A Recent Approach to Derive the Multinomial Logit Model for Choice Probability

A recent approach to derive the Multinomial Logit model for choice probability

Roberto Tadei, Guido Perboli, and Daniele Manerba

Abstract It is well known that the Multinomial Logit model for the choice probability can be obtained by considering a random utility model where the choice variables are independent and identically distributed with a Gumbel distribution. In this paper we organize and summarize existing results of the literature which show that using some results of the extreme values theory for i.i.d. random variables, the Gumbel distribution for the choice variables is not necessary anymore and any distribution which is asymptotically exponential in its tail is sufficient to obtain the Multinomial Logit model for the choice probability.
Keywords: random utility, extreme values theory, asymptotic approximation, Multinomial Logit model

1 Introduction

In this paper we consider a discrete choice model where a decision maker needs to select an alternative among a finite set of mutually exclusive alternatives. Each alternative is caracterized by a random utility. The decision maker will select the alternative with the greatest utility. Discrete choice models of this kind are called random utility models [13]. The aforementioned models are typical of several applications in operations management where decisions must be taken in advance with a

Roberto Tadei
Dept. of Control and Computer Engineering - Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino (Italy), e-mail: roberto.tadei@polito.it

Guido Perboli
Dept. of Control and Computer Engineering - Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino (Italy), e-mail: guido.perboli@polito.it

Daniele Manerba
Dept. of Control and Computer Engineering - Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino (Italy), e-mail: daniele.manerba@polito.it

limited knowledge of the alternatives, as in supply chain optimization, logistics, and transportation (see, e.g., [3], [4], [12], [15], [16], [17], [19], [22]).

It is well known that the Multinomial Logit model (MNL) for the choice probability can be derived assuming that the random utilities are independent and identical distributed (i.i.d.) across alternatives and that their common distribution is a Gumbel function ([11], [1], [2], [6]).

In ([9], [10]) an asymptotic derivation of the MNL is given. Using some results of the extreme values theory for i.i.d. random variables [7], it is shown that the Gumbel distribution for the random variables is not necessary anymore. A distribution that is asymptotically exponential in its tail is just required to obtain the MNL model for the choice probability. Similar derivations are obtained in many applications of location, routing, loading, and packing ([22], [18], [20], [21], [19]).

In this paper we want to organize and summarize all the above existing results by presenting a very simple and intuitive random utility model.

The remainder of the paper is organized as follows. In Section 2, we recall the well-known derivation of the MNL model for the choice probability when the random variable distribution is a Gumbel function. In Section 3, we show that we can relax the Gumbel distribution assumption and still derive the MNL model for the choice probability. Finally, the conclusions of our work are reported in Section 4.

2 Derivation of the MNL model when the random variable distribution is a Gumbel function

Let us consider

- $j = 1, \dots, n$: mutually exclusive choice alternatives
- v_{ij}: deterministic utility of alternative j for decision maker i
- $\tilde{x}_{ij}$: random utility of alternative j for decision maker i.

The decision maker i assigns a total utility to each alternative as follows

$$\tilde{u}_{ij} = v_{ij} + \tilde{x}_{ij} \tag{1}$$

In the following, we recall the main results from the literature, where the MNL model is derived under the assumption that the random variables $\tilde{x}_{ij}$ are i.i.d. over alternatives and their common distribution is a Gumbel function.

Following [23], to derive the MNL model we first consider the density for each random component of utility $\tilde{x}_{ij}$

$$f(\tilde{x}_{ij}) = e^{-\tilde{x}_{ij}} e^{-e^{-\tilde{x}_{ij}}}. \tag{2}$$

Its cumulative distribution is

$$F(\tilde{x}_{ij}) = e^{-e^{-\tilde{x}_{ij}}} \tag{3}$$

which is a Gumbel function.

Following [14], the probability that decision maker i chooses alternative j is

$$\begin{aligned} p_{ij} &= Pr\{v_{ij} + \tilde{x}_{ij} > v_{ik} + \tilde{x}_{ik} \ \ \forall k \neq j\} = \\ &= Pr\{\tilde{x}_{ik} < v_{ij} - v_{ik} + \tilde{x}_{ij} \ \ \forall k \neq j\}. \end{aligned} \tag{4}$$

If $\tilde{x}_{ij}$ is given, this expression is the cumulative distribution for each $\tilde{x}_{ik}$ evaluated at $v_{ij} - v_{ik} + \tilde{x}_{ij}$, which, according to (3), is $exp(-exp(-(v_{ij} - v_{ik} + \tilde{x}_{ij})))$. Since the $\tilde{x}'s$ are independent, this cumulative distribution over all $k \neq j$ is the product of the individual cumulative distributions

$$p_{ij}|\tilde{x}_{ij} = \prod_{k \neq j} e^{-e^{-(v_{ij} - v_{ik} + \tilde{x}_{ij})}}. \tag{5}$$

Of course, $\tilde{x}_{ij}$ is not actually given, then the choice probability is the integral of $p_{ij}|\tilde{x}_{ij}$ over all values of $\tilde{x}_{ij}$ weighted by its density (2), i.e.

$$p_{ij} = \int_{-\infty}^{+\infty} \left[\prod_{k \neq j} e^{-e^{-(v_{ij} - v_{ik} + \tilde{x}_{ij})}} \right] e^{-\tilde{x}_{ij}} e^{-e^{-\tilde{x}_{ij}}} d\tilde{x}_{ij}. \tag{6}$$

After some manipulation of this integral one gets the following expression for the choice probability

$$p_{ij} = \frac{e^{v_{ij}}}{\sum_{k=1}^{n} e^{v_{ik}}} \tag{7}$$

where n is the total number of alternatives. Equation (7) is a MNL model.

3 Derivation of the MNL model when the random variable distribution is not a Gumbel function

We want to show that, under some mild conditions, the Gumbel distribution assumption for the i.i.d. random variables is unjustified to derive the MNL model for choice probability. Thus, we reformulate the random utility choice theory in terms of the asymptotic extreme values theory [7]. This theory deals with the properties of maxima (or minima) of sequences of random variables with a large number of terms. In particular, we will show that when the number of alternatives becomes large the Gumbel distribution assumption is not necessary anymore. In the following, we derive all the results for the maximization case, but the same results can be adapted to the minimization case.

We assume that $F(x)$ is asymptotically exponential in its right tail, i.e. there is a constant $\beta > 0$ such that

$$\lim_{y \to +\infty} \frac{1 - F(x+y)}{1 - F(y)} = e^{-\beta x}. \tag{8}$$

Our research question then becomes: Under condition (8) and the assumption that the number of alternatives is large, can we still asymptotically get the MNL model for the choice probability?

Let us consider a set J of $N = |J|$ alternatives. We assume that J is partitioned into n nonempty disjoint subsets J_j, $j = 1, \ldots, n$, called clusters, of $N_j = |J_j|$ alternatives. The partition into clusters of the alternatives is faced by many choice processes. For instance, when a household is looking for a dwelling, first of all it will select the district where to live (this is the cluster) and, then, inside that district, it will choose the actual dwelling among all the alternatives.

Let $\tilde{u}_{ij}^z$ be the utility for decision maker i for choosing alternative $z \in J_j$.

As already stated, in a random utility model we assume that $\tilde{u}_{ij}^z$ is the sum of a deterministic variable v_{ij} and a random variable $\tilde{x}_{iz}$, i.e.

$$\tilde{u}_{ij}^z = v_{ij} + \tilde{x}_{iz}. \tag{9}$$

The deterministic variable v_{ij} of the utility includes variables which represent attributes of the cluster and the decision context. The random variable $\tilde{x}_{iz}$ represents aspects of utility that the researcher does not observe, e.g., idiosyncrasies of decision maker i.

The decision maker i will choose among all alternatives the one with the maximum utility.

Let us define the distribution of the maximum utility for decision maker i among all alternatives z in all clusters j as

$$\tilde{u}_i = \max_{j=1,\ldots,n;\, z\in J_j} \tilde{u}_{ij}^z = \max_{j=1,\ldots,n} (v_{ij} + \max_{z\in J_j} \tilde{x}_{iz}) = \max_{j=1,\ldots,n} (v_{ij} + \tilde{x}_i^j), \tag{10}$$

where

$$\tilde{x}_i^j = \max_{z\in J_j} \tilde{x}_{iz}. \tag{11}$$

Let

$$G_i(x) = Pr\{\tilde{u}_i < x\} \tag{12}$$

be the distribution of $\tilde{u}_i$ and

$$P_{ij}(x) = Pr\{\tilde{x}_i^j < x\} \tag{13}$$

be the distribution of $\tilde{x}_i^j$.

By the i.i.d. assumption of the random variables, the distribution $P_{ij}(x)$ becomes

$$P_{ij}(x) = \prod_{z\in J_j} Pr\{\tilde{x}_{iz} < x\} = [F(x)]^{N_j}. \tag{14}$$

Now, because of (10), (11), and (14), equation (12) becomes

$$G_i(x) = Pr\{\tilde{u}_i < x\} = Pr\{\max_{j=1,\ldots,n} (v_{ij} + \tilde{x}_i^j < x\} = \prod_{j=1,\ldots,n} Pr\{v_{ij} + \tilde{x}_i^j < x\} =$$

$$= \prod_{j=1,\dots,n} Pr\{\tilde{x}_i^j < x - v_{ij}\} = \prod_{j=1,\dots,n} P_{ij}(x - v_{ij}) =$$
$$= \prod_{j=1,\dots,n} [F(x - v_{ij})]^{N_j}. \tag{15}$$

Following [18] and [20], we will show that under assumption (8) the distribution $G_i(x)$ tends towards a Gumbel function as the total number of alternatives N becomes large. Then we will check if under these results the MNL model for the choice probability can be still derived.

First, let us consider that we can fix the origin for the utility scale arbitrarily, i.e., the choice probabilities are unaffected by a shift in the utility scale and any additive constant to the utilities can be ignored. Let us choose this constant as the root a_N of the equation

$$1 - F(a_N|N) = 1/N, \tag{16}$$

where we remind N is the total number of alternatives.

By replacing $\tilde{u}_i$ with $\tilde{u}_i - a_N$ in (15) one has

$$G_i(x|N) = \prod_{j=1,\dots,n} [F(x - v_{ij} + a_N|N)]^{N_j}. \tag{17}$$

Let us consider the ratio

$$\alpha_j = N_j/N \tag{18}$$

and assume that this ratio remains constant for each j while the values of $N = 1, 2, \dots$ vary, as needed later to compute the asymptotic behavior while N increases.

Because of (18), equation (17) can be written as

$$G_i(x|N) = \prod_{j=1,\dots,n} [F(x - v_{ij} + a_N|N)]^{\alpha_j N}. \tag{19}$$

Let us assume that N is large enough to use $\lim_{N\to+\infty} G_i(x|N)$ as an approximation of $G_i(x)$. Then, the following theorem holds.

Theorem 1. *Under condition (8), the probability distribution $G_i(x)$ becomes the following Gumbel distribution*

$$G_i(x) = \lim_{N\to+\infty} G_i(x|N) = \exp\left(-A_i e^{-\beta x}\right) \tag{20}$$

where

$$A_i = \sum_{j=1,\dots,n} \alpha_j e^{\beta v_{ij}} \tag{21}$$

is the accessibility in the sense of Hansen [8] to the overall set of alternatives.

Proof. By (17) and (18) one has

$$G_i(x) = \lim_{N\to+\infty} G_i(x|N) = \lim_{N\to+\infty} \prod_{j=1,\dots,n} [F(x - v_{ij} + a_N|N)]^{\alpha_j N} =$$

$$= \prod_{j=1,\ldots,n} \lim_{N\to+\infty} [F(x - v_{ij} + a_N|N)]^{\alpha_j N}. \quad (22)$$

From (16), $\lim_{N\to+\infty} a_N = +\infty$ only if $x \to +\infty$, since $\lim_{N\to+\infty} 1/N = 0$ and $1 - F(x) = 0$, or $F(x) = 1$.

From (8) one obtains

$$\lim_{N\to+\infty} \frac{1 - F(x - v_{ij} + a_N|N)}{1 - F(a_N|N)} = e^{-\beta(x-v_{ij})}. \quad (23)$$

By (23) and (16), it holds that

$$\lim_{N\to+\infty} F(x - v_{ij} + a_N|N) = \lim_{N\to+\infty} \left(1 - [1 - F(a_N|N)]e^{-\beta(x-v_{ij})}\right) = $$
$$= \lim_{N\to+\infty} \left(1 - \frac{e^{-\beta(x-v_{ij})}}{N}\right) \quad (24)$$

and, by reminding that $\lim_{n\to+\infty}(1+\frac{x}{n})^n = e^x$

$$\lim_{N\to+\infty} [F(x - v_{ij} + a_N|N)]^N = \lim_{N\to+\infty} \left(1 - \frac{e^{-\beta(x-v_{ij})}}{N}\right)^N = $$
$$= \exp\left(-e^{-\beta(x-v_{ij})}\right). \quad (25)$$

Substituting (25) into (22) and using (21), one finally has

$$G_i(x) = \prod_{j=1,\ldots,n} \exp\left(-\alpha_j e^{-\beta(x-v_{ij})}\right) = \exp\left(-A_i e^{-\beta x}\right), \quad (26)$$

which is a Gumbel distribution. □

By (25), the following approximation holds for large values of N

$$F(x - v_{ij})^N = \exp\left(-e^{-\beta(x-v_{ij}-a_N)}\right), \quad (27)$$

where a_N is a constant. We want to prove that under Theorem 1 the MNL model for the choice probability still holds.

The choice probability p_{ij} for decision maker i to choose cluster j can be determined as follows. Decision maker i chooses cluster j if and only if

$$v_{ij} + \tilde{x}_i^j \geq v_{ik} + \tilde{x}_i^k, \ \forall k \neq j \quad (28)$$

then

$$p_{ij} = Pr\{v_{ij} + \tilde{x}_i^j \geq v_{ik} + \tilde{x}_i^k, \ \forall k \neq j\} = Pr\{v_{ij} + \tilde{x}_i^j \geq \max_{k=1,\ldots n;\, k\neq j} (v_{ik} + \tilde{x}_i^k)\}. \quad (29)$$

Because $\{\tilde{x}_i^k\}$ are independent

$$Pr\{\max_{k=1,\dots n;\, k\neq j}(v_{ik}+\tilde{x}_i^k)<x\} = Pr\{\max_{k=1,\dots n;\, k\neq j}\tilde{x}_i^k < x-v_{ik}\} = \\ = \prod_{k=1,\dots,n;\, k\neq j} P_{ik}(x-v_{ik}) \qquad (30)$$

and

$$Pr\{v_{ij}+\tilde{x}_i^j<x\} = Pr\{\tilde{x}_i^j<x-v_{ij}\} = P_{ij}(x-v_{ij}). \qquad (31)$$

From the Total Probability Theorem [5], and the results in (30) and (31), equation (29) becomes

$$p_{ij} = \int_{-\infty}^{+\infty}\left[\prod_{k=1,\dots n;\, k\neq j} P_{ik}(x-v_{ik})\right] dP_{ij}(x-v_{ij}). \qquad (32)$$

The following theorem holds.

Theorem 2. *The choice probability p_{ij} for decision maker i to choose cluster j is given by*

$$p_{ij} = \frac{N_j e^{\beta v_{ij}}}{\sum_{k=1}^n N_k e^{\beta v_{ik}}}. \qquad (33)$$

Proof. From (32), by using (14) and (27), and setting $\gamma = e^{\beta a_N}$, one obtains

$$p_{ij} = \int_{-\infty}^{+\infty}\prod_{k=1,\dots,n;\, k\neq j}[F(x-v_{ik})]^{\alpha_k N}\, d\,[F(x-v_{ij})]^{\alpha_j N} = \\ = \int_{-\infty}^{+\infty}\prod_{k=1,\dots,n;\, k\neq j}\exp\left[-\alpha_k e^{-\beta(x-v_{ik}-a_N)}\right] d\exp\left[-\alpha_j e^{-\beta(x-v_{ij}-a_N)}\right] = \\ = \int_{-\infty}^{+\infty}\prod_{k=1,\dots,n;\, k\neq j}\exp\left[-\gamma\alpha_k e^{-\beta(x-v_{ik})}\right] d\exp\left[-\gamma\alpha_j e^{-\beta(x-v_{ij})}\right] = \\ = \gamma\alpha_j e^{\beta v_{ij}}\int_{-\infty}^{+\infty}\beta e^{-\beta x} exp(-\gamma A_i e^{-\beta x})dx = \\ = \gamma\alpha_j e^{\beta v_{ij}}\int_0^{+\infty} e^{-\gamma A_i t}dt = \frac{\alpha_j e^{\beta v_{ij}}}{A_i} = \frac{N_j e^{\beta v_{ij}}}{\sum_{k=1}^n N_k e^{\beta v_{ik}}},$$

where $t = e^{-\beta x}$. □

Note that the choice probability in (33) still represents a MNL model.

4 Conclusions

This paper has summarized the usefulness of reinterpreting random utility models by means of the asymptotic theory of extremes, which allows to derive the Multinomial Logit model for the choice probability.

Most well-known results on extreme values statistics concern sequences of i.i.d. random variables. This independency assumption, even if it provides a very convenient form for the choice probability, could be considered too restrictive. In order to relax this assumption, the independency assumption could be replaced by the asymptotic independency for many results.

References

1. Ben-Akiva, M., Lerman, S.R.: Disaggregate travel and mobility choice models and measures of accessibility. In: D. Hensher, P. Stopher (eds.) Behavioral Travel Modeling. Croom Helm, London (1979)
2. Ben-Akiva, M., Lerman, S.R.: Discrete choice analysis: theory and application to travel demand, vol. 9. MIT press (1985)
3. Beraldi, P., Bruni, M.E., Manerba, D., Mansini, R.: A stochastic programming approach for the traveling purchaser problem. IMA Journal of Management Mathematics **28**(1), 41–63 (2017)
4. Crainic, T.G., Gobbato, L., Perboli, G., Rei, W.: Logistics capacity planning: A stochastic bin packing formulation and a progressive hedging meta-heuristic. European Journal of Operational Research **253**(2), 404–417 (2016)
5. DeGroot, M., Schervish, M.: Probability and Statistics (Third Edition). Addison Wesley, Reading (2002)
6. Domencich, T., McFadden, D.: Urban travel dynamics: a behavioral analysis. North Holland, Amsterdam (1975)
7. Galambos, J.: The asymptotic theory of extreme order statistics. John Wiley, New York (1978)
8. Hansen, W.: How accessibility shapes land use. Journal of the American Institute of Planners **25**, 73–76 (1959)
9. Leonardi, G.: The structure of random utility models in the light of the asymptotic theory of extremes. In: M. Florian (ed.) Transportation planning models, pp. 107–133. Elsevier (1984)
10. Leonardi, G.: Asymptotic approximations of the assignment model with stochastic heterogeneity in the matching utilities. Environment and Planning A **17**, 1303–1314 (1985)
11. Luce, R.D.: Individual choice behavior: a theoretical analysis. John Wiley, New York (1959)
12. Manerba, D., Mansini, R., Perboli, G.: The capacitated supplier selection problem with total quantity discount policy and activation costs under uncertainty. International Journal of Production Economics **198**, 119–132 (2018)
13. Marschak, J.: Binary choice constraints on random utility indications. In: K. Arrow (ed.) Stanford Symposium on Mathematical Methods in the Social Sciences, pp. 312–329. Stanford University Press (1960)
14. McFadden, D.: Conditional logit analysis of qualitative choice behavior. In: P. Zarembka (ed.) Frontiers in Econometrics, pp. 105–142. Academic Press (1974)
15. Perboli, G., De Marco, A., Perfetti, F., Marone, M.: A new taxonomy of smart city projects. Transportation Research Procedia **3**, 470–478 (2014)
16. Perboli, G., Ghirardi, M., Gobbato, L., Perfetti, F.: Flights and their economic impact on the airport catchment area: an application to the italian tourist market. Journal of Optimization Theory and Applications **164**, 1109–1133 (2015)
17. Perboli, G., Gobbato, L., Maggioni, F.: A progressive hedging method for the multi-path traveling salesman problem with stochastic travel times. IMA Journal of Management Mathematics **28**, 65–86 (2015)
18. Perboli, G., Tadei, R., Baldi, M.: The stochastic generalized bin packing problem. Discrete Applied Mathematics **160**, 1291–1297 (2012)
19. Perboli, G., Tadei, R., Gobbato, L.: The multi-handler knapsack problem under uncertainty. European Journal of Operational Research **236**(3), 1000–1007 (2014)

20. Tadei, R., Perboli, G., Baldi, M.M.: The capacitated transshipment location problem with stochastic handling costs at the facilities. International Transactions in Operational Research **19**(6), 789–807 (2012)
21. Tadei, R., Perboli, G., Perfetti, F.: The multi-path traveling salesman problem with stochastic travel costs. EURO Journal on Transportation and Logistic **6**, 2–23 (2014). doi: 10.1007/s13676-014-0056-2
22. Tadei, R., Ricciardi, N., Perboli, G.: The stochastic p-median problem with unknown cost probability distribution. Operations Research Letters **37**, 135–141 (2009)
23. Train, K.E.: Discrete Choice Methods with Simulation. Cambridge University Press (2003)