Transportation Research Part B 115 (2018) 17–40

Contents lists available at ScienceDirect

Transportation Research Part B

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

Minorization-Maximization (MM) algorithms for semiparametric logit models: Bottlenecks, extensions, and comparisons

Prateek Bansal[a,*], Ricardo A Daziano[a], Erick Guerra[b]

[a] School of Civil and Environmental Engineering, Cornell University, USA
[b] City and Regional Planning, University of Pennsylvania, USA

ARTICLE INFO

Article history:
Received 27 March 2018
Revised 22 June 2018
Accepted 27 June 2018

JEL classification:
C13
C25
Q42

Keywords:
Discrete choice
Semiparametrics
Preference heterogeneity
Expectation-maximization
Minorization-maximization

ABSTRACT

Motivated by the promising performance of alternative estimation methods for mixed logit models, in this paper we derive, implement, and test minorization-maximization (MM) algorithms to estimate the semiparametric logit-mixed logit (LML) and mixture-of-normals multinomial logit (MON-MNL) models. In particular, we show that the reported computational efficiency of the MM algorithm is actually lost for large choice sets. Because the logit link that represents the parameter space in LML is intrinsically treated as a large choice set, the MM algorithm for LML actually becomes unfeasible to use in practice. We thus propose a faster MM algorithm that revisits a simple step-size correction. In a Monte Carlo study, we compare the maximum simulated likelihood estimator (MSLE) with the algorithms that we derive to estimate LML and MON-MNL models. Whereas in LML estimation alternative algorithms are computationally uncompetitive with MSLE, the faster-MM algorithm appears emulous in MON-MNL estimation. Both algorithms – faster-MM and MSLE – could recover parameters as well as standard errors at a similar precision in both models. We further show that parallel computation could reduce estimation time of faster-MM by 45% to 80%. Even though faster-MM could not surpass MSLE with analytical gradient (because MSLE also leveraged similar computational gains), parallel faster-MM is a competitive replacement to MSLE for MON-MNL that obviates computation of complex analytical gradients, which is a very attractive feature to integrate it into a flexible estimation software. We also compare different algorithms in an empirical application to estimate consumer's willingness to adopt electric motorcycles in Solo, Indonesia. The results of the empirical application are consistent with those of the Monte Carlo study.

1. Introduction

1.1. Background

With the increase in computation power during the last decade, the mixed multinomial logit (MMNL) model – a random parameter logit model with parametric and continuous heterogeneity distributions – is the most commonly used flexible

* Corresponding author.
E-mail address: pb422@cornell.edu (P. Bansal).

https://doi.org/10.1016/j.trb.2018.06.010

discrete-choice specification (Train, 2009; McFadden and Train, 2000). The problem of correctly specifying the heterogeneity (or mixing) distribution of the random parameters has received great attention (Hensher and Greene, 2003); however, there is still no consensus among researchers: restricting the shape of the mixing distributions can result into wrong signs and overestimation of welfare measures. Wrong (welfare) estimates can misguide policy and marketing decisions (Fosgerau, 2006; Cherchi and Polak, 2005). To overcome the problem of presuming the shape of the mixing distribution, differing specifications with semi- or nonparametric mixing distributions have been proposed. Vij and Krueger (2017) and Bhat and Lavieri (2017) provide a detailed review of advancements in parametric and semiparametric mixing distributions under extreme-value-distributed (logit kernel) and normally-distributed (probit kernel) error structures. In general, estimation of these flexible[1] models is complex and computationally expensive.

This study focuses on estimation of two state-of-the-art semiparametric logit models, namely the logit-mixed-logit (LML) and mixture-of-normals multinomial logit (MON-MNL) models, especially in the context of the promising performance of an alternative iterative optimization method with minimal coding, the minorization-maximization algorithm that will be introduced below, as reported for mixed logit (James, 2017).

The logit-mixed-logit model (Train, 2016) generalizes many previous semiparametric models including Bajari et al. (2007), Fosgerau and Bierlaire (2007), Train (2008), and Fox et al. (2011) (cf. Bhat, 1997). In LML, a finite parameter space is divided into a discrete multidimensional grid (cf. Train, 2008). Whereas Train (2008) considers the probability mass at each discrete point as a parameter of interest, LML reduces the number of parameters by specifying this probability using a logit link. In Monte Carlo studies, Bansal et al. (2018a) and Franceschinis et al. (2017) successfully tested flexibility of LML as the model could retrieve a series of continuous parametric mixing distributions (bi-modal, tri-modal, lognormal, and uniform) much better than parametric counterparts. The maximum simulated likelihood estimator (MSLE) of LML is much faster than that of parametric models, but computation of standard errors requires bootstrapping. Furthermore, the computational efficiency of point estimation is lost by a factor of 15 to 30 when fixed parameters are introduced (Bansal et al., 2018b), and computational efficiency becomes much worse when standard errors are derived.

The mixture-of-normals multinomial logit also offers a flexible representation of unobserved preference heterogeneity. The premise of MON-MNL[2] is that any continuous distribution can be approximated to a given degree of accuracy by a discrete mixture of normals (Ferguson, 1973). Prespecifying the number of mixture components (or classes[3]) imposes a heterogeneity structure, but unlike LML there is no need of predefining the parameter space. Resource-intensive bootstrapping to compute standard errors is not needed in MON-MNL either. In a Monte Carlo study, Fosgerau and Hess (2009) found that MON-MNL outperformed parametric specifications in all scenarios, ranging from retrieving the most trivial uniform distribution to the most complex multimodal distribution. Keane and Wasi (2013) further supported the superiority of MON-MNL in an extensive study of 10 stated preference datasets. However, only a handful of empirical studies have used MON-MNL, possibly due to the complexity of the analytical gradient of the loglikelihood and covergence problems when using numerical gradients. For instance, Fosgerau and Hess (2009) pointed out that MSLE led into troubles for more than 2 normal components in the mixing distribution. Whereas Keane and Wasi (2013) did not explicitly mention any such estimation problem, the authors set bounds on some parameters and also imposed hard constraints on the variance-covariance matrix of each component of the mixture.

Among frequentist methods to estimate logit models, researches have explored iterative optimation methods. Within this class of methods, the expectation-maximization (EM) algorithm has been reported (Bhat, 1997; Cherchi and Guevara, 2012; Sohn, 2017) to outperform MSLE in numerical stability (i.e., less sensitivity to initial values), empirical identification (i.e., avoiding an invertible Hessian matrix), and estimation simplicity. Whereas MSLE directly maximizes the loglikelihood function using quasi-Newton methods, the simplicity of EM stems from iteratively maximizing a simpler surrogate function and update parameters while maintaining monotonic improvements in the loglikelihood (Dempster et al., 1977; McLachlan and Krishnan, 2007). Furthermore, iterative parameter updates of the EM algorithm are either closed-form or straightforward econometric problems that can be solved using standard statistical packages (Train, 2008; Sohn, 2017). EM also provides a convenient parameterization of the complete-data likelihood function without worrying about over-identification (Ruud, 1991). In addition to these nice statistical properties, EM also converges quickly to the neighborhood of the optima. However, EM is plagued by slower convergence within the optimum neighborhood (Dempster et al., 1977). In fact, the computational performance of EM largely hinges upon the underlying data generating process and how well EM re-characterizes the objective function. More specifically, if the complete data model provides much more information about the parameter than the incomplete data model, then the EM algorithm is generally slow (Meilijson, 1989). Ruud (1991) suggested designing hybrid algorithms such that EM starts the maximization process and a Newton-type algorithm finishes it. In fact, Bhat (1997) could achieve computational efficiency and numerical stability in latent class logit estimation by shifting from EM to quasi-Newton methods when the difference in the loglikelihood of successive iterations achieved a given precision.

For some model specifications EM does not provide closed-form updates (the source of the EM benefits) for all parameters, making EM a rather slow method for estimation. For this reason, researchers have been exploring other alternative estimation methods. EM is actually nested in the minorization-maximization (MM) family of iterative optimization methods

[1] Model flexibility understood as the capacity to represent unobserved preference heterogeneity.

[2] MON-MNL was labeled *Mixed-Mixed Logit* by Keane and Wasi (2013) and Latent Class Mixed Multinomial Logit model by Greene and Hensher (2013).

[3] The mixture components can also be interpreted as classes as in a latent class logit model.