Introduction to disaggregate demand models

Michel Bierlaire * Virginie Lurkin *

Mai 08, 2017

Report TRANSP-OR 170508
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
transp-or.epfl.ch

*École Polytechnique Fédérale de Lausanne (EPFL), School of Architecture, Civil and Environmental Engineering (ENAC), Transport and Mobility Laboratory, Switzerland, {michel.bierlaire,virginie.lurkin}@epfl.ch

Abstract

Demand information is an input for a great deal of operations research models. Assumed as given in many problem instances addressed in the literature, demand data are difficult to generate. In this tutorial, we provide an introduction to disaggregate demand models that are designed to capture in detail the underlying behavioral mechanisms at the foundation of the demand.

1 Introduction

Most, if not all, operations research problems require demand data as input. For instance, the min-cost flow problem and the vehicle routing problem need the total amount of flow that is consumed/generated at each node. The multi-commodity flow problem uses as input, for each commodity, its origin, its destination, and the amount of flow of the commodity to be transported. The facility location problem requires a list of demand points. The management of the supply chain relies on the knowledge of downstream demand, and deals with the associated uncertainty.

Aggregate representations of demand are commonly used, typically in the form of flows. It is therefore rather common in practice to use aggregate statistical methods, such as time series analysis, to predict the demand data that is feeding the operations research models. Unfortunately, this aggregate modeling approach is not able to capture the deep causal mechanisms that generate the demand. Indeed, demand is actually the result of many decisions performed by individual actors in the system, typically customers who have decided to consume a specific good, or requested a specific service. The derivation of aggregate demand indicators from disaggregate demand models is the topic of this tutorial.

The state-of-the-art for the mathematical modeling of disaggregate demand relies on discrete choice models. Rooted in the theoretical foundations of microeconomics, these models are powerful operational tools to capture the causality between a vast set of explanatory variables and the choice itself. Many success stories have been reported in the scientific literature and in practice, especially in the fields of quantitative marketing and transportation planning. Still, choice models are rarely mentioned in the operations research literature. One reason may be that their mathematical properties are not always convenient for such a use. In addition to being probabilistic, they are nonlinear and non-convex in the variables of interest. Before introducing the methodology itself, we mention below some pieces of work from the operations research literature that have investigated the integration of choice models into operations research problems.

1.1 Choice Models in Operations Research

In the last decade there has been a growing body of literature on facility location problems that incorporate customer's choice behavior. In early publications, several authors address the problem of finding the optimal location of new facilities in a competitive market, using discrete choice theory to model the preferences of customers (e.g., Benati, 1999, Benati and Hansen, 2002, Haase, 2009). Applications on school location (e.g., Müller et al., 2009, Müller et al., 2012, Haase and Müller, 2013, Castillo-López and López-Ospina, 2015), health care facility location (e.g., Zhang et al., 2012, Haase and Müller, 2015), airline scheduling (e.g., Schön, 2007), retail facility location (e.g., Müller and Haase, 2014), and Park and Ride (P&R) facility location (e.g., Aros-Vera et al., 2013) have been proposed.

The integration of choice models in optimization is also present in the traffic assignment (e.g., Kant, 2008, Pel et al., 2009, Qian, 2011, Qian and Zhang, 2013) and network pricing (e.g., Gilbert et al., 2014a, Gilbert et al., 2014b) literatures.

Researchers from the revenue management (RM) community have also begun to investigate how discrete choice models can be integrated with optimization models. These models, referred to as "choice-based RM", aim at maximizing both revenue and customer satisfaction by deciding about pricing while controlling for product availability. Choice-based models for revenue management were introduced by Andersson, 1998. To date, several theoretical and empirical studies have appeared in the research community (e.g., Gallego and Phillips, 2004, Talluri and Van Ryzin, 2004, Ratliff et al., 2008, Bront et al., 2009, Bodea et al., 2009, Vulcano et al., 2010, Vulcano et al., 2012).

This tutorial has been motivated by a desire to see further interactions between operations research and discrete choice modeling researchers. Researchers from these two communities have generally different perspectives and research priorities but would benefit from deeper collaborations. Deeper collaborations require inevitably the ability of operations researchers to gain a greater appreciation for discrete choice models. Our primary objective is therefore to provide an introductory-level overview of discrete choice models to students, practitioners, faculty, and researchers with an operations research background. We hope that this tutorial can encourage them to venture into discrete choice modeling and can motivate them to exploit the power of these advanced mathematical methods.

The rest of the tutorial is structured as follows. In Section 2, we provide relevant background information from microeconomics and discrete choice theories. We then present the logit model in Section 3. In Section 4 we consider the problem of estimating the model parameters by maximum likelihood. This is followed by useful model applications in section 5. Finally, Section 6 summarizes the key points in the tutorial.

2 Foundations

Disaggregate demand models are rooted in microeconomics, the branch of economics that focuses on the decision-making behavior of economic actors. In this tutorial, we refer to these actors as *individuals*, although they can also be households or firms, for instance. In the next section, we show that the main concepts from microeconomics are derived from principles in optimization. We then continue by introducing the assumptions about choice modeling.

2.1 Traditional microeconomics

Consider a set X of goods, bundles, or actions. The objective is to determine what element(s) of X will be chosen/purchased by a given individual. The preferences of the individual are assumed to be characterized by a preference-indifference operator $\succeq$. Consider two goods a and b in X. Then $a \succeq b$ means that the individual either prefers a to b, or is indifferent between a and b. Other operators can be derived from the preference-indifference operator:

- $a \sim b$ is defined as ($a \preceq b$ and $b \preceq a$) and means that the individual is indifferent between a and b,

- $a \succ b$ is defined as (not $a \preceq b$) and means that the individual strictly prefers a to b.

Operators $\preceq$ and $\prec$ can be defined similarly.

A fundamental assumption is that each individual is *rational*. Formally, it means that her preferences must satisfy completeness and transitivity over the set X. *Completeness* means that, for each $a, b \in X$, it is possible to decide if $a \preceq b$ is true or false. *Transitivity* means that, for any $a, b, c \in X$, if $a \preceq b$ and $b \preceq c$, then $a \preceq c$.

It is possible to represent the preference structure of individuals using a *utility function* (see Debreu, 1954). Let $u : X \to \mathbb{R}$ be a function mapping the set of goods to the real numbers. We say that u represents $\preceq$ on X if

$$a \preceq b \iff u(a) \leq u(b).$$

The transitivity and completeness of the preferences guarantee the existence of a utility function. It can also be shown that it is unique, up to order preserving transformations[1]. Appropriate assumptions can also be made on the preference structure in order to obtain desirable properties of the utility function, such as continuity or differentiability. We refer the reader to textbooks in microeconomics such as Nicholson and Snyder, 2007, Pindyck and Rubinfeld, 2008, or Varian and Repcheck, 2010 for more details.

At this point, we have the mathematical tool that allows to introduce the second major assumption: each individual is a *utility maximizer*. It means that, when making a decision, each individual is solving the following optimization problem:

$$\max_{q} u(q) \quad (1)$$

subject to

$$q \in X. \quad (2)$$

Consider a concrete example, where the individual wants to purchase various items from a catalog of L products. Each item i is associated with a price p_i, and the individual has to decide, for each item i the quantity

[1] An equivalence relation on utility functions can be defined, where u and v are equivalent if v is the composition of u and a strictly increasing function g, that is $v(a) = g(u(a))$ for each $a \in X$. The uniqueness applies across equivalence classes.

q_i that she wants to purchase. Assuming that her total budget is I, the optimization problem to be solved is

$$\max_{q \in \mathbb{R}^L} u(q) \tag{3}$$

subject to

$$p^T q \leq I, \tag{4}$$
$$q \geq 0. \tag{5}$$

For instance, consider the classical Cobb-Douglas utility function (see Goldberger, 1968, among many references), parameterized with positive parameters θ_ℓ, $\ell = 0, \ldots, L$:

$$u(q) = \theta_0 \prod_{\ell=1}^{L} q_\ell^{\theta_\ell}. \tag{6}$$

Exercise: Derive the demand function for $u(q) = \sum_\ell \frac{q_\ell^\lambda}{\lambda}$, where $\lambda > 0$.

As the utility function is defined up to an order preserving transformation, it is equivalent to solve the following problem:

$$\max_{q \in \mathbb{R}^L} \ln u(q) = \ln \theta_0 + \sum_{\ell=1}^{L} \theta_\ell \ln(q_\ell) \tag{7}$$

subject to

$$p^T q \leq I, \tag{8}$$
$$q \geq 0, \tag{9}$$

where it is implicitly assumed that all items are purchased at the optimal solution, that is $q^* > 0$. This is done for the sake of simplifying the example, and does not really modify the generality of the results. Denote by $\mu_I \in \mathbb{R}$ the Lagrange multiplier of the budget constraint (8). At the optimal solution, that constraint must be binding as the utility increases with purchased quantities, so that there is no incentive not to spend the whole budget. Therefore, in order to verify the complementarity slackness conditions, we must have $\mu_I^* > 0$. The economic intuition behind this multiplier is that it represents the marginal increase in utility due to a marginal increase in total disposable income. As $q^* > 0$, the Lagrange multiplier μ_ℓ, associated with the ℓth non-negativity constraint $q_\ell \geq 0$, must be zero at the optimal solution, again from the complementarity slackness condition, that is $\mu_\ell^* = 0$.

Consequently, the first order necessary optimality conditions are

$$\mu_\ell^* = \mu_I^* p_\ell - \frac{\theta_\ell}{q_\ell^*} = 0, \ \forall \ell \tag{10}$$

or, equivalently,

$$\mu_I^* p_\ell q_\ell^* = \theta_\ell, \ \forall \ell. \tag{11}$$

Summing over all items, and using the fact that the budget constraint is binding, we obtain

$$\mu_I^* = \sum_k \theta_k / I. \tag{12}$$

Therefore, the multiplier can be eliminated from (11), that becomes

$$q_\ell^* = \frac{\theta_\ell}{\mu_I^* p_\ell} = \frac{I}{p_\ell} \frac{\theta_\ell}{\sum_k \theta_k}. \tag{13}$$

Equation (13), that provides the quantity of each product purchased by the individual as a function of the price of the item and the total budget is called a *demand function*. It is characterized by the optimality conditions of the utility maximization problem. Note that, in this example, the quantity of product ℓ depends on the price of the product ℓ only, and not on the price of other products. This is due to the specific form of the objective function, and is not a general property.

Now that the optimal quantity is known, it is possible to calculate the maximum possible utility that can be achieved, by inserting (13) into (6). We obtain

$$u(I, p; \theta) = \theta_0 \prod_{\ell=1}^{L} \left(\frac{I}{p_\ell} \frac{\theta_\ell}{\sum_k \theta_k} \right)^{\theta_\ell}. \tag{14}$$

This quantity is called the *indirect utility* function. It is the optimal value of the objective function at given prices and income levels. Its value is a function of the parameters of the problem.

Consider now that there are both continuous and discrete decisions to be made by the individual. For instance, choosing a car to purchase (discrete decision) and deciding about the total number of kilometers to drive every year (continuous decision). Or choosing a shopping center in the neighborhood (discrete decision) and purchasing a given bundle of products (continuous decision).

In addition to the continuous quantities $q \in \mathbb{R}^L$, there is now a set of binary decision variables y_i, $i = 1, \ldots, J$ that correspond to all possible discrete decisions that can be taken in the given context. The variable y_i takes the value one if decision i is taken, and zero otherwise. Each discrete decision i is associated with a value z_i and a cost c_i. The optimization problem solved by the individual is now the following:

$$\max_{q \in \mathbb{R}^L, y \in \{0,1\}^J} u(q, z^T y) \tag{15}$$

subject to

$$p^T q + c^T y \leq I, \tag{16}$$

$$q \geq 0. \tag{17}$$

This is a mixed integer optimization problem. Clearly, there is no optimality condition, so that demand functions cannot be directly derived.

Suppose that the discrete decisions are given, so that y is fixed. The problem (15)–(17) now contains only continuous variables. Consequently, demand functions can be derived as described above. These demand functions for the continuous products, conditional to the discrete decisions y, are called *conditional demand functions* and are denoted

$$q_{\ell|y}(I - c^T y, p, z^T y), \tag{18}$$

where $I - c^T y$ is the budget left after having paid the cost associated with the discrete decisions. Note that if this quantity happens to be negative, the individual cannot afford the set of discrete decisions characterized by y, and the conditional optimization problem is infeasible. The indirect utility is calculated by inserting the conditional demand (18) into the objective function (15), as illustrated by (14).

Therefore, the problem becomes a knapsack problem:

$$\max_{y \in \{0,1\}^J} u(q_y, z^T y) \tag{19}$$

subject to

$$c^T y \leq I - p^T q_y \tag{20}$$

where q_y is the vector of quantities provided by the conditional demand function (18), if y is the vector of discrete decisions.

This short introduction to microeconomics serves two purposes:

1. emphasize the role of optimization in microeconomic theory,

2. introduce the concepts of demand functions and indirect utility.

We now proceed with the introduction to discrete choice models that are the major ingredients of disaggregate demand models.

2.2 Discrete choice

Discrete choice models build on the above-mentioned theoretical derivation to represent choice behavior of individuals. In order to illustrate the concepts, we consider the (discrete) choice of a transportation mode to commute to work.

In the rest of this tutorial, we use the following notations:

- n: the individual, or the decision maker;

- $\mathcal{C}_n$: the choice set, that is the set of alternatives considered by the individual n, assumed finite and discrete (for example, the set of available transportation modes can be car, train, bus, walking, and biking),

- J_n: the number of alternatives in $\mathcal{C}_n$,

- U_{in}: the (indirect)[2] utility associated by individual n with alternative i, that is the objective function (19),

- z_{in}: the vector of attributes of alternative i for individual n (for example, the cost of the trip for each mode, the travel time for each mode),

- s_n: the vector of socio-economic characteristics of individual n (for example, the income, the age, the trip purpose, the level of education, the professional activity).

[2]We will drop this qualifier in the rest of the paper.

We also assume without loss of generality that exactly one alternative in $\mathcal{C}_n$ is selected. From the discussion in Section 2.1, alternative $i \in \mathcal{C}_n$ is chosen by individual n if

$$U_{in} \geq U_{jn}, \ \forall j \in \mathcal{C}_n. \tag{21}$$

In practice, the exact specification of U_{in} is unknown to the analyst. Moreover, any practical implementation of the model suffers from measurement errors. For these reasons, the utility U_{in} is modeled as a continuous random variable. Typically, it is defined as

$$U_{in} = V_{in}(z_{in}, s_n) + \varepsilon_{in}, \tag{22}$$

where $V_{in}(z_{in}, s_n)$ is a deterministic function of the attributes and the socio-economic characteristics, and ε_{in} is a continuous error term, capturing the specification and measurement errors. As a consequence, the choice model becomes probabilistic. The probability for individual n to choose alternative i is defined from (21) as

$$P_n(i|\mathcal{C}_n) = \text{Prob}(U_{in} \geq U_{jn}, \ \forall j \in \mathcal{C}_n). \tag{23}$$

Note that $\sum_{i \in \mathcal{C}_n} P_n(i|\mathcal{C}_n) = 1$. Also, as the error term is a continuous random variable, the probability of a tie is zero. It is also possible to write the model with the same choice set for everybody. Indeed, define

$$\mathcal{C} = \bigcup_n \mathcal{C}_n, \tag{24}$$

the set of J alternatives in choice set $\mathcal{C}$ and, for each individual, associate each alternative i in $\mathcal{C}$ with a binary variable a_{in}, that is one if $i \in \mathcal{C}_n$ and zero otherwise. It represents the availability of alternative i for individual n. Then, as the exponential is a strictly increasing function, (23) can be written

$$P_n(i) = \text{Prob}(a_{in} \exp(U_{in}) \geq a_{jn} \exp(U_{jn}), \ \forall j \in \mathcal{C}). \tag{25}$$

It is sometimes convenient to write it as

$$P_n(i) = \text{Prob}(U_{in} + \ln a_{in} \geq U_{jn} + \ln a_{jn}, \ \forall j \in \mathcal{C}), \tag{26}$$

understanding that $\ln(0) = -\infty$.

Concrete choice models can be derived from specific assumptions about the distribution of the error terms. For instance, if $\varepsilon_n = (\varepsilon_{1n}, \ldots, \varepsilon_{Jn})$ is a multivariate random variable with cdf

$$F_{\varepsilon_n}(\varepsilon_1, \ldots, \varepsilon_J) \tag{27}$$

it can be shown that

$$P_n(i|\mathcal{C}_n) = \int_{\varepsilon=-\infty}^{+\infty} \frac{\partial F_{\varepsilon_{1n},\varepsilon_{2n},\ldots,\varepsilon_J}}{\partial \varepsilon_i}(\ldots, V_{in} - V_{(i-1)n} + \varepsilon, \varepsilon, V_{in} - V_{(i+1)n} + \varepsilon, \ldots) d\varepsilon. \tag{28}$$

Exercise: Derive the model for $J = 2$ and $F_{\varepsilon_1,\varepsilon_2}(x_1, x_2) = \exp(-\exp(-x_1))\exp(-\exp(-x_2))$.

As the above form of the model is in general intractable, more specific assumptions can be used to derive simpler versions of the model, as illustrated in the next section.

3 The logit model

The logit model is the most widely used choice models in practice. As a complete review of all relevant models is beyond the scope of this paper, we focus only on logit. We refer the reader to the vast literature on the topic, and in particular to Ben-Akiva and Lerman, 1985, Train, 2003, Ben-Akiva and Bierlaire, 2003, Garrow, 2016, to cite just a few. We first present the case of binary choice situations, where a choice has to be made between two alternatives. We then generalize to larger choice sets.

3.1 The binary logit model

The *binary logit* model considers the special case where the choice set $\mathcal{C}$ contains exactly two alternatives, $\mathcal{C} = \{1, 2\}$, that are both available for every individual. Taking again the example of commuting to work, the two alternatives could be driving a car (alternative 1) and taking the train (alternative 2). In that case, (26) is written

$$P_n(1) = \text{Prob}(U_{1n} \geq U_{2n}), \tag{29}$$

where U_{in} is defined by (22). The probability of individual n choosing alternative 2 is trivially given by

$$P_n(2) = 1 - P_n(1). \tag{30}$$

Substituting (22) into (29), we obtain

$$P_n(1) = \text{Prob}(V_{1n} + \varepsilon_{1n} \geq V_{2n} + \varepsilon_{2n}) \tag{31}$$
$$= \text{Prob}(\varepsilon_{2n} - \varepsilon_{1n} \leq V_{1n} - V_{2n}) \tag{32}$$
$$= \text{Prob}(\varepsilon_n \leq V_{1n} - V_{2n}), \tag{33}$$

where $\varepsilon_n = \varepsilon_{2n} - \varepsilon_{1n}$.

As mentioned in Section 2.2, concrete choice models can be derived from specific assumptions about the distribution of the error terms[3] ε_{1n} and ε_{2n}. Consistently with the fact that we consider the maximum utility, we rely on the statistical theory of extreme values proposed by Gumbel (see Gumbel, 1962). It states that the maximum of many independent and identically distributed (*i.i.d.*) random variables approximately follows an *Extreme Value* distribution:

$$\varepsilon_{in} \sim EV(\eta, \mu), \tag{34}$$

where $\mu > 0$. The probability density function (pdf) and the cumulative distribution function (cdf) are reported in Appendix A.1. Note that the cdf has a closed form, contrarily to the cdf of a normal distribution. This is actually one additional motivation to prefer the above assumption to the more classical normality assumption.

We further assume that the error terms, ε_{1n} and ε_{2n}, are independent, and identically distributed across alternatives and individuals. This is why the parameters η and μ of the distribution do not carry any index i or n.

An important property of this distribution is that the difference $\varepsilon_n = \varepsilon_{2n} - \varepsilon_{1n}$ follows a logistic distribution (see property 5 in Appendix B). Therefore, the model (33) can be derived as

$$P_n(1) = \text{Prob}(\varepsilon_n \leq V_{1n} - V_{2n}) \tag{35}$$
$$= F_{\varepsilon_n}(V_{1n} - V_{2n}) \tag{36}$$
$$= \frac{1}{1 + e^{-\mu(V_{1n} - V_{2n})}} \tag{37}$$
$$= \frac{e^{\mu V_{1n}}}{e^{\mu V_{1n}} + e^{\mu V_{2n}}}, \tag{38}$$

[3]For the binary logit model, it would actually be sufficient to postulate an assumption about the distribution of the difference ε_n. But, anticipating the case of multiple alternatives, we treat each error term separately.

where F_{ε_n} is defined by (98).

Exercise: perform a similar derivation replacing (34) by the assumption that the random variables ε_{1n} and ε_{2n} are *i.i.d.* normal.

The binary logit model is illustrated in Figure 1 for two different values of μ. It can be seen that it is a sigmoid function.

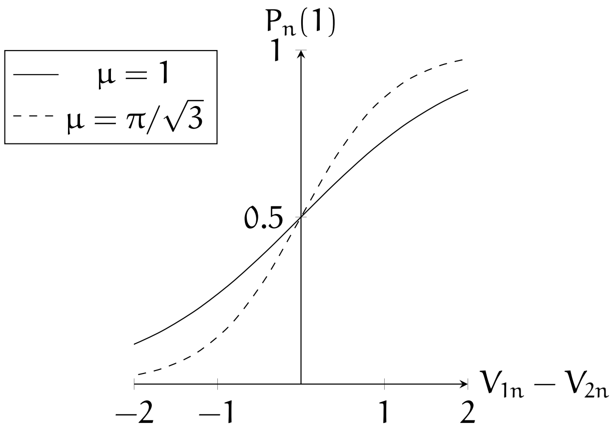


Figure 1: The binary logit model

Consider again the transportation mode choice example. Assume that the utility function associated by individual n with the "car" alternative (labeled 1) is

$$\begin{aligned} V_{1n} = 3.04 &- 0.0527 \cdot \text{cost}_{1n} - 2.66 \cdot \text{travelTime}_{1n} \cdot \text{work}_n \\ &- 2.22 \cdot \text{travelTime}_{1n} \cdot (1 - \text{work}_n) - 0.850 \cdot \text{male}_n \\ &+ 0.383 \cdot \text{mainEarner}_n - 0.624 \cdot \text{fixedArrivalTime}_n, \end{aligned} \tag{39}$$

where the attributes of the alternative are

- cost_{1n}: the travel cost of driving for individual n,
- travelTime_{1n}: the travel time (in hours) of driving for individual n,

and the socio-economic characteristics of individual n are

- work_n: 1 if the trip purpose is work, 0 otherwise,

- male_n: 1 if n is a male, 0 if n is a female,
- mainEarner_n: 1 if n is the main earner in the family, 0 otherwise,
- $\text{fixedArrivalTime}_n$: 1 if n has a fixed arrival time at the trip destination, 0 otherwise.

The utility function associated by individual n with the "train" alternative (labeled 2) is

$$V_{2n} = -0.0527 \cdot \text{cost}_{2n} - 0.576 \cdot \text{travelTime}_{2n} + 0.961 \cdot \text{firstClass}_n, \quad (40)$$

where the attributes of the alternative are

- cost_{2n}: the travel cost of the train for individual n,
- travelTime_{2n}: the travel time (in hours) by train for individual n,

and the socio-economic characteristic of individual n is firstClass_n, that is 1 if individual n prefers traveling first class, 0 otherwise. Note that these utility functions are not invented, and have been specified based on real data collected in the Netherlands.

Consider now three specific individuals who have to make a choice in a context characterized by the values of the variables reported in Table 1. The calculation of the choice model for each of them is reported in Table 2. Each row of this table corresponds to a variable in the model. The first column reports the name of the variable, and the second its coefficient in the utility functions (39) and (40). The other columns contain the information extracted from Table 1, positioned at the appropriate place. Therefore, the calculation of the utility functions V_{1n} and V_{2n} involves an inner product between the column corresponding to the coefficients, and the column corresponding to the variables. The results of these calculation is reported in the row label V_{in}. The last row reports the choice probabilities, as calculated by the binary logit model (38) with $\mu = 1$. It is seen that individual 1 as 94.7% probability to select the car to travel, while this probability is 7.58% for individual 2 and 77.5% for individual 3.

This example illustrates the flexibility of the model, in terms of the many variables that it can involve. Also, it should appear more clearly from this example why the model is called *disaggregate*. The choice of every individual is explicitly modeled, and the context of each different person is taken into account.

	Individual 1	Individual 2	Individual 3
Train cost	40.00	7.80	40.00
Car cost	5.00	8.33	3.20
Train travel time	2.50	1.75	2.67
Car travel time	1.17	2.00	2.55
Gender	M	F	F
Trip purpose	Not work	Work	Not work
Class	Second	First	Second
Main earner	No	Yes	Yes
Arrival time	Variable	Fixed	Variable

Table 1: Three hypothetical individuals for the choice of transportation mode

Exercise: If the deterministic parts of the utility, V_{1n} and V_{2n} are set, we consider two limiting cases for the scale parameter μ: $\mu \to \infty$ and $\mu \to 0$. What are the choice probabilities provided by the binary logit model for each of these special cases?

3.2 Multiple alternatives

The generalization of the binary logit model to more than two alternatives is relatively straightforward. As discussed above, the set of alternatives that is available to each individual is characterized by a choice set $\mathcal{C}_n$, or by availability binary variables a_{in}. Therefore, without loss of generality, we assume in the following that the choice set $\mathcal{C}$ is the same for all individuals.

In order to derive the model, we write the choice model 2.2 as a binary logit model, comparing alternative i with the best among all other alternatives:

$$P_n(i) = \text{Prob}(U_{in} \geq U_{jn}, \ \forall j \in \mathcal{C}) \tag{41}$$

$$= \text{Prob}(U_{in} \geq \max_{j \in \mathcal{C}, j \neq i} U_{jn}) \tag{42}$$

$$= \text{Prob}(V_{in} + \varepsilon_{in} \geq U^*_{-in}), \tag{43}$$

where

$$U^*_{-in} = \max_{j \in \mathcal{C}, j \neq i} U_{jn} = \max_{j \in \mathcal{C}, j \neq i} (V_{jn} + \varepsilon_{jn}) \tag{44}$$

		Individual 1		Individual 2		Individual 3	
Variables	Coef.	Car	Train	Car	Train	Car	Train
Car dummy	3.04	1	0	1	0	1	0
Cost	-0.0527	5.00	40.00	8.33	7.80	3.20	40.00
Tr. time by car (work)	-2.66	0	0	2	0	0	0
Tr. time by car (not work)	-2.22	1.17	0	0	0	2.55	0
Tr. time by train	-0.576	0	2.50	0	1.75	0	2.67
First class dummy	0.961	0	0	0	1	0	0
Male dummy	-0.850	1	0	0	0	0	0
Main earner dummy	0.383	0	0	1	0	1	0
Fixed arrival time dummy	-0.624	0	0	1	0	0	0
V_{in}		-0.6709	-3.5480	-2.9600	-0.4581	-2.4066	-3.6459
$P_n(i)$		0.947	0.0533	0.0757	0.924	0.775	0.225

Table 2: Application of the binary logit model

is the utility of the best alternative in the set $\mathcal{C} \setminus \{i\}$. As for the binary case, we assume that all the error terms ε_{in} are (1) independently, (2) identically, and (3) extreme value (EV) distributed with a location parameter η and a scale parameter $\mu > 0$. It can be assumed without loss of generality that $\eta = 0$. Indeed, from property 4 of the extreme value distribution described in Appendix B, if $\varepsilon_{in} \sim \text{EV}(\eta, \mu)$, then

$$U_{in} = V_{in} + \varepsilon_{in} = V_{in} + \eta + \varepsilon'_{in}, \tag{45}$$

where $\varepsilon'_{in} \sim \text{EV}(0, \mu)$. In any case, $U_{in} \sim \text{EV}(V_{in} + \eta, \mu)$. Consequently, we assume that η is integrated in V_{in} and

$$\varepsilon_{in} \overset{\text{iid}}{\sim} \text{EV}(0, \mu). \tag{46}$$

From property 6, we have that U^*_{-in} is EV distributed:

$$U^*_{-in} \sim \text{EV}\left(\frac{1}{\mu} \ln \sum_{j \in \mathcal{C}, j \neq i} e^{\mu V_{jn}}, \mu\right). \tag{47}$$

Using property 4 again, we can write $U^*_{-in} = V^*_{-in} + \varepsilon^*_n$, where

$$V^*_{-in} = \frac{1}{\mu} \ln \sum_{j \in \mathcal{C}, j \neq i} e^{\mu V_{jn}} \tag{48}$$

and ε^*_n is EV distributed with parameters $(0, \mu)$.

Therefore, (43) is a binary logit model, and

$$P_n(i) = \text{Prob}(V_{in} + \varepsilon_{in} \geq V^*_{-in} + \varepsilon^*_n) = \frac{e^{\mu V_{in}}}{e^{\mu V_{in}} + e^{\mu V^*_{-in}}}. \tag{49}$$

Using (48), we obtain the logit model

$$P_n(i) = \frac{e^{\mu V_{in}}}{\sum_{j \in \mathcal{C}} e^{\mu V_{jn}}}. \tag{50}$$

As noted in Section 2.2, if the availability of the alternatives are characterized by the indicators a_{in}, the model is written

$$P_n(i) = \frac{a_{in} e^{\mu V_{in}}}{\sum_{j \in \mathcal{C}} a_{jn} e^{\mu V_{jn}}}, \tag{51}$$

where a_{in} is one if $i \in \mathcal{C}_n$ and zero otherwise. This is a straightforward generalization of the binary logit model (38), where the sum at the denominator involves all alternatives available to the individual.

Exercise: In the example presented in Section 3.1, the commuter had the choice to use either the car or the train to commute to work. Assuming that the deterministic part V_{in} of the utility function is exactly the same for the two alternatives, what are the choice probabilities provided by the binary logit model? Now suppose that, for some reasons, some trains are painted in blue, and some trains are painted in red. The commuter now considers car, red train, and blue train as three of the available alternatives. Assuming again that the systematic part of the utilities are identical, what are the choice probabilities for this commuter? Do the ratios of choice probabilities change? Does it seem realistic?

4 Parameters estimation

As mentioned in the illustrative example in Section 3.1, the coefficients of the variables in the utility functions (39) and (40) have been estimated from data. We denote by $x_{in} = h(z_{in}, s_n)$ the vector of all explanatory variables involved in the model, as a function of both attributes and socio-economic characteristics, so that we can write

$$V_{in}(x_{in}; \beta) = \sum_{k=1}^{K} \beta_k x_{ink}. \tag{52}$$

In our example, we can write (39) as

$$V_{in}(x_{in}; \beta) = \sum_{i=1}^{7} \beta_k x_{ink}, \tag{53}$$

where $x_{in1} = 1$, $x_{in2} = \text{cost}_{in}$, $x_{in3} = \text{travelTime}_{in} \cdot \text{work}_n$, $x_{in4} = \text{travelTime}_{in} \cdot (1 - \text{work}_n)$, $x_{in5} = \text{male}_n$, $x_{in6} = \text{mainEarner}_n$ and $x_{in7} = \text{fixedArrivalTime}_n$. Also, $\beta_1 = 3.04$, $\beta_2 = -0.0527$, $\beta_3 = -2.66$, $\beta_4 = -2.22$, $\beta_5 = -0.850$, $\beta_6 = 0.383$ and $\beta_7 = -0.624$.

In this section, we show how maximum likelihood estimation (MLE) is used to solve the problem of estimating the values of the parameters

$\beta_1, ..., \beta_K$ from a sample of observations, assumed to be drawn at random from the population.

For each individual in the sample, we observe her choice, characterized by a binary choice variable y_{in} defined as

$$y_{in} = \begin{cases} 1 & \text{if individual } n \text{ chooses alternative } i, \\ 0 & \text{otherwise,} \end{cases} \tag{54}$$

as well as a vector containing all values of the relevant variables x_{ink}. The idea of maximum likelihood estimation is to select the β such that the probability that the model correctly predicts all observed choices (called the *likelihood*) is the highest possible. It therefore amounts to solve the following optimization problem:

$$\max_{\beta \in \mathbb{R}^K} \mathcal{L}^*(\beta) = \prod_{n=1}^{N} \prod_{i \in \mathcal{C}_n} P_n(i; \beta)^{y_{in}}, \tag{55}$$

where for logit

$$P_n(i; \beta) = \frac{e^{V_{in}(x_{in};\beta)}}{\sum_{j \in \mathcal{C}_n} e^{V_{jn}(x_{jn};\beta)}}. \tag{56}$$

The objective function as expressed in (55) is not easy to manipulate. As a product of N probabilities, it make take very small value, especially when working with large samples. It is convenient to maximize instead the logarithm of $\mathcal{L}^*$, called the *log likelihood* and denoted $\mathcal{L}$:

$$\max_{\beta \in \mathbb{R}^K} \mathcal{L}(\beta) = \log(\prod_{n=1}^{N} \prod_{i \in \mathcal{C}_n} P_n(i)^{y_{in}}) = \sum_{n=1}^{N} \sum_{i \in \mathcal{C}_n} y_{in} \ln(P_n(i)). \tag{57}$$

Substituting (56) into (57), and denoting $V_{in}(x_{in}; \beta)$ simply by V_{in}, we seek a maximum to

$$\max_{\beta \in \mathbb{R}^K} \mathcal{L}(\beta) = \sum_{n=1}^{N} \sum_{i \in \mathcal{C}_n} y_{in} \ln(\frac{e^{V_{in}}}{\sum_{j \in \mathcal{C}_n} e^{V_{jn}}}) \tag{58}$$

$$= \sum_{n=1}^{N} \sum_{i \in \mathcal{C}_n} y_{in} \left(\ln(e^{V_{in}}) - \ln(\sum_{j \in \mathcal{C}_n} e^{V_{jn}}) \right) \tag{59}$$

$$= \sum_{n=1}^{N} \sum_{i \in \mathcal{C}_n} y_{in} \left(V_{in} - \ln(\sum_{j \in \mathcal{C}_n} e^{V_{jn}}) \right) \tag{60}$$

$$= \sum_{n=1}^{N} \sum_{i \in \mathcal{C}_n} y_{in} V_{in} - \sum_{n=1}^{N} \sum_{i \in \mathcal{C}_n} \left(y_{in} \ln(\sum_{j \in \mathcal{C}_n} e^{V_{jn}}) \right) \tag{61}$$

$$= \sum_{n=1}^{N} \sum_{i \in \mathcal{C}_n} y_{in} V_{in} - \sum_{n=1}^{N} \left(\ln(\sum_{j \in \mathcal{C}_n} e^{V_{jn}}) \sum_{i \in \mathcal{C}_n} y_{in} \right) \tag{62}$$

$$= \sum_{n=1}^{N} \left(\sum_{i \in \mathcal{C}_n} y_{in} V_{in} - \ln(\sum_{i \in \mathcal{C}_n} e^{V_{in}}) \right). \tag{63}$$

This is a nonlinear continuous optimization problem. If a solution exists, it must satisfy the necessary first order conditions, that is

$$\frac{\partial \mathcal{L}}{\partial \beta_k} = 0, \quad \text{for} \quad k = 1, ..., K. \tag{64}$$

Applying the chain rule, we can calculate the vector of first derivatives of the log likelihood function with respect to the unknown parameters $k = 1, ..., K$ as follows:

$$\frac{\partial \mathcal{L}}{\partial \beta_k} = \sum_{n=1}^{N} \left(\sum_{i \in \mathcal{C}_n} y_{in} \frac{\partial V_{in}}{\partial \beta_k} - \frac{1}{\sum_{i \in \mathcal{C}_n} e^{V_{in}}} (\sum_{i \in \mathcal{C}_n} e^{V_{in}} \frac{\partial V_{in}}{\partial \beta_k}) \right) \tag{65}$$

$$= \sum_{n=1}^{N} \left(\sum_{i \in \mathcal{C}_n} y_{in} \frac{\partial V_{in}}{\partial \beta_k} - \frac{\sum_{i \in \mathcal{C}_n} e^{V_{in}} \frac{\partial V_{in}}{\partial \beta_k})}{\sum_{i \in \mathcal{C}_n} e^{V_{in}}} \right) \tag{66}$$

$$= \sum_{n=1}^{N} \left(\sum_{i \in \mathcal{C}_n} y_{in} \frac{\partial V_{in}}{\partial \beta_k} - \sum_{i \in \mathcal{C}_n} P_n(i) \frac{\partial V_{in}}{\partial \beta_k} \right) \tag{67}$$

$$= \sum_{n=1}^{N} \sum_{i \in \mathcal{C}_n} (y_{in} - P_n(i)) \frac{\partial V_{in}}{\partial \beta_k}). \tag{68}$$

Substituting (68) into (64), the necessary first order conditions become:

$$\sum_{n=1}^{N} \sum_{i \in \mathcal{C}_n} (y_{in} - P_n(i)) \frac{\partial V_{in}}{\partial \beta_k} = 0, \quad \text{for} \quad k = 1, \ldots, K. \tag{69}$$

If the model has a linear-in-parameters specification as in (56), we have that

$$\frac{\partial V_{in}}{\partial \beta_k} = x_{ink}, \tag{70}$$

and (69) is equivalent to

$$\sum_{n=1}^{N} \sum_{i \in \mathcal{C}_n} (y_{in} - P_n(i)) x_{ink} = 0, \quad \text{for} \quad k = 1, \ldots, K. \tag{71}$$

The first order optimality conditions of the optimization problem form therefore a system of K nonlinear equations in K unknowns $\beta_1, \beta_2, \ldots, \beta_K$, that can be solved using iterative algorithms such as Newton's method and its variants (Dennis Jr and Schnabel, 1996, Kelley, 2003).

Exercise: Show that the second derivatives of $\mathcal{L}$ are given by

$$\frac{\partial^2 \mathcal{L}}{\partial \beta_k \partial \beta_\ell} = -\sum_{n=1}^{N} \sum_{i \in \mathcal{C}_n} P_n(i) \left(x_{ink} - \sum_{j \in \mathcal{C}_n} x_{jnk} P_n(j) \right) \left(x_{in\ell} - \sum_{j \in \mathcal{C}_n} x_{jn\ell} P_n(j) \right). \tag{72}$$

Under relatively weak conditions, McFadden, 1974 showed that the log likelihood function for logit models with linear-in-parameters utility functions is globally concave, so the first order optimality conditions are sufficient. From a statistical point of view, the maximum likelihood estimates of $\beta_1, \ldots, \beta_K$ are consistent (that is, asymptotically unbiased), asymptotically efficient, and asymptotically normal.

Exercise : Consider again the mode choice described in Section 3.1, but consider a simple specification with only two parameters: the car dummy, and the generic coefficient of travel cost, as described in Table 2. Depict the value of the log likelihood as a function of the two parameters.

5 Model applications

A model predicting the choice of a given individual is often of little use in practice. In this section, we discuss some concrete uses of the model.

We assume that we have at our disposal a disaggregate choice model that has been estimated from data, tested and validated. We denote it

$$P_n(i|x_n), \tag{73}$$

where $i \in \mathcal{C}$, and x_n is a vector containing all explanatory variables of the model. We also assume that we are dealing with a population of N individuals, for which we have access to the value of x_n.

5.1 Demand forecasting

The demand for alternative i is defined as the total number of individuals choosing i. It is modeled as

$$D_i(x_1, \ldots, x_N) = \sum_{n=1}^{N} P_n(i|x_n). \tag{74}$$

Clearly, except for specific cases where N is small, the above calculation is intractable. Therefore, in practice, a sample of size N_s of the population is selected, and the values of x_n are observed for the sample. The demand for alternative i is therefore estimated as

$$D_i(x_1, \ldots, x_{N_s}) = \frac{N}{N_s} \sum_{n=1}^{N_s} P_n(i|x_n) = N W_i(x_1, \ldots, x_{N_s}), \tag{75}$$

where

$$W_i(x_1, \ldots, x_{N_s}) = \frac{1}{N_s} \sum_{n=1}^{N_s} P_n(i|x_n) \tag{76}$$

is usually called the market share of alternative i. Classical weighting methods to guarantee that the sample is representative of the population must be applied.

If we consider the mode choice example introduced in Section 3.1, and if we assume for the sake of the example that the population is composed of the three individuals reported in Table 1, the average number of individuals using the car to travel is 0.947+0.0758+0.775=1.8, with a market share of 1.8/3=60%.

These disaggregate demand functions can be embedded in a great deal of relevant operations research methods. In the following, we focus on a couple of specific applications.

5.2 Revenue maximization

We consider now the case of a market of competing products, each of them associated with a price. The objective is to analyze the impact of the change of the price of a product on the revenues it generates. Therefore, we write the choice model (73) as

$$P_n(i|p_n, x_n), \tag{77}$$

where p_n is the vector of prices for individual n, and x_n is the set of all other variables involved in the model. Therefore, the expected revenue generated by alternative i is

$$R_i(p_1, \ldots, p_{N_s}, x_1, \ldots, x_{N_s}) = \frac{N}{N_s} \sum_{n=1}^{N_s} p_{in} P_n(i|p_n, x_n). \tag{78}$$

Therefore, the actor in charge of pricing the product i must solve the following optimization problem:

$$\max_{p_i \in \mathbb{R}^+} R_i(p_1, \ldots, p_{N_s}, x_1, \ldots, x_{N_s}). \tag{79}$$

Note that it is implicitly assumed here that the costs of production are null and the prices of the competitors are fixed, which may not be the case in a competitive market.

It is interesting to investigate the shape of this revenue function on a simple example. Assume that we are considering a homogeneous population of 1000 individuals. There are two products 1 and 2, and the choice model is a logit model involving only the price variables:

$$V_{1n} = -0.65 p_1 - 0.5 \tag{80}$$

$$V_{2n} = -0.65 p_2, \tag{81}$$

so that

$$P_n(1; p_1, p_2) = \frac{e^{V_{1n}}}{e^{V_{1n}} + e^{V_{2n}}} = \frac{e^{-0.65 p_1 - 0.5}}{e^{-0.65 p_1 - 0.5} + e^{-0.65 p_2}}. \tag{82}$$

If the price of both products is 2.0, then $V_{1n} = -1.8$ and $V_{2n} = -1.3$, so that $P_n(1) = 37.8\%$ for all n. It means that the demand for product 1 is 378 individuals and the revenue is 755. If the price of product 1 varies from 0 to 10, say, the market share decreases monotonically from 69% to 0.33%, as illustrated by Figure 2(a). A zero price obviously generates no revenue. When the price increases, the revenue increases up to a point where the revenue starts decreasing. This value ($p_1 = 2.30$, here) corresponds to the maximum possible revenue. Note that the revenue function is unimodal, so that the local maximum is also a global maximum.

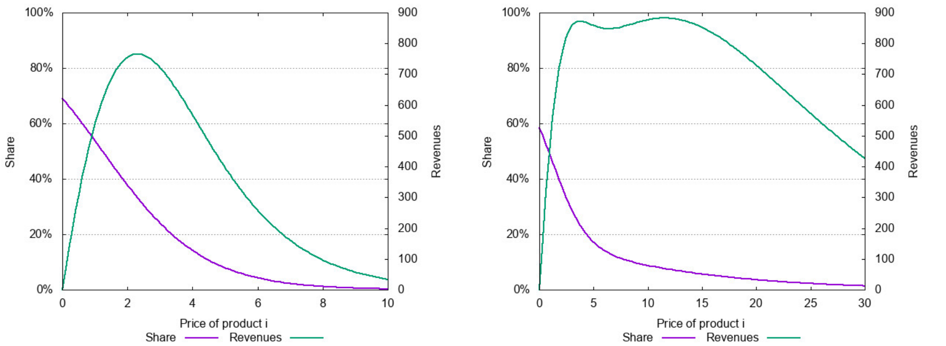
Consider now a heterogeneous population, with two segments. The first segment, composed of 600 individuals, say, has the same utility functions as above. The second segment, composed of 400 individuals, is less price sensitive. The utility functions are

$$V_{1n_2} = -0.1p_1 - 0.5, \tag{83}$$

$$V_{2n_2} = -0.1p_2. \tag{84}$$

If the price of product 1 is 0 and the price of product 2 is 2, then $V_{1n_1} = -0.5$ and $V_{2n_1} = -1.3$, so that $P_{n_1}(1) = 69\%$ for all n_1 in the first segment. Also, $V_{1n_2} = -0.5$ and $V_{2n_2} = -0.2$, so that $P_{n_2}(1) = 42.6\%$ for all n_2 in the second segment. Therefore, the total demand for product 1 at zero price is 584 individuals. The demand is strictly decreasing as the price monotonically increases, as expected, although with a different shape than for the homogeneous population (see Figure 2(b)). The revenue curve is particularly interesting. Contrarily to the homogeneous case, it is not unimodal anymore. It exhibits two local optima: one for $p_1 = 3.74$, with a revenue of 872, and one for $p_1 = 11.6$, with a revenue of 883. Note that the revenue function does not always exhibit multiple local optima.

This simplified example illustrates how accounting for the behavioral heterogeneity of the population complexifies the optimization of the revenues. From an algorithmic viewpoint, it stresses the need to use global optimization algorithms. From the application viewpoint, it justifies the calculation of different prices for different market segments, when it is possible.



(a) Homogeneous population

(b) Heterogeneous population (two segments)

Figure 2: Market shares and revenues

5.3 Satisfaction

Maximizing revenues is not the only relevant optimization problem involving demand models. In particular, local authorities invest in public services, where an important indicator is the satisfaction of the population. Disaggregate demand models provide a quantitative indicator of the satisfaction: the Expected Maximum Utility (EMU):

$$\bar{U}_n = E[\max_{i \in \mathcal{C}} a_{in} U_{in}]. \tag{85}$$

The exact formulation depends on the distribution of the U_{in}. For example, the expected maximum utility for the logit model (51) is

$$\bar{U}_n = \frac{1}{\mu} \ln \sum_{j \in \mathcal{C}} a_{jn} e^{\mu V_{jn}}. \tag{86}$$

Similarly to the revenue function, the functional form of the satisfaction function can be complex. Consider a simple example where two services 1 and 2 are characterized by the price, and the quality of the service. In practice, high levels of quality are associated with high prices as well. For our example, assume that it has been estimated that a level of quality $q(p) = 1 + \ln(p/10)$ can be achieved for a price p.

As above, we assume two segments in the population, with the following utility functions:

$$V_{1n_1} = -0.65p_1 + 1.5q_1 - 0.5 \tag{87}$$

$$V_{2n_1} = -0.65p_2 + 1.5q_2, \tag{88}$$

for the first segment, and

$$V_{1n_2} = -0.1p_1 + 1.5q_1 - 0.5 \quad (89)$$
$$V_{2n_2} = -0.1p_2 + 1.5q_2, \quad (90)$$

for the second one. It means that the population is homogeneous in terms of perception of quality, but heterogeneous in terms of sensitivity to price. Assuming that the first segment involves 60% of the population, the average expected maximum utility is

$$0.6\bar{U}_{n_1} + 0.4\bar{U}_{n_2}, \quad (91)$$

where $\bar{U}_{n_1}$ and $\bar{U}_{n_2}$ are the levels of satisfaction for individuals in the first and the second segment, respectively, as provided by (86). Assuming that the price of service 2 is 2, the functional form of the level of satisfaction as a function of the price of service 1 is illustrated in Figure 3. When the price is low, the quality is also low, and so is the satisfaction. It is observed that the function exhibits two local maxima: $p_1 = 4.74$, and $p_1 = 14.4$.

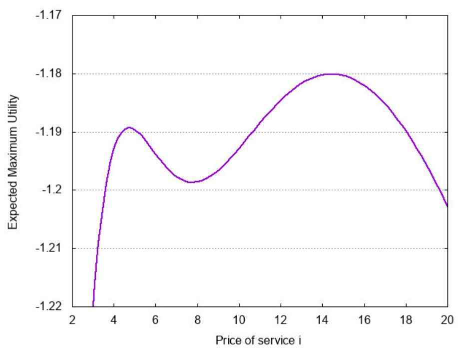


Figure 3: Expected Maximum Utility

6 Conclusion

Disaggregate demand models are powerful instruments that allow to capture the complexity of human behavior in great details. They provide key performance indicators particularly relevant in operations research.

A major drawback of this modeling framework is that these indicators are nonlinear and non-convex in the variables of interest, as it has been illustrated on some simple examples above. It significantly complicates their use in exact methods of optimization. However, as explained in Section 2.1, the theoretical foundations are based on optimization principles. As shown in Bierlaire and Azadeh, 2016 and Pacheco et al., 2016, this can be exploited to obtain mathematical formulations of the revenue maximization problems, and to solve the problem exactly.

The field of disaggregate demand models and discrete choice models is particularly active, and a vast literature is available. In this tutorial, we have touched upon the main principles, emphasizing the fundamental assumptions, and the close connections with optimization principles.

Appendices

We provide here some basic material relevant in this tutorial, for easy reference.

A Relevant statistical distribution

A continuous univariate random variable X is characterized by a *probability density function* (pdf), or by a *cumulative distribution function* (cdf). It is said that X has density f_X if for any real numbers $a \leq b$,

$$\text{Prob}(a \leq X \leq b) = \int_a^b f_X(t)dt. \tag{92}$$

The cdf F_X is defined as

$$F_X(x) = \text{Prob}(X \leq x). \tag{93}$$

Combining (92) and (93), we have

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \tag{94}$$

and f_X is the derivative of F_X. We provide here the pdf and the cdf of the distributions mentioned in this tutorial.

A.1 Extreme value distribution

The probability density function (pdf) of this distribution is given by

$$f(t) = \mu e^{-\mu(t-\eta)} e^{-e^{-\mu(t-\eta)}} \tag{95}$$

and the cumulative distribution function (cdf) by

$$F(x) = \int_{-\infty}^{x} f(t)dt = e^{-e^{-\mu(x-\eta)}}, \tag{96}$$

where $\eta \in \mathbb{R}$ is the location parameter, and $\mu \in \mathbb{R}$, $\mu > 0$ is the scale parameter.

A.2 Logistic distribution

The probability density function (pdf) of this distribution is given by

$$f(t) = \frac{\mu e^{-\mu(t-\eta)}}{(1 + e^{-\mu(t-\eta)})^2}, \tag{97}$$

and the cumulative distribution function (cdf) by

$$F(c) = \frac{1}{1 + e^{-\mu(c-\eta)}} \tag{98}$$

where $\eta \in \mathbb{R}$ is the location parameter, and $\mu \in \mathbb{R}$, $\mu > 0$ is the scale parameter.

B Properties of the extreme value distribution

The extreme value distribution EV(η,μ) has the following properties:

1. The mode is η.
2. The mean is $\eta + \frac{\gamma}{\mu}$,

   where
   $$\gamma = -\int_0^{+\infty} e^{-x} \ln x dx \simeq 0.5772 \tag{99}$$
   is Euler's constant.
3. The variance is $\frac{\pi^2}{6\mu^2}$.
4. If $\varepsilon \sim EV(\eta, \mu)$, then
   $$a\varepsilon + b \sim EV(a\eta + b, \frac{\mu}{a}),$$
   where $a, b \in \mathbb{R}, a > 0$.
5. If $\varepsilon_a \sim EV(\eta_a, \mu)$ and $\varepsilon_b \sim EV(\eta_b, \mu)$ are independent with the same scale parameter μ, then
   $$\varepsilon = \varepsilon_a - \varepsilon_b \sim \text{Logistic}(\eta_a - \eta_b, \mu).$$

6. If $\varepsilon_i \sim EV(\eta_i, \mu)$, for $i = 1, \ldots, J$, and ε_i are independent with the same scale parameter μ, then

$$\varepsilon = \max_{i=1,\ldots,J} \varepsilon_i \sim EV(\eta, \mu),$$

where

$$\eta = \frac{1}{\mu} \ln \sum_{i=1}^{J} e^{\mu \eta_i}. \tag{100}$$

References

Andersson, S.-E. (1998). Passenger choice analysis for seat capacity control: a pilot project in Scandinavian Airlines, *International Transactions in Operational Research* **5**(6): 471 – 486.

Aros-Vera, F., Marianov, V. and Mitchell, J. E. (2013). p-Hub approach for the optimal park-and-ride facility location problem, *European Journal of Operational Research* **226**(2): 277–285.

Ben-Akiva, M. and Bierlaire, M. (2003). Discrete choice models with applications to departure time and route choice, *Handbook of transportation science*, Springer, pp. 7–37.

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

Benati, S. (1999). The maximum capture problem with heterogeneous customers, *Computers & Operations Research* **26**(14): 1351 – 1367.

Benati, S. and Hansen, P. (2002). The maximum capture problem with random utilities: Problem formulation and algorithms, *European Journal of Operational Research* **143**: 518–530.

Bierlaire, M. and Azadeh, S. S. (2016). Demand-based discrete optimization, *Technical Report 160209*, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne.

Bodea, T., Ferguson, M. and Garrow, L. (2009). Data set–choice-based revenue management: Data from a major hotel chain, *Manufacturing & Service Operations Management* **11**(2): 356–361.

Bront, J. J. M., Méndez-Díaz, I. and Vulcano, G. (2009). A column generation algorithm for choice-based network revenue management, *Operations Research* **57**(3): 769–784.

Castillo-López, I. and López-Ospina, H. A. (2015). School location and capacity modification considering the existence of externalities in students school choice, *Computers & Industrial Engineering* **80**: 284–294.

Debreu, G. (1954). Representation of a preference ordering by a numerical function, *Decision Processes*, Wiley.

Dennis Jr, J. E. and Schnabel, R. B. (1996). *Numerical methods for unconstrained optimization and nonlinear equations*, Society for Industrial and Applied Mathematics (SIAM).

Gallego, G. and Phillips, R. (2004). Revenue management of flexible products, *Manufacturing & Service Operations Management* **6**(4): 321–337.

Garrow, L. A. (2016). *Discrete choice modelling and air travel demand: theory and applications*, Ashgate Publishing, Ltd.

Gilbert, F., Marcotte, P. and Savard, G. (2014a). Logit network pricing, *Computers & Operations Research* **41**: 291 – 298.

Gilbert, F., Marcotte, P. and Savard, G. (2014b). Mixed-logit network pricing, *Computational Optimization and Applications* **57**(1): 105–127.

Goldberger, A. S. (1968). The interpretation and estimation of Cobb-Douglas functions, *Econometrica* **36**(3/4): 464–472.

Gumbel, E. J. (1962). Statistical theory of extreme values, *Revue Belge de Statistique et de Recherche Opérationnelle* **3**(2): 3–11.

Haase, K. (2009). *Discrete location planning. Technical Report WP-09-07*, Institute of Transport and Logistics Studies, University of Sydney.

Haase, K. and Müller, S. (2013). Management of school locations allowing for free school choice, *Omega* **41**(5): 847–855.

Haase, K. and Müller, S. (2015). Insights into clients' choice in preventive health care facility location planning, *OR Spectrum* **37**(1): 273–291.

Kant, P. (2008). *Route choice modelling in dynamic traffic assignment*, Master's thesis, University of Twente.

Kelley, C. T. (2003). *Solving nonlinear equations with Newton's method*, Society for Industrial and Applied Mathematics (SIAM).

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior, *Frontiers in Econometrics*, P. Zarembka, pp. 105–142.

Müller, S. and Haase, K. (2014). Customer segmentation in retail facility location planning, *Business Research* **7**(2): 235–261.

Müller, S., Haase, K. and Kless, S. (2009). A multiperiod school location planning approach with free school choice, *Environment and Planning A* **41**(12): 2929–2945.

Müller, S., Haase, K. and Seidel, F. (2012). Exposing unobserved spatial similarity: evidence from German school choice data, *Geographical Analysis* **44**(1): 65–86.

Nicholson, W. and Snyder, C. M. (2007). *Microeconomic Theory: Basic Principles and Extensions*, South Western/Thomson.

Pacheco, M., Azadeh, S. S. and Bierlaire, M. (2016). A new mathematical representation of demand using choice-based optimization methods, *Proceedings of the 16th Swiss Transport Research Conference*, Ascona, Switzerland.

Pel, A., Bliemer, M. and Hoogendoorn, S. (2009). Hybrid route choice modeling in dynamic traffic assignment, *Transportation Research Record: Journal of the Transportation Research Board* (2091): 100–107.

Pindyck, R. S. and Rubinfeld, D. L. (2008). *Microeconomics*, 7th edn, Prentice Hall.

Qian, Z. (2011). *On Dynamic Traffic Assignment in Corridor Networks under Heterogeneous Travelers and Modes*, PhD thesis, University of California, Davis.

Qian, Z. S. and Zhang, M. H. (2013). A hybrid route choice model for dynamic traffic assignment, *Networks and Spatial Economics* **13**(2): 183–203.

Ratliff, R. M., Rao, B. V., Narayan, C. P. and Yellepeddi, K. (2008). A multi-flight recapture heuristic for estimating unconstrained demand from airline bookings, *Journal of Revenue and Pricing Management* **7**(2): 153–171.

Schön, C. (2007). Market-oriented airline service design, *Operations Research Proceedings 2006*, Springer, pp. 361–366.

Talluri, K. and Van Ryzin, G. (2004). Revenue management under a general discrete choice model of consumer behavior, *Management Science* **50**(1): 15–33.

Train, K. (2003). *Discrete choice methods with simulation*, Cambridge university press.

Varian, H. R. and Repcheck, J. (2010). *Intermediate microeconomics: a modern approach*, Vol. 6, WW Norton & Company New York.

Vulcano, G., Van Ryzin, G. and Chaar, W. (2010). On practice-choice-based revenue management: An empirical study of estimation and optimization, *Manufacturing & Service Operations Management* **12**(3): 371–392.

Vulcano, G., Van Ryzin, G. and Ratliff, R. (2012). Estimating primary demand for substitutable products from sales transaction data, *Operations Research* **60**(2): 313–334.

Zhang, Y., Berman, O. and Verter, V. (2012). The impact of client choice on preventive healthcare facility network design., *OR Spectrum* **34**: 349–370.