Transportation Research Part B 91 (2016) 1–20

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

A class of RUM choice models that includes the model in which the utility has logistic distributed errors

J.M. del Castillo*

Department of Material Science and Transportation Engineering, University of Seville, Escuela Superior de Ingenieros, Seville 41092, Spain

ARTICLE INFO

Article history:
Received 20 August 2015
Revised 27 April 2016
Accepted 28 April 2016
Available online 24 May 2016

Keywords:
Random utility
Choice probability generating function
Closed-form choice probability
Logistic distribution
Heteroscedastic model
Divided difference

ABSTRACT

A class of random utility maximization (RUM) models is introduced. For these RUM models the utility errors are the sum of two independent random variables, where one of them follows a Gumbel distribution. For this class of RUM models an integral representation of the choice probability generating function has been derived which is substantially different from the usual integral representation arising from the RUM theory. Four types of models belonging to the class are presented. Thanks to the new integral representation, a closed-form expression for the choice probability generating function for these four models may be easily obtained. The resulting choice probabilities are fairly manageable and this fact makes the proposed models an interesting alternative to the logit model. The proposed models have been applied to two samples of interurban trips in Japan and some of them yield a better fit than the logit model. Finally, the concavity of the log-likelihood of the proposed models with respect to the utility coefficients is also analyzed.

1. Introduction

Random utility maximization (RUM) discrete choice models are based on the assumption that the utility of the alternative i of the choice set $C = \{1, 2\dots, m\}$ is the sum of a deterministic component or systematic utility V_i and a random component ε_i that accounts for the error in the perception of the utility:

$$U_i = V_i + \varepsilon_i \quad i = 1, 2\dots, m.$$

The above decomposition of the utility is more precisely referred to as the additive random utility model (ARUM). It is well known that the multinomial logit model (MNL) can be derived by assuming that the errors ε_k, $k \in C$ are independent and identically distributed (i.i.d.) Gumbel random variables. This assumption for the random component allows obtaining a simple and closed-form analytical expression for the choice probabilities. Namely, the probability that an individual chooses the alternative i is given by

$$p_i = \frac{e^{V_i}}{\sum_{k=1}^{m} e^{V_k}} \tag{1}$$

where m is the number of alternatives.

Fosgerau et al. (2013) show that the choice probabilities of any RUM choice model can be expressed as the derivatives of the logarithm of a function called the "choice probability generating function." The choice probability generating function

* Tel.: +34 954 486082.
E-mail address: delcastillo@us.es

http://dx.doi.org/10.1016/j.trb.2016.04.022
0191-2615/

is a nonnegative function $G : [0, +\infty]^m \to [0, +\infty]$ that should satisfy a set of properties (see definition 5 in Fosgerau et al., 2013. The choice probabilities can be derived from the choice probability generating function as follows:

$$p_i = \frac{\partial \log G(e^{\mathbf{V}})}{\partial V_i} \tag{2}$$

For the MNL model, the choice probability generating function is

$$G(e^{\mathbf{V}}) = G(e^{V_1}, e^{V_2}, \dots e^{V_m}) = \sum_{k=1}^{m} e^{V_k} \tag{3}$$

and then the MNL choice probabilities are given by (1).

In particular, if the error components ε_i of the utilities are independent random variables whose distribution function is $F_i(x)$, we have, according to Theorem 1 of Fosgerau et al. (2013):

$$\log G(e^{\mathbf{V}}) = \int_0^{\infty} \left(1 - \prod_{i=1}^{m} F_i(x - V_i)\right) dx - \int_{-\infty}^{0} \prod_{i=1}^{m} F_i(x - V_i)\, dx. \tag{4}$$

The choice probability of alternative i is obtained by differentiating (4) with respect to V_i:

$$p_i = \frac{\partial \log G(e^{\mathbf{V}})}{\partial V_i} = \int_{-\infty}^{\infty} \prod_{\substack{k=1,\dots m \\ k \neq i}} F_k(x - V_k) f_i(x - V_i)\, dx \tag{5}$$

where $f_k(x) = F_k'(x)$ is the probability density function of the utility error ε_k. If we specify the distribution functions of the errors, we might find a closed-form expression for the choice probability generating function by solving the integrals in (4), or if such an expression cannot be obtained, we could try to solve the integrals in (5) to get at least a closed-form expression for the choice probabilities. Unfortunately, the integrals in (4) and (5) turn out to be an almost insurmountable obstacle that limits drastically the availability of RUM discrete choice models with closed-from choice probabilities.

A recent review of discrete choice models with closed form expressions for the choice probabilities can be found in Koppelman and Sethi (2008). Basically all these models belong to the family of Generalized Extreme Value RUM models. There are a limited number of RUM models outside of the GEV family. A very appealing deviation from the assumption of identically distributed errors is the inclusion of a scale parameter θ_i for the utility error of the i alternative. This is the assumption of Bhat (1995), assuming that the utility has the form $U_i = V_i + \theta_i \varepsilon_i$, where the ε_i are i.i.d. Gumbel variables. Although this model is fairly suitable for capturing the heteroscedasticity of the errors, the choice probabilities have no closed form and they need to be evaluated using Gaussian quadrature.

An earlier departure from the assumption of Gumbel errors is due to Recker (1995). He introduced the so-called oddball RUM model as a model in which the utility error of one alternative is the sum of two independent Gumbel variables and the rest of the alternatives have Gumbel errors. For this model, Recker derived the choice probabilities and studied their properties. Another interesting model is the reversed multinomial logit model of Anderson and de Palma (1999). In the reversed multinomial logit model, the utility is the difference of a deterministic component and a Gumbel random variable: $U_i = V_i - \varepsilon_i$ where the ε_i are independent Gumbel random variables. The choice probabilities for this model can also be obtained.

An interesting alternative for the GEV models is to consider a mixed discrete/continuous probability distribution for the utility error, namely, the two-condition model was defined by Swait (2009) as an RUM model in which the utility of alternative i may take one of the following two expressions:

$$U_i = \begin{cases} V_i + \varepsilon_i & \text{with probability } p_i \\ -\infty & \text{with probability } 1 - p_i \end{cases}$$

and the three-condition model, when it may take one of the three following expressions:

$$U_i = \begin{cases} \infty & \text{with probability } p_i \\ V_i + \varepsilon_i & \text{with probability } q_i \\ -\infty & \text{with probability } r_i \end{cases}$$

where ε_i follows a Gumbel distribution.

Another type of RUM model whose choice probabilities may obtained is that introduced by Swait (2001) and improved by Cantillo and Ortúzar (2005). In those models, and unlike the conventional choice models that consider a pre-specified choice set, a given alternative belong to the choice set of an individual only when all its attributes are within their corresponding thresholds. The expression for the choice probabilities is then obtained by including a correction that takes into account the probability of the alternative's being in the choice set.

Within the framework of an assignment problem for which the route costs follow a Weibull distribution, Castillo et al. (2008) consider the case of independent utility error terms having a Weibull distribution. They obtained the route choice probabilities, which turn out to be equivalent to those of the logit model, as the authors explain: "From a practical modelling point of view, there is no more flexibility in the Weibull model. Any Weibull choice model could also be represented as a logit model. So, one can enter costs linearly or in logarithmic form in a logit model, and its counterpart in the