Intern. J. of Research in Marketing 31 (2014) 35–48

Contents lists available at ScienceDirect

Intern. J. of Research in Marketing

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

A simple method for estimating preference parameters for individuals

Bart D. Frischknecht [a,*], Christine Eckert [a,b], John Geweke [a,c], Jordan J. Louviere [a,b]

[a] Centre for the Study of Choice, University of Technology, Sydney, Australia
[b] Marketing Discipline Group, University of Technology, Sydney, Australia
[c] Economics Discipline Group, University of Technology, Sydney, Australia

ARTICLE INFO

Article history:
First received in 9 March 2012 and was under review for 11 months
Available online 1 October 2013

Area Editor: Bruce G.S. Hardie

Keywords:
Choice modeling
Bayesian estimation
Discrete choice experiment
Conjugate prior
Individual level choice model

ABSTRACT

This paper demonstrates a method for estimating logit choice models for small sample data, including single individuals, that is computationally simpler and relies on weaker prior distributional assumptions compared to hierarchical Bayes estimation. Using Monte Carlo simulations and online discrete choice experiments, we show how this method is particularly well suited to estimating values of choice model parameters from small sample choice data, thus opening this area to the application of choice modeling. For larger sample sizes of approximately 100–200 respondents, preference distribution recovery is similar to hierarchical Bayes estimation of mixed logit models for the examples we demonstrate. We discuss three approaches for specifying the conjugate priors required for the method: specifying priors based on existing or projected market shares of products, specifying a flat prior on the choice alternatives in a discrete choice experiment, or adopting an empirical Bayes approach where the prior choice probabilities are taken to be the average choice probabilities observed in a discrete choice experiment. We show that for small sample data, the relative weighting of the prior during estimation is an important consideration, and we present an automated method for selecting the weight based on a predictive scoring rule.

1. Introduction

Populations of individuals are heterogeneous in their choices. There are many ways to model heterogeneity in quantitative models of human choice. Estimating a different choice model for each person is a conceptually attractive approach, but it is often thwarted by data separation. The investigator is then left with the more demanding procedure of pooling the data across individuals and setting up a formal distribution of preferences in a population. In the common paradigm of heterogeneous choice model estimation, the investigator specifies a prior over the preference space (Allenby, Arora, & Ginter, 1995; Rossi, Allenby, & McCulloch, 2005). The specification of a prior over the preference space requires the specification of the distributional form of the prior as well as parameter values for the distribution. The approach proposed in this article overcomes the issue of data separation without requiring a prior distribution on the preference parameters or pooling of the data across respondents. The method is based on maximum likelihood estimation of a logit model and thus provides a simple way to estimate a choice model for a single individual. Applied over a sample of individuals, the proposed method shows a similar performance in recovering the sample distribution of preferences and reduced computational complexity compared to hierarchical Bayes (HB) estimation.

1.1. Data separation

A critical drawback in estimating a different choice model for each person by maximum likelihood is that a single individual's data often exhibit data separation, whereby the responses of the individual can be perfectly classified by a linear combination of the covariates described in Albert and Anderson (1984), and updated by Santner and Duffy (1986). Complete separation occurs when a combination of explanatory variables classifies responses without error according to a strict inequality. Quasi-complete separation occurs when a combination of explanatory variables classifies responses without error up to a non-strict inequality. All other cases are considered to exhibit data overlap. Cases of complete or quasi-complete separation are more likely in small samples or when a particular alternative is chosen with low probability, which has been recognized in biostatistics literature with application to clinical trials (Heinze, 2006) and in econometrics (Beggs, Cardell, & Hausman, 1981) and marketing (Chapman, 1984) literature with previous efforts to estimate choice models using a small sample of data from a single individual.[1]

* Corresponding author.

[1] Encountering data separation in choice data can be interpreted in two ways. One interpretation is that the underlying choice behavior is stochastic and that multinomial logistic regression is a suitable model to describe the observed choice behavior. Here, data separation is an artifact of a relatively small number of observations. The second interpretation is that the data separation is evidence of a deterministic choice process such as lexicographic decision-making. The second interpretation would indicate that multinomial logistic regression is inappropriate for classifying the data at hand. In that case, we expect data separation to persist as the number of observations increases. In this article, we adopt the first interpretation.

0167-8116/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijresmar.2013.07.005

In the case of complete or quasi-complete data separation (Albert & Anderson, 1984), maximum likelihood estimates for multinomial logistic regression (Train, 2003), one of the most commonly applied choice models in marketing and economics, do not exist. Maximum likelihood estimation in these cases implies that the estimates of the parameters are unbounded. Data overlap alone does not guarantee a sufficiently small bias that would result in satisfactory estimates, as shown by King and Ryan (2002), who investigated a case of near separation in which data overlap existed but was relatively small.

1.2. Maximum penalized likelihood estimation

Although the complete separation problem has been encountered previously in econometrics and marketing (Beggs et al., 1981; Chapman, 1984; Savin & Wurtz, 1999), it has likely attracted limited attention because of the extensive use of data pooling across respondents, leading to large samples and data overlap rather than separation. Various means have been proposed to overcome the data separation challenge, especially for biostatistics applications, due to the small sample sizes and low incidence rates of many clinical trials (Bull, Mak, & Greenwood, 2002; Cardell, 1993; Clogg, Rubin, Schenker, Schultz, & Weidman, 1991; Firth, 1993; Heinze, 2006). The proposed approaches, including the approach used in this article, rely on the principle of shrinkage as described by Stein (1956) and James and Stein (1961).

In our case, the shrinkage is accomplished by estimation of parameters based on the maximum penalized likelihood. The penalty function that we adopt is to augment the limited data with prior beliefs about the behavior of the data (Geweke, 2005), which corresponds to a Bayesian approach designed to overcome the challenge of using finite samples. The relevance of a penalty approach to econometric or marketing choice problems as a method for capturing sample heterogeneity appears not to have been recognized by the choice modeling community (see, for example, section 1 of Allenby & Rossi, 1999) until Evgeniou, Pontil, and Toubia (2007).

To date, maximum penalized likelihood approaches can be classified as either fixed penalty methods or updated penalty methods, according to the penalty function adopted.[2] We first discuss fixed penalty methods, to which the approach used in this paper belongs. Fixed penalty methods add a carefully considered fixed set of artificial observations to the data, thereby ensuring data overlap for the extended sample. Haldane (1955), motivated by reducing parameter estimate bias in a binomial logit case, suggests a change to the likelihood formulation for an estimation that adds an artificial observation to the data for each binary outcome. Each artificial observation is given half the weight of one of the original observations in the log likelihood function. Both Clogg et al. (1991) and Cardell (1993), motivated by data separation (see also Beggs et al., 1981), propose artificially generating sets of observations (or chosen and unchosen alternatives) coupled with specific explanatory variables that are generated in a particular way. Clogg et al. (1991) illustrate their approach only for a binomial case. They consider the relative outcome frequency observed in the data and the number of estimation parameters to determine the number of artificial observations.

In general, the fixed penalty methods that add artificial observations to the data are examples of applying a conjugate prior to the data, i.e., a prior that has the same distributional form as the likelihood function. We discuss the priors employed in this paper later in the manuscript. The Cardell (1993) approach can be applied to a binomial or multinomial case and is intended to be applied to choice data rather than clinical trials or census demographics. It adds J artificial choice task observations where J is the total number of unique outcome alternatives (e.g., car, bus, train). The chosen alternative in each artificial choice task is represented by the average of the explanatory variables associated with the alternative from the choice tasks when the alternative was not chosen in the original data set. Overlap is ensured in this way by adding artificial observations that are opposite to the observed data. Because the artificial observations are composed based on the design and the choice responses for a particular alternative, the Cardell (1993) approach appears most appropriate for alternative specific choice models. Even in this case, the interpretation of the artificial observations as a conjugate prior is dependent on the specific values of the explanatory variables.

A more complex alternative to fixed penalty methods is to derive an updated penalty, which is a penalty function that is a function of the estimated model itself. Firth (1993), initially motivated by the goal of reducing parameter estimate bias, illustrates an approach for updating the penalty function at each iteration of a numerical procedure for maximizing the log likelihood function. Heinze and Schemper (2002) for binary and Bull et al. (2002) for multinomial logistic regression recognize that Firth's technique can be applied to the case of separated data and expand on his approach. Evgeniou et al. (2007) develop an updating penalty method for maximum penalized likelihood estimation and applies this method to discrete choice data. Gilbride, Lenk, and Brazell (2008) and Evgeniou et al. (2007) find that this method produces point estimates and predictions very similar to those of hierarchical Bayes estimation.

1.3. Proposed approach

For reasons we discuss in more detail below, we propose the use of an approach in the tradition of fixed penalty methods. This approach is similar to those of Clogg et al. (1991) and Cardell (1993), except Clogg et al. (1991) present their method only for a binomial case and for repeated observations of the vectors of explanatory variables, and the prior employed in Cardell's method is not interpretable for generic, i.e., unlabeled-alternative, choice models.

We present our method for a multinomial case in choice model format and for explanatory variables that vary between alternatives, such that our approach is readily applied to data collected from a single individual completing an unlabeled discrete choice experiment. This article further differs from similar methods presented previously in economics and statistics literature that focus on data sets of sufficient size in the following ways: the relative weight given to the prior during estimation does not have a large impact, and the methods present a specific prior weighting strategy without regard to its effect on the prediction performance of the estimated model. We show that for small sample data, the relative weighting of the prior during estimation is an important consideration, and we present an automated method for selecting a prior weight based on a predictive scoring rule.

We discuss three ways to formulate a conjugate prior for application to data from discrete choice experiments. First, the investigator can specify beliefs about the choice shares of the alternatives presented in the discrete choice experiment. For example, a flat prior, or probability $\pi_j = 1 / J$, for $j = 1,\ldots,J$ alternatives shrinks parameter estimates towards zero, implying that each alternative is equally likely. It thus pulls the maximum likelihood estimates away from $\pm\infty$ in the case of separated data. Similarly, an alternative to a flat prior for alternatives in the discrete choice experiment is to adopt an empirical Bayes approach (Carlin & Louis, 2000), where the prior implies the aggregate choice shares of the discrete choice experiment alternatives observed in the sample population. Finally, given a specific set of product alternatives available in the market, an investigator can specify the observed market share of each alternative or her beliefs about the relative market share of each alternative, as in the case of a new product introduction. These three approaches are illustrated in subsequent examples.

[2] Apart from fixed and updated penalty methods, exact logistic regression (Mehta & Patel, 1995) has been proposed as an alternative to maximum penalized likelihood estimation when data are separated. However, its application is limited in many practical cases because the method is computationally more intense than the maximum penalized likelihood, continuous explanatory variables are not handled well and confidence intervals can be overly conservative (Heinze, 2006). Additionally, Heinze and Schemper (2002) and Heinze (2006) compare exact logistic and penalized likelihood approaches for logistic regression with separated data and conclude that the penalty method is superior in most instances.