Operations Research Letters 44 (2016) 234–237

Contents lists available at ScienceDirect

Operations Research Letters

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

Discrete choice models based on random walks

Gerardo Berbeglia

Melbourne Business School, University of Melbourne, Australia

ARTICLE INFO

Article history:
Received 13 May 2015
Received in revised form
20 January 2016
Accepted 20 January 2016
Available online 29 January 2016

Keywords:
Discrete choice models
Markov chain
Random utility
Revenue management
Stochastic preferences
Random walk

ABSTRACT

We show that a large class of discrete choice models which contain the Markov chain model introduced by Blanchet, Gallego, and Goyal (2013) belong to the class of discrete choice models based on random utility.

1. Introduction

Discrete choice models (DCM) have been widely used to model the choices individuals make when they are offered a set of alternatives (or options). DCM have played an important role in several research areas such as psychology, economics, marketing, and, more recently, in the field of revenue management.

In the standard discrete choice model setting, there is a universe of alternatives or product types $\mathcal{C} = \{1, \ldots, N\}$. Individuals are then exposed to offer sets $S \subseteq \mathcal{C}$, and they must select one element from S or nothing at all. Each subset S offered to individuals is known as the choice set. In this setting, a discrete choice model is characterized by a function $\mathcal{P} : \mathcal{C} \times 2^{\mathcal{C}} \rightarrow \mathbb{R}$ such that for all $i \in \mathcal{C} \cup \{0\}$ and all $S \subseteq \mathcal{C}$, $\mathcal{P}(i, S)$ denotes the probability of selecting element i given that the offer set is S. The function $\mathcal{P}$ is often called the system of choice probabilities that characterizes a specific discrete choice model.

The practical advantage of using a particular discrete choice model depends on two characteristics. The first one is the degree upon which it is able to model the choice behaviour of individuals in different settings. The better a discrete choice model can approximate complex choice behaviour, the *richer* is said to be. The second one is by the degrees of freedom it has. A model with a large number of degrees of freedom would typically require a large historical data-set in order to find the right values of its parameters and it is more prone to over-fitting problems. These two dimensions are usually in conflict, i.e., discrete choice models that can capture complex choice behaviours generally have a higher number of degrees of freedom. Finding the "right" discrete choice model for a given application involves searching a model with a proper trade-off between the two mentioned dimensions (see, e.g., [11]).

One of the most simple and also most used models in discrete choice is the Multinomial Logit Model (MNL) [15]. Because of its simplicity, the MNL fails to predict complex behaviour and that has motivated the creation of richer models such as the Probit Model [1], the Nested Logit Model [14], the Mixed Multinomial Logit [17] and distance-based models [18] which contain the Mallows model [16] as a special case. All these models belong to an important and large class of discrete choice models based on random utility known as *Random Utility Models* (RUMs) which are defined for completeness in Section 2. Recently, Webb [22] showed that a predominant class of discrete choice models to understand neural decision making, known as *Bounded Accumulation models* [20], are as well a subclass of RUMs.

There exist nevertheless classes of discrete choice models that fall outside the class of RUMs. We briefly mention the most relevant ones. Echenique et al. [8] have considered a discrete choice model in which agents select an alternative following a perception priority order. The authors showed that their model, called *Perception-Adjusted Luce Model* (PALM), is able to explain recent experiments carried out with consumers [7] that cannot be explained by any random utility model. Another choice model, initially proposed to understand sales diversity under monopolistic competition [21,6], is the *Representative Agent Model* (RAM). In a

E-mail address: g.berbeglia@mbs.edu.

http://dx.doi.org/10.1016/j.orl.2016.01.009
0167-6377/

RAM, the choice among the alternatives is decided by a specifically constructed agent that is a representative of the whole population. Hofbauer and Sandholm [12] proved that when the number of alternatives is at least 4, there is always a RAM which is not a RUM. Natarajan et al. [19] have presented a strict generalization of the RUM, known as *Semi-parametric Choice Model* (SCM), and showed how it can be used to make predictions from a real-life transportation data set that outperform several RUM models. The question of what is the relationship between the classes RAM and SCM was answered only recently by Feng et al. [10]. The authors provide a detailed classification of several discrete choice models and proposed a new class which they called welfare-based choice models. Their main contribution is a proof that the RAM, the SCM, and their new proposed class are essentially the same. (It follows from their result that welfare-based choice models include RUMs as a proper subset.)

Recently, another class of discrete choice models have been proposed by Blanchet et al. [3]. In their model, individual preferences are built using a Markov chain in which states are the alternatives or products. The model prescribes how would an individual decide which alternative to select when she is offered a subset of alternatives S. Specifically, with probability p_i an individual has alternative i as her most favourite. If the alternative i is not available, with probability p_{ij} the individual who preferred i *walks* to select the alternative j. If again, the alternative j is not available, with probability p_{jk} she walks to alternative k, etc. Since it is assumed that for all $i \in \mathcal{C}$, $\sum_{j \in \mathcal{C}\setminus\{i\}} p_{ij} < 1$, this walking process will end either when the individual walks to the no-choice option (in which case selects nothing) or when she arrives to an alternative from the set S (and selects it). Blanchet et al. [3] showed how this Markov chain model is an appealing model in the area of revenue management, which consists of a set of methodologies firm use to decide on the availability and/or the price of their products and/or services. First, they showed that the model generalizes the widely used MNL, and second, the assortment problem under this new model can be solved efficiently. The assortment problem under the Markov chain model was recently studied in more depth by Feldman and Topaloglu [9] and Désir et al. [5] where the authors analysed different extensions such as when there is limited inventory, or when there is an upper bound on the size of the assortments that can be shown to individuals.

1.1. Our results

Blanchet et al. [3], Feldman and Topaloglu [9] and Désir et al. [5] left open a fundamental question about the Markov chain model: how rich is this class? More specifically, how does this class compares to RUMs? In their analysis of discrete choice models, Feng et al. [10] argued that the Markov chain model is not related to other discrete choice models they studied such as the RUMs, SCM and RAM, and they omitted the Markov model from their analysis and classification. (To the best of our knowledge, the Markov chain model is the only discrete choice model studied in the literature whose connection with the RUM remained open.) In this paper, we prove that every Markov chain model belongs to the class of RUMs. Thus, despite their appealing features, Markov chain models are not able to replicate every choice behaviour that can be modelled by other discrete choice models such as the PALM and the RAM (or equivalently, the SCM or the welfare-based models). On the positive side, it follows from our result that the three performance guarantees for the simple and well-known revenue-ordered assortment heuristic for the assortment problem that hold for all RUMs [2] are valid as well for every Markov chain model.

A natural way to enrich the class of Markov models is by providing more 'memory' to the individual along her walk over the alternatives. Suppose now that the probability that an individual who is currently in alternative i would walk to the alternative j depends not only on i and j, but also on the alternative that was visited right before alternative i. In other words, the individual now considers the previous two alternatives (instead of only the last one) in order to probabilistically walk to the following alternative. Is it possible that this model extension becomes richer than RUMs? If not, what about the class of models where individuals remember the last three, or the last k alternatives with $k \in \mathbb{N}$? We answer this question negatively. Specifically, we introduce, in Section 3, a new class of discrete choice models called *Discrete choice Models Based on Random Walks* where the walking probabilities depend on the *whole* sequence of the alternatives previously visited. We then prove that for every such a model, there always exists a random utility model that is equivalent to it. In other words, regardless on how much memory the individuals are endowed with about the different alternatives that they walked through (this included arbitrarily large memory as sequences's length are unbounded), the model will fail to explain behaviours that cannot be explained by RUMs.

2. Preliminaries

In a Random Utility Model, each alternative i (including the no-choice option) has associated a random real variable (utility) u_i. These $N+1$ variables are jointly distributed over $\mathbb{R}^{N+1}$, with a certain probability measure $\mathbb{P}$ with $\mathbb{P}(U_i = U_j) = 0$ for all $i, j \in \mathcal{C} \cup \{0\}$, $i \neq j$. Then, the probability of selecting alternative $x \in S \cup \{0\}$ with $S \subseteq \{1, 2 \dots, N\}$ is equal to the probability that alternative x has the highest utility among those in $S \cup \{0\}$. The system of choice probabilities is then characterized as follows:

$$\mathcal{P}(x, S) = \mathbb{P}(U_x = \max\{U_c : c \in (S \cup \{0\})\}).$$

An alternative way to think about random utility models is using what is known as *stochastic preference*. Given $k \in \mathbb{N}$, let S_k denote the symmetric group which is the set of all the permutations of the elements in $\{0, 1, 2, \dots, k-1\}$. A stochastic preference model is described by a probability distribution Pr over the set S_{N+1} (i.e. a probability distribution over all the rankings of the elements in $\mathcal{C} \cup \{0\}$). Given two elements $i, j \in \mathcal{C} \cup \{0\}$ and a permutation $\preceq \in S_{N+1}$, we say that $i \preceq j$ whenever i appears before j in $\preceq$ or when $i = j$. A system of choice probabilities $\mathcal{P}$ characterizes a stochastic preference Pr if

$$\mathcal{P}(x, S) = Pr(\preceq \in S_{N+1} : x \preceq y, \text{ for all } y \in (S \cup \{0\})) \tag{1}$$

for all $S \subseteq \mathcal{C}$ and $x \in S \cup \{0\}$.

Thus $\mathcal{P}(x, S)$ represents the probability of selecting a ranking from the probability space Pr in which element x appears before the no-choice option 0, and any other element in S.

The equivalence of stochastic preference models and random utility models is expressed in the following theorem.

Theorem 1. *A system of choice probabilities $\mathcal{P}$ characterizes a stochastic preference model if and only if it characterizes a RUM.*

For a proof of this result see, e.g., [4,13].

3. Discrete choice models based on random walks

Let Ω denote the set of all finite sequences of elements in $\mathcal{C}$ (including the empty sequence $s = \emptyset$). A discrete choice model based on random walks can be defined based on a series of probability distribution functions $\mathsf{P}_\alpha : \mathcal{C} \cup \{0\} \to [0, 1]$, one for each $\alpha \in \Omega$. We begin by describing how a random walk is constructed given the PDF's P_α and a choice set S.

Suppose we offer to an individual a subset of alternatives S. With probability $\mathsf{P}_{\emptyset}(i_1)$, the first alternative traversed in the random walk by the individual is the element $i_1 \in (\mathcal{C} \cup \{0\})$. If $i_1 \in (S \cup \{0\})$, then the individual selects alternative i_1 (note that if $i_1 = 0$ the individual selects the no-choice option). In any other case, the individual *walks* from current alternative i_1 to alternative $i_2 \in \mathcal{C} \cup \{0\}$ with probability $\mathsf{P}_{s=(i_1)}(i_2)$. Following the same logic, if $i_2 \in (S \cup \{0\})$, the individual selects alternative i_2, otherwise he/she *walks* from i_2 to alternative $i_3 \in \mathcal{C} \cup \{0\}$ with probability $\mathsf{P}_{s=(i_1,i_2)}(i_3)$. In general, assuming the individual has walked along a sequence ($s = i_1, i_2, \ldots, i_k$) (note that an alternative may appear multiple times in the sequence s), the probability to move to alternative $i_{k+1} \in \mathcal{C} \cup \{0\}$ is $\mathsf{P}_{s=(i_1,i_2,\ldots,i_k)}(i_{k+1})$. In order to guarantee that such process will always finish in finite time, we will impose the constraint that the probability that the individual would ever arrive back to an alternative already visited is always less than one. We call this property *infinite-loop avoidance* and we now state it formally. Let $\alpha = (x_1, x_2, \ldots, x_k) \in \Omega$ and denote by α^ℓ ($\ell \in \{1, \ldots, k\}$) the sub-sequence of α that contains the first ℓ elements, i.e. $\alpha^\ell = (x_1, x_2, \ldots, x_\ell)$. The property *infinite-loop avoidance* then states that for any $\alpha \in \Omega$ and any pair of indices i, j such that $i < j$ and $x_i = x_j$, it must be the case that

$$\mathsf{P}_{\alpha^i}(x_{i+1}) \cdot \mathsf{P}_{\alpha^{i+1}}(x_{i+2}) \cdot \ldots \cdot \mathsf{P}_{\alpha^{j-1}}(x_j) < 1.$$

Now that we have described the individual walk process along the alternatives when given an arbitrary choice set S, we are ready to define the class of discrete choice models based on random walks. To do so, it suffices to specify the system of choice probabilities $\mathcal{P} : \mathcal{C} \times 2^{\mathcal{C}} \to \mathbb{R}$. We define $\mathcal{P}(i, S)$, i.e. the probability that the individual would pick alternative i when is given an offer set S in a discrete choice model based on random walks, as the probability that i is the first alternative that belongs to $S \cup \{0\}$ in the random walk constructed based on the PDF's P_α.

It worth to note that most discrete choice models based on random walks that might be used in practice can be characterized with a finite number of PDF's $\mathsf{P}_\alpha : \mathcal{C} \cup \{0\} \to [0, 1]$. As an example one could provide the function P_β for all $\beta \in \Omega$ where $|\beta| \le k$ and then state that for all $\alpha \in \Omega$ such that $|\alpha| > k$, $\mathsf{P}_\alpha = \mathsf{P}_\beta$ where β is the sequence of the last k elements of α. In particular, applying this idea with $k = 1$, we recover the class of Markov chain models.

Theorem 2. *Every discrete choice model based on random walks is a random utility model.*

Proof. Let $\{\mathsf{P}_\alpha : \alpha \in \Omega\}$ denote the countable set of PDF's that characterize a given discrete choice model based on random walks. We consider the random walk that would take place by an individual in the case where the choice set offered is the empty set. Now for every $h \in \Omega$, let $\mathbb{P}(h)$ denote the probability that the random walk traversed by the individual before reaching the alternative 0 is the sequence h (where $\mathbb{P}(h = \emptyset) = \mathsf{P}_\emptyset(0)$). Formally, suppose that $h = (x_1, x_2, \ldots, x_k)$, then

$$\mathbb{P}(h) = \mathsf{P}_\emptyset(x_1) \cdot \mathsf{P}_{(x_1)}(x_2) \cdot \mathsf{P}_{(x_1,x_2)}(x_3) \cdot \ldots \\ \cdot \mathsf{P}_{(x_1,x_2,\ldots,x_{k-1})}(x_k) \cdot \mathsf{P}_{(x_1,x_2,\ldots,x_k)}(0).$$

Because of the *infinite loop avoidance* property and by the fact that the individual was offered the empty set as the choice set, we know that every possible random walk has a finite length before arriving to alternative 0. Thus, it holds that

$$\sum_{h\in\Omega} \mathbb{P}(h) = 1.$$

Let $h \in \Omega$, $i \in \mathcal{C}$, and $S \subseteq \mathcal{C}$. We denote by $top(h, S)$ the alternative in S that appears first in the sequence h (in case h does not contain any alternative from S, $top(h, S)$ returns -1). We define the function $\omega : \Omega \times \mathcal{C} \times 2^{\mathcal{C}} \to \{0, 1\}$ as follows

$$\omega(h, i, S) := \begin{cases} 1 & \text{if } top(h, S) = i, \\ 0 & \text{otherwise.} \end{cases}$$

We can now use $\mathbb{P}(h)$ and ω to calculate $\mathcal{P}(i, S)$, i.e., the probability of selecting the alternative i when the choice set S is being offered as follows:

$$\mathcal{P}(i, S) = \sum_{h\in\Omega} \mathbb{P}(h)\omega(h, i, S).$$

Suppose $h \in \Omega$ has an alternative i appearing more than once, and let h' be the sequence which is exactly the same as h but in which the element i is removed all the times it appears except the first one. It is then clear that

$$\omega(h, i, S) = \omega(h', i, S).$$

Given a finite sequence $h \in \Omega$, let $sk(h)$ be the sequence obtained from h by removing all but the first appearances of each of the elements contained in h. Clearly, $sk(h)$ would contain each alternative in $\mathcal{C}$ at most once and it must hold that

$$\mathcal{P}(i, S) = \sum_{h\in\Omega} \mathbb{P}(h)\omega(sk(h), i, S).$$

Finally, let ϕ be the function that assigns to each sequence s without repetitions of elements in $\mathcal{C}$, the permutation of $\pi \in S_{N+1}$ of the elements in $\mathcal{C} \cup \{0\}$ which consists of the concatenation of the sequence s, joined with the no-choice option 0, and joined in the end with the elements that are in $\mathcal{C}$ but which are not in s sorted in increasing order. For example, if $\mathcal{C} = (1, 2, 3, 4)$ and $s = (3, 2)$, then $\phi(s) = (3, 2, 0, 1, 4)$. Because in $\phi(s)$ the alternative 0 appears before any alternative which is not in s, we can write $\mathcal{P}(i, S)$ as follows.

$$\mathcal{P}(i, S) = \sum_{h\in\Omega} \mathbb{P}(h)\omega(sk(h), i, S) \tag{2}$$

$$= \sum_{h\in\Omega} \mathbb{P}(h)\omega(\phi(sk(h)), i, S \cup \{0\}) \tag{3}$$

which in turn can be re-written as

$$\mathcal{P}(i, S) = \sum_{\pi\in S_{N+1}} \left(\sum_{\substack{h\in\Omega,\\ \phi(sk(h))=\pi}} \mathbb{P}(h) \right) \omega(\pi, i, S \cup \{0\}). \tag{4}$$

By defining a probability distribution Pr over the $(N + 1)!$ rankings of the elements in $\mathcal{C} \cup \{0\}$ as

$$Pr(\pi) = \sum_{\substack{h\in\Omega,\\ \phi(sk(h))=\pi}} \mathbb{P}(h), \tag{5}$$

we have that

$$\begin{aligned} \mathcal{P}(i, S) &= \sum_{\pi\in S_{N+1}} Pr(\pi)\omega(\pi, i, S \cup \{0\}) \\ &= \sum_{\substack{\pi\in S_{N+1},\\ \omega(\pi,i,S\cup\{0\})=1}} Pr(\pi) \cdot 1 + \sum_{\substack{\pi\in S_{N+1},\\ \omega(\pi,i,S\cup\{0\})=0}} Pr(\pi) \cdot 0 \\ &= \sum_{\substack{\pi\in S_{N+1},\\ \omega(\pi,i,S\cup\{0\})=1}} Pr(\pi) \\ &= \sum_{\substack{\pi\in S_{N+1},\\ top(\pi,S\cup\{0\})=i}} Pr(\pi) \\ &= Pr(\preceq \in S_{N+1} : i \preceq y, \ \text{for all } y \in (S \cup \{0\})). \end{aligned}$$

This proves that every discrete choice model based on random walks can be represented by a stochastic preference. By Theorem 1 it follows that every discrete choice model based on random walks is also a Random Utility Model. □

The following corollary easily follows.

Corollary 1. *The Markov chain model is a Random Utility Model.*

Proof. The Markov chain model is a special case of the class of discrete choice models based on random walks in which for every $\alpha = (x_1, \ldots, x_k) \in \Omega$ where $|\alpha| > 0$, $\mathsf{P}_\alpha(i) = \mathsf{P}_{(x_k)}(i)$ for all $i \in \mathcal{C}$. □

We end this note by making the simple remark that the inverse of Theorem 2 is also true. Indeed every random utility model can be characterized with a stochastic preference, which in turn can be described with a series of probability distribution functions P_α for $\alpha \in \Omega$ such that resulting discrete choice model based on random walks coincides with the original random utility model.

Acknowledgements

The author would like to thank Federico Echenique and Andrés Abeliuk for helpful discussions. Thanks are also due to two anonymous referees for reading the draft carefully and providing relevant comments.

References

[1] Moshe Ben-Akiva, Michel Bierlaire, Discrete choice methods and their applications to short term travel decisions, in: Handbook of Transportation Science, Springer, 1999, pp. 5–33.
[2] Gerardo Berbeglia, Gwenaël Joret, Assortment optimisation under a general discrete choice model: A tight analysis of revenue-ordered assortments. Available at: SSRN 2620165, 2015.
[3] Jose Blanchet, Guillermo Gallego, Vineet Goyal, A Markov chain approximation to choice modeling, in: ACM Conference on Electronic Commerce, 2013, pp. 103–104.
[4] Henry David Block, Jacob Marschak, et al., Random orderings and stochastic theories of responses, Contrib. Probab. Statist. 2 (1960) 97–132.
[5] Antoine Désir, Vineet Goyal, Danny Segev, Chun Ye, Capacity constrained assortment optimization under the Markov chain based choice model, Oper. Res. (2016) forthcoming.
[6] Avinash K. Dixit, Joseph E. Stiglitz, Monopolistic competition and optimum product diversity, Amer. Econ. Rev. (1977) 297–308.
[7] John R. Doyle, David J. O'Connor, Gareth M. Reynolds, Paul A. Bottomley, The robustness of the asymmetrically dominated effect: buying frames, phantom alternatives, and in-store purchases, Psychol. Mark. 16 (3) (1999) 225–243.
[8] Federico Echenique, Kota Saito, Gerelt Tserenjigmid, The perception-adjusted luce model. Technical report, 2013.
[9] Jacob B. Feldman, Huseyin Topaloglu, Revenue management under the Markov chain choice model, 2014.
[10] Guiyun Feng, Xiaobo Li, Zizhuo Wang, Analysis of discrete choice models: A welfare-based framework. Available at: SSRN 2574466, 2015.
[11] John R. Hauser, Testing the accuracy, usefulness, and significance of probabilistic choice models: An information-theoretic approach, Oper. Res. 26 (3) (1978) 406–421.
[12] Josef Hofbauer, William H. Sandholm, On the global convergence of stochastic fictitious play, Econometrica 70 (6) (2002) 2265–2294.
[13] Ruud H. Koning, Geert Ridder, Discrete choice and stochastic utility maximization, Econom. J. 6 (1) (2003) 1–27.
[14] Hongmin Li, Woonghee Tim Huh, Pricing multiple products with the multinomial logit and nested logit models: Concavity and implications, Manuf. Serv. Oper. Manage. 13 (4) (2011) 549–563.
[15] D. Luce, Individual Choice Behavior, John Wiley and Sons, 1965.
[16] Colin L. Mallows, Non-null ranking models. I, Biometrika (1957) 114–130.
[17] Daniel McFadden, Kenneth Train, Mixed mnl models for discrete response, J. Appl. Econometrics 15 (5) (2000) 447–470.
[18] Thomas Brendan Murphy, Donal Martin, Mixtures of distance-based models for ranking data, Comput. Statist. Data Anal. 41 (3) (2003) 645–655.
[19] Karthik Natarajan, Miao Song, Chung-Piaw Teo, Persistency model and its applications in choice modeling, Manage. Sci. 55 (3) (2009) 453–469.
[20] Roger Ratcliff, A theory of memory retrieval, Psychol. Rev. 85 (2) (1978) 59.
[21] Michael Spence, Product selection, fixed costs, and monopolistic competition, Rev. Econom. Stud. (1976) 217–235.
[22] Ryan Webb, Dynamic constraints on the distribution of stochastic choice: Drift diffusion implies random utility. Available at: SSRN 2226018, 2013.