

Extending R package idfix

Research proposal

Daniel Gerardo GIL SANCHEZ

Supervisor: Prof. Martina Vandebroek

Academic year 2018-2019

Contents

Contents	i
1 Introduction	1
2 Multinomial logit model	2
3 Individually Adaptive Sequential Bayesian for Mixed Multinomial Logit	3
4 Modified Fedorov Algorithm	3
5 Kullback-Leibler criterion	4
6 Coordinate Exchange Algorithm	5
7 Research objective	5
References	6

1 Introduction

Discrete choice models are widely used in applied sciences such as transportation, marketing, medicine, among others. These kind of models are characterized by a discrete choice response, that can be nominal or ordinal. To collect information of this nature, fundamental research has focused on the development of the theory of discrete choice experiments, where the respondents are asked to select an alternative, composed by different combinations of attribute levels, from a series of choice sets.

As in experimental design for linear models, in discrete choice experiments the number of possible profiles or treatments explodes as the number of attributes and levels increase. So the use of optimal designs, also known as efficient designs, is widely used in several scenarios because it allows to give reliable conclusions without the use of factorial designs. Consequently, there are different optimality criteria to choose a final design, such as D, A, G and V optimal, each using different methodologies (see Kessels et al., 2006). Since the focus of the majority of applications is to obtain precise estimates from a specified model, the D-optimal criterion is commonly used.

In regard to software implementation, there were no available packages to generate discrete choice designs in R. What people used to do, and still do, is to generate an optimal design for a linear model and use it to collect the data and analyze it with a discrete choice model. This approach can be improved in the sense that instead of using the information matrix of a linear model (usually OLS) to maximize its determinant, the information matrix of non-linear discrete choice models can be used, such as the multinomial logit.

The *idefix* package, already available in CRAN, generates D-optimal (or DB-optimal) designs for discrete choice experiments based on the Multinomial Logit model assuming homogeneity in the respondents and implements individually adapted designs for the mixed multinomial model when the assumption of homogeneity does not hold. The current state of this package generates optimal designs using the Modified Fedorov Algorithm and sequential modified Fedorov algorithm for the Multinomial Logit and Mixed Multinomial Logit model, respectively. This algorithm is computationally expensive and makes difficult its application in real life situations. For this reason, the purpose of this study is to implement other algorithms to create optimal designs and make a simulation study to compare results from different approaches.

In the following, an introduction to the Multinomial Logit model is given, continued by an explanation of the Individually Adaptive Sequential Bayesian (IASB) approach to generate individual designs for each respondent. Then, the modified Fedorov algorithm is presented, as well as the Kullback-Leibler criterion and Coordinate Exchange Algorithm, which are part of the research proposal. Finally, a description of the research objective is presented.

2 Multinomial logit model

This model is derived from the random utility model, where it is assumed that a respondent makes its final choice by measuring the utility of each of the alternatives presented (see Train, 2009, chap. 3). This utility can be expressed as:

$$U_j = \mathbf{x}_j^\top \boldsymbol{\beta} + \epsilon_j \quad (1)$$

Where j represents the different alternatives shown within a choice set, \mathbf{x}_j is the corresponding row of the j -th alternative/profile in the design matrix \mathbf{X} , $\boldsymbol{\beta}$ denotes the importance of each attribute level and ϵ_j is an i.i.d extreme value error term. Now, the multinomial logit probability that a respondent chooses the j -th profile in the s -th choice set is:

$$p_{js}(\boldsymbol{\beta}) = \frac{\exp(\mathbf{x}_{js}^\top \boldsymbol{\beta})}{\sum_{i=1}^J \exp(\mathbf{x}_{is}^\top \boldsymbol{\beta})} \quad (2)$$

Because the errors are assumed to be independent, the choices of N respondents represent independent draws from a multinomial distribution. Therefore, the log-likelihood can be written as:

$$LL(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta}) = \sum_{s=1}^S \sum_{j=1}^J \sum_{n=1}^N y_{jsn} \ln(p_{js}(\boldsymbol{\beta})) \quad (3)$$

Where \mathbf{Y} is the matrix of choices from N respondents with elements y_{jsn} . Since the *idefix* package chooses the best design based on the D-optimal criterion, i.e., the choice that minimizes the determinant of the variance-covariance matrix of the parameter estimators or conversely maximizes the determinant of the Information Matrix, it is of special interest the expression of the Fisher Information matrix:

$$\mathbf{M}(\mathbf{X}, \boldsymbol{\beta}) = N \sum_{s=1}^S \mathbf{X}_s^\top (\mathbf{P}_s - \mathbf{p}_s \mathbf{p}_s^\top) \mathbf{X}_s \quad (4)$$

Where $\mathbf{p}_s = [p_{1s}, \dots, p_{Js}]$ and $\mathbf{P}_s = \text{diag}[p_{1s}, \dots, p_{Js}]$. It is important to notice that the information matrix depends on the unknown parameters through the probabilities, so initial parameter values are required before computing this matrix. Kessels et al. (2006) show that the Bayesian design approach works well in this scenario, specifying a prior normal distribution for $\boldsymbol{\beta}$, $N(\boldsymbol{\beta}|\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0)$.

As a consequence of the Bayesian approach, a modification in the calculation of the D-optimal criterion has to be done, including the prior distribution of the coefficients. The Bayesian D-optimal criterion, or D_B criterion, is defined as:

$$D_B = \int_{\mathcal{R}^k} \{ \det(\mathbf{M}^{-1}(\mathbf{X}, \boldsymbol{\beta})) \}^{1/k} \pi(\boldsymbol{\beta}) d\boldsymbol{\beta} \quad (5)$$

Where k is the number of unknown parameters in the model, i.e., the number of columns in the design matrix \mathbf{X} ; and $\pi(\boldsymbol{\beta})$ is the prior distribution of $\boldsymbol{\beta}$.

In the *idefix* package, these D_B optimal designs are found using the Modified Fedorov Algorithm.

3 Individually Adaptive Sequential Bayesian for Mixed Multinomial Logit

This approach, proposed by Yu et al. (2011), is useful when the assumption of homogeneity between respondents does not hold. In this scenario, an extension of the Multinomial Logit model is considered, called the Mixed Multinomial Logit model (MIXL) (see Train, 2009, chap. 6). It considers that when the respondents are heterogeneous, each of them has its own preference structure. This structure is assumed to behave according to a Multinomial Logit Model. In this sense, the MIXL model can be seen as a combination of different Multinomial Logit models, one for each respondent. As a consequence, the Individually Adaptive Sequential Bayesian generates an individual optimal design for each of the respondents involved in the study, instead of applying the same design for everyone.

According to Yu et al. (2011), this procedure consists of two stages: an initial static stage, where the prior distribution of β is the same for all respondents and it is used to generate an initial design. And the fully adaptive sequential stage, where the prior information for the design is sequentially updated after each choice made by a respondent. As a result, each respondent will have different choice sets.

In the *idexif* package, the D_B criterion is used in the search of individual optimal designs through the Modified Fedorov Algorithm.

4 Modified Fedorov Algorithm

Proposed by Cook et al. (1980), it begins with a random initial design, with its corresponding optimality criterion. The algorithm exchanges an entire row/profile from the initial design matrix with each of the rows from a candidate set of profiles; usually, this candidate set is just the list of all possible combinations of attribute levels. As a result, there will be different modifications of the initial design to which the optimality criterion is computed. Then, the initial design matrix is updated by replacing the existing profile with the profile from the candidate set that improves the optimality criterion. This process is repeated for each of the rows/profiles in the initial design and stops when no exchanges are performed.

This algorithm is computationally expensive because in each iteration, the information matrix and its determinant have to be calculated for each exchange in the initial design. An easy illustration of how time consuming it can be, suppose that the initial design has 12 profiles, corresponding to 6 choice sets with 2 alternatives each. The candidate set has, say, 32 profiles. Then, the algorithm has to calculate $12 * 32 = 384$ information matrices and their respective determinants to decide which is the best optimal design.

Furthermore, on top of this number of calculations, remember that the information matrix needs prior values for β . Assuming just 10 draws from the normal prior distribution, then the number of information matrices and determinants to compute will be 3840.

And taking into account that the optimal design depends on the random initial design, it is desirable to have as many initial designs as possible to avoid local optima. So let's assume that 100 initial designs are enough. In this sense, the final number of calculations of information matrices and determinants to obtain the best possible optimal design is 3,840,000.

For this reason, different algorithms and criteria have been proposed that are not as computationally expensive as this one. In the following the Kullback-Leibler criterion is presented as well as the Coordinate Exchange Algorithm.

5 Kullback-Leibler criterion

In the context of individual designs for each of the respondents in a study (MIXL model), Crabbe et al. (2014) present the Kullback-Leibler criterion to obtain designs that estimate model coefficients as precise as possible. They claim that designs generated with this approach are equally efficient to those generated by the D_B criterion and at the same time are much faster to calculate. It consists in the Kullback-Leibler distance (KL), which is a measure to compare two density functions. The KL increases when these densities become more divergent. This distance is defined as:

$$KL(f, g) = E_f \left[\log \frac{f(x)}{g(x)} \right] \quad (6)$$

where f and g are densities of a continuous variable X and E_f is the expected value with respect to f .

Now, the selection of the best next choice set for a specific respondent is based on the maximization of the KL distance between the current posterior of the coefficients and the updated that one will obtain with the additional response information from the next choice (Crabbe et al., 2014). Since in a choice set there are multiple possible alternatives, the expectation over all possible alternatives is maximized. This expression is as follows:

$$KLP = \sum_{j=1}^J \pi(y_{jsn} | \mathbf{y}_n^{s-1}) KL [f(\boldsymbol{\beta})_n | \mathbf{y}_n^{s-1}, f(\boldsymbol{\beta})_n | \mathbf{y}_n^{s-1}, y_{jsn}] \quad (7)$$

Where s is the next choice set, n is a particular respondent and j is the chosen alternative. The densities $f(\boldsymbol{\beta})_n | \mathbf{y}_n^{s-1}$ and $f(\boldsymbol{\beta})_n | \mathbf{y}_n^{s-1}, y_{jsn}$ are the updated posteriors and $\pi(y_{jsn} | \mathbf{y}_n^{s-1})$ is the posterior weighted choice probabilities for the alternatives in the choice set s , given the previous responses.

According to Crabbe et al. (2014), since the KLP only involves the calculation of the posterior weighted choice probabilities, it is much faster than the D-optimal criterion, where not only the next choice set, but also the previous sets in the design are used in the calculation of the information matrix.

In the *idefix* package, this algorithm has already been implemented; however, simulations in R are not consistent with the results presented by Crabbe et al. (2014).

6 Coordinate Exchange Algorithm

Proposed by Meyer and Nachtsheim (1995) as an alternative to profile exchange algorithms, where the whole row of the design matrix is exchanged by a profile from a candidate set (such as Modified Fedorov algorithm). This algorithm exchange one attribute at the time instead of the full profile from the design matrix, resulting in coordinate exchanges in the initial design as its name suggests¹.

There are some benefits in the search for an optimal design with this algorithm: First, there is no need to give a list of candidate sets because it does not use any candidate profiles to improve the initial design. Second, discrete and continuous design spaces are easily handled because for continuous attributes only one coordinate/column is changed and for discrete attributes, changing one level results in a simultaneous change of all coordinates/columns related to the same attribute. Third and maybe the most important, a reduction in computing time of one or two orders of magnitude in large problems is found when compared with profile exchange algorithms (Meyer and Nachtsheim, 1995).

The algorithm begins with a random initial design and its corresponding optimality criterion, such as D_B or KL criterion. Then, for every profile/row in the initial design, each attribute level is exchanged with all possible levels in the attribute and the initial design is updated by keeping the level that presents the best value of the optimality criterion. Having done this in each element of the design matrix, the process is repeated again until no elements have changed or until the algorithm reaches a prespecified number of iterations. Since the algorithm starts with a random design, it is advisable to have several initial designs to avoid local optima.

7 Research objective

Currently, the *idefix* package uses the Modified Fedorov algorithm to create optimal designs for the Multinomial Logit model and Mixed Multinomial Logit model. As it has been shown above, there are different algorithms and optimality criteria that can be implemented to improve the current processing time to generate optimal designs. The purpose of this study can be listed as follows:

- Optimize the processing time of the Modified Fedorov algorithm by implementing some parts of the algorithm in C++.
- Implement the KL criterion and compare it with the function that is already available in the package.
- Implement the Coordinate Exchange algorithm to create optimal designs.

¹Note that when an attribute is continuous, there is just one coordinate or column in the design matrix. When an attribute is discrete with a levels, there are $a - 1$ coordinates/columns in the design matrix.

- Make a simulation study to compare processing times and optimality of designs between the Modified Fedorov algorithm, the Coordinate Exchange algorithm and the use of D_B and KL criteria.
- Reorganize some functions inside the package, remove possible redundancy in code and implement parts of the code in C++.
- If there is enough time, implement an algorithm to generate optimal design with partial profiles. This approach is useful when the number of attributes or levels is large.

References

- Cook, R. D., Nachtsheim, C. J., and Cook, R. D. (1980). A Comparison of Algorithms for Constructing Exact D-Optimal Designs Linked references are available on JSTOR for this article : A Comparison of Algorithms for Constructing Exact D-Optimal Designs. *American Society for Quality*, 22(3):315–324.
- Crabbe, M., Akinc, D., and Vandebroek, M. (2014). Fast algorithms to generate individualized designs for the mixed logit choice model. *Transportation Research Part B: Methodological*, 60:1–15.
- Kessels, R., Goos, P., and Vandebroek, M. (2006). A comparison of criteria to design efficient choice experiments. *Journal of Marketing Research*, 43(3):409–419.
- Meyer, R. K. and Nachtsheim, C. J. (1995). The Coordinate-Exchange Algorithm for Constructing Exact Optimal Experimental Designs. *Technometrics*, 37(1):60–69.
- Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge.
- Yu, J., Goos, P., and Vandebroek, M. (2011). Individually adapted sequential Bayesian conjoint-choice designs in the presence of consumer heterogeneity. *International Journal of Research in Marketing*, 28(4):378–388.