Transportation Research Part B 43 (2009) 984–994

Contents lists available at ScienceDirect

Transportation Research Part B

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

Sampling of alternatives for route choice modeling ☆

E. Frejinger [a,*], M. Bierlaire [b], M. Ben-Akiva [c]

[a] Royal Institute of Technology, Centre for Transport Studies, Teknikringen 78B, SE-100 44 Stockholm, Sweden
[b] École Polytechnique Fédérale de Lausanne, Transport and Mobility Laboratory, Station 18, CH-1015 Lausanne, Switzerland
[c] Massachusetts Institute of Technology, Room 1-181, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

ARTICLE INFO

Article history:
Received 21 November 2007
Received in revised form 11 March 2009
Accepted 11 March 2009

Keywords:
Route choice
Choice set generation
Sampling of alternatives
Path size

ABSTRACT

This paper presents a new paradigm for choice set generation in the context of route choice model estimation. We assume that the choice sets contain all paths connecting each origin–destination pair. Although this is behaviorally questionable, we make this assumption in order to avoid bias in the econometric model. These sets are in general impossible to generate explicitly. Therefore, we propose an importance sampling approach to generate subsets of paths suitable for model estimation. Using only a subset of alternatives requires the path utilities to be corrected according to the sampling protocol in order to obtain unbiased parameter estimates. We derive such a sampling correction for the proposed algorithm.

Estimating models based on samples of alternatives is straightforward for some types of models, in particular the multinomial logit (MNL) model. In order to apply MNL for route choice, the utilities should also be corrected to account for the correlation using, for instance, a path size (PS) formulation. We argue that the PS attribute should be computed based on the full choice set. Again, this is not feasible in general, and we propose a new version of the PS attribute derived from the sampling protocol, called Expanded PS.

Numerical results based on synthetic data show that models including a sampling correction are remarkably better than the ones that do not. Moreover, the Expanded PS shows good results and outperforms models with the original PS formulation.

1. Introduction

Route choice modeling is complex for various reasons and involves several steps before the actual model estimation. We start by giving an overview of the modeling process in Fig. 1. In a real network a very large set of paths connect an origin s_o and a destination s_d. This set, referred to as the universal choice set $\mathscr{U}$, cannot be explicitly generated. In order to estimate a route choice model, a subset of paths needs to be defined and path generation algorithms are used for this purpose. There exist deterministic and stochastic approaches for generating paths.

Deterministic methods always generate the same set $\mathscr{M}$ of paths for a given origin–destination pair. Most of them are based on some form of repeated shortest path search. This type of approach is computationally appealing thanks to the efficiency of shortest path algorithms. Examples are link elimination (Azevedo et al., 1993), link penalty (de la Barra et al., 1993) and labeled paths (Ben-Akiva et al., 1984). Instead of performing repeated shortest path searches, a constrained enumeration

☆ This research is supported by the Swiss National Science Foundation Grant 200021-107777/1.

* Corresponding author. Tel.: +46 8 7906843; fax: +46 87906002.
E-mail addresses: emma.frejinger@infra.kth.se (E. Frejinger), michel.bierlaire@epfl.ch (M. Bierlaire), mba@mit.edu (M. Ben-Akiva).

0191-2615/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.trb.2009.03.001

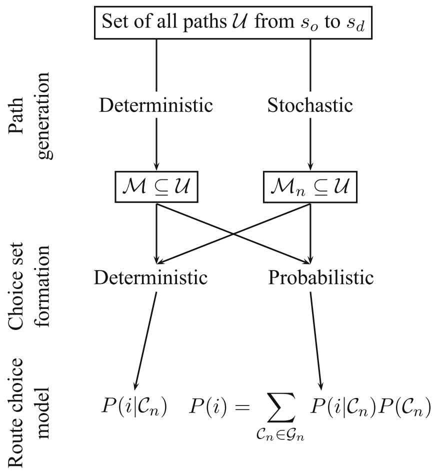

Fig. 1. Choice set generation overview.

approach referred to as branch-and-bound has recently been proposed. Friedrich et al. (2001) present an algorithm for public transport networks, Hoogendoorn-Lanser (2005) for multi-modal networks and Prato and Bekhor (2006) for route networks.

Stochastic methods generate an individual (or observation) specific subset $\mathscr{M}_n$. Actually, most of the deterministic approaches can be made stochastic by using random generalized cost for the shortest path computations. Ramming (2001) proposes a simulation method that produces alternative paths by drawing link costs from different probability distributions. The shortest path according to the randomly distributed generalized cost is calculated and introduced in the choice set. Recently, Bovy and Fiorenzo-Catalano (2006) proposed the doubly stochastic choice set generation approach. It is similar to the simulation method but the generalized cost functions are specified like utilities and both the parameters and the attributes are stochastic. They also propose to use a filtering process such that, among the generated paths, only those satisfying some constraints are kept in the choice set.

Once $\mathscr{M}$ (or $\mathscr{M}_n$) has been generated, a choice set $\mathscr{C}_n$ can be defined in either a deterministic way by including all feasible paths, $\mathscr{C}_n = \mathscr{M}$ (or $\mathscr{C}_n = \mathscr{M}_n$), or by using a probabilistic model $P(\mathscr{C}_n)$ where all non-empty subsets $\mathscr{G}_n$ of $\mathscr{M}$ (or $\mathscr{M}_n$) are considered. Defining choice sets in a probabilistic way is complex due to the size of $\mathscr{G}_n$ and has never been used in a real size application. See Manski (1977), Swait and Ben-Akiva (1987), Ben-Akiva and Boccara (1995) and Morikawa (1996) for more details on probabilistic choice set models. Cascetta and Papola (2001) and Cascetta et al. (2002) propose to simplify the complex probabilistic choice set models by viewing the choice set as a fuzzy set in a implicit availability/perception of alternatives model.

Several route choice models $P(i|\mathscr{C}_n)$ exist in the literature. Multinomial logit based models; C-logit (Cascetta et al., 1996) and path size logit (Ben-Akiva and Ramming, 1998; Ben-Akiva and Bierlaire, 1999) are the most frequently used models in practice due to their simple structure. In these models, the utilities are deterministically corrected with an attribute that accounts for correlation. More complex models explicitly capturing the correlation among paths have been proposed in the literature. The link-nested logit (Vovsha and Bekhor, 1998) model has a cross-nested logit structure but is difficult to estimate because of the large number of nesting parameters. Error Component (Bekhor et al., 2002; Frejinger and Bierlaire, 2007) and multinomial probit (Yai et al., 1997) models have also been proposed which require simulated maximum likelihood estimation.

The formal evaluation of the relevance and realism of generated choice sets is difficult in practice since the actual choice sets in general are unknown to the modeler. Several researchers, including Ramming (2001), Hoogendoorn-Lanser (2005), Bekhor et al. (2006), Bovy and Fiorenzo-Catalano (2006), Prato and Bekhor (2006, 2007), Van Nes et al. (2006), Bovy (2007) and Fiorenzo-Catalano (2007), have proposed various measures of quality of the generated sets. Empirical analysis shows that no choice set generation algorithm is able to fully reproduce observed paths. Namely, Ramming (2001) finds at best 91% of the observations by combining various algorithms and Prato and Bekhor (2006) find 91% of the observations using their branch-and-bound algorithm.

In the context of our new paradigm based on sampling from the universal choice set, these measures do not apply as all possible paths belong to the choice set. Moreover, the observed path is always in the sample by design. Unlike existing choice set generation approaches which aim at generating actual choice sets, we focus on obtaining unbiased parameter estimates using samples of alternatives. The validation of our approach consists in comparing estimated parameters with their true values.

In the following section we give an introduction to sampling of alternatives. We derive the sampling correction in Section 3 and continue by describing the proposed algorithm in Section 4. The expanded path size attribute is presented in Section 5 and numerical results based on synthetic data in Section 6. Finally we give conclusions and discuss issues for future research.