Parameter identification in Markov chain choice models

Arushi Gupta [a,*], Daniel Hsu [b]

[a] *Princeton University, Princeton, NJ, United States of America*
[b] *Columbia University, New York, NY, United States of America*

A B S T R A C T

This work studies the parameter identification problem for the Markov chain choice model of Blanchet, Gallego, and Goyal used in assortment planning. In this model, the product selected by a customer is determined by a Markov chain over the products, where the products in the offered assortment are absorbing states. The underlying parameters of the model were previously shown to be identifiable from the choice probabilities for the all-products assortment, together with choice probabilities for assortments of all-but-one products. Obtaining and estimating choice probabilities for such large assortments is not desirable in many settings. The main result of this work is that the parameters may be identified from assortments of sizes two and three, regardless of the total number of products. The result is obtained via a simple and efficient parameter recovery algorithm.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

In assortment planning, the seller's goal is to select a subset of products (called an *assortment*) to offer to a customer so as to maximize the expected revenue. This task can be formulated as an optimization problem given the revenue generated from selling each product, along with a probabilistic model of the customer's preferences for the products. Such a *discrete choice model* must capture the customer's *substitution behavior* when, for instance, the offered assortment does not contain the customer's most preferred product.

Our focus in this paper is the Markov chain choice model (MCCM) proposed by Blanchet et al. [2]. This MCCM chooses to model the product selected by the customer as a Markov chain over products where the products in the offered assortment are absorbing states. The current state represents the desired product; if that product is not offered, the customer transitions to another product according to the Markov chain probabilities, and the process continues until the desired product is offered or the customer leaves. MCCM generalizes widely-used discrete choice models such as the multinomial logit model [6,8], as well as other generalized attraction models [5]; it also well-approximates other random utility models found in the literature such as mixed multinomial logit models [7]. At the same time, the MCCM permits computationally efficient unconstrained assortment optimization as well as efficient approximation algorithms in the constrained case [2,4]; this stands in contrast to some richer models such as mixed multinomial logit models [9] and the nested logit model [3] for which assortment optimization is generally intractable. This combination of expressiveness and computational tractability makes MCCM very attractive for use in assortment planning.

A crucial step in this overall enterprise—e.g., before assortment optimization may take place—is the estimation of the choice model's parameters from observational data. Parameter estimation for MCCM is only briefly considered in the original work of Blanchet et al. [2]. In that work, it is shown that the parameters can be determined from the choice probabilities

---

* Corresponding author.
  E-mail address: djhsu@cs.columbia.edu (D. Hsu).

for the all-products assortment, together with the assortments comprised of all-but-one product. This is not satisfactory because it may be unrealistic or unprofitable to offer assortments of such large cardinality. Therefore, it is desirable to be able to determine the parameters from choice probabilities for smaller cardinality assortments. We note that this is indeed possible for simpler choice models such as the multinomial logit model (see, e.g., [10]), but these simpler models are limited in expressiveness—for example, they cannot express heterogeneous substitution behavior.

In this paper, we show that the MCCM parameters can be identified from the choice probabilities for assortments of sizes as small as two and three, independent of the total number of products.[1] We also give a simple and efficient algorithm for reconstructing the parameters from these choice probabilities.

2. Model and notation

In this section, we describe the Markov chain choice model (MCCM) of Blanchet et al. [2], along with notations used for choice probabilities and model parameters.

The set of n products in the system is denoted by $\mathcal{N} := \{1, 2, \ldots, n\}$. The "no purchase" option is denoted by product 0. Upon offering an assortment $S \subseteq \mathcal{N}$, the set of possible outcomes is $S_+ := S \cup \{0\}$: either some product in S is purchased, or no product is purchased.

Underlying the MCCM is a Markov chain with state space $\mathcal{N}_+$. The (true) parameters of the model are the initial state probabilities $\boldsymbol{\lambda} = (\lambda_i)_{i \in \mathcal{N}_+}$ and the transition probabilities $\boldsymbol{\rho} = (\rho_{i,j})_{(i,j) \in \mathcal{N}_+ \times \mathcal{N}_+}$ (a row stochastic matrix). The transition probabilities satisfy the following properties:

1. $\rho_{0,0} = 1$ and $\rho_{0,j} = 0$ for $j \in \mathcal{N}$ (i.e., the "no purchase" state is absorbing).
2. $\rho_{i,i} = 0$ for $i \in \mathcal{N}$ (i.e., no self-loops in product states).
3. The submatrix $\tilde{\boldsymbol{\rho}} := (\rho_{i,j})_{(i,j) \in \mathcal{N} \times \mathcal{N}}$ is irreducible.

We use $\boldsymbol{\rho}_i = (\rho_{i,j})_{j \in \mathcal{N}_+}$ to denote the i-th row of $\boldsymbol{\rho}$.

In MCCM, the customer arrives at a random initial state X_1 chosen according to $\boldsymbol{\lambda}$. At time $t = 1, 2, \ldots$:

- If $X_t = 0$, the customer leaves the system without purchasing a product.
- If the product X_t is offered (i.e., $X_t \in S$), the customer purchases X_t and leaves.
- If the product X_t is not offered (i.e., $X_t \notin S$), the customer transitions to a new random state X_{t+1} chosen according to $\boldsymbol{\rho}_{X_t}$ and the process continues in time step $t + 1$ as if the customer had initially arrived at X_{t+1}.

Another way to describe this process is that the Markov chain distribution is temporarily modified so that the states S_+ are absorbing, and the customer purchases the product upon reaching such a state (or makes no purchase if the state is 0). The irreducibility of $\tilde{\boldsymbol{\rho}}$ ensures that the customer eventually leaves the system (i.e., an absorbing state is reached). Note that only the identity of the final (absorbing) state is observed, as it corresponds to either a purchase or non-purchase. The $(X_t)_{t=1,2,\ldots}$ themselves do not correspond to observable customer behavior, and hence the model parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$ cannot be directly estimated.

The choice probabilities are denoted by $\pi(j, S)$ for $S \subseteq \mathcal{N}$ and $j \in S_+$: this is the probability that j is the final state in the aforementioned process. Blanchet et al. [2] relate the choice probabilities and the parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$ as follows:

$$\lambda_j = \pi(j, \mathcal{N}), \qquad \rho_{i,j} = \begin{cases} 1 & \text{if } i = 0 \text{ and } j = 0, \\ \dfrac{\pi(j, \mathcal{N} \setminus \{i\}) - \pi(j, \mathcal{N})}{\pi(i, \mathcal{N})} & \text{if } i \in \mathcal{N}, j \in \mathcal{N}_+, \text{ and } i \neq j, \\ 0 & \text{otherwise}. \end{cases} \qquad (1)$$

The relations in Equation (1) show that the parameters may be identified from choice probabilities for the assortments $S = \mathcal{N}$ and $S = \mathcal{N} \setminus \{i\}$ for $i \in \mathcal{N}$. These choice probabilities may be directly estimated from observations upon offering such assortments to customers.

3. Main result

The following theorem establishes identifiability of the MCCM parameters from choice probabilities for assortments of sizes as small as two and three.

Theorem 1. *There is an efficient algorithm that, for any $r \in \{2, 3, \ldots, n-1\}$, when given as input the choice probabilities $(\pi(j, S))_{j \in S_+}$ for all assortments $S \subseteq \mathcal{N}$ of cardinality r and $r + 1$ for a Markov chain choice model, returns the parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$ of the model.*

---

[1] We focus on identifiability because estimation of choice probabilities from observational data is fairly straightforward, especially when the assortments have small cardinality. However, this issue is revisited in Section 6 in the context of sample complexity.

The number of assortments for which the algorithm actually requires choice probabilities is $O(n^2)$ when $r \le n/2$, which is far fewer than $\binom{n}{r} + \binom{n}{r+1}$, the total numbers of assortments of sizes r and $r+1$. The details of this bound are shown following the proof of Theorem 1. However, to simplify the presentation, we describe our parameter recovery algorithm as using choice probabilities for all assortments of sizes r and $r+1$.

The main steps of our algorithm, shown as Algorithm 1, involve setting up and then solving systems of linear equations that (as we will prove) determine the unknown parameters $\boldsymbol{\lambda}$ and $(\boldsymbol{\rho}_i)_{i \in \mathcal{N}}$. (Note that $\boldsymbol{\rho}_0$ is already known.) The coefficients of the linear equations are determined by the given choice probabilities via *conditional choice probabilities* $\pi(j, S \mid i)$ for $S \subseteq \mathcal{N}$ and $i, j \in \mathcal{N}_+$, defined as follows:

$$\pi(j, S \mid i) := \Pr(\text{state } j \text{ is reached before any state in } S_+ \setminus \{j\} \mid \text{initial state is } i) . \tag{2}$$

Algorithm 1 Parameter recovery algorithm for Markov chain choice model.

---

input For some $r \in \{2, 3, \ldots, n-1\}$, choice probabilities $(\pi(j, S))_{j \in S_+}$ for all assortments $S \subseteq \mathcal{N}$ of sizes r and $r+1$.
output Parameters $\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\rho}}$.
1: **for** $i \in \mathcal{N}$ **do**
2:     Solve the following system of linear equations for $\hat{\boldsymbol{\rho}}_i = (\hat{\rho}_{i,k})_{k \in \mathcal{N}_+}$:

$$\sum_{k \in \mathcal{N}_+} \pi(j, S \mid k) \cdot \hat{\rho}_{i,k} = \pi(j, S \mid i) \qquad \text{for all } S \in \binom{\mathcal{N}}{r} \text{ s.t. } i \notin S \text{ and } j \in S_+, \tag{3}$$

    where $\binom{\mathcal{N}}{r}$ denotes the family of subsets of $\mathcal{N}$ of size r, and $\pi(j, S \mid k)$ is defined in Equation (5).
3: **end for**
4: Solve the following system of linear equations for $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_i)_{i \in \mathcal{N}_+}$:

$$\sum_{k \in \mathcal{N}_+} \pi(j, S \mid k) \cdot \hat{\lambda}_k = \pi(j, S) \qquad \text{for all } S \in \binom{\mathcal{N}}{r} \text{ and } j \in S_+ \tag{4}$$

5: **return** $\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\rho}}$.

Note that the initial state in the MCCM is not observed, so these conditional probabilities cannot be directly estimated. Nevertheless, they can be indirectly estimated via the following relationship between the conditional choice probabilities and the (unconditional) choice probabilities.

Lemma 1. *For any $S \subseteq \mathcal{N}$ and $i, j \in S_+$,*

$$\pi(j, S \mid i) = \begin{cases} 1 & \text{if } i = j, \\ \dfrac{\pi(j, S) - \pi(j, S \cup \{i\})}{\pi(i, S \cup \{i\})} & \text{if } i \in \mathcal{N} \setminus S, \\ 0 & \text{if } i \in S_+ \setminus \{j\}. \end{cases} \tag{5}$$

Proof. The cases where $i = j$ ($\Rightarrow \pi(j, S \mid i) = 1$) and $i \in S_+ \setminus \{j\}$ ($\Rightarrow \pi(j, S \mid i) = 0$) are clear from the definition in Equation (2). It remains to handle the case where $i \in \mathcal{N} \setminus S$. Fix such a product i, and observe that

$$\begin{aligned} \pi(j, S) &= \Pr(j \text{ is reached before } S_+ \setminus \{j\}) \\ &= \Pr(j \text{ is reached before } S_+ \setminus \{j\} \wedge i \text{ is not reached before } S_+) \\ &\quad + \Pr(j \text{ is reached before } S_+ \setminus \{j\} \wedge i \text{ is reached before } S_+) \\ &= \Pr(j \text{ is reached before } (S_+ \cup \{i\}) \setminus \{j\}) \\ &\quad + \Pr(j \text{ is reached before } S_+ \setminus \{j\} \mid i \text{ is reached before } S_+) \\ &\quad\quad \cdot \Pr(i \text{ is reached before } S_+) \\ &= \Pr(j \text{ is reached before } (S \cup \{i\})_+ \setminus \{j\}) \\ &\quad + \Pr(j \text{ is reached before } S_+ \setminus \{j\} \mid \text{initial state is } i) \\ &\quad\quad \cdot \Pr(i \text{ is reached before } (S \cup \{i\})_+ \setminus \{i\}) \\ &= \pi(j, S \cup \{i\}) + \pi(j, S \mid i) \cdot \pi(i, S) . \end{aligned}$$

The penultimate step uses the Markov property and the case condition that $i \in \mathcal{N} \setminus S$. Rearranging the equation gives the relation claimed by the lemma in this case. $\square$

Lemma 1 shows that the conditional choice probabilities for assortments S of size r can be determined from the unconditional choice probabilities of assortments of size r and $r + 1$. The systems of linear equations used in Algorithm 1 (Equations (3) and (4)) are defined in terms of these conditional choice probabilities and hence are ultimately defined in terms of the unconditional choice probabilities provided as input to Algorithm 1.

It is clear that the true MCCM parameters λ and ρ satisfy the systems of linear equations in Equations (3) and (4). However, what needs to be proved is that they are uniquely determined by these linear equations; this is the main content of the proof of Theorem 1.

4. Proof of Theorem 1

In this section, we give the proof of Theorem 1.

4.1. The case without the "no purchase" option

For sake of clarity, we first give the proof in the case where the "no purchase" option is absent. This can be regarded as the special case where $\lambda_0 = 0$ and $\rho_{i,0} = 0$ for all $i \in \mathcal{N}$. So here we just regard $\lambda = (\lambda_j)_{j \in \mathcal{N}}$ and each $\rho_i = (\rho_{i,j})_{j \in \mathcal{N}}$ as probability distributions on $\mathcal{N}$. The general case will easily follow from the same arguments with minor modification.

4.1.1. Proof strategy

We make use of the following result about M-matrices, i.e., the class of matrices A that can be expressed as $A = sI - B$ for some $s > 0$ and non-negative matrix B with spectral radius at most s. (Here, I denotes the identity matrix of appropriate dimensions.) In particular, the matrix $I - \rho$ is a (singular) M-matrix that is also irreducible.

Lemma 2 *(See, e.g., Theorems 6.2.3 & 6.4.16 in [1]). If $A \in \mathbb{R}^{p \times p}$ is an irreducible M-matrix (possibly singular), then every principal submatrix[2] of A, other than A itself, is non-singular. If A is also singular, then it has rank $p - 1$.*

For each $S \in \binom{\mathcal{N}}{r}$ and $j \in S$, define the vector

$$\boldsymbol{h}_{j,S} := (\pi(j, S \mid k))_{k \in \mathcal{N}}.$$

For each $i \in \mathcal{N}$, the collection of the vectors $\{\boldsymbol{h}_{j,S} : S \not\ni i \wedge j \in S\}$ provide the left-hand side coefficients in Equation (3) for $\hat{\rho}_i$. We'll show that the span of these vectors (in fact, a particular subset of them) has dimension at least $n - 1$. This is sufficient to conclude that ρ_i is the unique solution to the system of equations in Equation (3) because it has at most $n - 1$ unknown variables, and it is clear that ρ_i satisfies the system of equations. (In fact, there are really only $n - 2$ unknown variables, because we can force $\hat{\rho}_{i,i} = 0$ and $\hat{\rho}_{i,n} = 1 - \sum_{k=1}^{n-1} \hat{\rho}_{i,k}$.) For the same reason, it is also sufficient to conclude that λ is the unique solution to the system of equations in Equation (4) (where, in fact, we may use all vectors $\{\boldsymbol{h}_{j,S} : S \in \binom{\mathcal{N}}{r} \wedge j \in S\}$).

4.1.2. Rank of linear equations from a single assortment

We begin by characterizing the space spanned by $\{\boldsymbol{h}_{j,S} : j \in S\}$ for a fixed $S \in \binom{\mathcal{N}}{r}$. We claim, by Lemma 1, that the vectors in $\{\boldsymbol{h}_{j,S} : j \in S\}$ are linearly independent. Indeed, if this collection of vectors is arranged in a matrix $[\boldsymbol{h}_{j,S} : j \in S]$, then the submatrix obtained by selecting rows corresponding to $j \in S$ is the $S \times S$ identity matrix. Thus we have proved

Lemma 3. *For any $S \in \binom{\mathcal{N}}{r}$, $\dim\left(\text{span}\{\boldsymbol{h}_{j,S} : j \in S\}\right) = |S| = r$.*

Note that in the case $r = n - 1$, we are done. But when $r < n - 1$, the linear equations given by the $\{\boldsymbol{h}_{j,S} : j \in S\}$ may not uniquely determine the ρ_i for $i \in \mathcal{N} \setminus S$. To overcome this, we need to be able to combine linear equations derived from multiple assortments. However, for a sum of subspaces V and W,

$$\dim(V + W) \neq \dim(V) + \dim(W)$$

unless V and W are orthogonal. In our case, the subspaces $\text{span}\{\boldsymbol{h}_{j,S} : j \in S\}$ and $\text{span}\{\boldsymbol{h}_{j,S'} : j \in S'\}$ for different assortments S and S' are *not* necessarily orthogonal (even if S and S' are disjoint). So a different argument is needed.

4.1.3. Rank of linear equations from multiple assortments

Our aim is to show that the intersection of subspaces $V := \text{span}\{\boldsymbol{h}_{j,S} : j \in S\} \cap \text{span}\{\boldsymbol{h}_{j,S'} : j \in S'\}$ for different assortments S and S' cannot have high dimension. We do this by showing that the intersection is orthogonal to a subspace of high dimension.

---

[2] Recall that a *principal submatrix* of a $p \times p$ matrix A is a submatrix obtained by removing from A the rows and columns indexed by some set $I \subseteq [p]$.

For each $i \in \mathcal{N}$, let $\boldsymbol{a}_i$ denote the i-th row of the matrix $\boldsymbol{A} := \boldsymbol{I} - \boldsymbol{\rho}$ (which is an M-matrix). That is, $\boldsymbol{a}_i := \boldsymbol{e}_i - \boldsymbol{\rho}_i$, where $\boldsymbol{e}_i$ is the i-th coordinate basis vector. Recall that if $i \in \mathcal{N} \setminus S$, then $\boldsymbol{\rho}_i$ satisfies Equation (3). This fact can be written in our new notation as

$$\boldsymbol{h}_{j,S}^\top \boldsymbol{e}_i - \boldsymbol{h}_{j,s}^\top \boldsymbol{\rho}_i = \boldsymbol{a}_i^\top \boldsymbol{h}_{j,S} = 0, \quad j \in S.$$

In other words,

Lemma 4. *For any $S \in \binom{\mathcal{N}}{r}$, $\operatorname{span}\{\boldsymbol{h}_{j,S} : j \in S\} \perp \operatorname{span}\{\boldsymbol{a}_i : i \in \mathcal{N} \setminus S\}$.*

Now consider two assortments S and S', and the intersection of their respective subspaces. It follows from Lemma 4 that

$$\operatorname{span}\{\boldsymbol{h}_{j,S} : j \in S\} \cap \operatorname{span}\{\boldsymbol{h}_{j,S'} : j \in S'\} \perp \operatorname{span}\{\boldsymbol{a}_i : i \in \mathcal{N} \setminus (S \cap S')\}.$$

This orthogonality is the key to lower-bounding the dimension of the sum of these subspaces, which we capture in the following general lemma.

Lemma 5. *Let $\mathcal{S}$ be a family of subsets of $\mathcal{N}$, S' be a subset of $\mathcal{N}$, and $\mathcal{S}' := \mathcal{S} \cup \{S'\}$. Define the subspaces*

$$V_{\mathcal{S}} := \operatorname{span}\{\boldsymbol{h}_{j,S} : S \in \mathcal{S}, j \in S\},$$
$$V_{S'} := \operatorname{span}\{\boldsymbol{h}_{j,S'} : j \in S'\},$$
$$V_{\mathcal{S}'} := V_{\mathcal{S}} + V_{S'}.$$

Then

$$\dim(V_{\mathcal{S}'}) \geq \dim(V_{\mathcal{S}}) + |S'| - \max\left\{1, \left|\left(\bigcup_{S \in \mathcal{S}} S\right) \cap S'\right|\right\}.$$

Proof. Let $S_0 := \bigcup_{S \in \mathcal{S}} S$. Fix any $\boldsymbol{v} \in V_{\mathcal{S}} \cap V_{S'}$. Then, by Lemma 4, $\boldsymbol{a}_i^\top \boldsymbol{v} = 0$ for all $i \in (\mathcal{N} \setminus S_0) \cup (\mathcal{N} \setminus S') = \mathcal{N} \setminus (S_0 \cap S')$. In other words,

$$V_{\mathcal{S}} \cap V_{S'} \perp W,$$

where $W := \operatorname{span}\{\boldsymbol{a}_i : i \in \mathcal{N} \setminus (S_0 \cap S')\}$, and

$$\dim(V_{\mathcal{S}} \cap V_{S'}) \leq \dim(W^\perp) = n - \dim(W).$$

To determine $\dim(W)$, observe that W is the span of rows of certain rows of the M-matrix $\boldsymbol{A}$. By Lemma 2, the principal submatrix of $\boldsymbol{A}$ corresponding to $\mathcal{N} \setminus (S_0 \cap S')$ is either non-singular (when $S_0 \cap S' \neq \emptyset$) or is $\boldsymbol{A}$ itself; in either case, it has rank $n - \max\{1, |S_0 \cap S'|\}$. Hence,

$$\dim(W) = n - \max\{1, |S_0 \cap S'|\}$$

as well. Combining the dimension formula with the last two equation displays gives

$$\begin{aligned}
\dim(V_{\mathcal{S}'}) &= \dim(V_{\mathcal{S}} + V_{S'}) \\
&= \dim(V_{\mathcal{S}}) + \dim(V_{S'}) - \dim(V_{\mathcal{S}'} \cap V_{S'}) \\
&\geq \dim(V_{\mathcal{S}}) + \dim(V_{S'}) - \max\{1, |S_0 \cap S'|\}.
\end{aligned}$$

The claim now follows from Lemma 3. $\quad\square$

4.1.4. Choice of assortments

We now choose a collection of assortments and argue, via Lemma 3 and Lemma 5, that they define linear equations of sufficiently high rank. Specifically, for each $i \in \mathcal{N}$, we need a collection $\mathcal{S} \subset \binom{\mathcal{N}}{r}$ such that each $S \in \mathcal{S}$ does not contain i, and

$$\dim\left(\operatorname{span}\bigcup_{S \in \mathcal{S}}\{\boldsymbol{h}_{j,S} : j \in S\}\right) \geq n - 1. \tag{6}$$

Lemma 6. *Suppose the assortments $S_1, S_2, \ldots, S_T \in \binom{\mathcal{N}}{r}$ have a pairwise common intersection $S_\cap = S_t \cap S_{t'}$ for all $t \neq t'$, and $|S_\cap| = r - 1$. Then $\dim\left(\operatorname{span}\bigcup_{t=1}^T \{\boldsymbol{h}_{j,S_t} : j \in S_t\}\right) \geq T + r - 1$.*

Proof. Let $d_\tau := \dim\left(\text{span}\bigcup_{t=1}^{\tau}\{\boldsymbol{h}_{j,S_t} : j \in S_t\}\right)$ for $\tau \in \{1, 2, \ldots, T\}$. By Lemma 3, we know that $d_1 = r$. Now, assume $d_\tau \geq \tau + r - 1$, and use the fact $r \geq 2$ and Lemma 5 to conclude that $d_{\tau+1} \geq d_\tau + r - (r-1) = d_\tau + 1 \geq \tau + r$. The claim now follows by induction. $\square$

Fix any $i \in \mathcal{N}$ and $S_\cap \in \binom{\mathcal{N}\setminus\{i\}}{r-1}$, and observe that $|\mathcal{N}\setminus(S_\cap \cup \{i\})| = n - r$. Consider the collection of size-r assortments given by

$$\mathscr{S} := \{S_\cap \cup \{k\} : k \in \mathcal{N}\setminus(S_\cap \cup \{i\})\}. \tag{7}$$

These assortments do not contain i, they have the common intersection $S_\cap$, with $|S_\cap| = r-1$, and there are $n-r$ assortments in total. So by Lemma 6, the collection $\mathscr{S}$ satisfies the dimension bound in Equation (6).

As was already argued in the proof strategy, this suffices to establish the uniqueness of the $\boldsymbol{\rho}_i$ and $\boldsymbol{\lambda}$ as solutions to the respective systems of linear equations in Equation (3) and Equation (4).

This concludes the proof of Theorem 1 without the "no purchase" option. $\square$

4.2. The general case with the "no purchase" option

We now consider the general case, where the "no purchase" option is present. The main difference relative to the previous subsection is that $\boldsymbol{\rho}$ is no longer irreducible, as the "no purchase" state 0 is absorbing. However, the submatrix $\tilde{\boldsymbol{\rho}} = (\rho_{i,j})_{(i,j) \in \mathcal{N} \times \mathcal{N}}$ is irreducible, so $\boldsymbol{I} - \tilde{\boldsymbol{\rho}}$ is an irreducible M-matrix.

The definition of $\boldsymbol{h}_{j,S}$, for $S \subseteq \mathcal{N}$ and $j \in S_+$, is now taken to be

$$\boldsymbol{h}_{j,S} := (\pi(j, S \mid k))_{k \in \mathcal{N}_+}.$$

Because the indexing starts at 0, we still define $\boldsymbol{a}_i$ to be the i-th row of $\boldsymbol{A} = \boldsymbol{I} - \boldsymbol{\rho}$, so $\boldsymbol{a}_i = \boldsymbol{e}_i - \boldsymbol{\rho}_i$. (In particular, $\boldsymbol{a}_0$ is the all-zeros vector.)

With these definitions, we have the following analogue of Lemma 3 and Lemma 4:

Lemma 7. *For any $S \in \binom{\mathcal{N}}{r}$,*

$$\dim(\text{span}\{\boldsymbol{h}_{j,S} : j \in S_+\}) = |S_+| = r + 1,$$
$$\text{span}\{\boldsymbol{h}_{j,S} : j \in S_+\} \perp \text{span}\{\boldsymbol{a}_i : i \in \mathcal{N}\setminus S\}.$$

Here, the key difference is that the dimension is $r + 1$, rather than just r.

We now establish an analogue of Lemma 5 (which is typographically nearly identical).

Lemma 8. *Let $\mathscr{S}$ be a family of subsets of $\mathcal{N}$, S' be a subset of $\mathcal{N}$, and $\mathscr{S}' := \mathscr{S} \cup \{S'\}$. Define the subspaces*

$$V_{\mathscr{S}} := \text{span}\{\boldsymbol{h}_{j,S} : S \in \mathscr{S}, j \in S_+\},$$
$$V_{S'} := \text{span}\{\boldsymbol{h}_{j,S'} : j \in S'_+\},$$
$$V_{\mathscr{S}'} := V_{\mathscr{S}} + V_{S'}.$$

Then

$$\dim(V_{\mathscr{S}'}) \geq \dim(V_{\mathscr{S}}) + |S'| - \max\left\{1, \left|\left(\bigcup_{S \in \mathscr{S}} S\right) \cap S'\right|\right\}.$$

Proof. The proof is nearly the same as that of Lemma 5. Define $S_0 := \bigcup_{S \in \mathscr{S}} S$ and take $\boldsymbol{v} \in V_{\mathscr{S}} \cap V_{S'}$. By Lemma 7, $V_{\mathscr{S}} \cap V_{S'} \perp W$, where $W := \text{span}\{\boldsymbol{a}_i : i \in \mathcal{N}\setminus(S_0 \cap S')\}$, and

$$\dim(V_{\mathscr{S}} \cap V_{S'}) \leq \dim(W^\perp) = n + 1 - \dim(W).$$

We now use the fact that $\boldsymbol{I} - \tilde{\boldsymbol{\rho}}$, which is a submatrix of $\boldsymbol{A}$, is an irreducible M-matrix. By Lemma 2, the principal submatrix of $\boldsymbol{A}$ corresponding to $\mathcal{N}\setminus(S_0 \cap S')$ is either non-singular (when $S_0 \cap S' \neq \emptyset$) or is $\boldsymbol{I} - \tilde{\boldsymbol{\rho}}$; in either case, it has rank $n - \max\{1, |S_0 \cap S'|\}$. So we have

$$\dim(W) = n - \max\{1, |S_0 \cap S'|\} \quad \text{and} \quad \dim(W^\perp) = 1 + \max\{1, |S_0 \cap S'|\}.$$

Finishing the proof as in Lemma 5, we have

$$\dim(V_{\mathscr{S}'}) \geq \dim(V_{\mathscr{S}}) + \dim(V_{S'}) - 1 - \max\{1, |S_0 \cap S'|\}$$
$$\geq \dim(V_{\mathscr{S}}) + r - \max\{1, |S_0 \cap S'|\}$$

where the second inequality uses Lemma 7 (instead of Lemma 3). $\square$

The choice of assortments demonstrating the subspace of required dimension is the same as before, except now we show that the dimension is at least n. Again, fix some $i \in \mathcal{N}$, and choose the collection of $n - r$ assortments $\mathscr{S} \subseteq \binom{\mathcal{N}}{r}$ as before (described in and directly before Equation (7)). Following the inductive argument in the proof of Lemma 6, but now using Lemma 7 and Lemma 8 (instead of Lemma 3 and Lemma 5), we have

$$\dim\left(\operatorname{span} \bigcup_{S \in \mathscr{S}} \{\boldsymbol{h}_{j,S} : j \in S\}\right) \geq (r+1) + (n-r-1) \cdot (r - \max\{1, r-1\}) = n.$$

Since each of the systems of linear equations from Equation (3) and Equation (4) have (at most) n unknown variables, we conclude that the $\boldsymbol{\rho}_i$ and $\boldsymbol{\lambda}$ are unique as solutions to their respective systems of linear equations.

This concludes the proof of Theorem 1. □

4.3. Total number of assortments required

We now show that the number of assortments for which we need the choice probabilities is $O(n^2)$ for $r \leq n/2$. Indeed, the construction given above based on Lemma 6 can be used to avoid using all assortments of size r (and $r+1$) in Algorithm 1.

We choose two sets $S_\cap, S'_\cap \in \binom{\mathcal{N}}{r-1}$, which shall serve as "common intersection sets" (in the sense used in Section 4.1.4), as follows. The first set $S_\cap \in \binom{\mathcal{N}}{r-1}$ is chosen arbitrarily; it serves as the common intersection set for all $i \in \mathcal{N} \setminus S_\cap$. The second set $S'_\cap \in \binom{\mathcal{N} \setminus S_\cap}{r-1}$ is chosen arbitrarily as long as it is disjoint from $S_\cap$ (which is possible because $r \leq n/2$); it serves as the common intersection set for $i \in S_\cap$.

For each $i \in \mathcal{N} \setminus S_\cap$, we need the equations for the assortments $S_\cap \cup \{k\}$ for all $k \in \mathcal{N} \setminus (S_\cap \cup \{i\})$. Obtaining the equations for one such $S_\cap \cup \{k\}$ requires choice probabilities for assortments $S_\cap \cup \{k\}$ and $S_\cap \cup \{k\} \cup \{j\}$ for $j \in \mathcal{N} \setminus (S_\cap \cup \{k\})$ as per Lemma 1. In total, for all $i \in \mathcal{N} \setminus S_\cap$, we need choice probabilities for $O(n^2)$ assortments. For the remaining $i \in S_\cap$, we use the same argument for the disjoint common intersection set $S'_\cap$, and thus require the choice probabilities for at most another $O(n^2)$ assortments.

5. Sample complexity

Estimating all the purchase probabilities for an assortment of size r involves inferring the parameters of a Multinomial random variable, $MN(m, \pi_1, ..., \pi_k)$ where m is the number of customers and π_i is the probability for the i-th category. Using larger assortments for the algorithm requires estimating as few as $n + 1$ of these random variables. Using smaller assortments could require estimating as many as $O(n^2)$ random variables, but these would be binomial or multinomial with fewer categories, which could require fewer samples overall. We discuss how many samples would be needed to limit the percentage difference of the maximum values between the calculated and actual parameters. We present the following lemma which describes how many samples are needed to estimate a multinomial random variable with k categories.

Lemma 9. *For a multinomial random variable with category probabilities π_i $i \in \{1, ..., k\}$ and observed probability p_i, a sample size of*

$$n = \max_{k \geq 1} \left\{ \frac{(1/k)(1 - 1/k)\chi^2_{(1,\alpha/k)}}{d^2} \right\} \tag{8}$$

is sufficient to satisfy that

$$P[\cap_{i=1}^{k} |p_i - \pi_i| \leq d_i] \geq 1 - \alpha \; \forall i = 1, ..., k \tag{9}$$

if all the $d_i = d$.

Lemma 10. *Suppose we have an original system $Ax = b$ which is perturbed to $(A + \Delta A)x = (b + \Delta b)$. Then*

$$\frac{\|\Delta x\|}{\|x\|} \leq K(A)\left(\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta b\|}{\|b\|}\right) \tag{10}$$

where $K(A)$ is the condition number of A.

The maximum values for n in Lemma 9 actually occur for small values of k, so the size of the assortments itself per se may not require more samples. However, Lemma 9 deals with the case where all the $d_i = d$. We are not requiring that all parameters be within a percentage of their true values. If this is what we require, then using larger assortments could require more samples than using smaller assortments, because the smallest probability $p_S = \min_{i \in S} \pi_i$ for a large assortment may be smaller than that for a small assortment. There is no sample size that can guarantee that the parameters are within

a percentage of their true value for all possible parameters, but if the smallest category probability is p_s, then we may choose $d = c \cdot p_s$, where c is some small positive constant, .1 for example, to guarantee that all parameters are within 10% of their true value. This means that when calculating the $\pi(j, |S, i)$, the numerator in Lemma 1 is guaranteed to be no more than $\Delta_s = c \cdot p_S + c \cdot p_{S \cup \{i\}}$ away from its true value. Since $p_{S \cup \{i\}} \leq p_S$ we can just write $\Delta_s = 2c \cdot p_S$, and since the denominator is guaranteed to be within 10% of its true value $\pi(j|S, i)$ is guaranteed to be within

$$\frac{c}{1-c} \pi(j|s, i) + \frac{1}{1-c} \Delta_s \tag{11}$$

of its true value. Letting the ρ_i function as x in Lemma 10, we can see that we need only to figure out the norm of ΔA and Δb to get a bound on the ρ_i. We pick the ∞ norm and note that at least one of the $\pi(j|S, i) = 1$ for any nonempty S. Therefore taking the sum across the row,

$$\|A\| = \max_{i \in [n]} \left(1 + \sum \pi(\cdot|S_i, \cdot) \right) \tag{12}$$

The smallest value $\|b\|$ can take is 1. The largest value $\|\Delta A\|$ can take is

$$\max_{i \in [n]} \left(\sum \left(\frac{c}{1-c} \pi(\cdot|S_i, \cdot) + \frac{1}{1-c} \Delta_s \right) \right) \tag{13}$$

and the maximum value that Δb can take is

$$\frac{c}{1-c} \pi(\cdot|\cdot, \cdot) + \frac{1}{1-c} \Delta_s \tag{14}$$

So the final expression becomes

$$\frac{\|\Delta \rho_i\|}{\|\rho_i\|} \leq K(A) \left(\frac{\max_{i \in [n]} \left(\sum \left(\frac{c}{1-c} \pi(\cdot|S_i, \cdot) + \frac{2c}{1-c} \cdot p_{S_i} \right) \right)}{\max_{i \in [n]} \left(1 + \sum \pi(\cdot|S_i, \cdot) \right)} + \frac{\frac{c}{1-c} \pi(\cdot|\cdot, \cdot) + \frac{2c}{1-c} \cdot p_{S_i}}{1} \right) \tag{15}$$

Although p_S appears in the numerator of the previous equation, a smaller p_S requires that the number of samples be much larger in Lemma 9.

6. Discussion

Our main result establishes the identifiability of MCCM parameters from choice probabilities for assortments of sizes different from $n - 1$ and n. This is important because real systems often have cardinality constraints on the assortment sizes. While such constraints are typically considered in the context of assortment optimization (see, e.g., [4]), it is also important in the context of parameter estimation.

One complication of using small size assortments to estimate the MCCM parameters is that the number of different assortments required may be as large as $O(n^2)$. In contrast, only $n + 1$ assortments are needed when the sizes are $n - 1$ and n. On the other hand, the statistical difficulty of estimating choice probabilities for large assortments may be higher than the same task for smaller assortments. This work assumed throughout that no noise is present in the inputs, and determining how the addition of noise affects the accuracy of the parameter estimation is an interesting question we leave to future work.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

References

[1] Abraham Berman, Robert J. Plemmons, Nonnegative Matrices in the Mathematical Sciences, SIAM, 1994.
[2] Jose Blanchet, Guillermo Gallego, Vineet Goyal, A Markov chain approximation to choice modeling, Oper. Res. 64 (4) (2016) 886–905.
[3] James M. Davis, Guillermo Gallego, Huseyin Topaloglu, Assortment optimization under variants of the nested logit model, Oper. Res. 62 (2) (2014) 250–273.

[4] Antoine Désir, Vineet Goyal, Danny Segev, Chun Ye, Capacity constrained assortment optimization under the markov chain based choice model, https://dx.doi.org/10.2139/ssrn.2626484, 2015.
[5] Guillermo Gallego, Richard Ratliff, Sergey Shebalov, A general attraction model and sales-based linear program for network revenue management under customer choice, Oper. Res. 63 (1) (2014) 212–232.
[6] R. Duncan Luce, Individual Choice Behavior: A Theoretical Analysis, Wiley, 1959.
[7] Daniel McFadden, Kenneth Train, Mixed MNL models for discrete response, J. Appl. Econom. (2000) 447–470.
[8] Robin L. Plackett, The analysis of permutations, J. R. Stat. Soc., Ser. C, Appl. Stat. (1975) 193–202.
[9] Paat Rusmevichientong, David Shmoys, Huseyin Topaloglu, Assortment Optimization with Mixtures of Logits, Technical report, Tech. rep., School of IEOR, Cornell University, 2010.
[10] Kenneth E. Train, Discrete Choice Methods with Simulation, Cambridge University Press, 2009.