Available online at www.sciencedirect.com
ScienceDirect
European Journal of Operational Research 176 (2007) 1778–1794
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

Decision Support

Classification methods for random utility models with i.i.d. disturbances under the most probable alternative rule

N. Kalouptsidis [a,*], K. Koutroumbas [b,1], V. Psaraki [c,2]

[a] *Department of Informatics and Telecommunications, Division of Communications and Signal Processing, University of Athens, Panepistimiopolis, T.Y.P.A. Buildings, 15781 Athens, Greece*

[b] *National Observatory of Athens, Institute for Space Applications and Remote Sensing, Metaxa & V. Pavlou, 15236 Palaia Penteli, Athens, Greece*

[c] *Faculty of Civil Engineering, National Technical University of Athens, 5 Iroon Politexniou Street, 15773 Athens, Greece*

Received 11 May 2004; accepted 9 November 2005
Available online 25 January 2006

Abstract

In this paper random utility maximization based on maximization of correct classification of the choice decisions over a given data set is considered. It is shown that if the disturbance vector in the random utility model is independent and identically distributed, then preference determination based on the most probable alternative reduces to deterministic utility maximization. As a consequence of the above equivalence, the form of the error distribution (normal, Weibull, uniform etc.) plays no role in the determination of the preferred alternative. Parameter estimation under the most probable alternative rule is carried out using two methods. The first is based on the solution of an appropriately defined system of linear inequalities and the second one is based on the function optimization of a newly proposed function, whose optimum is achieved when the number of correctly classified individuals is maximized. The ability to use these algorithms in the framework of pattern recognition and machine learning is pointed out. Simulations and a real case study involving intercity travel behavior are employed to assess the proposed methods.

Keywords: Decision analysis; Discrete choice models; Random utility model; Logit model; Pattern recognition

* Corresponding author. Tel.: +30 210 7275304.
E-mail addresses: kalou@di.uoa.gr (N. Kalouptsidis), koutroum@space.noa.gr (K. Koutroumbas), vpsaraki@central.ntua.gr (V. Psaraki).
[1] Tel.: +30 210 8109189.
[2] Tel.: +30 210 7721204.

0377-2217/$ - see front matter
doi:10.1016/j.ejor.2005.11.004

1. Introduction

Many popular discrete choice models are founded on the random utility maximization (RUM) model. An extensive account of this theory is given in (Domencich and McFadden, 1975; Ben-Akiva and Lerman, 1985; Ortuzar and Willumsen, 1996; McFadden, 2000, 1973, 1978; Greene, 2003). The family of RUM models including the multinomial logit and probit models, generalized extreme value models, nested models, random parameter or mixed models, latent class models, has been applied in a wide range of application areas too numerous to be listed here. Indicative relatively recent examples include recreation behavior (Provencher and Bishop, 2004), brand choice in consumer decision making incorporating state dependence (Seetharam, 2003) and multiple brands (Baltas, 2004), competitive location of facilities (Benati and Hansen, 2002).

Consider an individual n choosing a single option among k alternatives. Preferences for such discrete alternatives are determined by the corresponding utilities. These in turn depend on characteristics of the alternative and characteristics and/or tastes of the individual. The RUM model assumes that the utility of the individual n, U_{ni}, for the alternative $1 \leqslant i \leqslant k$ is formed by the sum of a deterministic component V_{ni} and a random component e_{ni}

$$U_{ni} = V_{ni} + e_{ni}. \tag{1}$$

The random error e_{ni} is due to interpersonal and intrapersonal variation in preferences (McFadden, 1973). The systematic utility V_{ni} is determined by a combination of alternative specific factors $\boldsymbol{w}_{\boldsymbol{ni}}^{\mathrm{T}} = [w_{ni}^1, w_{ni}^2, \ldots, w_{ni}^p]$ and alternative invariant factors $\boldsymbol{x}_{\boldsymbol{n}}^{\mathrm{T}} = [x_n^1, x_n^2, \ldots, x_n^r]$ via the rule

$$V_{ni} = g(\boldsymbol{w}_{\boldsymbol{ni}}, \boldsymbol{x}_{\boldsymbol{n}}, \boldsymbol{\beta}), \tag{2}$$

where “T” denotes transpose. All factors $\boldsymbol{w}_{\boldsymbol{ni}}$ and $\boldsymbol{x}_{\boldsymbol{n}}$ are measurable and g is a known function. Hence the only unknown parameter that needs to be determined for the specification of the systematic utilities is the finite dimensional vector $\boldsymbol{\beta}$.

Let $P_n(i)$ denote the probability that individual n picks alternative i. The RUM model postulates that $P_n(i)$ is given by

$$P_n(i) = P[U_{ni} \geqslant U_{nj}, \forall j \neq i]. \tag{3}$$

The estimation of the parameter vector $\boldsymbol{\beta}$ is then accomplished by optimizing a suitable objective function that combines the basic Eqs. (1)–(3). The first and most popular such estimation procedure relies on the multinomial logit (MNL) model formulation which establishes the connection between the RUM framework as an organizing concept for model development and the specification of empirical demand models (McFadden, 1973, 1978, 2000; Erlander, 1998; Manrai, 1995).

The operability of the MNL model rests upon three assumptions. The first assumption considers that the systematic utility V_{ni} depends linearly on the unknown parameters so that Eq. (2) takes the form

$$V_{ni} = \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{z}_{ni}, \tag{4}$$

where $\boldsymbol{z}_{ni} = [\boldsymbol{x}_{\boldsymbol{n}}^{\mathrm{T}}, \boldsymbol{w}_{\boldsymbol{ni}}^{\mathrm{T}}]^{\mathrm{T}}$.

The second assumption states that the error variables e_{ni} are independent and Weibull distributed over the alternatives. In this case the probabilities defined by Eq. (3) are simply expressed in the logit form

$$P_n(i) = \frac{\exp(V_{ni})}{\sum_{j=1}^{k} \exp(V_{nj})}. \tag{5}$$

The third assumption employs the maximization of the log-likelihood function

$$\hat{\boldsymbol{\beta}} : \max_{\boldsymbol{\beta}} L(\boldsymbol{\beta}) = \max_{\boldsymbol{\beta}} \sum_{n=1}^{N} \sum_{i=1}^{k} d_{ni} \log P_n(i) \tag{6}$$