A New Approach to Modeling Choice

Vivek Farias • Srikanth Jagabathula • Devavrat Shah

MIT Sloan, Massachusetts Institute of Technology, Cambridge MA, USA
EECS, Massachusetts Institute of Technology, Cambridge MA, USA
EECS, Massachusetts Institute of Technology, Cambridge MA, USA
vivekf@mit.edu • jskanth@mit.edu • devavrat@mit.edu

Abstract

A central push in operations models over the last decade has been the incorporation of models of customer choice. Real world implementations of many of these models face the formidable stumbling block of simply identifying the ‘right’ model of choice to use. Thus motivated, we visit the following problem: For a ‘generic’ model of consumer choice (namely, distributions over preference lists) and a limited amount of data on how consumers actually make decisions (such as marginal preference information), how may one predict revenues from offering a particular assortment of choices? We present a framework to answer such questions and design a number of tractable algorithms from a data and computational standpoint for the same. This paper thus takes the very first step towards ‘automating’ the crucial task of selecting an appropriate choice model.
1. Introduction

How does the set of Plasma TVs displayed in response to a query on Amazon.com influence the shoppers ultimate purchase? In today’s data-driven world, customer choice models form a crucial component in strategic and tactical decision making in applications ranging from retail to inventory management in the face of substitution. The OM approach to such problems is typically:

1. Posit a parametric, structural model of demand.
2. Fit the model using historical data.
3. Optimize the relevant objective.

Parametric models have gained widespread use for the simple reason that the data available to fit a choice model is typically limited. However, such models implicitly make structural assumptions that may not hold in practice. In addition a parametric model particularly suited to a particular type of product may not be applicable across a broad swathe of product types making the approach above not easily scalable to contexts such as electronic retail.

As such, there is a natural need for non-parametric models and procedures that while imposing minimal structural restrictions are still capable of recovering models of choice with useful predictive power from limited data. This work takes a promising step towards filling this need. In particular, we propose a generic nonparametric model of customer choice that subsumes the vast majority of parametric choice models used in practice. We address fundamental questions related to its use:

1. We show that identification of a model from within our family is possible with surprisingly limited data. In doing so, we extend the theory of compressive sensing; these developments are also of independent interest in non-parametric statistics.
2. We show that when identification is not possible, we can still recover ‘worst-case’ models with effective predictive power. This entails an innovative mathematical programming formulation of model estimation and the development of approximate solution techniques for the same.
3. We carry out a computational study with Amazon.com DVD data that not only demonstrates the surprising efficacy of our approach, but illustrates the potential costs of using a choice model that makes assumptions inaccurate for the setting.

2. Model and main results

In this section, we describe the non-parametric choice model we propose and summarize our main theoretical results.

**Model:** Suppose $\mathcal{N}$ is a universe of $N$ products; for any assortment of products $\mathcal{M} \subseteq \mathcal{N}$ and any product $i \in \mathcal{M}$, a choice model specifies the probability that $i$ is purchased when $\mathcal{M}$ is offered. We model customer choice decisions as follows: each customer has a preference list of all the products\(^1\). When offered an assortment of products, the customer purchases her most preferred\(^2\) of the offered products. The customer choice model is determined by a distribution over preference lists, seen as follows. Assume that there are $K$ customer types in the population with each customer type associated with a preference list. The choice model is completely specified by the distribution that assigns to each preference list a weight equal the fraction of customers that belong to the

---

\(1\)The customer, however, need only be aware of her preferences over the offered products.

\(2\)We implicitly assume here that a ‘no-purchase’ option is available.
corresponding customer type. Now, the probability that a product \( i \in \mathcal{M} \) is purchased when offered \( \mathcal{M} \) is equal to the sum of the weights of all the preference lists that result in the purchase of \( i \) when offered \( \mathcal{M} \). Essentially any reasonable ‘parametric’ model for customer decisions can be cast in this form.

**Data:** In order to learn this choice model, we assume we have access to limited data in the form of ‘marginal’ information about the distribution over preference lists. For example, one such type of data is comparison data in which we have, for every pair of products \( i, j \), access to the probability that \( i \) (\( j \)) is purchased when offered the assortment \( \{i, j\} \). Any data available from transactional information can be cast as marginal information about our distribution over preference lists.

**Identification:** For various types of limited data (i.e. marginal information), we identify a family of choice models that can be inferred exactly. The characterization is primarily in terms of the sparsity of the true model, i.e. the number of customer types \( K \) the model assumes. The main conclusion one can draw from our characterization is that, essentially, the ‘complexity’ or sparsity of the choice model that can be fit scales with the ‘amount of information’ that is available in a precise way. To show that this characterization is useful, we show that, in terms of revenue predictions, an arbitrary choice model is well approximated by a sparse model; in other words, it is potentially reasonable to assume that many real world choice models are indeed sparse, i.e. have small \( K \). Further, we propose a simple iterative algorithm, the sparsest-fit algorithm, that either identifies the underlying model if it belongs to the family or terminates with a certificate that the underlying choice model does not belong to the family.

**Robust Model Selection:** If identification from the available limited data is not possible, we consider selecting a model from among all those consistent with the observed data that is ‘worst-case’ in the sense that it would result in minimal revenues (among all consistent choice models) for some offer set. As it turns out, this problem can be cast as a linear program with exponentially many (in \( N \)) variables. We propose a procedure that provides sequentially improving approximate solutions to this program, producing the exact solution in a finite number of steps. The efficiency of our procedure depends on the type of limited data available. For certain kinds of limited data, we prove that our procedure produces the exact solution in just one step (and is thus efficient). For other practically relevant types of data, like the comparison data described above, our empirical results show that the approximations produced after just the first step are surprisingly good.

### 3. Computational Study

In order to demonstrate the efficacy of our approach, we performed a data-driven computational study based on the sales data of 25 high-volume DVDs on Amazon during a three month period. Specifically, we fit a multinomial logit (MNL) model to the DVD sales data and treated this model as ‘ground truth’. We used this ground truth model to generate synthetic transaction data. We considered several experiments:

1. We predicted revenues for several random assortments using our robust approach (MIN) on the synthetic transaction data, and compared these predictions to ground truth (i.e. those determined by the MNL model generating the transaction data). Fig. 1(a) shows the quality of these predictions. We repeated a similar set of experiments using a similarly constructed nested MNL model to generate transaction data (see Fig. 1(b)).

2. We force-fit an MNL model to transaction data generated from the nested MNL model alluded to above. We used this force fit model to guide the selection of a revenue maximizing

---

\(^3\)Note that these probabilities don’t add up to 1 because customers also have the ‘no-purchase’ option.
Figure 1: When parametric model is a good-fit, little is lost in using our approach. Further, fitting an incorrect model can result in highly suboptimal decisions.

We draw two main conclusions from our computational study: First, in cases when a parametric model is a good fit, little is lost using our approach. Indeed, the fact that our estimates under the structurally distinct MNL and nested MNL models are very good indicates that our approach is robust to the underlying structural assumptions. Second, as illustrated by our experiments with the force-fit model, in cases when the parametric model is a poor-fit, fitting an incorrect model can lead to highly suboptimal decisions. In such scenarios, using our model can have substantial benefits.