Transportation Research Part B 60 (2014) 1–15

Contents lists available at ScienceDirect

Transportation Research Part B

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

Fast algorithms to generate individualized designs for the mixed logit choice model

Marjolein Crabbe [a,*], Deniz Akinc [a], Martina Vandebroek [a,b]

[a] Faculty of Economics and Business, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium
[b] Leuven Statistics Research Center, KU Leuven, Celestijnenlaan 200B, 3001 Leuven, Belgium

ARTICLE INFO

Article history:
Received 29 October 2012
Received in revised form 16 November 2013
Accepted 21 November 2013

Keywords:
Discrete choice
Mixed logit
Individualized design
$\mathcal{D}$-efficiency
Kullback–Leibler information

ABSTRACT

The mixed logit choice model has become the common standard to analyze transport behavior. Moreover, more and more transport studies start to make use of stated preference data to obtain precise knowledge on travelers' preferences. Accounting for the individual-specific coefficients in the mixed logit choice model, this research advocates an individualized design approach to generate these stated choice experiments. Individualized designs are sequentially generated for each person separately, using the answers from previous choice sets to select the next best set in a survey. In this way they are adapted to the specific preferences of an individual and therefore more efficient than an aggregate design. In order for individual sequential designs to be practicable, the speed of designing an additional choice set in an experiment is obviously a key issue. This paper introduces three design criteria used in optimal test design, based on Kullback–Leibler information, and compares them with the well known $\mathcal{D}$-efficiency criterion to obtain individually adapted choice designs for the mixed logit choice model. Being equally efficient to $\mathcal{D}$-efficiency and at the same time much faster, the Kullback–Leibler criteria are well suited for the design of individualized choice experiments.

1. Introduction

Discrete choice is a popular and widely used methodology to study preferences in transportation (Axhausen et al., 2008; Bhat, 2012; Hess and Hensher, 2010; Hess et al., 2008; Rose and Bliemer, 2009). Although revealed preference is still most applied, more and more transport studies start to make use of stated preference data to analyze the transport behavior. In revealed preference adequate information concerning the attribute levels of the real-life choice options faced might be missing, which complicates the analysis of the choice behavior but which can be avoided by shifting to the use of stated preference choice experiments (Hess et al., 2007). It is then however beneficial to efficiently design these experiments to obtain precise estimates for the coefficients in the choice models. By selecting those choice sets that are most informative on the choice behavior, a higher level of estimation accuracy can be achieved for a given sample size, reducing the cost of the empirical study. In most researches, a single (or aggregate) design is used, which is equal for all respondents in the choice experiment. This study however continues on the recent developments in efficient individualized discrete choice design (Bliemer and Rose, 2010b (for the conditional logit choice model); Toubia et al., 2004; Yu et al., 2011 (for the mixed logit choice model)).

In transportation research, the mixed (or random coefficients) logit choice model has been used since the nineties (Ben-Akiva et al., 1993; Bhat, 1998; Revelt and Train, 1998) and it is still popular to analyze travelers' preferences (Bliemer and Rose, 2010a; Greene et al., 2006; Hess and Hensher, 2010; Hess and Train, 2011). This model extends and improves the conditional logit choice model in which parameters are only estimated at an aggregate level. The strength of the mixed logit

* Corresponding author. Tel.: +32 16 326961; fax: +32 16 326624.
E-mail address: marjolein.crabbe@kuleuven.be (M. Crabbe).

0191-2615/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.trb.2013.11.008

choice model is that, in addition to explaining preference heterogeneity with covariates for instance, the model is able to take unexplained heterogeneity into account. The mixed logit choice model assumes individual-specific coefficients following a heterogeneity distribution in the population and as such mirrors real choice behavior better. Moreover, as the individual coefficients are assumed constant throughout the choice experiment, the model accounts for the correlation between the respondents' successive choices.

However, computing aggregate efficient designs for the mixed logit choice model is a lot more complicated. Sándor and Wedel (2002) and Yu et al. (2009) obtained respectively locally and Bayesian aggregate $\mathcal{D}$-efficient designs for the cross-sectional mixed logit choice model which neglects the panel structure of the data. Bliemer and Rose (2010a) were the first to construct aggregate $\mathcal{D}$-efficient designs for the model considering the correlations between the individuals' choices, but only succeeded in obtaining locally efficient designs assuming specific prior values for the coefficients. Generating Bayesian aggregate designs for the panel version of the mixed logit choice model, taking uncertainty about the model parameters into account, appeared infeasible in a reasonable amount of time. To circumvent the computational burden, Yu et al. (2011) introduced individualized Bayesian $\mathcal{D}$-efficient designs to elicit choice data for the mixed logit choice model. Note however that this alternative design approach is not only sensible because of technical boundaries. As the mixed logit choice model assumes individual-specific preferences and therefore individual-specific parameters, it is more in line with the underlying model assumptions to design individually adapted choice experiments instead of an aggregate design. Individualized choice designs are sequentially generated for each person separately by summarizing the answers to previous choice sets as prior information to efficiently select the next best set. By taking previous choices into account in the design process, the designs are tailored to the specific preferences of an individual. Therefore, individualized designs yield higher quality choice data and yield more efficient estimates for the mixed logit choice model than an aggregate design optimized for a simpler model (Danthurebandara et al., 2011; Yu et al., 2011).

As one cannot let respondents wait for minutes, even seconds, obviously, sequential designs are only practicable if each additional set in the choice experiment is generated sufficiently fast. Despite the increasing computational capacity of modern computers, it thus remains necessary to search for methods that reduce the computation time of the design procedure. In this line, this research explores new design criteria that have been used in optimal test design to construct individually adapted choice designs for the mixed logit choice model and compares them with the $\mathcal{D}$-efficiency criterion that has often been used is this context.

In item response (or test) design, the individualized design approach has been generally accepted and successfully applied for years. It has become common practice to customize tests to the aptitude of a specific individual by incorporating a test taker's answers from previous test items to select the next best item in the test. Items too hard or too easy, adding hardly any information about an individual's ability, are in this way discarded from his/her test. Many of the test studies also apply $\mathcal{D}$-efficiency as optimality criterion, which is feasible in this context because of the simpler models involved. Recently however, three novel item selection rules, based on Kullback–Leibler information, have been introduced in the test design literature. The first maximizes the expected Kullback–Leibler divergence between subsequent posteriors of the individual-specific coefficients (Mulder and van der Linden, 2010; Wang and Chang, 2011). The other two criteria are derived from respectively mutual information (Mulder and van der Linden, 2010; Wang and Chang, 2011; Weissman, 2007) and entropy (Cheng, 2009; Wang and Chang, 2011), but are in essence also Kullback–Leibler distances.

Kullback–Leibler divergence was originally introduced in adaptive test design by Chang and Ying (1996) in a search for more global design criteria, as at that point Fisher information was only applied as a local criterion by constructing efficient designs at intermediate estimates of the model coefficients. In the beginning of a test, with few data available, this could however be problematic. The interest in Kullback–Leibler divergence as design criterion continued to grow due to its easy generalization to a multidimensional setting and its straightforward interpretation as a distance measure (Mulder and van der Linden, 2010). The main goal of using Kullback–Leibler design criteria remained however the same: generating designs in an efficient way to estimate model coefficients as accurate as possible with a given amount of data.

For individualized test design, the new criteria have been shown to be very useful. Moreover, also in fields completely different to test theory these criteria appear to have great potential compared to traditional design approaches. Some examples are paired comparison designs for tournament scheduling (Glickman and Jensen, 2005), space-filling designs for computer experiments (Jourdan and Franco, 2010) and plasma diagnostics (Dreier et al., 2006). Encouraged by the positive results from the test design studies, we apply the Kullback–Leibler criteria to design individualized choice experiments. Their implementation in a discrete choice setting is shown to be efficient and very fast.

The remainder of this paper is organized as follows. The following section discusses the mixed logit choice model and the individualized design algorithms either employing the $\mathcal{D}$-error or the Kullback–Leibler information. Section 3 comprises an extensive simulation study comparing the efficiency and practicality of the design criteria. A final section closes the study with some conclusions.

2. Methodology

2.1. The mixed logit choice model

In a discrete choice experiment respondents must choose their preferred travel option in a series of choice sets contrasting multiple alternatives. Each alternative or profile in a set is characterized by a number of attributes, like for instance the