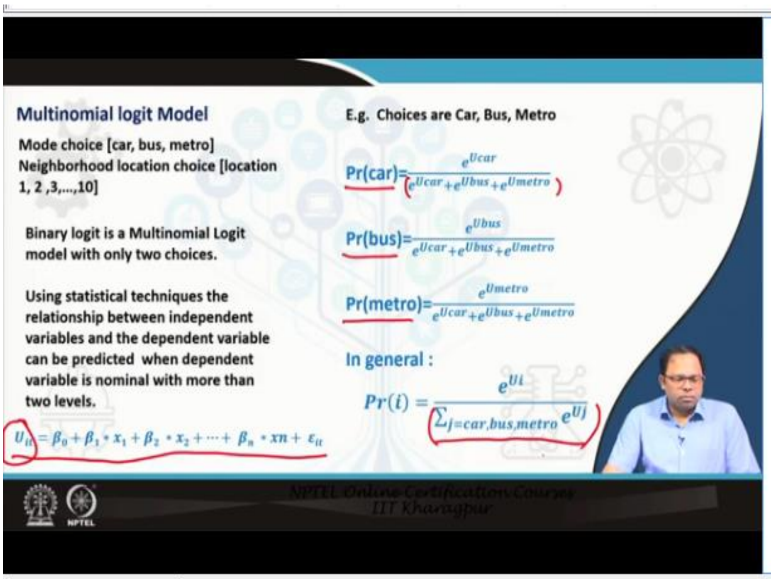
(Refer Slide Time: 21:24)



MNL probabilities takes the S shape or a sigmoid curve. Because in MNL model the output is transformed using a logit function, therefore, the probability of each alternative or choice ranges from 0 to 1. If the utility of an alternative relative to other alternatives is high, then probability tends to 1, and if the utility of an alternative relative to other alternative is low then probability tends to 0.

At extreme values, the curve has a gradual slope, which implies that when the utility of an alternative is very low or very high, a change in utility will not have a substantial change in the choice probability. Even though utility changes by a large factor, there would not be much change in the probability. At maximum slope along the curve, the utility of an alternative equals some of the utility of other alternatives, then the small change in utility of an alternative can induce a substantial change in the probability of being chosen. So, the change is very abrupt and the chance of other alternatives being chosen is much higher.

(Refer Slide Time: 23:10)



MNL model can be developed for mode choice (like car, bus or metro), or neighbourhood location choice, and so on. Multinomial logit model deals with multi class problem which means that choice set has multiple alternatives. It is important to note that, when the choice set has only two alternatives, it is a binary logit model, and when choice set has three or more alternatives it is a multinomial logit model. But the mathematical formulation for both the model is same. The relationship between independent variables and the dependent variable can be predicted, when the dependent variable is nominal with two or more than two levels, using logistic regression statistical techniques.

As discussed earlier, the functional form to determine the utility of an alternative i for individual t can be written as:

$$U_{it} = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \ldots\ldots \; + \beta_n * x_n + \varepsilon_{it}$$

and the probability of choosing alternative ‘i’ (i= 1 to J) from a set of J alternatives is written as:

$$\Pr(i) = \frac{e^{U_i}}{\sum_{j=1}^{J} e^{U_j}}$$

Let us consider a mode choice problem, where the probability of choosing a car, bus and metro are to be determined. Based on the probability equation, the choice probability of car, bus and metro can be written as:

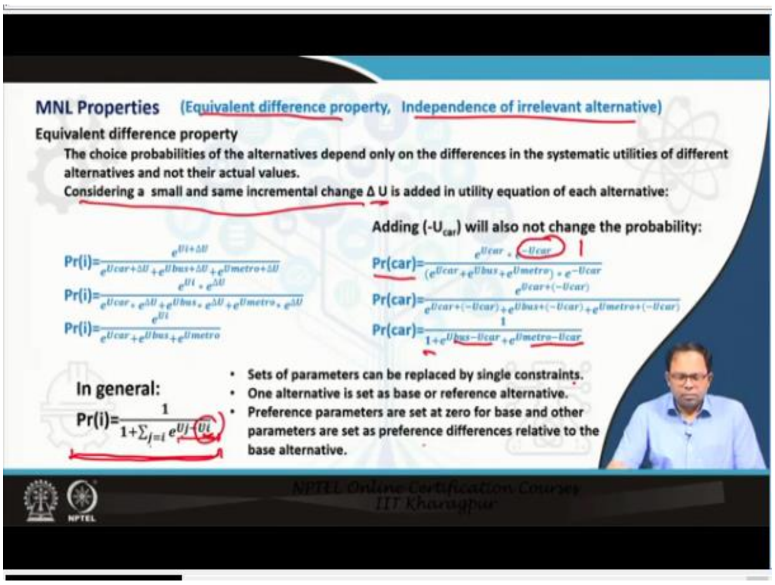
$$\Pr(car) = \frac{e^{Ucar}}{(e^{Ucar} + e^{Ubus} + e^{Umetro})}$$

$$Pr(bus) = \frac{e^{Ubus}}{(e^{Ucar} + e^{Ubus} + e^{Umetro})}$$

$$Pr(metro) = \frac{e^{Umetro}}{(e^{Ucar} + e^{Ubus} + e^{Umetro})}$$

Where U_{car}, U_{bus}, and U_{metro} are the utility equation of car, bus, and metro respectively. The log of the denominator i.e. log of sum of exponential of all the different modes is termed as logsum. Logsum is the combined utility of all the alternatives considered in a choice scenario.

(Refer Slide Time: 25:14)



Multinomial Logit Model properties

In addition to Sigmoid or S curve property of MNL probabilities, the two other primary properties of the MNL model are the equivalent difference property and the independence of irrelevant alternatives. These properties have been used in many models that are developed.

The equivalent difference property states that the choice probabilities of alternatives depend only on the differences in the systematic utilities of different alternatives and not their actual values. It has been already discussed that adding, multiplying or dividing utility values by a value θ does not have any effect, because the difference between the utility values of alternatives is important. To explain this property further, let us consider three modes of travel (car, bus and metro), and a small and same incremental change ∆U is added to the utility equation of each alternative.

The earlier probabilities of car, bus, and metro can be written as:

$$Pr(car) = \frac{e^{Ucar}}{e^{Ucar} + e^{Ubus} + e^{Umetro}}$$

$$\Pr(bus) = \frac{e^{Ubus}}{e^{Ucar} + e^{Ubus} + e^{Umetro}}$$

$$\Pr(metro) = \frac{e^{Umetro}}{e^{Ucar} + e^{Ubus} + e^{Umetro}}$$

Or in general form:

$$\Pr(i) = \frac{e^{Ui}}{e^{Ucar} + e^{Ubus} + e^{Umetro}}$$

Where 'i' is the alternative for which probability is being computed. After the addition of ΔU in the utility equation of car, bus and metro, the probability can be written as:

$$\Pr(i) = \frac{e^{Ui+\Delta U}}{e^{Ucar+\Delta U} + e^{Ubus+\Delta U} + e^{Umetro+\Delta U}}$$

$$\Pr(i) = \frac{e^{Ui} * e^{\Delta U}}{e^{Ucar} * e^{\Delta U} + e^{Ubus} * e^{\Delta U} + e^{Umetro} * e^{\Delta U}}$$

$$\Pr(i) = \frac{e^{Ui} * \cancel{e^{\Delta U}}}{(e^{Ucar} + e^{Ubus} + e^{Umetro}) * \cancel{e^{\Delta U}}}$$

So, the probability before and after the addition of ΔU are similar. The probability equation can also be written in a different form which clearly explains equivalent difference property. Similar to adding ΔU, -Ucar is added to the utility equation of all three alternatives. The equation obtained are:

$$\Pr(car) = \frac{e^{Ucar} * e^{-Ucar}}{(e^{Ucar} + e^{Ubus} + e^{Umetro}) * e^{-Ucar}}$$

$$\Pr(car) = \frac{e^{Ucar+(-Ucar)}}{e^{Ucar+(-Ucar)} + e^{Ubus+(-Ucar)} + e^{Umetro+(-Ucar)}}$$

$$\Pr(car) = \frac{1}{1 + e^{Ubus+(-Ucar)} + e^{Umetro+(-Ucar)}}$$

The equation has now changed into a different form. So, the probability equation could be written in this format:

$$\Pr(i) = \frac{1}{1 + \sum_{j \neq i} e^{Uj-Ui}} \qquad \nabla\, i \in J$$

It clearly shows that the probability of choosing an alternative car (or 'i') is equal to the difference in the utility between car and bus, and car and metro.

If the utility equations are developed in this form, it would help in the development of the multinomial logit model. The utility equation of alternative 'i' and alternative 'j' will have beta coefficients for each attributes i.e.

$U_i = \beta_{i0} + \beta_{i1} * x_1 + \beta_{i2} * x_2 + \dots \dots . \ + \beta_{in} * x_n$

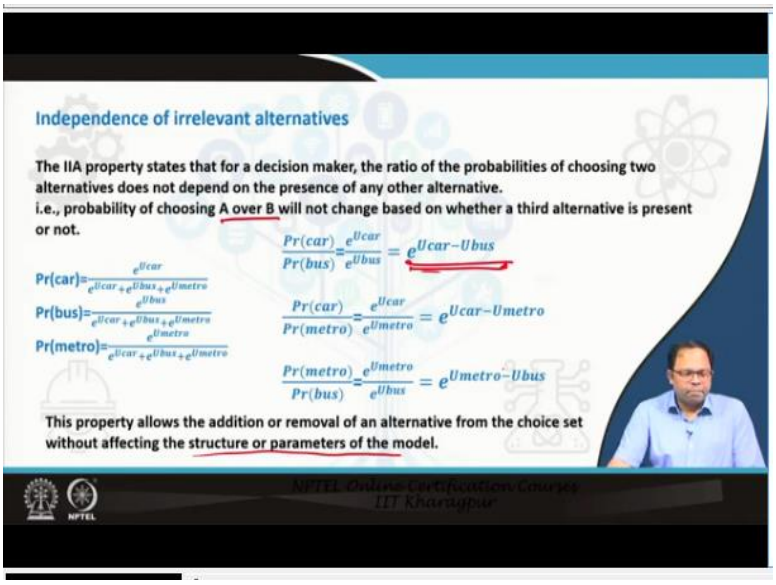
$U_j = \beta_{j0} + \beta_{j1} * x_1 + \beta_{j2} * x_2 + \dots \dots . \ + \beta_{jn} * x_n$

The difference between the alternatives becomes:

$U_j - U_i = (\beta_{j0} - \beta_{i0)} + (\beta_{j1} - \beta_{i1}) * x_1 + \dots \dots . \ + (\beta_{jn} - \beta_{in}) * x_n)$

It is not possible to estimate all the parameters coefficients. So, the sets of parameters are replaced by a single constraint. The easiest constraint is to set one alternative as base or reference alternative, for example, in this particular case 'car' is the base alternative, and the preference parameters are set at 0 for base and other parameters are set as preference differences relative to the base alternative. So, everything is determined based on the base alternative.

(Refer Slide Time: 29:34)



Independence of irrelevant alternatives or IIA property states that for a decision maker, the ratio of probabilities of choosing two alternatives does not depend on the presence of any other alternative. For example, if a choice set constitutes three alternatives (A, B and C) then the probability of choosing A over B will not change based on whether C is present or not.

In order to understand further, let us assume three modes of travel i.e. car, bus and metro. The probability of choosing a car, bus and metro can be written as follows respectively.

$$\Pr(car) = \frac{e^{Ucar}}{e^{Ucar} + e^{Ubus} + e^{Umetro}}$$

$$\Pr(bus) = \frac{e^{Ubus}}{e^{Ucar} + e^{Ubus} + e^{Umetro}}$$

$$Pr(metro) = \frac{e^{Umetro}}{e^{Ucar} + e^{Ubus} + e^{Umetro}}$$

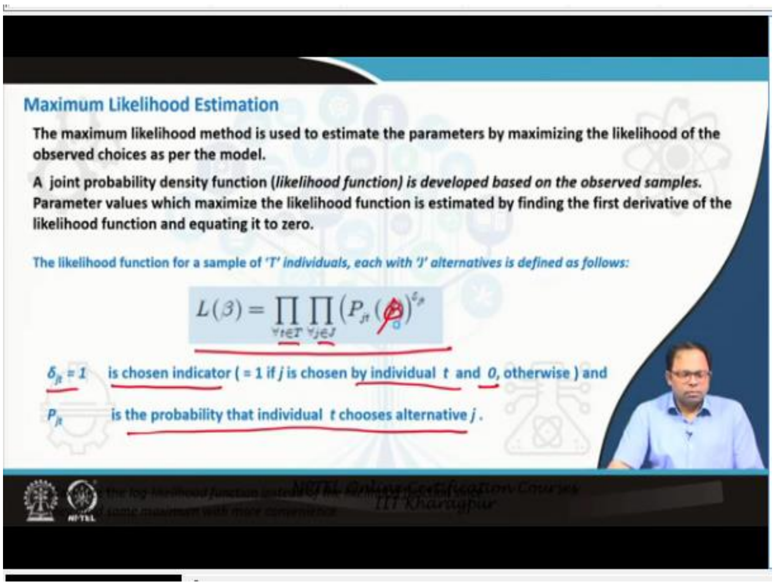
The probability of choosing a car over a bus can be derived from the above expression and written as:

$$\frac{Pr(car)}{Pr(bus)} = \frac{e^{U_{car}}}{e^{U_{bus}}} = e^{U_{car} - U_{bus}}$$

So, the probability of choosing a car over bus depends only on the attributes of car and bus, and not on the attributes of the metro. Therefore, the ratio of the probability of car and bus is independent of metro in the choice set.

This property allows the addition or removal of an alternative from the choice set without affecting the structure or parameters of the model. So, this property provides flexibility to apply MNL model in choice scenarios where different individual faces different choice sets i.e. mode choice modelling, location choice modelling etc. While developing such models an alternative is taken as the reference category.

(Refer Slide Time: 31:26)



Maximum Likelihood Estimation

The most important part of a model development process is the estimation of parameters of the attributes. The estimation of parameters is accomplished by the maximum likelihood method. This method is used to estimate the parameters by maximizing the likelihood of the observed choices as per the model. Firstly, the joint probability density function of the likelihood function is developed based on the observed samples. Secondly, the parameter values which maximizes the likelihood function is estimated by finding the first derivative of the likelihood function and equating it to zero. The likelihood function for a sample of 'T' individual, each with 'J' alternatives is defined as follows:

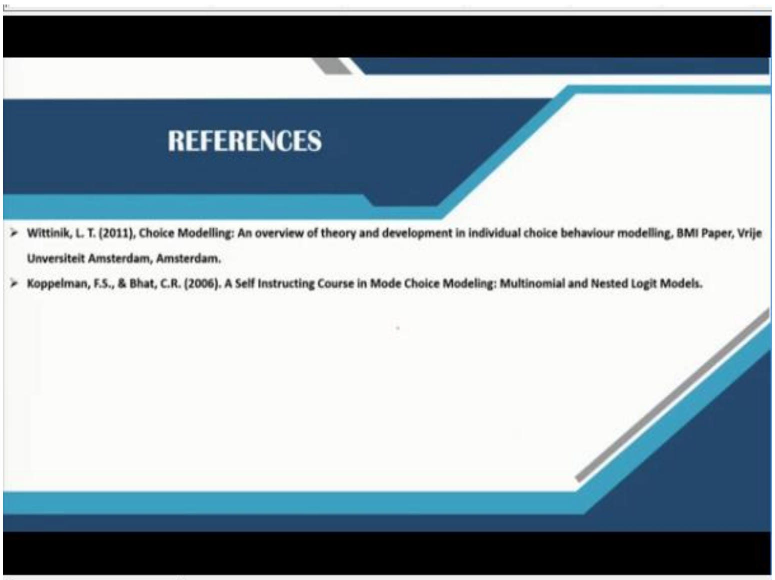
$$L(\beta) = \prod_{\forall t \epsilon T} \prod_{\forall j \epsilon J} (P_{jt}(\beta))^{\delta_{jt}}$$

Where L(β) is the likelihood function, Pjt is the probability that the individual 't' chooses alternative 'j', and δjt is equal to 1, if alternative 'j' is chosen by individual and 0 otherwise.

The log transformation of the given equation gives the same maximum as the likelihood function. Therefore, the parameter is determined by differentiating the log likelihood function instead of likelihood function. So, the first derivative of the log likelihood function is taken, the maximum likelihood is yielded by equating it to zero. This gives the best values of beta parameters and would be adopted for the considered utility equation.

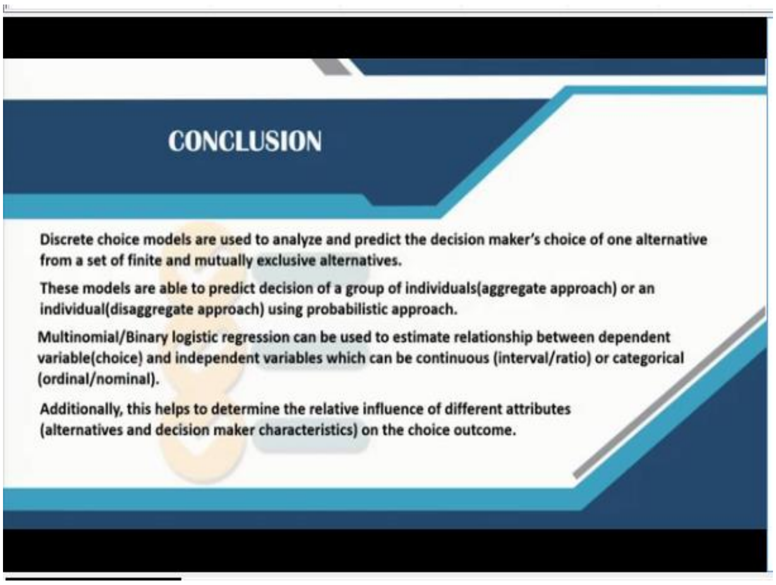
Generally, software are used to determine the beta parameters i.e., the software uses the maximum likelihood method to determine the coefficient values for each of the attributes included in the equation. These coefficient values explain the relative influence of the attributes (or independent variable) on the dependent variable.

(Refer Slide Time: 34:00)



The references are mentioned in the above slide.

(Refer Slide Time: 34:03)



Conclusion

Discrete choice models are used to analyze and predict the decision makers' choice of one alternative from a set of finite and mutually exclusive alternatives. These models can predict the decision of a group of individuals like in aggregate approach, or an individual in a disaggregate approach, using a probabilistic approach. Multinomial or binary logistic regression can be used to estimate the relationship between dependent variable (choice) and independent variables which can be continuous variable (interval/ratio) or categorical variables (ordinal or nominal). So, additionally, this helps to determine the relative influence of different attributes on the choice outcome. These attributes are either alternative or decision maker characteristics.