2 DISCRETE CHOICE METHODS AND THEIR APPLICATIONS TO SHORT TERM TRAVEL DECISIONS

Moshe Ben-Akiva and Michel Bierlaire

2.1 Introduction

Modeling travel behavior is a key aspect of demand analysis, where aggregate demand is the accumulation of individuals' decisions. In this chapter, we focus on "short-term" travel decisions. The most important short-term travel decisions include choice of destination for a non-work trip, choice of travel mode, choice of departure time and choice of route. It is important to note that short-term decisions are conditional on long-term travel and mobility decisions such as car ownership and residential and work locations.

The analysis of travel behavior is typically disaggregate, meaning that the models represent the choice behavior of individual travelers. Discrete choice analysis is the methodology used to analyze and predict travel decisions. Therefore, we begin this chapter with a review of the theoretical and practical aspects of discrete choice models. After a brief discussion of general assumptions, we introduce the random utility model, which is the most common theoretical basis of discrete choice models. We then present the alternative discrete choice model forms such as Logit, Nested Logit, Generalized Extreme Value and Probit, as well as more recent developments such as Hybrid Logit and the Latent Class choice model. Finally, we elaborate on the applications of these models to two specific short term travel decisions: route choice and departure time choice.

2.2 Discrete Choice Models

We provide here a brief overview of the general framework of discrete choice models. We refer the reader to Ben-Akiva and Lerman (1985) for the detailed developments.

General Modeling Assumptions

The framework for a discrete choice model can be presented by a set of general assumptions. We distinguish among assumptions about the:

1. decision-maker -- defining the decision-making entity and its characteristics;
2. alternatives -- determining the options available to the decision-maker;
3. attributes -- measuring the benefits and costs of an alternative to the decision-maker; and
4. decision rule -- describing the process used by the decision-maker to choose an alternative.

Decision-maker Discrete choice models are also referred to as disaggregate models, meaning that the decision-maker is assumed to be an individual. The "individual" decision-making entity depends on the particular application. For instance, we may consider that a group of persons (a household or an organization, for example) is the decision-maker. In doing so, we may ignore all internal interactions within the group, and consider only the decisions of the group as a whole. We refer to "decision-maker" and "individual" interchangeably throughout this chapter. To explain the heterogeneity of preferences among decision-makers, a disaggregate model must include their characteristics such as the socio-economic variables of age, gender, education and income.

Alternatives Analyzing individual decision making requires not only knowledge of what has been chosen, but also of what has not been chosen. Therefore, assumptions must be made about available options, or alternatives, that an individual considers during a choice process. The set of considered alternatives is called the choice set.

A discrete choice set contains a finite number of alternatives that can be explicitly listed. The choice of a travel mode is a typical example of a choice from a discrete choice set. The identification of the list of alternatives is a complex process usually referred to as *choice set generation*. The most widely used method for choice set generation uses deterministic criteria of alternative availability. For example, the possession of a driver's license determines the availability of the auto drive option.

The universal choice set contains all potential alternatives in the application's context. The choice set is the subset of the universal choice set considered by, or available to, a particular individual. Alternatives in the universal choice set that are not available to the individual are therefore excluded from the choice set.

In addition to availability, the decision-maker's awareness of the alternative could also affect the choice set. The behavioral aspects of awareness introduce uncertainty in modeling the choice set generation process and motivate the use of probabilistic choice set generation models that predict the probability of each feasible choice set within the universal set. A discrete choice model with a probabilistic choice set generation model is described later in this chapter as a special case of the latent class choice model.

Attributes Each alternative in the choice set is characterized by a set of attributes. Note that some attributes may be generic to all alternatives, and some may be alternative-specific.

An attribute is not necessarily a directly measurable quantity. It can be any function of available data. For example, instead of considering travel time as an attribute of a transportation mode, the logarithm of the travel time may be used, or the effect of out-of-pocket cost may be represented by the ratio between the out-of-pocket cost and the income of the individual. Alternative definitions of attributes as functions of available data must usually be tested to identify the most appropriate.

Decision Rule

The decision rule is the process used by the decision-maker to evaluate the attributes of the alternatives in the choice set and determine a choice. Most models used for travel behavior applications are based on *utility theory*, which assumes that the decision-maker's preference for an alternative is captured by a value, called utility, and the decision-maker selects the alternative in the choice set with the highest utility.

This concept, employed by consumer theory of micro-economics, presents strong limitations for practical applications. The underlying assumptions of this approach are often violated in decision-making experiments. The complexity of human behavior suggests that the decision rule should include a probabilistic dimension.

Some models assume that the decision rule is intrinsically probabilistic, and even complete knowledge of the problem would not overcome the uncertainty. Others consider the individuals' decision rules as deterministic, and motivate the uncertainty from the limited capability of the analyst to observe and capture all the dimensions of the choice process, due to its complexity.

Specific families of models can be derived depending on the assumptions about the source of uncertainty. Models with probabilistic decision rules, like the model proposed by Luce (1959), or the "elimination by aspects" approach proposed by Tversky (1972), assume a deterministic utility and a probabilistic decision process. Random utility models, used intensively in econometrics and in travel behavior analysis, are based on deterministic decision rules, where utilities are represented by random variables.

Random Utility Theory

Random utility models assume, as does the economic consumer theory, that the decision-maker has a perfect discrimination capability. However, the analyst is assumed to have incomplete information and, therefore, uncertainty must be taken into account. Manski (1977) identifies four different sources of uncertainty: unobserved alternative attributes; unobserved individual characteristics (also called "unobserved taste variations"); measurement errors; and proxy, or instrumental, variables.

The utility is modeled as a random variable in order to reflect this uncertainty. More specifically, the utility that individual n associates with alternative i in the choice set C_n is given by

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

where V_{in} is the deterministic (or systematic) part of the utility, and ε_{in} is the random term, capturing the uncertainty. The alternative with the highest utility is chosen. Therefore, the probability that alternative i is chosen by decision-maker n from choice set C_n is

$$P(i|\, C_n) = P[\, U_{in} \geq U_{jn}\ \forall\, j \in C_n] = P[U_{in} = \max_{j \in C_n} U_{jn}].$$

In the following we introduce the assumptions necessary to make a random utility model operational.

Location and scale parameters Considering two arbitrary real numbers α and μ, where $\mu > 0$, we have that

$$P[\, U_{in} \geq U_{jn}\ \ \forall\, j \in C_n] =$$

$$P[\mu U_{in}+\alpha \geq \mu U_{jn}+\alpha\ \ \forall\, j \in C_n] =$$

$$P[\, U_{in} - U_{jn} \geq 0\ \forall\, j \in C_n].$$

The above illustrates the fact that only the signs of the *differences* between utilities are relevant here, and not utilities themselves. The concept of ordinal utility is relative and not absolute. In order to estimate and use a specific model arbitrary values have to be selected for α and μ. The selection of the scale parameter μ is usually based on a convenient normalization of one of the variances of the random terms. The location parameter α is usually set to zero. See also the discussion below of Alternative Specific Constants.

Alternative specific constants The means of the random terms can be assumed to be equal to any convenient value c (usually zero, or the Euler constant γ for Logit models). This is not a restrictive assumption. If we denote the mean of the error term of alternative i by $m_i = E[\varepsilon_{in}]$, we can define a new random variable $e_{in} = \varepsilon_{in} - m_i + c$ such that $E[e_{in}]=c$. We have

$$P[\, U_{in} \geq U_{jn}\ \forall\, j \in C_n] = P[\, V_{in} + m_i + e_{in} \geq V_{jn} + m_j + e_{jn}\ \forall\, j \in C_n],$$

a model in which the deterministic part of the utilities are $V_{in} + m_i$ and the random terms are e_{in} (with mean c). The terms m_i are then included as Alternative Specific

Constants (ASC) that capture the means of the random terms. Therefore, we may assume without loss of generality that the error terms of random utility models have a constant mean c by including alternative specific constants in the deterministic part of the utility functions.

As only differences between utilities are relevant, only differences between ASCs are relevant as well. It is common practice to define the location parameter α as the negative of one of the ASCs. This is equivalent to constraining that ASC equal zero. From a modeling viewpoint, the choice of the particular alternative whose ASC is constrained is arbitrary. However, Bierlaire, Lotan and Toint (1997) have shown that the estimation process may be affected by this choice. In the context of the Multinomial Logit Model, they show that constraining the sum of ASCs to 1 is optimal for the speed of convergence of the estimation process. This result is also generalized for the Nested Logit Model.

The deterministic term of the utility The deterministic term V_{in} of each alternative is a function of the attributes of the alternative itself and the characteristics of the decision-maker. That is

$$V_{in} = V(z_{in}, S_n)$$

where z_{in} is the vector of attributes as perceived by individual n for alternative i, and S_n is the vector of characteristics of individual n.

This formulation is simplified using any appropriate vector valued function h that defines a new vector of attributes from both z_{in} and S_n, that is

$$x_{in} = h(z_{in}, S_n).$$

The choice of h is very general, and several forms may be tested to identify the best representation in a specific application. It is usually assumed to be continuous and monotonic in z_{in}. For a linear in the parameters utility specification, h must be a fully determined function (meaning that is does not contain unknown parameters). Then we have

$$V_{in} = V(x_{in}).$$

A linear in the parameters function is denoted as follows

$$V_{in} = \sum_k \beta_k x_{ink} \ .$$

The deterministic term of the utility is therefore fully specified by the vector of parameters β.

The random part of the utility Among the many potential models that can be derived for the random parts of the utility functions, we describe below the most popular. The models within the Logit family are based on a probability distribution function of the maximum of a series of random variables, introduced by Gumbel (1958). Probit and Probit-like models are based on the Normal distribution motivated by the Central Limit Theorem.

The main advantage of the Probit model is its ability to capture all correlations among alternatives. However, due to the high complexity of its formulation, very few applications have been developed. The Logit model has been much more popular, because of its tractability, but it imposes restrictions on the covariance structure. They may be unrealistic in some contexts. The derivation of other models in the "Logit family" is aimed at relaxing restrictions, while maintaining tractability.

We discuss here the specification and properties of the models from the Logit family (the Multinomial Logit model, the Nested Logit model, the Cross-Nested model and the Generalized Extreme Value model). After presenting the Probit model, we introduce more advanced models. The Generalized Factor Analytical Representation and the Hybrid Logit models are designed to bridge the gap between Logit and Probit models. The Latent Class Choice model is a further extension designed to explicitly include in the model discrete unobserved factors.

The LOGIT Family

Logit-based models have been widely used for travel demand analysis. Practitioners and researchers have used, refined and extended the original Binary Logit Model to obtain a class of models based on similar assumptions. We refer to this class as the Logit-family.

Multinomial logit model The Logistic Probability Unit, or the Logit Model, was first introduced in the context of binary choice where the logistic distribution is used. Its generalization to more than two alternatives is referred to as the Multinomial Logit Model. The Multinomial Logit Model is derived from the assumption that the error terms of the utility functions are independent and identically Gumbel distributed (or Type I extreme value). That is, ε_{in} for all *i,n* is distributed as:

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

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

where η is a location parameter and μ is a strictly positive scale parameter. The mean of this distribution is

$$\eta + \gamma / \mu$$

where

$$\gamma = \lim_{k\to\infty} \sum_{i=1}^{k} \frac{1}{i} - \ln(k) \cong 0.5772$$

is the Euler constant. The variance of the distribution is

$$\pi^2/6\mu^2.$$

The probability that a given individual n chooses alternative i within the choice set C_n is given by

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

An important property of the Multinomial Logit Model is Independence from Irrelevant Alternatives (IIA). This property can be stated as follows: *The ratio of the probabilities of any two alternatives is independent of the choice set.* That is, for any choice sets C_1 and C_2 such that $C_1 \subseteq C_n$ and $C_2 \subseteq C_n$, and for any alternatives i and j in both C_1 and C_2, we have

$$\frac{\mathrm{P}(i|C_1)}{\mathrm{P}(j|C_1)} = \frac{\mathrm{P}(i|C_2)}{\mathrm{P}(j|C_2)}.$$

An equivalent definition of the IIA property is: *The ratio of the choice probabilities of any two alternatives is unaffected by the systematic utilities of any other alternatives.*

The IIA property of Multinomial Logit Models is a limitation for some practical applications. This limitation is often illustrated by the red bus/blue bus paradox in the modal choice context. We use here instead the following path choice example.

Consider a commuter traveling from origin O to destination D. He/she is confronted with the path choice problem described in Figure 2-1, where the choice set is {1,2a,2b} and the only attribute considered for the choice is travel time. We assume furthermore that the travel time for any alternative is the same, that is $V(1) = V(2a) = V(2b) = T$, and that the travel time on the small sections a and b is δ.

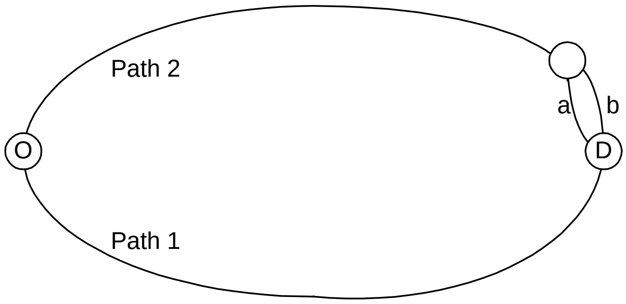

Figure 2-1. Path Choice Problem

The probability of each alternative provided by the Multinomial Logit Model for this example is

$$P(1|\{1,2a,2b\}) = P(2a|\{1,2a,2b\}) = P(2b|\{1,2a,2b\}) = \frac{e^{\mu T}}{\sum_{j\in\{1,2a,2b\}} e^{\mu T}} = \frac{1}{3}$$

Clearly, this result is independent of the value of δ. However, when δ is significantly smaller than the total travel time T, we expect the probabilities to be close to 50%/25%/25%. The Multinomial Logit Model is not consistent with this intuitive result. This situation appears in choice problems with significantly correlated random utilities, as it is clearly the case in the path choice example. Indeed, alternatives 2a and 2b are so similar that their utilities share many unobserved attributes of the path and, therefore, the assumption of independence of the random parts is not valid in this context.

Nested logit model The Nested Logit Model, first proposed by Ben-Akiva (1973 and 1974), is an extension of the Multinomial Logit Model designed to capture some correlations among alternatives. It is based on the partitioning of the choice set C_n into M nests C_{mn} such that

$$C_n = \bigcup_{m=1}^{M} C_{mn}$$

and

$$C_{mn} \cap C_{m'n} = \varnothing \quad \forall m \neq m'.$$

The utility function of each alternative is composed of a term specific to the alternative and a term associated with the nest. If i is an alternative from nest C_{mn}, we have

$$U_{in} = \tilde{V}_{in} + \tilde{\varepsilon}_{in} + \tilde{V}_{C_{mn}} + \tilde{\varepsilon}_{C_{mn}} .$$

The error terms $\tilde{\varepsilon}_{in}$ and $\tilde{\varepsilon}_{C_{mn}}$ are supposed to be independent. As in the Multinomial Logit Model, the error terms $\tilde{\varepsilon}_{in}$ are assumed to be independent and identically Gumbel distributed, with scale parameter μ_m (it can be different for each nest). The distribution of $\tilde{\varepsilon}_{C_{mn}}$ is such that the random variable $\max_{j \in C_{mn}} U_{jn}$ is Gumbel distributed with scale parameter μ. Each nest within the choice set is associated with a *composite utility*

$$V_{C_{mn}} = \tilde{V}_{C_{mn}} + \frac{1}{\mu_m} \ln \sum_{j \in C_{mn}} e^{\mu_m \tilde{V}_{jn}} .$$

The second term is called *expected maximum utility*, *LOGSUM*, *inclusive value* or *accessibility* in the literature. The probability for individual n to choose alternative i within nest C_{mn} is given by

$$P(i|C_n) = P(C_{mn}|C_n) P(i|C_{mn})$$

where

$$P(C_{mn} \mid C_n) = \frac{e^{\mu V_{C_{mn}}}}{\sum_{l=1}^{M} e^{\mu V_{C_{ln}}}} ,$$

and

$$P(i|C_{mn}) = \frac{e^{\mu_m \tilde{V}_{in}}}{\sum_{j \in C_{mn}} e^{\mu_m \tilde{V}_{jn}}} .$$

Parameters μ and μ_m reflect the correlation among alternatives within the nest C_{mn}. The covariance between the utility of two alternatives i and j in nest C_{mn} is

$$\mathrm{Cov}(U_{in}, U_{jn}) = \begin{cases} \mathrm{var}(\tilde{\varepsilon}_{C_{mn}}) & \text{if } i \text{ and } j \in C_{mn} \\ 0 & \text{otherwise} \end{cases}$$

and the correlation is

$$\mathrm{Corr}(U_{in}, U_{jn}) = \begin{cases} 1 - \dfrac{\mu^2}{\mu_m^2} & \text{if } i \text{ and } j \in C_{mn} \\ 0 & \text{otherwise} \end{cases} .$$

Therefore, as the correlation is non negative, we have

$$0 \leq \frac{\mu}{\mu_m} \leq 1,$$

and

$$\frac{\mu}{\mu_m} = 1 \iff \mathrm{corr}(U_{in}, U_{jn}) = 0.$$

The parameters μ and μ_m are closely related in the model. Actually, only their ratio is meaningful. It is not possible to identify them separately. A common practice is to arbitrarily constrain one of them to a specific value (usually 1).

As an example, we apply now the Nested Logit Model to the route choice problem described in Figure 1. We partition the choice set C_n={1,2a,2b} into C_{1n}={1} and C_{2n}={2a,2b}. The probability of choosing path 1 is given by

$$P(1 \mid \{1,2a,2b\}) = \frac{1}{1+2^{\frac{\mu}{\mu_2}}},$$

where μ_2 is the scale parameter of the random term associated with C_{2n}, and μ is the scale parameter of the choice between C_{1n} and C_{2n}. Note that we require $0 \leq \mu/\mu_2 \leq 1$. The probability of the two other paths is

$$P(2a \mid C_n) = P(2b \mid C_n) = \frac{1}{2}\frac{2^{\frac{\mu}{\mu_2}}}{1+2^{\frac{\mu}{\mu_2}}}.$$

In this example, we need to normalize either μ or μ_2 to 1. In the latter case we have

$$P(1|\{1,2a,2b\}) = \frac{1}{1+2^{\mu}}$$

and

$$P(2a|C_n) = P(2b|C_n) = \frac{1}{2}\left(\frac{2^{\mu}}{1+2^{\mu}}\right)$$

and we require that $0 \leq \mu \leq 1$. Note that for μ=1 we obtain the MNL result. For μ approaching zero, we obtain the expected result when paths 2a and 2b fully overlap. A model where the scale parameter μ is normalized to 1 is said to be "normalized from the top."

A model where one of the parameters μ_m is normalized to 1 is said to be "normalized from the bottom." The latter may produce a simpler formulation of the model. We illustrate it using the following example.

In the context of a mode choice with C_n={bus, metro, car, bike}, we consider a model with two nests: C_{1n}={bus,metro} contains the public transportation modes and C_{2n}={car,bike} contains the private transportation modes. For the example's sake, we consider the following deterministic terms of the utility functions:

$$V_{\text{bus}}=\beta_1\, t_{\text{bus}};\ V_{\text{metro}}=\beta_1\, t_{\text{metro}};\ V_{\text{car}}=\beta_2\, t_{\text{car}};\ V_{\text{bike}}=\beta_2\, t_{\text{bike}}$$

where t_i is the travel time using mode i and β_1 and β_2 are parameters to be estimated. Note that we have one parameter for private and one for public transportation, and we have not included the alternative specific constants in order to keep the example simple.

Applying the Nested Logit Model, we obtain

$$P(\text{bus})=\left(\frac{e^{\mu_1\beta_1 t_{\text{bus}}}}{e^{\mu_1\beta_1 t_{\text{bus}}}+e^{\mu_1\beta_1 t_{\text{metro}}}}\right)\frac{e^{\frac{\mu}{\mu_1}\ln\left(e^{\mu_1\beta_1 t_{\text{bus}}}+e^{\mu_1\beta_1 t_{\text{metro}}}\right)}}{e^{\frac{\mu}{\mu_1}\ln\left(e^{\mu_1\beta_1 t_{\text{bus}}}+e^{\mu_1\beta_1 t_{\text{metro}}}\right)}+e^{\frac{\mu}{\mu_2}\ln\left(e^{\mu_2\beta_2 t_{\text{car}}}+e^{\mu_2\beta_2 t_{\text{bike}}}\right)}}$$

Define $\theta_1=\mu/\mu_1$, $\theta_2=\mu/\mu_2$, $\beta_1^*=\mu_1\beta_1$ and $\beta_2^*=\mu_2\beta_2$ to obtain

$$P(\text{bus})=\frac{e^{\beta_1^* t_{\text{bus}}}}{e^{\beta_1^* t_{\text{bus}}}+e^{\beta_1^* t_{\text{metro}}}}\frac{e^{\theta_1\ln\left(e^{\beta_1^* t_{\text{bus}}}+e^{\beta_1^* t_{\text{metro}}}\right)}}{e^{\theta_1\ln\left(e^{\beta_1^* t_{\text{bus}}}+e^{\beta_1^* t_{\text{metro}}}\right)}+e^{\theta_2\ln\left(e^{\beta_2^* t_{\text{car}}}+e^{\beta_2^* t_{\text{bike}}}\right)}},$$

with $0\le\theta_1,\theta_2\le1$.

This formulation simplifies the estimation process. For this reason, it has been adopted by the Ben-Akiva and Lerman (1985) textbook and in estimation packages like ALOGIT (Daly, 1987) and HieLoW (Bierlaire, 1995, Bierlaire and Vandevyvere, 1995). We emphasize here that these packages should be used with caution when the same parameters are present in more than one nest. Specific techniques inspired from artificial trees proposed by Bradley and Daly (1991) must be used to obtain a correct specification of the model. In the above example, if $\mu_1=\mu_2$, then imposing the restriction $\beta_1=\beta_2$ is straightforward. However, for the case of $\mu_1\neq\mu_2$ and $\beta_1=\beta_2=\beta$, we define $\beta^*=\mu_1\mu_2\beta$ and create artificial nodes below each alternative, with a scale μ_2 for the first nest and scale μ_1 for the second. We refer the reader to Koppelman and Chen (1998) for further discussion.

A direct extension of the Nested Logit Model consists in partitioning some or all nests into sub-nests which can in turn, be divided into sub-nests. The model described above is valid at every layer of the nesting, and the whole model is generated recursively. Because of the complexity of these models, their structure is usually represented as a tree. Clearly, the number of potential structures reflecting the correlation among alternatives can be very large. No technique has been proposed thus far to identify the most appropriate correlation structure directly from the data.

The Nested Logit Model is designed to capture choice problems where alternatives within each nest are correlated. No correlation across nests can be captured by the Nested Logit Model. When alternatives cannot be partitioned into well separated nests to reflect their correlation, the Nested Logit Model is not appropriate.

Cross-nested logit model The Cross-Nested Logit Model is a direct extension of the Nested Logit Model, where each alternative may belong to more than one nest. Similar to the Nested Logit Model, the choice set C_n is partitioned into M nests C_{mn}. Moreover, for each alternative i and each nest m, parameters α_{im} ($0 \leq \alpha_{im} \leq 1$) representing the degree of "membership" of alternative i in nest m are defined. The utility of alternative i is given by

$$U_{imn} = \tilde{V}_{in} + \tilde{\varepsilon}_{in} + \tilde{V}_{C_{mn}} + \tilde{\varepsilon}_{C_{mn}} + \ln \alpha_{im}.$$

The error terms $\tilde{\varepsilon}_{in}$ and $\tilde{\varepsilon}_{C_{mn}}$ are independent. The error terms $\tilde{\varepsilon}_{in}$ are independent and identically Gumbel distributed, with unit scale parameter (this assumption is not the most general, but simplifies the derivation of the model). The distribution of $\tilde{\varepsilon}_{C_{mn}}$ is such that the random variable $\max_{j \in C_{mn}} U_{jmn}$ is Gumbel distributed with scale parameter μ. The probability for individual n to choose alternative i is given by

$$P(i|C_n) = \sum_{m=1}^{M} P(C_{mn}|C_n) P_n(i|C_{mn})$$

where

$$P(C_{mn} \mid C_n) = \frac{e^{\mu V_{C_{mn}}}}{\sum_{l=1}^{M} e^{\mu V_{C_{ln}}}},$$

$$P(i \mid C_{mn}) = \frac{\alpha_{im} e^{\tilde{V}_{in}}}{\sum_{j \in C_{mn}} \alpha_{jm} e^{\tilde{V}_{jn}}},$$

and $V_{C_{mn}} = \tilde{V}_{C_{mn}} + \ln \sum_{j \in C_{mn}} \alpha_{jm} e^{\tilde{V}_{jn}}$.

This model was first presented by McFadden (1978) as a special case of the GEV model that is presented below. It was applied by Small (1987) for departure time choice and by Vovsha (1998) for route choice.

Generalized extreme value model The Generalized Extreme Value (GEV) model has been derived from the random utility model by McFadden (1978). This general model consists of a large family of models that include the Multinomial Logit and the Nested Logit models. The probability of choosing alternative i within C_n is

$$P(i \mid C_n) = \frac{e^{V_{in}} \frac{\partial G}{\partial e^{V_{in}}}\left(e^{V_{1n}},\ldots,e^{V_{J_n}}\right)}{\mu G\left(e^{V_{1n}},\ldots,e^{V_{J_n}}\right)}.$$

J_n is the number of alternatives in C_n and G is a non-negative differentiable function defined on $\mathrm{IR}_+^{J_n}$ with the following properties:

1. G is homogeneous of degree $\mu > 0$[1],
2. $\lim_{x_i \to \infty} G(x_1,\ldots,x_i,\ldots,x_{J_n}) = \infty, \forall i = 1,\ldots,J_n$
3. the kth partial derivative with respect to k distinct x_i is non-negative if k is odd, and non-positive if k is even, that is, for any distinct $i_1,\ldots i_k \in \{1,\ldots J_n\}$ we have

$$(-1)^k \frac{\partial^k G}{\partial x_{i_1} \ldots \partial x_{i_k}}(x) \le 0 \;\; \forall x \in \mathrm{IR}_+^{J_n}.$$

The Multinomial Logit Model, the Nested Logit Model and the Cross-Nested Logit Model are GEV models, with

$$G(x) = \sum_{i=1}^{J_n} x_i^{\mu}$$

for the Logit model,

$$G(x) = \sum_{m=1}^{M} \left(\sum_{i \in C_{mn}} x_i^{\mu_m} \right)^{\frac{\mu}{\mu_m}}$$

for the Nested Logit model and

[1] McFadden's original formulation with μ=1 was generalized to μ>0 by Ben-Akiva and François (1983).

$$G(x) = \sum_{m=1}^{M} \left(\sum_{j \in C_n} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}$$

for the Cross-nested Logit model.

Multinomial Probit Model

The Probability Unit (or Probit) model should have been called Normit, for Normal Probability Unit model. It is derived from the assumption that the error terms of the utility functions are normally distributed. The Probit model captures explicitly the correlation among all alternatives. Therefore, we adopt a vector notation for the utility functions:

$$\boldsymbol{U}_n = \boldsymbol{V}_n + \varepsilon_n,$$

where $\boldsymbol{U}_n$, $\boldsymbol{V}_n$ and ε_n are $(J_n \times 1)$ vectors. The vector of error terms $\varepsilon_n = [\varepsilon_{1n}, \varepsilon_{2n}, ..., \varepsilon_{Jn}]^T$ is multivariate normal distributed with a vector of means $\mathbf{0}$ and a $J_n x J_n$ variance-covariance matrix Σ_n.

The probability that a given individual n chooses alternative i from the choice set C_n is given by

$$P(i|C_n) = P(U_{jn} - U_{in} \leq 0 \quad \forall j \in C_n) .$$

Denoting Δ_i the $(J_n\text{-}1 \times J_n)$ matrix such that

$$\Delta_i \boldsymbol{U}_n = [U_{1n}\text{-}U_{in}, \ldots, U_{(i-1)n}\text{-}U_{in}, U_{(i+1)n}\text{-}U_{in}, \ldots, U_{J_n n} \text{-}U_{in}]^T,$$

we have that

$$\Delta_i \boldsymbol{U}_n \sim N(\Delta_i \boldsymbol{V}_n, \Delta_i \Sigma_n \Delta_i^T).$$

The density function is given by

$$f_i(x) = \lambda \exp\left[-\frac{1}{2} (x - \Delta_i \boldsymbol{V}_n)^T (\Delta_i \Sigma_n \Delta_i^T)^{-1} (x - \Delta_i \boldsymbol{V}_n) \right]$$

where

$$\lambda = (2\pi)^{-\frac{J_n - 1}{2}} \, | \Delta_i \Sigma_n \Delta_i^T |^{-1/2}$$

and

$$P(i \mid C_n) = P(\Delta_i \boldsymbol{U}_n \leq 0) = \int_{-\infty}^{0} \ldots \int_{-\infty}^{0} f_i(x) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_{J_n} \; .$$

The matrix Δ_i is such that the i*th* column contains -1 everywhere. If the i*th* column is removed, the remaining (J_n–1×J_n–1) matrix is the identity matrix. For example, in the case of a trinomial choice model, we have

$$\Delta_2 = \begin{pmatrix} 1 & -1 & 0 \\ 0 & -1 & 1 \end{pmatrix}.$$

We note that he multifold integral becomes intractable even for a relatively low number of alternatives. Moreover, the number of unknown parameters in the variance-covariance matrix grows with the square of the number of alternatives. We refer the reader to McFadden (1989) for a detailed discussion of multinomial Probit models. We present below the Generalized Factor Analytic formulation designed to decrease the degree of complexity of Probit models.

Generalized Factor Analytic Specification of the Random Utility

The general formulation of the factor analytic formulation is

$$\boldsymbol{U}_n = \boldsymbol{V}_n + \varepsilon_\mathrm{n} = \boldsymbol{V}_n + \mathrm{F_n}\, \zeta_{\mathrm{n}},$$

where $\boldsymbol{U}_n$ is a (J_n×1) vector of utilities, $\boldsymbol{V}_n$ is a (J_n×1) vector of deterministic utilities, ε_n is a (J_n×1) vector of random terms, ζ_n is a (M×1) vector of factors which are IID standard normal distributed, and $\mathrm{F_n}$ is a $J_n \times M$ matrix of loadings that map the factors to the random utility vector. This specification is very general. If $M = J$, the number of alternatives in the universal set, we can define the matrix F as the Cholesky factor of the variance-covariance matrix Σ, that is $\Sigma = \mathrm{F\ F^T}$. $\mathrm{F_n}$ is then obtained by removing the rows associated with unavailable alternatives. We describe here special cases of factor analytical representations. They are discussed in more details by Ben-Akiva and Bolduc (1996).

Heteroscedasticity A heteroscedastic[2] model is obtained when $\mathrm{F_n}$ is a J_n×J_n diagonal matrix. Let T be a diagonal matrix containing the alternative specific standard deviations σ_i. $\mathrm{F_n}$ is obtained by removing the rows and columns of the unavailable alternatives. We obtain the following model, in scalar form:

$$U_{in} = V_{in} + \sigma_i \zeta_{in}.$$

[2] Heteroscedasticity here refers to different variances among the alternatives. We use it in this context to refer to a diagonal variance-covariance matrix with potentially different terms on the diagonal.

Factor analytic In this model, the general matrix F_n is divided into a matrix of loadings Q_n and a diagonal matrix T containing the factor specific standard deviations. We obtain the following model,

$$\boldsymbol{U}_n = \boldsymbol{V}_n + Q_n T \zeta_n.$$

Or, in scalar form:

$$U_{in} = V_{in} + \sum_{m=1}^{M} q_{imn} \sigma_m \zeta_{mn} ,$$

where q_{imn} are the elements of Q_n and σ_m are the diagonal elements of T. The matrix Q_n is normalized so that

$$\sum_{i,m} q_{imn}^2 = 1 \quad \forall n \cdot$$

When the matrix Q_n is known and does not need to be estimated the model is referred to as the Error Component Formulation.

General autoregressive process We consider the case where the error term ε_n is generated from a first-order autoregressive process:

$$\varepsilon_n = \rho W_n \varepsilon_n + T \zeta_n,$$

where W_n is a $(J_n \times J_n)$ matrix of weights describing the influence of each component of the error terms on the others, and $\zeta_n \sim N(0, I_{jn})$, as above. Then we have

$$\varepsilon_n = (I - \rho W_n)^{-1} T \zeta_n,$$

which is a special case of the factor analytic representation with

$$Q_n = (I - \rho W_n)^{-1}.$$

Hybrid Logit Model

The Multinomial Probit with a Logit kernel, or Hybrid Logit[3], model has been introduced by Ben-Akiva and Bolduc (1996). It is intended to bridge the gap between Logit and Probit models by combining the advantages of both of them. It is based on the following utility functions:

$$U_{in} = V_{in} + \xi_{in} + \upsilon_{in},$$

where ξ_{in} are normally distributed and capture correlation between alternatives, and υ_{in} are independent and identically distributed Gumbel variables. If the ξ_{in} are given, the model corresponds to a Multinomial Logit formulation:

[3] Sometimes called mixed logit

$$P(i|C_n,\xi_n)=\frac{e^{V_{in}+\xi_{in}}}{\sum_{j\in C_n}e^{V_{jn}+\xi_{in}}},$$

where $\xi_n=[\xi_1,..., \xi_J]^T$ is the vector of unobserved random terms. Therefore, the probability to choose alternative i is given by

$$P(i\mid C_n)=\int_{\xi_n}P(i\mid C_n,\xi_n)f(\xi_n)d\xi_n$$

where $f(\xi_n)$ is the probability density function of ξ_n. This model is a generalization of the Multinomial Probit Model when the distribution $f(\xi_n)$ is a multivariate normal. Other distributions may also be used. The earliest application of this model to capture random coefficients in the Logit Model (see below) was by Cardell and Dunbar (1980). More recent results highlighted the robustness of Hybrid Logit (see McFadden and Train, 1997).

Hybrid logit with factor analytical representation The Hybrid Logit Model can be combined with the factor analytical representation presented above to allow practical estimation using a simulated maximum likelihood procedure (see Ben-Akiva and Bolduc, 1996). The “Probit” error term is transformed using any appropriate factor analytical representation to obtain the following choice probability:

$$P(i|C_n)=\int_{\zeta_n}P(i|C_n,\zeta_n)N(0,I_M)d\zeta_n.$$

This formulation of the multinomial Probit is especially useful when the number of alternatives is so high that the use of probability simulators is required.

Random coefficients We conclude our discussion of the Hybrid Logit model with a formulation of the Multinomial Logit Model with randomly distributed coefficients:

$$\boldsymbol{U}_n=\boldsymbol{V}_n+\upsilon_n = X_n\beta_n+\upsilon_n.$$

Assume that β_n~$N(\beta,\Omega)$. If Γ is the Cholesky factor of Ω such that $\Gamma\Gamma^T=\Omega$, we replace β_n by $\beta+\Gamma\zeta_n$ to obtain

$$\boldsymbol{U}_n= X_n\,\beta+ X_n\,\Gamma\zeta_n+\upsilon_n.$$

It is an Hybrid Logit model with a factor analytic representation with $F_n= X_n\,\Gamma$.

Latent Class Choice Model

Latent class choice models are also designed to capture unobserved heterogeneity. The underlying assumption is that the heterogeneity is generated by discrete constructs. These constructs are not directly observable and therefore are represented by latent classes. For example, heterogeneity may be produced by taste variations across segments of the population, or when choice sets considered by individuals vary (latent choice set).

The latent class choice model is given by:

$$P(i|X_n) = \sum_{s=1}^{S} P(i|X_n;\beta_s,C_s)\,P(s|X_n;\theta)$$

where S is the number of latent classes, X_n is the vector of attributes of alternatives and characteristics of decision-maker n, β_s are the choice model parameters specific to class s, C_s is the choice set specific to class s, and θ is an unknown parameter vector.

The model

$$P(s|X_n;\theta)$$

is the class membership model, and

$$P(i|X_n;\beta_s,C_s)$$

is the class-specific choice model.

Special case: latent choice sets A special case is the choice model with latent choice sets:

$$P(i,n) = \sum_{C_n \in G} P(i \mid C_n)P(C_n)$$

where G is the set of all non-empty subsets of the universal choice set M, and $P(i|C_n)$ is a choice model. We note here that the size of G grows exponentially with the size of the universal choice set.

The latent choice set can be modeled using the concept of alternative availability. Then, a list of constraints or criteria are used to characterize the availability of alternatives. For each alternative i, a binary random variable A_{in} is defined such that A_{in}=1 if alternative i is available to individual n, and 0 otherwise. A list of K_{in} constraints is defined as follows:

$$A_{in} = 1 \text{ if } H_{ink} \geq 0,\ \forall k=1,\ldots,K_{in}.$$

For example, in a path choice context, one may consider that a path is not available is the ratio between its length and the shortest path length is above some threshold, represented by a random variable. The associated constraint for path i would then be:

$$L_i / L^* \geq 2+\varepsilon$$

where L^* is the length of the shortest path, L_i is the length of path i and ε a random variable with zero mean. It means that, on average, paths longer than twice the length of the shortest path are rejected.

The probability for an alternative to be available is given by

$$\mathrm{P}(A_{in} = 1) = \mathrm{P}(H_{ink} \geq 0\ \forall k=1,\ldots,K_{in}).$$

The latent choice set probability is then:

$$P(C_n) = \frac{P(A_{in} = 1, \forall i \in C_n \text{ and } A_{jn} = 0, \forall j \notin C_n)}{1 - P(A_{\mathrm{ln}} = 0, \forall l \in M)}.$$

If the availability criteria are assumed to be independent, we have

$$P(C_n) = \frac{\prod_{i \in C} P(A_{in} = 1) \prod_{j \notin C} P(A_{in} = 0)}{1 - \prod_{l \in M} P(A_l = 0)}.$$

Swait and Ben-Akiva (1987) estimate a latent choice set model of mode choice in a Brazilian city.

2.3 Route Choice Applications

The route choice problem can be stated as follows. Given a transportation network composed of nodes, links, origins and destinations; and given an origin *o*, a destination *d* and a transportation mode *m*, what is the chosen route between *o* and *d* on mode *m*. This discrete choice problem has specific characteristics. First, the universal choice set is usually very large. Second, not all physically feasible alternatives are considered by the decision-maker. Third, the alternatives are usually correlated, due to overlapping paths.

We now describe typical assumptions associated with route choice models.

Decision-Maker

The traveler's characteristics most often used for route choice applications are:

- Value-of-time. Obviously, travel time is a key attribute of alternative routes. Its influence on behavior, however, may vary across individuals. A Wall Street broker is likely to perceive and evaluate travel time differently from a retired Floridian. The sensitivity of an individual to travel time is usually referred to as the value-of-time. It can be represented by a continuous variable (e.g., the dollar-value equivalent of a minute spent traveling) or by a discrete variable identifying the decision-maker's value-of-time as low, medium or high.

- Access to information. Information about network conditions may significantly influence route choice behavior. Therefore, it may be important that a route choice model explicitly differentiates travelers with access to such information from those without access. It may be modeled by a single binary attribute (access/no access) or by several binary variables identifying the type of information available to the traveler (pre-trip information, on-board computer, etc.)

- Trip purpose. The purpose of the trip may significantly influence the route choice behavior. For example, a trip to work may be associated with a penalty for late arrival, while a shopping trip would usually have no such penalty. However, note that the trip purpose may be highly correlated with the value-of-time.

Alternatives

Identifying the choice set in a route choice context is a difficult task. Two main approaches can be considered.

First, it may be assumed that each individual can potentially choose any path between her/his origin and destination. The choice set is easy to identify, but the number of alternatives can be very large, causing operational problems in estimating and applying the model. Moreover, this assumption is behaviorally unrealistic.

Second, a restricted number of paths may be considered in the choice set. The choice set generation can be deterministic or stochastic, depending on the analyst's knowledge of the problem.

Dial (1971) proposes to include in the choice set "reasonable" paths composed of links that would not move the traveler farther away from her/his destination. The labeling approach (proposed by Ben-Akiva *et al*., 1984) includes paths meeting specific criteria, such as shortest paths, fastest paths, most scenic paths, paths with fewest stop lights, paths with least congestion, paths with greatest portion of freeways, paths with no left turns, etc.

An application of an implicit probabilistic choice set generation model has been proposed by Cascetta and Papola (1998), where the utility function associated with path i by individual n is defined as

$$U_{in} = V_{in} + \ln q_{in} + \varepsilon_{in},$$

where q_{in} is a random variable with mean

$$\bar{q}_{in} = \frac{1}{1 + e^{\sum_k -\gamma_k X_{ink}^A}}.$$

X_{ink}^A are the attributes for availability and perception of the path and γ_k are parameters to be estimated.

Some recent models (Nguyen and Pallottino, 1987, Nguyen, Pallottino and Gendreau, 1988) consider hyperpaths instead of paths as alternatives. An hyperpath is a collection of paths with associated strategies at decision nodes. This technique is particularly appropriate for a public transportation network.

Attributes

In describing the attributes of the alternatives to be included in the utility function, we need to distinguish between link-additive and non-link-additive attributes.

If i is a path composed of links $a \in \Gamma_i$, x_i is a link-additive attribute of i if

$$x_i = \sum_{a \in \Gamma_i} x_a \ ,$$

where x_a is the corresponding attribute of link a. For example, the travel time on a path is the sum of the travel times on links composing the path. Qualitative attributes are in general non-link-additive. For example, a binary variable x_i equal to one if the path is an habitual path and 0 otherwise, is non-additive. In the context of public transportation, variables like transfers and fares are usually not link-additive. The distinction is important because some models, designed to avoid path enumeration, use link attributes and not path attributes.

Among the many attributes that can potentially be included in a utility function, travel time is probably the most important. But what does travel time mean for the decision-maker? How does she/he perceive travel time? Many models are based on the assumption that most travelers are sufficiently experienced and knowledgeable about usual network conditions and, therefore, are able to estimate travel times accurately. This assumption may be satisfactory for planning applications using static models. With the emergence of Intelligent Transportation Systems, models that are able to predict the impact of real-time information have been developed. In this context, the "perfect knowledge" assumption is contradictory with the ITS services that provide information. Several approaches can be used to capture perceptions of travel times. One approach represents travel time as a random variable in the utility function. This idea was introduced by Burrell (1968) and is captured by a random utility model. Also, the uncertainty or the variability of travel time along a given path can be explicitly included as an attribute of the path.

In addition to travel time, the following attributes are usually included.

- <u>Path length</u>. The length of the path is likely to influence the decision maker's choice. Also, this attribute is easy to measure. Note that it may be highly correlated with travel time, especially in uncongested networks.

- <u>Travel cost</u>. In addition to the obvious behavioral motivation, including travel cost in the utility function is necessary to forecast the impact of tolls and congestion pricing, for example. It is common practice to distinguish the so-called out-of-pocket costs (like tolls), which are directly associated with a specific trip, from other general costs (like car operating costs).

- <u>Transit specific</u>. Attributes specific to route choice in transit networks include number of transfers, waiting and walking time and service frequency.

- <u>Others.</u> Traffic conditions (e.g. level of congestion, volume of conflicting traffic streams or pedestrian movements), obstacles (e.g. number of stop signs, number of traffic lights, number of left turns against traffic), road types (e.g. dummy variable capturing preference for freeways) and road condition (e.g. surface quality, number of lanes, safety, scenery) are some of the other attributes that may be considered. Whether to include them in the utility function depends on their behavioral pertinence in a specific context, and on data availability.

Decision Rules

Shortest path The simplest possible decision rule in the route choice context assumes that each individual chooses the path with the highest utility. Models based on deterministic utility maximization are supported by efficient algorithms to compute shortest paths in a graph (e.g. Dijkstra, 1959, and Dial, 1969). However, the behavioral limitations of this approach have motivated the development of stochastic models based on the random utility model.

Logit route choice A Multinomial Logit Model with an efficient algorithm for route choice has been proposed by Dial (1971). Using the concept of "reasonable paths" to define the choice set and assuming the paths attributes to be link-additive, this algorithm avoids explicit path enumeration.

As described earlier, the IIA property of the Multinomial Logit Model is the major weakness of Dial's algorithm in the context of highly overlapping routes. Therefore, its use is limited to networks with specific topologies. A Logit model may also be used with a choice set generation model, such as the Labeling approach, that results in a small size choice set with limited overlap.

Probit route choice Given the shortcomings of the Logit route choice model, Probit models have been proposed in the context of stochastic network loading by Burrell (1968) and Daganzo and Sheffi (1977). The two problems in this case are (i) the complexity of the variance-covariance matrix and (ii) the lack of an analytical formulation for the probabilities. The covariance structure can be simplified when path utilities are link-additive, the variance of link utility is proportional to the utility itself, and the covariance of utilities of two different links is zero. A Monte-Carlo simulation is often used to circumvent the absence of a closed analytical form.

C-Logit The C-Logit model, proposed by Cascetta *et al.* (1996) in the context of route choice, is a Multinomial Logit Model which captures the correlation among alternatives in a deterministic way. They add to the deterministic part of the utility function a term, called "commonality factor", that captures the degree of similarity between the alternative and all other alternatives in the choice set.

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

Cascetta *et al.* (1996) propose the following specification for the commonality factor

$$\mathrm{CF}_{in} = \beta_{\mathrm{CF}} \ln \sum_{j \in C_n} \left(\frac{L_{ij}}{\sqrt{L_i L_j}} \right)^{\gamma}$$

where L_{ij} is the length[4] of links common to paths i and j, and L_i and L_j are the overall length of paths i and j, respectively. β_{CF} is a coefficient to be estimated. The parameter γ may be estimated or constrained to a convenient value, often 1 or 2.

Considering the path choice example in Figure 1, the commonality factor for path 1 is zero because it does not overlap with any other path. The commonality factor for paths 2a and 2b is

$$\beta_{CF} \ln(1 + [(T\text{-}\delta)/T]^{\gamma}).$$

Note that the commonality factor of an alternative is not one of its attributes. It can be viewed as a measure of how the alternative is perceived within a choice set.

PS-Logit Path-Size Logit is an application of the notion of elemental alternatives and size variables. See Ben-Akiva and Lerman (Chapter 9) for details about models with elemental and aggregate alternatives. In the route choice context, we assume that an overlapping path may not be perceived as a distinct alternative. Indeed, a path contains links which may be shared by several paths. Hence, the size of a path with one or more shared links may be less than one. We include a size variable in the utility of a path to obtain the following model :

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

where the size S_{in} is defined by

$$S_{in} = \sum_{a \in \Gamma_i} \frac{l_a}{L_i} \frac{1}{\sum_{j \in C_n} \delta_{aj} \frac{L^*_{C_n}}{L_j}}$$

and Γ_i is the set of links in path i; l_a and L_i are the length of link a and path i, respectively; δ_{aj} is the link-path incidence variable that is one if link a is on path j and 0 otherwise; and $L^*_{C_n}$ is the length of the shortest path in C_n.

Considering again the path choice problem from Figure 1, the size of path 1 is 1, and the size of paths 2a and 2b is $(T+\delta)/2T$. It is interesting to note that the size variable formulation is equivalent to the commonality factor formulation for the

[4] or any other link-additive attribute

extreme cases where δ=0 or δ=T, assuming that β_{CF}=1 and for any value of γ. However, the two models are different for intermediary values.

2.4 Departure Choice Applications

Modeling the choice of departure time appears in the context of dynamic traffic assignment as an extension of the route choice problem. It is important to distinguish the departure time choice itself and the choice of changing departure time. The latter appears usually in the context of Traveler Information Systems, where individuals may revisit a previous choice using additional information. We now describe typical modeling assumptions associated with the departure time choice model.

Decision-Maker

The relevant traveler's socioeconomic characteristics are similar to those of route choice models. Additional characteristics important for departure time choice are desired arrival time and penalties for early and late arrival.

In the context of departure time change, the individual's "habitual" or "historical" departure time must also be known.

Alternatives

The choice set specification for departure time models is an intricate problem. First, the continuous time must be discretized. A reasonable compromise must be found between a fine temporal resolution and the model complexity. Indeed, there is a potentially large number of alternatives, particularly for realistic dynamic traffic applications. Second, the correlation among alternatives cannot be ignored, especially when time intervals are short. Choosing between the 7:45-7:50 and 7:50-7:55 time intervals differs from choosing between 7:45-7:50 and 8:45-8:50. In the first case, the two alternatives are likely to share unobserved attributes. Third, the perception of the alternatives depends on trip travel time. Most individuals round time and the rounding may depend on the travel time and travel time variability. For short trips, 7:52 may be rounded to 7:50, whereas for long trips it may be approximated by 8:00.

The choice set generation consists of defining an acceptable range of departure time intervals considered by an individual n. A common procedure is based on the desired arrival time AT^*_n. Let [$AT_{n,min}$; $AT_{n,max}$] be the feasible arrival time interval, and let [$TT_{n,min}$; $TT_{n,max}$] be the range of travel times. Then the interval of acceptable departure times is [$DT_{n,min}$; $DT_{n,max}$] = [$AT_{n,min}$-$TT_{n,max}$; $AT_{n,max}$-$TT_{n,min}$]. Small (1987) analyzed the impact of truncating the departure time choice set. He concluded that there is no problem if the true model is a Multinomial Logit Model. Some adjustments are needed if a Cross-Nested Logit with ordered alternatives is assumed.

In the context of departure time change, the alternatives may be described in a relative way. Antoniou *et al.* (1997) propose a choice set with five alternatives: do not change, switch to an earlier or a later departure, by one or two time intervals.

Attributes

Travel time is a key attribute of departure time alternatives. Other important attributes are the early and late schedule delays. Given a desired arrival time AT^*_n, a penalty-free interval is defined: $[AT^*_{n,min};AT^*_{n,max}]$. It is assumed that the individual suffers no penalty if the arrival times lies within the interval. The actual arrival time AT_n is equal to $DT_n+TT(DT_n)$,where $TT(DT_n)$ is the travel time if the trip starts at time DT_n. The early schedule delay is defined as

$$\text{Max}\ [AT^*_{n.min}-AT_n\ ,\ 0]$$

and the late schedule delay is defined as

$$\text{Max}\ [AT_n-AT^*_{n.max}\ ,\ 0].$$

In the context of departure time change, a penalty can also be associated with departure times very different from the habitual choice, capturing the inertia associated with habits.

Decision Rules

Small (1982) and Cascetta *et al*. (1992) use Multinomial Logit Models for departure time choice. However, the intrinsic aforementioned correlation among alternatives is not captured by such models. Small (1987) proposed an Ordered Generalized Extreme Value model. It is a Cross-Nested Logit Model, where *m* adjacent departure time intervals are nested together, capturing their intrinsic correlation. A single departure time interval belongs to *m* different nests, source of the cross-nested structure.

In the context of departure time change, Antoniou *et al*. (1997) propose a Nested Logit Model for joint choice of departure time and route. Liu and Mahmassani (1998) propose a Probit model where day-to-day correlation is assumed.

2.5 Conclusion

Discrete choice methods are constantly evolving to accommodate the requirements of specific applications. This is an exciting field of research, where a deep understanding of the underlying theoretical assumptions is necessary both to apply the models and develop new ones. In this Chapter, we have summarized the fundamental aspects of discrete choice theory, and we have introduced recent model developments, illustrating their richness. A discussion on route choice and departure time choice applications have shown how specific aspects of real applications must be addressed.

Acknowledgment

We wish to thank Scott Ramming for his useful input and suggestions. We also benefited from comments by Joan Walker, Andrea Papola and Julie Bernardi.

References

Anderson, S. P., de Palma, A. and Thisse, J.-F. (1992). *Discrete Choice Theory of Product Differentiation*, MIT Press, Cambridge, Ma.

Antoniou, C., Ben-Akiva, M.E., Bierlaire, M., and Mishalani, R. (1997) Demand Simulation for Dynamic Traffic Assignment. Proceedings of the 8th IFAC/IFIP/IFORS symposium on transportation systems.

Ben-Akiva, M. and François, B. (1983). μ homogeneous generalized extreme value model, *Working paper*, Department of Civil Engineering, MIT, Cambridge, Ma.

Ben-Akiva, M. E. (1973). *Structure of passenger travel demand models*, PhD thesis, Department of Civil Engineering, MIT, Cambridge, Ma.

Ben-Akiva, M. E. (1974). Structure of passenger travel demand models, *Transportation Research Record* **526**.

Ben-Akiva, M. E. and Boccara, B. (1995). Discrete choice models with latent choice sets, *International Journal of Research in Marketing* **12**: 9–24.

Ben-Akiva, M. E. and Bolduc, D. (1996). Multinomial probit with a logit kernel and a general parametric specification of the covariance structure, presented at the 3rd International Choice Symposium, Columbia University.

Ben-Akiva, M. E. and Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, Ma.

Ben-Akiva, M. E., Bergman, M. J., Daly, A. J. and Ramaswamy, R. (1984). Modeling inter-urban route choice behaviour, *in* J. Volmuller and R. Hamerslag (eds), *Proceedings from the ninth international symposium on transportation and traffic theory*, VNU Science Press, Utrecht, Netherlands, pp. 299–330.

Ben-Akiva, M. E., Cyna, M. and de Palma, A. (1984). Dynamic model of peak period congestion, *Transportation Research B* **18**(4–5): 339–355.

Bierlaire, M. (1995).A robust algorithm for the simultaneous estimation of hierarchical logit models, *GRT Report 95/3*, Department of Mathematics, FUNDP.

Bierlaire, M. (1998). Discrete choice models, *in* M. Labbé, G. Laporte, K.Tanczos and Ph. Toint (eds), *Operations Research in Traffic and Transportation Management*, Vol. 166 of *NATO ASI Series, Series F: Computer and Systems Sciences*, Springer Verlag, pp. 203-227.

Bierlaire, M. and Vandevyvere, Y. (1995).*HieLoW: the interactive user's guide*, Transportation Research Group - FUNDP, Namur.

Bierlaire, M., Lotan, T. and Toint, Ph. L. (1997). On the overspecification of multinomial and nested logit models due to alternative specific constants, *Transportation Science* **31**(4): 363–371.

Bolduc, D., Fortin, B. and Fournier, M.-A. (1996). The effect of incentive policies on the practice location of doctors: A multinomial Probit analysis, *Journal of labor economics* **14**(4): 703.

Bradley, M. A. and Daly, A. (1991). Estimation of logit choice models using mixed stated preferences and revealed preferences information, *Methods for understanding travel behaviour in the 1990's*, International Association for Travel Behaviour, Québec, pp.~116--133. 6th international conference on travel behaviour.

Burrell, J. E. (1968). Multipath route assignment and its application to capacity restraint. *Proceedings of the 4th international symposium on the theory of road traffic flow*, Karslruhe, Germany.

Cardell and Dunbar (1980).Measuring the societal impacts of automobile downsizing, *Transportation Research A* **14**(5–6): 423–434.

Cascetta, E. and Papola, A. (1998). Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand, *Technical report*, Universita degli Studi di Napoli Federico II.

Cascetta, E., Nuzzolo, A. and Biggiero, L. (1992). Analysis and Modeling of Commuters' Departure Time and Route Choice in Urban Networks. *Proceedings of the Second International CAPRI Seminar on Urban Traffic Networks.*

Cascetta, E., Nuzzolo, A., Russo, F. and Vitetta, A. (1996). A modified logit route choice model overcoming path overlapping problems. Specification and some calibration results for interurban networks, *Proceedings of the 13th International Symposium on the Theory of Road Traffic Flow (Lyon, France).*

Chang, G. L. and Mahmassani, H. S. (1986). Experiments with departure time choice dynamics of urban commuters, *Transportation Research B* **20**(4): 297–320.

Chang, G. L. and Mahmassani, H. S. (1988). Travel time prediction and departure time adjustment dynamics in a congested traffic system, *Transportation Research B* **22** (3): 217–232.

Daganzo, C. F. and Sheffi, Y. (1977). On stochastic models of traffic assignment. *Transportation Science* **11**(3): 253–274.

Daly, A. (1987). Estimating "tree" logit models, *Transportation Research B* **21**(4): 251–268.

de Palma, A., Khattak, A. J. and Gupta, D. (1997). Commuters' departure time decisions in Brussels, Belgium, *Transportation Research Record* **1607**: 139–146.

Dial, R. B. (1969). Algorithm 360: shortest path forest with topological ordering., *Communications of ACM* **12**: 632–633.

Dial, R. B. (1971). A probabilistic multipath traffic assignment algorithm which obviates path enumeration, *Transportation Research* **5**(2): 83–111.

Dijkstra, E. W. (1959). A note on two problems in connection with graphs, *Numerische Mathematik* **1**: 269–271.

Gumbel, E. J. (1958). *Statistics of Extremes*, Columbia University Press, New York.

Hendrickson, C. and Kocur, G. (1981). Schedule delay and departure time decisions in a deterministic model, *Transportation Science* **15**: 62–77.

Hendrickson, C. and Plank, E. (1984). The flexibility of departure times for work trips, *Transportation Research A* **18**: 25–36.

Hensher, D. A. and Johnson, L. W. (1981). *Applied discrete choice modeling*, Croom Helm, London.

Horowitz, J. L., Koppelman, F. S. and Lerman, S. R. (1986). *A self-instructing course in disaggregate mode choice modeling*, Technology Sharing Program, US Department of Transportation, Washington, D.C. 20590.

Khattak, A. J. and de Palma, A. (1997). The impact of adverse weather conditions on the propensity to change travel decisions: a survey of Brussels commuters, *Transportation Research A* **31**(3): 181–203.

Koppelman, F. S. and Wen, C.-H. (1997). The paired combinatorial logit model: properties, estimation and application, Transportation Research Board, 76th Annual Meeting, Washington DC. Paper #970953.

Koppelman, F. S. and Wen, C.-H. (1998). Alternative nested logit models: Structure, properties and estimation, *Transportation Research B*. (forthcoming).

Liu, Y.-H. and Mahmassani, H. (1998). Dynamic aspects of departure time and route decision behavior under ATIS: modeling framework and experimental results, presented at the 77th annual meeting of the Transportation Research Board, Washington DC.

Luce, R. (1959). *Individual choice behavior: a theoretical analysis*, J. Wiley and Sons, New York.

Luce, R. D. and Suppes, P. (1965). Preference, utility and subjective probability, *in* R. D. Luce, R. R. Bush and E. Galanter (eds), *Handbook of Mathematical Psychology*, J. Wiley and Sons, New York.

Manski, C. (1977). The structure of random utility models, *Theory and Decision* **8**: 229–254.

McFadden, D. (1978). Modeling the choice of residential location, *in* A. K. et al. (ed.), *Spatial interaction theory and residential location*, North-Holland, Amsterdam, pp. 75–96.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration, *Econometrica* **57**(5): 995–1026.

McFadden, D. and Train, K. (1997). Mixed multinomial logit models for discrete response, *Technical report*, University of California, Berkeley, Ca.

Nguyen, S. and Pallottino, S. (1987). Traffic assignment for large scale transit networks, *in* A. Odoni (ed.), *Flow control of congested networks*, Springer Verlag.

Nguyen, S., Pallottino, S. and Gendreau, M. (1998). Implicit Enumeration of Hyperpaths in a Logit Model for Transit Networks. *Transportation Science* **32**(1).

Small, K. (1982). The scheduling of consumer activities: work trips, *The American Economic Review* pp. 467–479.

Small, K. (1987). A discrete choice model for ordered alternatives, *Econometrica* **55**(2): 409–424.

Swait, J.D. and Ben-Akiva, M. (1987). Incorporating Random Constraints in Discrete Models of Choice Set Generation, *Transportation Research B* **21**(2).

Tversky, A. (1972). Elimination by aspects: a theory of choice, *Psychological Review* **79**: 281–299.

Vovsha, P. (1997). Cross-nested logit model: an application to mode choice in the Tel-Aviv metropolitan area, Transportation Research Board, 76th Annual Meeting, Washington DC.Paper #970387.

Whynes, D., Reedand, G. and Newbold, P. (1996). General practitioners' choice of referral destination: A Probit analysis, *Managerial and Decision Economics* **17**(6): 587.

Yai, T., Iwakura, S. and Morichi, S. (1997). Multinomial Probit with structured covariance for route choice behavior, *Transportation Research B* **31**(3): 195–208.