



## UvA-DARE (Digital Academic Repository)

### Socio-dynamic discrete choice: Theory and application

Dugundji, E.R.

**Publication date**  
2013

[Link to publication](#)

#### **Citation for published version (APA):**

Dugundji, E. R. (2013). *Socio-dynamic discrete choice: Theory and application*. [Thesis, fully internal, Universiteit van Amsterdam].

#### **General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

#### **Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

## BACKGROUND

This chapter reviews the necessary notation and convention for the classic multinomial logit model and the nested logit model which rest at the heart of this thesis, and draws a few mathematical parallels in statistical physics.

## 2.1 MULTINOMIAL LOGIT MODEL

Discrete choice theory allows prediction based on computed individual choice probabilities for heterogeneous agents' evaluation of alternatives. Individual choice probabilities are aggregated for policy forecasting. In accordance with notation and convention in Ben-Akiva and Lerman (1985), the so-called multinomial logit model well known in econometrics is specified as follows. Assume a sample of  $N$  decision-making entities indexed  $(1, \dots, n, \dots, N)$  each faced with a choice among  $J_n$  alternatives indexed  $(1, \dots, j, \dots, J_n)$  in subset  $C_n$  of some universal choice set  $C$ . The choice alternatives are assumed to be mutually exclusive (a choice for one alternative excludes the simultaneous choice for another alternative, that is, an agent cannot choose two alternatives at the same moment in time) and collectively exhaustive within  $C_n$  (an agent must make a choice for one of the options in the agent's choice set). In general the composite choice set  $C_n$  will vary in size and content across agents: not all elemental alternatives in the universal choice set may be available to all agents. See Figure 2.1.

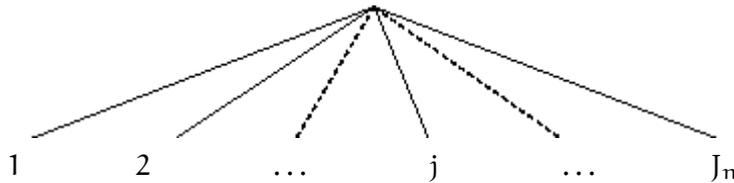


Figure 2.1: Multinomial choice structure for a decision-making entity  $n$

Let  $U_{in} = V_{in} + \varepsilon_{in}$  be the utility that a given decision-making entity  $n$  is presumed to associate with a particular alternative  $i$  in its choice set  $C_n$ , where  $V_{in}$  is the deterministic (to the modeler) or so-called "systematic" utility and  $\varepsilon_{in}$  is an error term. The error term  $\varepsilon_{in}$  represents unobserved heterogeneity. Such unobserved heterogeneity may arise due to unobserved attributes of the choice alternatives, unobserved characteristics of the decision-making entities or simply measurement errors in observed attributes and/or characteristics; also in the case where instrumental variables are used as a proxy

for variables which are not observable, the error term is relevant for capturing unobserved heterogeneity (Manski 1973). Under the assumption of independent and identically Gumbel distributed disturbances  $\varepsilon_{in}$ , the probability  $P_n(i|C_n)$  that the individual decision-making entity  $n$  chooses alternative  $i$  within the choice set  $C_n$  has a convenient closed form expression given by:

$$\begin{aligned} P_{in} &\equiv P_n(i|C_n) = \Pr(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}, \forall j \in C_n) \\ &= \Pr\left[V_{in} + \varepsilon_{in} \geq \max_{j \in C_n} (V_{jn} + \varepsilon_{jn})\right] \\ &= \frac{e^{\mu V_{in}}}{\sum_{j \in C_n} e^{\mu V_{jn}}} \end{aligned} \quad (2.1)$$

where  $\mu$  is a strictly positive scale parameter which we often normalize to 1. The assumption that the disturbances are Gumbel distributed can be defended as an approximation to the normal density.<sup>1</sup>

This so-called *multinomial logit* model has been a "workhorse" (Walker, Ben-Akiva and Bolduc, ) of discrete choice theory owing to its tractable analytic form, and can be derived in many different ways. Its original formulation in the field of individual choice behavior is due to mathematical psychologist Luce (1959), notably, resulting not from assumptions about the disturbances, but rather on choice probabilities themselves. His choice axiom is now better known as the Independence from Irrelevant Alternatives property (IIA). As explained in Ben-Akiva and Lerman, "the IIA property holds that for a specific individual the ratio of the choice probabilities of any two alternatives is entirely unaffected by the systematic utilities of any other alternatives."

$$\frac{P_{in}}{P_{kn}} = \frac{e^{\mu V_{in}} / \sum_{j \in C_n} e^{\mu V_{jn}}}{e^{\mu V_{kn}} / \sum_{j \in C_n} e^{\mu V_{jn}}} = \frac{e^{\mu V_{in}}}{e^{\mu V_{kn}}} \quad (2.2)$$

This axiom was seen to be desirable as it greatly simplifies experimental collection of choice data by allowing the probabilities for choice between multiple alternatives to be inferred from binary choice experiments. Marschak (1960) brought Luce's work and fellow mathematical psychologist Thurstone's earlier "Law of Comparative Judgement" (1927) yielding the binary probit model, to the field of economics via his *random utility maximization* model.

The multinomial logit model bears mathematical similarity to the so-called *canonical distribution* well known in statistical mechanics. Instead of considering a decision-making entity in a population faced with making a choice from some set of possible choices, where we ask

<sup>1</sup> For derivation and discussion with application in transportation, the reader is referred to Domencich and McFadden (1975) and Ben-Akiva and Lerman (1985).

under conditions of equilibrium what is the probability  $P_{i_n}$  of finding the entity has made a particular choice  $i$  with systematic utility  $V_{i_n}$ , let us consider a small macroscopic system  $A$  in thermal interaction with a large heat reservoir, or a microscopic system if particles can be clearly distinguished, such as an atom at a particular lattice site in a solid with the solid acting as a heat reservoir. Following Reif (1965, Chapter 6), the probability  $P_i$  of finding the small system, or lattice atom, in any one particular microstate  $i$  of energy  $E_i$  is given by:

$$P_i = \frac{e^{-\beta E_i}}{\sum_j e^{-\beta E_j}} \quad (2.3)$$

with temperature parameter  $\beta = 1/kT$  where  $k$  is the Boltzmann constant<sup>2</sup> and  $T$  is the temperature at thermal equilibrium. The exponential factor  $e^{-\beta E_i}$  in the numerator of (2.3) is termed the "Boltzmann factor." The sum over all microstates of these factors, appearing in the denominator of (2.3), is termed the "partition function" and is sometimes denoted by the letter  $Z$  due to the German name *Zustandsumme*. A statistical ensemble of systems all of which having probabilities governed by (2.3), is termed a "canonical ensemble." Note that the denominators in (2.1) and in (2.3) are proportionality constants, independent respectively of the particular alternative  $i$  or particular microstate  $i$ . These are obtained simply from the normalization conditions for the probabilities:

$$\sum_{j \in C_n} P_{j_n} = 1 \quad (2.4)$$

and

$$\sum_j P_j = 1 \quad (2.5)$$

Thus at least superficially, mathematically we have a direct mapping between: (i) the decision-making entity  $n$  in the population and the atom at a particular lattice site in the solid; (ii) finding the decision-making entity has chosen a particular alternative  $i$  described by its systematic utility  $V_{i_n}$  and finding the atom is in a particular microstate  $i$  described by its energy  $E_i$ ; and (iii) the scale parameter  $\mu$  and the temperature parameter  $\beta$ .

If we consider a small macroscopic system  $A$  brought to some final macrostate as a result of interaction with a large number of other macroscopic systems, the physical role of the heat reservoir (where the reservoir is assumed to be so large compared to  $A$  that its known temperature  $T$  is negligibly effected by the small amount of energy given to  $A$ ) is played instead by the totality of all the other systems; instead of knowing the energy of each system in the ensemble, we know

<sup>2</sup> Named in honor of the Austrian physicist who contributed so significantly to the development of kinetic theory and statistical mechanics (Reif, 1965, p. 137).

here information only about the mean energy of the final macrostate of  $A$ . The canonical distribution (2.3) is still valid, however the temperature parameter  $\beta$  does not have any immediate physical significance in terms of a known temperature  $T$  of a real heat bath; instead we calculate the temperature parameter  $\beta$  from the known value of the mean energy, which must satisfy:

$$\bar{E} = \frac{\sum_j E_j e^{-\beta E_j}}{\sum_j e^{-\beta E_j}} \quad (2.6)$$

The expression for the mean energy (2.6), regardless of whether we consider a small macroscopic system in thermal interaction with a heat reservoir or in interaction with a large number other macroscopic systems, takes a particularly convenient and compact form by noting that the partial derivative with respect to  $\beta$  of the partition function  $Z$  in the denominator is simply the negative of the numerator. That is:

$$\bar{E} = \frac{-(\partial Z / \partial \beta)}{Z} = -\frac{\partial \ln Z}{\partial \beta} \quad (2.7)$$

In fact as shown in Reif (1965, Section 6.6), all the important macroscopic physical quantities in thermodynamics can be expressed completely in terms of  $\ln Z$ , which is calculable from microscopic information about the system. Of particular note are the entropy  $S$  given by:

$$S \equiv k(\ln Z + \beta \bar{E}) \quad (2.8)$$

and the Helmholtz free energy  $F$ :

$$F \equiv \bar{E} - TS = -kT \ln Z \quad (2.9)$$

## 2.2 NESTED LOGIT MODEL

Interestingly enough, the natural logarithm of the denominator in (2.6) also plays an important role in discrete choice theory when we advance from the basic multinomial logit model to the so-called nested logit model. Suppose for the moment that the choice set  $C_n$  faced by decision-making entity  $n$  is in fact partitioned into  $M$  mutually exclusive and collectively exhaustive nests  $C_{mn}$  indexed  $(1, \dots, m, \dots, M_n)$ :

$$\begin{aligned} C_n &= \{C_{1n}, \dots, C_{mn}, \dots, C_{Mn}\} \\ C_{mn} \cap C_{m'n} &= \varnothing, \forall m \neq m' \\ \bigcup_{m=1}^M C_{mn} &= C_n \end{aligned} \quad (2.10)$$

Each decision-making entity  $n$  is faced with a single choice among the mutually exclusive elemental alternatives  $i$  in the composite choice set  $C_n$ . Such a "nested" choice structure is depicted schematically in Figure 2.2.

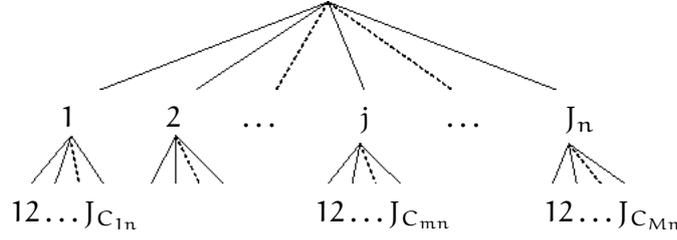


Figure 2.2: Nested choice structure for a given decision-making entity  $n$ .

Let  $U_{in} = V_{in} + \varepsilon_{in}$  be the utility that a given decision-making entity  $n$  is presumed to associate with a particular elemental alternative  $i$  in its choice set  $C_n$ , where  $V_{in}$  is the deterministic (to the modeler) or so-called "systematic" utility and  $\varepsilon_{in}$  is an error term. Similarly, let  $U_{mn} = V_{mn} + \varepsilon_{mn}$  be the composite utility that a given decision-making entity  $n$  is presumed to associate with a particular choice subset  $C_{mn}$ . As derived in Ben-Akiva and Lerman (1985) and Ben-Akiva and Bierlaire (1999), the joint probability that the decision-making entity  $n$  chooses alternative  $i$  within the nest  $C_{mn}$  among all possible alternatives in its choice set  $C_n$  is given by:

$$P_n(i|C_n) = P_n(i|C_{mn}) \cdot P_n(C_{mn}|C_n) \quad (2.11)$$

where the probability of choosing alternative  $i$  within nest  $C_{mn}$ , conditional on having chosen that nest is:

$$P_n(i|C_{mn}) = \frac{e^{\mu_m V_{in}}}{\sum_{j \in C_{mn}} e^{\mu_m V_{jn}}} \quad (2.12)$$

and the probability of choosing nest  $C_{mn}$  among the set of  $M$  nests is:

$$P_n(C_{mn}|C_n) = \frac{e^{\mu V_{mn}}}{\sum_{m' \in M_n} e^{\mu V_{m'/n}}} \quad (2.13)$$

Each nest  $C_{mn}$  within the choice set  $C_n$  is associated with a "composite systematic utility" given by:

$$V_{mn} = \tilde{V}_{mn} + \frac{1}{\mu_m} I_{mn} \quad (2.14)$$

where we have the inclusive value  $I_{mn}$ , otherwise known in diverse literature as the "logsum" or the "accessibility"<sup>3</sup>:

$$I_{mn} = \ln \sum_{j \in C_{mn}} e^{\mu_m V_{jn}} \quad (2.15)$$

<sup>3</sup> The term accessibility has its origins in the early applications of nested logit models to travel demand analysis.

The second term in the expression (2.14) gives summary information from the lower nest. If we recall that temperature parameter  $\beta = 1/kT$  in the macroscopic thermodynamic system, we note a striking mathematical similarity with the Helmholtz free energy  $F$  in (2.9).

### 2.3 ECONOMETRICS VERSUS STATISTICAL PHYSICS

Despite the apparent similarities between the econometric case of decision-making entity  $n$  in a sample population having chosen a particular alternative  $i$  described by its systematic utility  $V_{in}$  and the physical case of an atom at a particular lattice site in the solid being in a particular microstate  $i$  described by its energy  $E_i$ , there are however some important distinctions. Perhaps the most significant of these, simply said, is the fact that in our econometric case the probability is subscripted by  $i$  as well as  $n$ , specific to the particular choice alternative as well as each individual decision-making entity; in the physical case the probability is only subscripted by  $i$ , specific to the particular microstate and not to each individual atom.

For our econometric case, we generally have micro-level information about a vector of individual characteristics  $S_n$  of each decision-making entity  $n$  in the sample, as well as micro-level information about a vector of the attributes  $z_{in}$  for each alternative  $i$  faced by each decision-making entity  $n$ . To give an example relevant to this thesis, the travel time to work by car and the travel time by train may in general not only be different from each other, but also different for different decision-making entities residing in different locations. For the econometric case, we generally have:

$$V_{in} = V(z_{in}, S_n) \quad (2.16)$$

For the physical case however, we generally do not have micro-level information about each individual atom  $A$  in the solid. Instead we have aggregate level or structurally hypothesized information about the system.

Furthermore, in a general econometric context, decision-making entities need not all necessarily be faced with the same alternatives. That is, the choice set  $C_n$  may be of variable size from one decision-making entity to another, and some alternatives may not be available or may be otherwise unknown to various decision-making entities. We can model this by allowing  $A_{in}$  to be an alternative availability indicator variable defined as 1 if elemental alternative  $i$  is available to decision-making entity  $n$  and defined as 0 otherwise.<sup>4</sup> The proba-

<sup>4</sup> The explicit definition of availability at the nest level is superfluous as an entire nest will become unavailable if all of the elemental alternatives within that nest are given to be unavailable.

bility  $P_n(i|C_{m,n})$  that agent  $n$  chooses alternative  $i$  within nest  $C_{m,n}$  conditional on having chosen that nest is then given by:

$$P_n(i|C_{m,n}) = \frac{A_{in} e^{\mu_m V_{in}}}{\sum_{j \in C_{m,n}} A_{jn} e^{\mu_m V_{jn}}} \quad (2.17)$$

How a decision-making entity determines the list of alternatives that it considers is typically referred to as choice set generation and is most often modeled via application of simple deterministic rules. For example spatio-temporal constraints as defined from a fixed reference point such as the location of employment might determine geographically the subset of municipalities under consideration for choice of residential location, or vice versa. Another yet simpler example might be the presence or not of a carpool meeting point in the municipality where a household resides as a deterministic rule for whether carpooling is considered to be an alternative for choice of commuter transport mode. Of course, a household may not realize that a particular municipality is within easy commute of their workplace, or may not realize that there is a carpool meeting point in the municipality where they live. To capture aspects of differing awareness of alternatives from one decision-making entity to the next, probabilistic choice set generation models have been applied. One such example in the discrete choice framework is the so-called latent choice set model (Swait and Ben-Akiva, 1987; Ben-Akiva and Boccara, 1995), a special case of the latent class choice model.

We might also hypothesize that interactions between decision-making units situated in a given exogenous or endogenous social or spatial network might impact the choice set generation process via various specific mechanisms: (i) information transmission in the network about availability of an alternative; (ii) congestion affects – some number of decision-making entities choosing the same alternative makes an alternative not a considerable option for another decision-making entity; (iii) agglomeration affects – some number of decision-making units choosing the same alternative makes an alternative then an option to another decision-making entity where it wouldn't have been an option otherwise; (iv) network proximity between different decision-making entities as either a positive or negative determinant of the availability of an alternative. Just as discussed earlier in the introduction with regard to interactions influencing an outcome, identifiable versus aggregate interactions may also play a role in influencing the choice set generation in the first place. Let us consider the congestion effects in case (ii). An example of aggregate interactions could be the aggregate level existence of so much road congestion or so little parking, that a household doesn't even consider taking the car. An example of identifiable interactions could be an elderly person ruling out public transit as an option for fear of neighborhood

youth. With regard to agglomeration effects, on an aggregate level, perhaps a woman traveling by herself at night would rule out traveling by public transit if she had to stand at a lonely subway station for an extended period, but if the same subway were filled with plenty of other people she wouldn't perceive any threat and transit would be a perfectly viable option. With regard to agglomeration effects at the identifiable level, if two parties commuting to work at the same general time and to the same general location are interested in sharing rides, then carpooling may be an option, otherwise it isn't even if one party is still interested.

Finally, depending on the application, a decision-making entity may be in fact an individual or may be a group of individuals such as a household that takes a single unified decision. For the purposes of this thesis and simplicity of exposition, we will presently consider the decision-making entities as individuals. This is however a non-trivial point worthy of an entire research of its own, for example in considering the more complex problem of double-income earner households faced with residential location choice where both income-earners are faced with multi-modal transport commuting choices, or even individual mode choice but with communal household resources, constraints and activity patterns in mind. Lerman (1975) has touched upon mobility issues in multi-worker households using revealed preference data. There are various reasons why we might expect a priori that intra-household interaction would be important in travel demand behavior. For example, Vovsha, Bradley and Bowman (2004) and Vovsha et al (2004) distinguish between coordination of individual daily activity patterns, joint participation in activities and travel and intra-household mechanisms for allocation of maintenance activities. There exists a growing stream of research in these three areas as cited at the outset in chapter 1. Other interesting examples of the application of intra-household interactions include residential choice behavior (Timmermans et al 1992; Borgers and Timmermans 1993; Abraham and Hunt 1997; Molin 1999; Molin, Oppewal and Timmermans 2002) as suggested above.