DEMAND INFERENCE FOR FREE-FLOATING MICRO-MOBILITY: ACCESSIBILITY AND AVAILABILITY

FINAL PROJECT REPORT

by

Chiwei Yan
University of Washington Seattle

Ang Xu
University of Washington Seattle

Sponsorship
Pacific Northwest Transportation Consortium (PacTrans)
College of Engineering, University of Washington
Department of Industrial and Systems Engineering, University of Washington

for

Pacific Northwest Transportation Consortium (PacTrans)
USDOT University Transportation Center for Federal Region 10
University of Washington
More Hall 112, Box 352700
Seattle, WA 98195-2700

In cooperation with U.S. Department of Transportation,
Office of the Assistant Secretary for Research and Technology (OST-R)

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The Pacific Northwest Transportation Consortium, the U.S. Government and matching sponsor assume no liability for the contents or use thereof.

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
	01872733	

4. Title and Subtitle	5. Report Date
Demand Inference for Free-Floating Micro-Mobility: Accessibility and Availability	June 2023
	6. Performing Organization Code

7. Author(s) and Affiliations	8. Performing Organization Report No.
Chiwei Yan, 0000-0001-5770-2079; Assistant Professor, University of Washington Ang Xu, PhD student, University of Washington	2022-S-UW-1

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
PacTrans Pacific Northwest Transportation Consortium University Transportation Center for Federal Region 10 University of Washington More Hall 112 Seattle, WA 98195-2700	**11. Contract or Grant No.** 69A3551747110

12. Sponsoring Organization Name and Address	13. Type of Report and Period Covered
United States Department of Transportation Research and Innovative Technology Administration 1200 New Jersey Avenue, SE Washington, DC 20590	**14. Sponsoring Agency Code**

15. Supplementary Notes

Report uploaded to: www.pactrans.org

16. Abstract

Micro-mobility systems (such as bike-sharing or scooter-sharing) have been widely adopted across the globe as sustainable modes of urban transportation. To efficiently plan and operate such systems, it is crucial to understand the underlying rider demand --- where riders come from and the rates of arrivals into the service area. Estimating rider demand is not trivial as most systems only track trip data, which are a biased representation of underlying demand. In this report, we describe development of a locational demand model to estimate rider demand based on only trip and vehicle location and status data. We established conditions under which our estimators are identifiable and consistent. In addition, we devised an expectation-maximization (EM) algorithm with closed-form updates for efficient estimation. To scale the estimation procedures, this EM algorithm is complemented with a location-discovery procedure that gradually adds new locations in the service region that make the largest improvements to the log-likelihood. Experiments using both synthetic data and real data from a dockless bike-sharing system in the Seattle area demonstrated the accuracy and scalability of the model and its estimation algorithm.

17. Key Words	18. Distribution Statement
locational demand model, bike-sharing, expectation-maximization, location discovery	

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified.	Unclassified.	40	N/A

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

.

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS				
Symbol	**When You Know**	**Multiply By**	**To Find**	**Symbol**
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
	NOTE: volumes greater than 1000 L shall be shown in m^3			
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	$candela/m^2$	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS				
Symbol	**When You Know**	**Multiply By**	**To Find**	**Symbol**
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	$candela/m^2$	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.
(Revised March 2003)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The authors thank the New Mobility Program at the Seattle Department of Transportation (SDOT) for providing support for this project.

LIST OF ACRONYMS

BIC:	Bayesian Information Criteria
CTMC:	Continuous-time Markov chain
EM:	Expectation-maximization
FW:	Frank-Wolfe
GPS:	Global Positioning System
KKT:	Karush-Kuhn-Tucker
LCL:	Latent-class logit
MAP:	Maximum a posteriori
MLE:	Maximum likelihood estimator
MM:	Minorization-maximization
MNL:	Multinomial logit
WMAPE:	Weighted mean absolute percentage error

EXECUTIVE SUMMARY

Micro-mobility systems (such as bike-sharing and scooter-sharing) have been widely adopted across the globe as sustainable modes of urban transportation. To efficiently plan, monitor, and operate such systems, it is crucial for agencies to understand the underlying rider demand—where riders come from and rates of arrivals into the service area. Estimating rider demand is not trivial, as most systems only keep track of trip data, which are a biased representation of underlying demand. In this report, we describe the development of a locational demand model to estimate rider demand based on only trip and vehicle location and status data. We established conditions under which our estimators are identifiable and consistent. In addition, we devised an expectation-maximization (EM) algorithm with closed-form updates for efficient estimation. To scale the estimation procedures, this EM algorithm is complemented with a location-discovery procedure that gradually adds new locations in the service region that provide the largest improvements to the log-likelihood. Experiments using both synthetic data and real data from a dockless bike-sharing system in the Seattle area demonstrated the accuracy and scalability of the model and its estimation algorithm.

CHAPTER 1. INTRODUCTION

Bike-sharing services have been widely adopted to provide a sustainable mode of urban transportation. In the United States, as of July 2022, 61 docked bike-sharing systems operated 8,473 docking stations (Bureau of Transportation Statistics 2023). As of July 2022, dockless bike-sharing systems served 35 cities and e-scooters served 158 cities in the United States (Bureau of Transportation Statistics 2023). Regionally, since its debut in Seattle in 2017, dockless bike-sharing and e-scooter systems have quickly expanded their coverage and gained popularity because of their increased accessibility.

One critical aspect of monitoring and operating bike-sharing systems is understanding rider demand and how accessible the current service is to different communities in the service region. For example, the city of Seattle has been actively monitoring the usage and accessibility of these services by using trip data (see Seattle Department of Transportation 2022 for a comprehensive online dashboard). In addition, these systems often experience supply and demand imbalances that require careful bike allocation and rebalancing (Transportation Research Board 2010). The success of these operations crucially depends on the operator's ability to estimate ridership demand accurately over the planning horizon. These issues motivated the topic of our study.

Demand estimation for bike-sharing systems is not trivial because of the following difficulties. First, the operator often does not know the exact location of riders but observes only the booking data, which indicate the location and time a bike is reserved and picked up. Although technically speaking rider location data can be accessed through the GPS on riders' phones, to the best of the authors' knowledge, it is not standard industry practice to record these data (Open Mobility Foundation 2022). These rider location data can be hard to collect and are subject to errors themselves; for example, the place that a rider checks her phone can be different from where she departs to pick up a bike. Second, observed trip locations and trip counts can form biased estimates for underlying demand. This is because the booking location is likely not the actual rider location, and a rider may choose not to book or to switch to other transportation services if no bikes are available near her location. In such a case, this demand will be censored in the observed trip data.

A standard approach for modeling such demand is to use a choice model, which specifies the likelihood that a customer will make a certain choice when presented with a set of

alternatives. Analogous to the retail settings to which such models are commonly applied, we can view each bike in the system as a product to be considered among a set of available bikes. One characteristic of a bike-sharing service is that products are horizontally differentiated; that is, they are the same in terms of price and quality and are differentiated largely by their proximity to the rider.

We primarily focused on this locational feature to build our demand model, which is applicable to both docked and dockless systems. In particular, we considered a model in which riders arrive within a set of locations inside the service region according to a Poisson process. Upon arrival, a rider makes a choice of which bike to pick (or leaves the system) based on her walking distance to all available bikes, governed by a general choice model. Our goal was to estimate the set of rider locations and their corresponding arrival rates by using only booking and vehicle location and status data.

Our key contributions are summarized as follows.

- First, we studied the statistical properties of the demand model. Assuming that the rider choice model is known, we determined conditions under which the maximum likelihood estimator (MLE) of the location weights is identifiable and consistent. We also determined identifiability properties when riders make bike choices according to specific choice models, including a multinomial logit model and a model based on distance ranking, under both docked and dockless settings.
- Second, we derived an efficient expectation-maximization (EM) algorithm with closed-form updates for the estimation problem. To scale this algorithm and handle situations when the set of potential rider locations is large or even not known a priori, we developed a location-discovery procedure that iteratively explores and adds new rider locations in the service region.
- Lastly, we implemented our algorithms with a set of synthetic data and real dockless bike-sharing data in the Seattle area. These experiments demonstrated the scalability and accuracy of our proposed demand model and its estimation algorithm.

Our report is organized as follows. In Chapter 2, we discuss relevant literature. In Chapter 3, we describe the data generative process, our demand model, and the identifiability and consistency of the MLE. In Chapter 4, we introduce the EM algorithm and the location-discovery procedure. We report on the performance of our demand model and its estimation

algorithm applied to an extensive set of numerical experiments in Chapter 5. The data and code to reproduce all experiments in the report can be found here. All proofs of technical results stated in this report, as well as additional technical results, can be found here.

CHAPTER 2. LITERATURE REVIEW

this chapter briefly reviews related work. Numerous works have analyzed bike sharing usage by incorporating data from heterogeneous sources (see, e.g., Rixey 2013, Singhvi et al. 2015, El-Assi et al. 2017). Among them are demographic characteristics, built environment factors, weather, and usage data from other connecting transportation services (see El-Assi et al. 2017 for a summary of the recent literature and references therein). In contrast, our work was more closely related to the literature about structurally understanding demand censoring and substitution due to the service (un)availability of bike-sharing systems. More closely related to our work were O'Mahony and Shmoys (2015), Mellou and Jaillet (2019), Kabra et al. (2019), Freund et al. (2019) and He et al. (2021). Although the main focuses of O'Mahony and Shmoys (2015), Freund et al. (2019), and Mellou and Jaillet (2019) were on improving the operations of bike allocation and transshipment, they all emphasized the importance of demand correction. Specifically, O'Mahony and Shmoys (2015) and Freund et al. (2019) filtered out time periods when the station ran out of bikes to correct for demand censoring.

Kabra et al. (2019) focused primarily on the question of how the accessibility and availability of a docked bike-sharing service impact ridership. To that end, the authors proposed a structural demand model in which the rider arrival rate at a location was assumed to vary with several covariates, such as the local population density and metro usage. The pick-up model was constructed with a logit choice model in which the utility was a piecewise linear function of walking distance with a break-point at 300 meters.

He et al. (2021) studied the network effect of bike-sharing demand, i.e., riders choosing to pick up a bike because both the origin and destination stations were attractive. They developed an instrumental variable method to tackle the endogeneity issue of choice set in estimating demand, and they applied the method to a London bike share system to estimate the rider demand for network products.

While the existing literature on bike-sharing systems has often explored the impacts of station locations on riders' choices, our study dove deeper into this aspect by concentrating on the dynamics of how the distance between a rider's origin and bike locations influences riders' decision-making process. This allowed us to provide a more systematic understanding of locational choices, which distinguishes our study from others. We looked at a parsimonious and operational setting in which rider demands are inferred solely on the basis of the trip and vehicle

status data that fleet operators collect in their daily operations. We revealed the statistical properties of estimating rider arrival location and intensities and developed scalable estimation procedures that are applicable to both docked and dockless systems. These add to the toolkit of fleet operators and municipal agencies to effectively plan, operate, and monitor micro-mobility systems.

Our work was also closely related to the abundantly available literature on estimating the demand for substitutable products by using choice models based on possibly censored transaction data (Anupindi et al. 1998, Talluri and van Ryzin 2004, Vulcano et al. 2012, Newman et al. 2014, Abdallah and Vulcano 2020). Assuming that customer demand follows a multinomial logit (MNL) choice model, Vulcano et al. (2012) proposed an expectation-maximization (EM) method to estimate customers' demand for substitutable products from (censored) sales transaction data. Newman et al. (2014) developed a two-step strategy that could serve as an alternative to the EM method for parameter estimation. Their approach entailed breaking down the log-likelihood function into separate marginal and conditional components. To improve upon these two methods, Abdallah and Vulcano (2020) proposed a minorization-maximization (MM) algorithm that could achieve a unique global maximum of the log-likelihood function under some data requirements of the transaction data. However, it was less efficient in our setting because of a lack of closed-form updates. Van Ryzin and Vulcano (2014) proposed a market discovery algorithm to estimate customer demand based on ranking preferences. The algorithm started with a small set of customer types and enlarged the set by iteratively generating new customer types (ranking preference), which increased the likelihood value. We drew inspiration from this algorithm in developing our location-discovery procedure, although the underlying generative process and the structure of the subproblem to generate new customer types (rider location in our context) were vastly different.

Structurally speaking, our demand model and the EM algorithm share some similarities with the latent-class logit (LCL) model discussed by Bhat (1997) and Greene and Hensher (2003). A latent class in our setting corresponds to a particular rider location, but unlike the LCL, we do not restrict rider choice behaviors to follow MNL models. Another important distinguishing factor of our model is that the features of different alternatives (walking distances to different bikes) depend on the latent class (rider location), which is assumed to be invariant across latent classes in the LCL. This renders existing identifiability results (e.g., Grün and

Leisch 2008) or enhanced estimation algorithms (e.g., Jagabathula et al. 2020) developed for the LCL inapplicable in our setting.

CHAPTER 3. MODEL AND PRELIMINARIES

In this chapter, we introduce our model and preliminaries. We first discuss our data generative process and describe the observed data from an operator's perspective. Then we describe deriving the likelihood function of our statistical model. We conclude the chapter by discussing various identifiability and consistency results regarding our estimators.

3.1. Generative Process and Data

3.1.1. Generative Process

We consider a set of rider locations $\mathcal{L} := \{1, \cdots, L\}$ distributed over a bounded space $\mathcal{P} \subset \mathbb{R}^2$ from which potential riders come. We assume that riders arrive at the area according to a Poisson process with a total rate λ. For each arriving rider, its arrival location follows a multinomial distribution with probabilities $\mathbf{w} = \{w_1, \cdots, w_L\}$ satisfying $\sum_{l \in \mathcal{L}} w_l = 1$. One can think of this arrival process as one that is dedicated to a particular hour of day or week. We consider a total length of arrival period T, and time runs continuously. We define $\mathcal{B} := \{1, \cdots, B\}$ as the set of bikes in the system. At time $t \in [0, T]$, we define the coordinate of bike $b \in \mathcal{B}$ as $(x_{b,t}, y_{b,t})$. Let $z_{b,t} = 1$ if bike b is available to be booked at time t, and $z_{b,t} = 0$ if bike b is not available at time t because it is booked or occupied. Let $\mathcal{B}_t := \{j \in \mathcal{B} : z_{b,t} = 1\} \subset \mathcal{B}$ be the set of bikes that are available for booking at time t.

When a rider arrives at a location $l \in \mathcal{L}$ at time t, she is presented with a *bike pattern* $S_t := \{(x_{b,t}, y_{b,t})\}_{b \in \mathcal{B}_t}$, which contains the coordinates of all available bikes at time t. This captures both dock-based and dockless systems. Let p_{l,b,S_t} be the probability that a rider at location l chooses bike $b \in \mathcal{B}_t$ at time t. In addition, let $p_{l,0,S_t}$ denote the probability that a rider at location l leaves without choosing a bike at time t. We require that $\sum_{b \in \mathcal{B}_t \cup \{0\}} p_{l,b,S_t} = 1$. One example of such riders' choice behavior is a multinomial logit (MNL) choice model. Walking distance to the bike is arguably the most important feature. Let d_{l,b,S_t} be the distance from rider location $l \in \mathcal{L}$ to bike $b \in \mathcal{B}$ in bike pattern S_t. For each rider location $l \in \mathcal{L}$ and available bike $b \in \mathcal{B}_t$,

$$p_{l,b,S_l} = \frac{\exp\left(\beta_{0,l} + \beta_{1,l} d_{l,b,S_l}\right)}{1 + \sum_{b \in \mathcal{B}_l} \exp\left(\beta_{0,l} + \beta_{1,l} d_{l,b,S_l}\right)}, \; p_{l,0,S_l} = \frac{1}{1 + \sum_{b \in \mathcal{B}_l} \exp\left(\beta_{0,l} + \beta_{1,l} d_{l,b,S_l}\right)},$$

Equation 3.1. Choice probability function for the MNL model

where $\beta_{0,l} \in \mathbb{R}, \beta_{1,l} \in \mathbb{R}_{<0}$ are parameters measuring rider tolerance for walking distance at different locations.

Another example is a distance-ranking choice model in which the preference is determined by the ranking of distances to available bikes. Specifically,

$$p_{l,b,S_t} = \frac{1\{d_{l,b,S_t} \le d_{l,b',S_l}, \forall b' \in \mathcal{B}_t, d_{l,b,S_t} \le \bar{r}_l\}}{\left|\left\{b' \in \mathcal{B}_t : d_{l,b',S_t} = min\{d_{l,b'',S_t} : b'' \in \mathcal{B}_t\}\right\}\right|}, p_{l,0,S_t} = 1\{d_{l,b,S_t} > \bar{r}_l, \forall b \in \mathcal{B}_t\},$$

Equation 3.2. Choice probability function for the ranking model

where $1(\cdot)$ is the indicator function and $\bar{r}_l$ is the consideration radius at rider location l, which specifies how far a rider is willing to walk to a bike. In other words, a rider at location l chooses to pick up an available bike $b \in \mathcal{B}_t$ if and only if bike b is the closest available bike to rider location l and its distance does not exceed $\bar{r}_l$. When a tie occurs, each bike with the shortest distance within the consideration radius has the same probability to be booked by a rider.

Although the model is mainly applicable for a dockless system, it can be well adapted to a dock-based system or a hybrid system in which customers may focus on selecting docks rather than bikes. We accommodate this by assuming a riders' dock-choice model instead. In this choice model, a rider makes a booking choice among docks with at least one available bike according to the distance between the rider and the docks. Additionally, incorporating the number of bikes available at each dock as a covariate in the model may further refine our approach.

3.1.2. Data and Observations

The data available to the operator consists of records of bookings and returns, as well as the statuses and locations of all bikes in real time. The statuses of a bike include "available" or "occupied." When a rider books a bike, the status changes from "available" to "occupied." When she drops off the bike at her destination, the status changes back to "available." As a key feature of our problem, the operator cannot observe riders' arriving locations. As a consequence, the operator cannot distinguish between no arrival and a rider arriving without choosing a bike.

3.2. Model Formulation

Given the above generative process and observations, our goal was to estimate the following two quantities: (1) riders' total arrival rate λ into the service region and (2) the probability/weight vector $\mathbf{w}$ distributed over the set of rider locations $\mathcal{L}$. We derive the likelihood of a given set of observations. To simplify the notation, the *observed* arrival rate is defined as

$$\tilde{\lambda}(t) := \lambda\left(1 - \textstyle\sum_{l \in \mathcal{L}} w_l p_{l,0,S_l}\right)$$

This quantity removes the portion of riders who choose not to pick up any bike upon arrival from the total arrival rate λ. Suppose there are N bookings in total during the arrival period $[0, T]$. We denote the sequence $\mathbf{t} := \{t_n\}_{n=1}^N$, where $t_1, \dots, t_N \in [0, T],\ t_1 < t_2 < \cdots < t_N$ as the time epochs that bookings occur and $\mathbf{b} := \{b_n\}_{n=1}^N$, where $b_1, \dots, b_N \in \mathcal{B}$ as the bikes booked by the riders in the corresponding booking times. Define $t_0 = 0$ and $t_{N+1} = T$. Consider a short enough time period $\delta > 0$ around each booking time $\{t_n\}_{n=1}^N$ such that bike patterns S_t do not change during these intervals $t \in [t_n, t_n + \delta],\ n = 1, \cdots, N$. The incomplete data log-likelihood function is then given by

$$l_I(\boldsymbol{w}, \lambda) = -\int_0^T \tilde{\lambda}(t)dt + \sum_{n=1}^N \log\, \tilde{\lambda}(t_n) + \sum_{n=1}^N \log \frac{\sum_{l\in\mathcal{L}} w_l p_{l,b_n,S_{l_n}}}{1 - \sum_{l\in\mathcal{L}} w_l p_{l,0,S_{l_n}}}$$

The above equation holds because $\int_{t_n}^{t_n+\delta} \tilde{\lambda}\,(t)dt = \tilde{\lambda}(t_n)\delta$, as S_t does not change within $[t_n, t_n + \delta]$. Taking the first-order condition with respect to λ, it can be seen that the total arrival rate λ has a unique closed-form maximizer.

$$-\int_0^T \left(1 - \sum_{l\in\mathcal{L}} w_l p_{l,0,S_l}\right)dt + \sum_{n=1}^N \frac{1 - \sum_{l\in\mathcal{L}} w_l p_{l,0,S_{l_n}}}{\tilde{\lambda}(t_n)} = 0 \Rightarrow \lambda = \frac{N}{\int_0^T \left(1 - \sum_{l\in\mathcal{L}} w_l p_{l,0,S_l}\right)dt}$$

Plugging the closed form of λ into the likelihood function, we can rewrite the incomplete log-likelihood function as a function of $\mathbf{w}$ only,

$$l_I(\boldsymbol{w}) = -N + N\log\, N - N\log \int_0^T \left(1 - \sum_{l\in\mathcal{L}} w_l p_{l,0,S_l}\right)dt + \sum_{n=1}^N \log \left(\sum_{l\in\mathcal{L}} w_l p_{l,b_n,S_{l_n}}\right)$$

Equation 3.3. Log-likelihood function under the generative process

A key quantity involved in the above equation is $\int_0^T \left(1 - \sum_{l\in\mathcal{L}} w_l\, p_{l,0,S_t}\right) dt$. Remarkably, $\int_0^T \left(1 - \sum_{l\in\mathcal{L}} w_l\, p_{l,0,S_t}\right) dt/T$ measures the average percentage of riders who enter the system and pick up a bike. Typically, bike pattern S_t changes at events such as (1) a rider books a bike; (2) a rider drops off her bike at her destination; (3) the operator relocates bikes. On the other hand, for the likelihood function to be valid, there is no requirement for how bike patterns S_t change. Note that because S_t does not change continuously over time, the integral over t can be reorganized as a finite sum. Assume that the pattern changes at time epochs $t'_1, \cdots, t'_Q$, where Q is the total number of changes within $[0, T]$. Then we have $\int_0^T \left(1 - \sum_{l\in\mathcal{L}} w_l\, p_{l,0,S_t}\right) dt = \sum_{q=0}^Q \left(1 - \sum_{l\in\mathcal{L}} w_l\, p_{l,0,S_{t_q}}\right) \left(t'_{q+1} - t'_q\right)$ with $t'_0 = 0$ and $t'_{Q+1} = T$.

The log-likelihood function is non-concave in general. The non-concavity stems from the fact that the operator is not able to observe rider arrival locations. Although it can be rearranged into a difference of two concave functions and further solved by the minorization-maximization (MM) algorithm, numerical experiments (see Chapter 5.1.1) showed that it is computationally expensive in comparison to the EM algorithm. In the case of complete data that allow the operator to observe every rider's arrival and their arriving locations, the likelihood function becomes concave, as shown in Chapter 4.

3.3. Consistency of the Location Weights Estimator

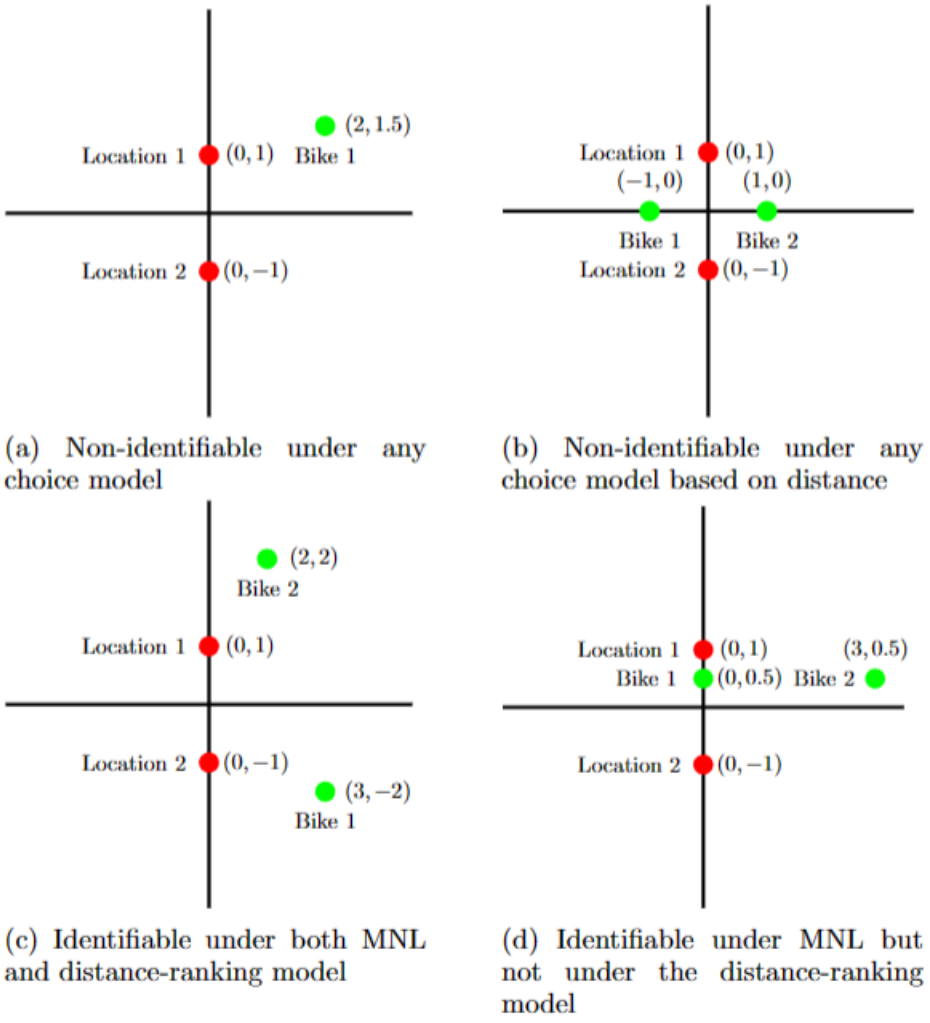
Let $\widehat{\mathbf{w}} \in \operatorname{argmax}_{\mathbf{w}\in\Delta L}\, l_I(\mathbf{w})$ be the MLE. In this subsection, we investigate the consistency of $\widehat{\mathbf{w}}$ — whether the estimator $\widehat{\mathbf{w}}$ converges to true values in an asymptotic limit when the length of the arrival period $T \to \infty$. Note that the consistency of the arrival rate estimator $\hat{\lambda}$ follows directly from the consistency of the location weights.

We first consider the following asymptotic regime when the length of the arrival period T gets large. We assume that there are K bike patterns (K is finite), which is denoted by $\mathcal{S} := \{S_1, \cdots, S_K\}$. We assume that $S_k \neq \emptyset,\ \forall k \in \{1, \cdots, K\}$. The long-run average fraction of time that we observe bike pattern S_k follows $\lim_{T\to\infty} \int_0^T \mathbf{1}\,(S_t = S_k)dt/T = \alpha_k > 0$ for all $k \in \{1, \cdots, K\}$ with $\alpha_1 + \cdots + \alpha_K = 1$. This setting typically models dock-based systems. For example, O'Mahony (2015) and Banerjee et al. (2022) used a continuous-time Markov chain (CTMC) to model state evolution in which each state (bike pattern) was defined as the number of bikes in each station. Then $\{\alpha_1, \cdots, \alpha_K\}$ could be thought of as the steady-state distribution of this CTMC. Let $\mathbf{w}^*$ denote the underlying true weight vector. We first establish the identifiability of our estimator $\widehat{\mathbf{w}}$ (i.e., different parameter values correspond to different data-generating distributions), which is a necessary condition for consistency. Its proof relies on showing that the long-run average expected likelihood function $\lim_{T\to\infty} \mathbb{E}[l_I(\mathbf{w})]$ (the expectation is taken over bookings) has a unique maximizer at $\mathbf{w} = \mathbf{w}^*$, an equivalent condition for identifiability (see Lemma 5.35 in van der Vaart 2000).

3.3.1. Theorem 1

The location weights $\mathbf{w}$ are identifiable if the set of vectors $\{[p_{1,b,S_k}, \cdots, p_{L,b,S_k}]:\ b \in \mathcal{B},\ k \in \{1, \cdots, K\}\}$ spans the vector space $\mathbb{R}^L$. Moreover, the condition becomes necessary and sufficient when $w_l^* = 0$ for at most one $l \in \mathcal{L}$.

Theorem 1 shows that a sufficient condition for identifiability is to have L linearly independent vectors from the vectors of riders' choice probabilities originating from different locations. This condition becomes necessary when the operator has relatively precise prior knowledge of where rider locations are — at most one location can be redundant in the candidate set $\mathcal{L}$. This result also implies that if riders' choice behaviors depend only on distances, a minimum of L distinct bike locations across all bike patterns is required to possibly have identifiability of the estimator. A few concrete examples are provided in Figure 3.1 to illustrate this result.



(a) Non-identifiable under any choice model

(b) Non-identifiable under any choice model based on distance

(c) Identifiable under both MNL and distance-ranking model

(d) Identifiable under MNL but not under the distance-ranking model

Figure 3.1. Examples of identifiability of location weights with only one bike pattern. In the graphs, red circles represent arrival locations and green circles represent bike locations.

Suppose that we have two rider arrival locations, shown in Figure 3.1 (red circles). We consider a case with only one bike pattern $\mathcal{S} = \{S\}$. We adopt Euclidean distance to calculate

walking distances. In Figure 3.1(a), we begin by examining a scenario in which a single bike (green circle) remains fixed at a specific location throughout the arrival period until it is booked. In this scenario, the location weights cannot be consistently estimated with *any* choice model that depends only on distances. Intuitively, riders have only two options: book the bike or leave. Therefore, the only observation here is the booking time, which results in the weights being non-identifiable. On the other hand, when there are two bikes fixed at two distinct locations, then the identification depends on the specification of the choice model. By Theorem 1, the weights can be identified if and only if $p_{1,1,S}p_{2,2,S} \neq p_{1,2,S}p_{2,1,S}$, i.e., vectors $[p_{1,1,S}, p_{2,1,S}]$ and $[p_{1,2,S}, p_{2,2,S}]$ are linearly independent. For the MNL model specified in Equation 3.1, this can be simplified to $\beta_{1,1}(d_{1,1,S} - d_{1,2,S}) \neq \beta_{1,2}(d_{2,1,S} - d_{2,2,S})$, where $d_{l,b,S}$ is the distance from location l to bike b under pattern S. For the distance-ranking choice model specified in Equation 3.2, the necessary and sufficient conditions for the location weights to be identifiable are (i) the closer bike to locations 1 and 2 are different; and (ii) for both locations 1 and 2, at least one bike is within its consideration radius. To ease the presentation, we assume that $\beta_{0,1} = \beta_{0,2}$ and $\beta_{1,1} = \beta_{1,2}$ in the MNL choice model, and the consideration radius r is infinite in the distance-ranking choice model.

Figure 3.1(b) illustrates a scenario in which non-identifiability of location weights occurs under any choice model because bikes 1 and 2 both have the same distance to locations 1 and 2. Figure 3.1(c) shows a scenario in which location weights are identifiable under both the MNL and distance-ranking choice models. Finally, in Figure 3.1(d), location weights are identifiable under the MNL model but not the distance-ranking choice model. This is because in the latter, a rider always chooses bike 1 over bike 2 regardless of where she arrives. It is not hard to prove that in this example of two rider locations and two bike locations, if location weights are identifiable under the distance-ranking model, they must be identifiable under the MNL model as well. Interestingly, this is *not* true in general.

In general, establishing consistency requires not only identifiability but also the uniform convergence of the log-likelihood function. Below are two sufficient conditions that ensure strong consistency of the MLE, that is $\widehat{\mathbf{w}} \to \mathbf{w}^*$ with probability one as $T \to \infty$.

3.3.2. Proposition 1

If the location weights $\mathbf{w}$ are identifiable, then the MLE $\widehat{\mathbf{w}}$ converges to $\mathbf{w}^*$ with a probability 1 if one of the following conditions holds:

1. choice probability $p_{l,b,S_k} > 0$ for all $l \in \mathcal{L},\ b \in \mathcal{B}_k,\ k \in \{1, \cdots, K\}$; or

2. there exists $\epsilon > 0$ such that $w_l^* \geq \epsilon$ for all $l \in \mathcal{L}$.

In Proposition 1, the first condition holds under the MNL choice model specified in Equation 3.1. The second condition holds when the operator has precise knowledge of the set of true locations. Either assumption guarantees a finite log-likelihood value for all $\mathbf{w}$ in a compact parameter space that contains the true location weights $\mathbf{w}^*$, where the log-likelihood function is dominated by an integrable function. Then by the uniform law of large numbers, the uniform convergence of the log-likelihood can be established by the dominance condition, together with the continuity of the log-likelihood function (see Theorem 7.48 in Shapiro et al. 2009).

In most dockless systems, bikes are located at any place where parking is allowed. We thus provide a generalization for scenarios in which bike patterns are continuously distributed over space. Let $(\mathcal{S}, \mathcal{F})$ be a measurable space of bike patterns. Let π and μ both be measures of $\mathcal{F}$. Furthermore, π is an invariant probability measure that is absolutely continuous with respect to μ, which measures the long-run average portion of each bike pattern, $\pi(A) = \lim_{T\to\infty}(1/T)\int_{t=0}^{T}\int_{\mathcal{S}}\mathbb{I}_A\,(S_t)d\mu(S_t)dt,\ \forall A \in \mathcal{F}$. Thus, we have the following corollary regarding the identifiability of the MLE. Consistency can be established in a fashion to that in Proposition 1. We omit the discussion here.

3.3.3. Corollary 1

The location weights $\boldsymbol{w}$ are identifiable if the set of vectors $\{(p_{1,b,S}, \cdots, p_{L,b,S}) : b \in \mathcal{B}, S \in \mathcal{S}'\}$ spans the vector space $\mathbb{R}^L$ for any $\mathcal{S}' \subseteq \mathcal{S}$ such that $\pi(\mathcal{S}') = 1$. The condition becomes necessary and sufficient when we have $w_l^* = 0$ for at most one $l \in \mathcal{L}$.

CHAPTER 4. ESTIMATION PROCEDURES

This chapter shows how we can obtain the set of rider locations and the MLE of their weights $\widehat{\mathbf{w}}$. First, given a set of candidate rider locations, we show how we can computationally approach the MLE by using an expectation-maximization (EM) algorithm with closed-form updates on $\mathbf{w}$. We also show that the parameters in the MNL choice model can be estimated by solving concave programming. We then introduce a location-discovery procedure to iteratively explore new rider locations until convergence.

4.1. Estimation of Location Weights

Given a candidate set of rider locations $\mathcal{L}$, we consider the optimization problem $\max_{\mathbf{w}} l_I(\mathbf{w})$. In this subsection, we develop an EM algorithm with closed-form updates to optimize the weight vector $\mathbf{w}$. The EM algorithm, proposed by Dempster et al. (1977), is an iterative algorithm that consists of an *expectation* step (E-step) and a *maximization* step (M-step) in each iteration. The algorithm computes the conditional expected log-likelihood with respect to unobserved data (E-step) and then updates the estimates by maximizing the expected log-likelihood in the E-step (M-step). This procedure is repeated until convergence. In our case, the *observed* data contain sequences of booking times $\mathbf{t}$, booked bikes $\mathbf{b}$, and bike patterns S_t at any time $t \in [0, T]$. The *unobserved* data include (1) the arrival times of riders who leave without choosing a bike (unobserved riders) and (2) the arrival locations of riders. The total number of arrivals is defined as $\widetilde{N} := N + N'$, where N' is the number of unobserved riders. We first consider the *complete data log-likelihood function*, which is the log-likelihood function derived under the full observation of all arrival times, arrival locations, and booked bikes. Let $\tilde{\mathbf{l}} := \{\tilde{l}_n\}_{n=1}^{\widetilde{N}}$ be the arrival locations of all riders in the sequence, $\tilde{\mathbf{t}} := \{\tilde{t}_n\}_{n=1}^{\widetilde{N}}$ be the arrival times, and $\tilde{\mathbf{b}} := \{\tilde{b}_n\}_{n=1}^{\widetilde{N}}$, $\tilde{b}_n \in \mathcal{B} \cup \{0\}$ be the sequence of bikes booked by the riders. Here, $\tilde{b}_n = 0$ means that the n^{th} rider arrives without picking up a bike. Using these notations, we can write down the complete data log-likelihood function by following a similar procedure as in Equation 3.3.

$$
\begin{aligned}
l_C(\boldsymbol{w}, \lambda) = \quad & \lim_{\delta \downarrow 0} \log \Big(\prod_{n=0}^{\tilde{N}} \mathbb{P}(\text{no rider books bikes from } t_n + \delta \text{ to } t_{n+1}) \\
& \cdot \prod_{n=1}^{\tilde{N}} \frac{\mathbb{P}\big(\text{ a rider arriving at } \tilde{l}_n \text{ books bike } \tilde{b}_n \text{ (or leave) from } t_n \text{ to } t_n + \delta\big)}{\delta} \Big) \\
= \quad & \log\left(\exp\left(-\lambda T\right)\right) + \sum_{n=1}^{\tilde{N}} \log\left(\lambda w_{\tilde{l}_n} p_{\tilde{l}_n, \tilde{b}_n, S_{\tilde{l}_n}}\right) \\
= \quad & -\lambda T + \tilde{N} \log\, \lambda + \sum_{n=1}^{\tilde{N}} \log\left(w_{\tilde{l}_n}\right) + \sum_{n=1}^{N} \log\left(p_{\tilde{l}_n, \tilde{b}_n, S_{\tilde{l}_n}}\right).
\end{aligned}
$$

Similarly, the arrival rate λ can be substituted with the MLE $\hat{\lambda} = \tilde{N}/T$. This simplifies the log-likelihood function to $l_C(\mathbf{w})$, which depends only on the location weights $\mathbf{w}$,

$$
l_C(\boldsymbol{w}) = -\tilde{N} + \tilde{N} \log\left(\tilde{N}/T\right) + \sum_{n=1}^{\tilde{N}} \log\left(w_{\tilde{l}_n}\right) + \sum_{n=1}^{\tilde{N}} \log\left(p_{\tilde{l}_n, \tilde{b}_n, S_{\tilde{l}_n}}\right)
$$

4.1.1. E-Step

In the E-step, we compute the expectation of the complete data log-likelihood conditional on the observed data $\mathbf{b}$ and $\mathbf{t}$ and the current estimate $\mathbf{w}^{(m)}$ at the m^{th} iteration. The expectation is taken over the randomness of unobserved data, which include (1) the number of unobserved riders N', (2) the arrival times of unobserved riders $\tilde{\mathbf{t}} \setminus \mathbf{t}$, and (3) the arrival locations of all riders $\tilde{\mathbf{l}}$.

$$
\mathbb{E}\left[l_C(\boldsymbol{w}) \mid \boldsymbol{b}, \boldsymbol{t}, \boldsymbol{w}^{(m)}\right] = \mathbb{E}\left[-\tilde{N} + \tilde{N} \log\left(\frac{\tilde{N}}{T}\right) + \sum_{n=1}^{\tilde{N}} \log\left(w_{\tilde{l}_n}\right) + \sum_{n=1}^{\tilde{N}} \log\left(p_{\tilde{l}_n, \tilde{b}_n, S_{\tilde{l}_n}}\right) \mid \boldsymbol{b}, \boldsymbol{t}, \boldsymbol{w}^{(m)}\right]
$$

Equation 4.1. The expectation of the complete data log-likelihood

Notice that the only part of the expectation in Equation 4.1 that depends on $\mathbf{w}$ is $\sum_{n=1}^{\tilde{N}} \log\left(w_{\tilde{l}_n}\right)$. Therefore, it is sufficient to focus solely on the conditional expectation $\mathbb{E}\left[\sum_{n=1}^{\tilde{N}} \log\left(w_{\tilde{l}_n}\right) \mid \mathbf{b}, \mathbf{t}, \mathbf{w}^{(m)}\right]$. We denote by $\{l_1, l_2, \cdots, l_N\}$ the sequence of the arrival locations of the observed riders, which is a subsequence of $\{\tilde{l}_1, \cdots, \tilde{l}_{\tilde{N}}\}$, the sequence of arrival locations of all riders. For the sake of exposition, we consider two quantities, which are $\mathbb{P}\left(l_n = l \mid \mathbf{b}, \mathbf{t}, \mathbf{w}^{(m)}\right)$ and $\mathbb{E}\left[N' \mid \mathbf{b}, \mathbf{t}, \mathbf{w}^{(m)}\right]$. The first quantity refers to the probability that the n^{th} observed rider arrives at location $l \in \mathcal{L}$, whereas the second quantity refers to the expected number of unobserved riders. We now separate the conditional expectation $\mathbb{E}\left[\sum_{n=1}^{\tilde{N}} \log\left(w_{\tilde{l}_n}\right) \mid \mathbf{b}, \mathbf{t}, \mathbf{w}^{(m)}\right]$ into two parts, which consist of observed and unobserved data. Let the sequence $\{l'_1, l'_2, \cdots, l'_{N'}\}$ denote the arrival locations of the unobserved riders.

$$
\begin{aligned}
& \mathbb{E}\left[\sum_{n=1}^{\tilde{N}} \log\left(w_{\tilde{l}_n}\right) \mid \boldsymbol{b}, \boldsymbol{t}, \boldsymbol{w}^{(m)}\right] \\
= \; & \mathbb{E}\left[\sum_{n=1}^{N} \log\left(w_{l_n}\right) \mid \boldsymbol{b}, \boldsymbol{t}, \boldsymbol{w}^{(m)}\right] + \mathbb{E}\left[\sum_{n=1}^{N'} \log\left(w_{l'_n}\right) \mid \boldsymbol{b}, \boldsymbol{t}, \boldsymbol{w}^{(m)}\right]
\end{aligned}
$$

$$= \sum_{l\in\mathcal{L}} \left(\sum_{n=1}^{N} \mathbb{P}\left(l_n = l \mid \boldsymbol{b}, \boldsymbol{t}, \boldsymbol{w}^{(m)}\right) + \mathbb{E}\left[N' \mid \boldsymbol{b}, \boldsymbol{t}, \boldsymbol{w}^{(m)}\right] \frac{\int_0^T w_l^{(m)} p_{l,0,s_t} dt}{\int_0^T \sum_{l'\in\mathcal{L}} w_{l'}^{(m)} p_{l',0,s_l} dt} \right) \log\left(w_l\right)$$

$$= \sum_{l\in\mathcal{L}} \left(\sum_{n=1}^{N} \mathbb{P}\left(l_n = l \mid \boldsymbol{b}, \boldsymbol{t}, \boldsymbol{w}^{(m)}\right) + \mathbb{E}\left[N' \mid \boldsymbol{b}, \boldsymbol{t}, \boldsymbol{w}^{(m)}\right] \frac{\int_0^T w_l^{(m)} p_{l,0,s_t} dt}{\int_0^T \sum_{l'\in\mathcal{L}} w_{l'}^{(m)} p_{l',0,s_l} dt} \right) \log\left(w_l\right)$$

$$= \sum_{l\in\mathcal{L}} \underbrace{\left(\sum_{n=1}^{N} \frac{p_{l,b_n,s_{l_n}} w_l^{(m)}}{\sum_{l'\in\mathcal{L}} w_{l'}^{(m)} p_{l',b_n,s_{l_n}}} + N \frac{\int_0^T w_l^{(m)} p_{l,0,s_l} dt}{\int_0^T \left(1 - \sum_{l'\in\mathcal{L}} w_{l'}^{(m)} p_{l',0,s_l}\right) dt} \right)}_{c_l} \log\left(w_l\right)$$

$$= \sum_{l\in\mathcal{L}} c_l \log\left(w_l\right).$$

Equation 4.2. Simplification of the conditional expectation

We have thus simplified the conditional expectation into a weighted sum of logarithms, which can be maximized in closed form in the M-step described below.

4.1.2. M-Step

Let $c := \sum_{l\in\mathcal{L}} c_l$ be the sum of c_l for all $l \in L$. By weighted AM–GM inequality, we have

$$0 = \log\left(\sum_{l\in\mathcal{L}} \frac{c_l}{c} \left(\frac{c w_l}{c_l} \right) \right) \geq \log\left(\prod_{l=1}^{L} \left(\frac{c w_l}{c_l} \right)^{\frac{c_l}{c}} \right) = \sum_{l\in\mathcal{L}} \frac{c_l}{c} \left(\log\left(c w_l\right) - \log\ c_l \right)$$

$$\Leftrightarrow \sum_{l\in\mathcal{L}} c_l \log\left(w_l\right) \leq \sum_{l\in\mathcal{L}} c_l \log\left(c_l\right) - \sum_{l\in\mathcal{L}} c_l \log\left(c\right).$$

The equality holds if and only if $w_l = c_l/c$, which shows that the optimal $\mathbf{w}$ has a closed-form solution $w_l = c_l/(\sum_{l'\in\mathcal{L}} c_{l'})$ for $l \in \mathcal{L}$. We summarize our EM algorithm in Figure 4.1. To simplify the exposition, define $s = \int_0^T (1 - \sum_{l\in\mathcal{L}} w_l\, p_{l,0,s_t}) dt$. Recall that s/T measures the average percentage of riders who arrive and book a bike.

Algorithm 1 EM Algorithm for Location Weights Estimation

Given a set of candidate locations $\mathcal{L}$ with coordinates $(x_1, y_1), \cdots, (x_L, y_L)$.

Initialize the location weights $\boldsymbol{w} \leftarrow \boldsymbol{w}_0 \in \Delta^L$.

while $\boldsymbol{w}$ does not converge **do**

$s \leftarrow \int_0^T \left(1 - \sum_{l\in\mathcal{L}} w_l p_{l,0,s_l}\right) dt.$

$c_l \leftarrow \sum_{n=1}^{N} \left(p_{l,b_n,s_{l_n}} w_l / \left(\sum_{l'=1}^{L} w_{l'} p_{l',b_n,s_{l_n}} \right) \right) + N(T-s)/s$ for all $l \in \mathcal{L}$.

$w_l \leftarrow c_l / \sum_{l'=1}^{L} c_{l'}$ for all $l \in \mathcal{L}$.

end while

Output location weights $\boldsymbol{w}$.

Figure 4.1. EM algorithm for location weights estimation.

Since the closed-form solution $\widehat{\mathbf{w}}$ derived in the M-step is a unique maximizer, in the EM algorithm, we know the M-step generates a sequence of vectors $\{\mathbf{w}^{(m)}, m = 1,2,\cdots\}$ where $\mathbf{w}^{(m)}$ is the unique maximizer for the expected complete log-likelihood function $\mathbb{E}\left[l_C(\mathbf{w}) \mid \mathbf{b}, \mathbf{t}, \mathbf{w}^{(m-1)}\right]$ for $m = 1,2,\cdots$. Here, $\mathbf{w}^{(0)}$ is the initial weight vector to start the EM algorithm. Moreover, from Equation 4, it is clear that the expected complete log-likelihood function $\mathbb{E}\left[l_C(\mathbf{w}) \mid \mathbf{b}, \mathbf{t}, \mathbf{w}^{(m)}\right]$ is continuous in both $\mathbf{w}$ and $\mathbf{w}^{(m)}$. Then by Theorem 2 in Wu (1983), we can infer that the sequence $\{l\left(\mathbf{w}^{(1)}\right), l\left(\mathbf{w}^{(2)}\right), \cdots\}$ converges monotonically to $l(\mathbf{w}^*)$ for some stationary point $\mathbf{w}^*$. Although our EM method does not promise to converge to global maxima, as we will discuss in Chapter 5, the sequence of points estimated by our EM method always converges to a limit point and empirically gives a close approximation to the ground truth.

In some circumstances, the operator may have some prior information about $\mathbf{w}$ (e.g., a rough estimation derived from population density). To incorporate this beforehand, we can use maximum a posteriori (MAP) estimation instead of an MLE estimation. The key difference here is that MAP estimation uses an augmented optimization objective that brings in a prior distribution over the quantity we are looking to estimate. Specifically, let $g(\mathbf{w})$ be the corresponding prior distribution, and we assume that $\mathbf{w} \sim \text{Dirichlet}(\gamma_1, \ldots, \gamma_L)$. In our model, we want to maximize $\widehat{w}_{\text{MAP}} = \text{argmax}_{\mathbf{w}} l_I(\mathbf{w}) + \log g(\mathbf{w}) = \text{argmax}_{\mathbf{w}} l_I(\mathbf{w}) + \sum_{l \in \mathcal{L}} \gamma_l \log w_l$. A key highlight here is that our EM algorithm still has a closed-form update $w_l = (c_l + \gamma_l)/\sum_{l' \in \mathcal{L}}($ $c_{l'} + \gamma_{l'})$, and the update is a simple tweak from the original form, in which we replace c_l with $c_l + \gamma_l$.

Note that a broader class of algorithms for maximizing the log-likelihood used in the related literature is the MM algorithm (Hunter and Lange 2004). MM algorithms maximize simple concave surrogate functions that minorize the log-likelihood function. Different MM algorithms differ in the way the surrogate function is constructed. EM can be viewed as a special case of an MM algorithm using a surrogate function that is constructed by Jensen's inequality (see Chapter 3.1 of Hunter and Lange 2004). In Chapter 5, we compare the performance of our EM algorithm with another MM algorithm that is constructed by viewing the log-likelihood function Equation 3.3 as a difference of two concave functions, and constructing surrogate functions by using supporting hyperplanes of the second concave function. We show that our EM algorithm has superior performance because of its simple closed-form updates.

4.1.3. Estimation of Model Parameters

Our EM algorithm can be adapted to the case in which the rider follows the MNL choice model described in Equation 3.1, and the model parameter $\boldsymbol{\beta}$ is unknown to the operator. In this case, we can jointly estimate $\mathbf{w}$ and $\boldsymbol{\beta}$. Our result below shows that the M-step has the same updating form for $\mathbf{w}$ as that in Algorithm 1, and optimizing $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ in the M-step is a well-behaved concave maximization problem.

4.1.4. Proposition 2

Assume that rider choice behaviors are governed by an MNL model with parameters $\boldsymbol{\beta} := (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)$. Then in the M-step of the EM algorithm, $\mathbf{w}$ has the same closed-form update, and the expected likelihood in the E-step is concave with respect to $\boldsymbol{\beta}$.

4.2. Estimation of Location Coordinates

In practice, the operator often does not have full knowledge of the possible rider arrival locations. As a result, the set of candidate rider locations $\mathcal{L}$ is usually underdetermined and has to be estimated from data. A naïve approach is to enumerate a large set of *all possible* candidate locations, for example, all of the residential buildings in a service area. We then implement our EM algorithm on this set to estimate each location's weight. We refer to this algorithm as the *all-in* algorithm. If the underlying arrival locations are entirely contained in the initial set, then by various results shown in in section 3.2, the MLE can be identifiable and consistent under certain conditions. In such cases, the MLE converges to the ground truth for candidate locations that coincide with the true underlying locations, and 0 for all other locations. However, this algorithm has the following drawbacks because of a large cardinality of all possible rider locations: (1) it significantly increases the number of iterations necessary for convergence, resulting in a computationally demanding task; and (2) it renders the data requirement for identifiability (e.g., Theorem 1) harder to satisfy, making the estimates less accurate.

4.2.1. A Location-Discovery Procedure

Motivated by van Ryzin and Vulcano (2014), who proposed a market-discovery algorithm to estimate ranking-based customer preferences, we develop a *location-discovery* algorithm to address the aforementioned drawbacks of the all-in algorithm. The algorithm iteratively and adaptively explores new locations until convergence.

Suppose $\mathcal{L}$ is the set of all possible rider locations. We begin with a parsimonious set of locations $\mathcal{L}_0 \subset \mathcal{L}$, and a discovery procedure is employed to gradually enlarge our location set.

The main idea is to maximize the log-likelihood by iteratively adding new locations with the largest potential for improvements and then executing the EM algorithm (Algorithm 1) to update their weights.

To start the derivation, we define the restricted Lagrangian function $\Theta^{\mathcal{L}_0}(\mathbf{w}, \mu)$ with location set $\mathcal{L}_0$ by relaxing the equality constraint $\sum_{l\in\mathcal{L}_0} w_l = 1$ with a Lagrangian multiplier μ.

$$\Theta^{\mathcal{L}_0}(\boldsymbol{w}, \mu) = -N \log \int_0^T \left(1 - \sum_{l\in\mathcal{L}_0} w_l p_{l,0,S_l}\right) dt + \sum_{n=1}^{N} \log \sum_{l\in\mathcal{L}_0} w_l p_{l,b_n,S_{l_n}} + \mu\left(1 - \sum_{l\in\mathcal{L}_0} w_l\right).$$

Consider a local optimum $\bar{\mathbf{w}}$ of the problem $\max_{\mathbf{w}\in\Delta^{|\mathcal{L}_0|}} l_I(\mathbf{w})$ with location set $\mathcal{L}_0$. Recall that $s = \int_0^T (1 - \sum_{l\in\mathcal{L}_0} w_l\, p_{l,0,S_t}) dt$. Suppose that $(\bar{\mathbf{w}}, \bar{\mu})$ satisfies Karush-Kuhn-Tucker (KKT) conditions of the restricted Lagrangian function $\Theta^{\mathcal{L}_0}(\mathbf{w}, \mu)$. We have for all $l \in \mathcal{L}_0$ such that $\bar{w}_l > 0$,

$$\left.\frac{\partial \Theta^{\mathcal{L}_0}(\boldsymbol{w}, \mu)}{\partial w_l}\right|_{\boldsymbol{w}=\bar{\boldsymbol{w}}, \mu=\bar{\mu}} = \frac{N}{s}\int_0^T p_{l,0,S_l} dt + \sum_{n=1}^{N} \frac{p_{l,b_n,S_{l_n}}}{\sum_{l\in\mathcal{L}_0} w_l p_{l,b_n,S_{l_n}}} - \bar{\mu} = 0,$$

which gives that

$$\bar{\mu} = \frac{N}{s}\int_0^T p_{l,0,S_l} dt + \sum_{n=1}^{N} \frac{p_{l,b_n,S_{l_n}}}{\sum_{l\in\mathcal{L}_0} w_l p_{l,b_n,S_{l_n}}}, \text{ for any } l \in \mathcal{L}_0 \text{ such that } \bar{w}_l > 0.$$

The essence of our location-discovery procedure is to discover a new location that potentially defines the largest improvement direction for the log-likelihood value. To do so, we consider the full Lagrangian function $\Theta^{\mathcal{L}}(\mathbf{w}, \mu)$ with the full set of all possible rider locations $\mathcal{L}$.

Define $\bar{\bar{\mathbf{w}}} \in \Delta^{|\mathcal{L}|}$ such that $\bar{\bar{w}}_l = \bar{w}_l, \forall l \in \mathcal{L}_0$ and $\bar{\bar{w}}_l = 0, \forall l \in \mathcal{L}\backslash\mathcal{L}_0$. Suppose that

$$\left.\frac{\partial \Theta^{\mathcal{L}}(\boldsymbol{w}, \mu)}{\partial w_l}\right|_{\boldsymbol{w}=\bar{\boldsymbol{w}}, \mu=\bar{\mu}} = \frac{N}{s}\int_0^T p_{l,0,S_l} dt + \sum_{n=1}^{N} \frac{p_{l,b_n,S_{l_n}}}{\sum_{l\in\mathcal{L}_0} w_l p_{l,b_n,S_{l_n}}} - \bar{\mu} \le 0, \forall l \in \mathcal{L} \setminus \mathcal{L}_0$$

then $(\bar{\bar{\mathbf{w}}}, \bar{\mu})$ satisfies the KKT conditions of the full Lagrangian function $\Theta^{\mathcal{L}}(\mathbf{w}, \mu)$ as well. If not, then there exists some $l \in \mathcal{L}\backslash\mathcal{L}_0$ such that $\partial\Theta^{\mathcal{L}}(\mathbf{w}, \mu)/\,\partial w_l > 0$ evaluated at $\mathbf{w} = \bar{\bar{\mathbf{w}}}$ and $\mu = \bar{\mu}$. Note that because the log-likelihood function $l_I(\mathbf{w})$ is not concave in $\mathbf{w}$, KKT conditions are not sufficient for local optimum. This means that including such a new location l is not guaranteed but may lead to an improved estimate with greater likelihood. Nevertheless, this provides a principled procedure for gradually including new locations. We summarize the estimation algorithm with a location discovery procedure in Algorithm 2 in Figure 4.2.

Algorithm 2 Estimation Algorithm with a Location-Discovery Procedure

Initialize a parsimonious location set $\mathcal{L}_0 \subset \mathcal{L}$.

Initialize the location weights vector $\boldsymbol{w}$ by running Algorithm 1 with location set $\mathcal{L}_0$.

while stopping criteria are not met **do**

$s \leftarrow \int_0^T \left(1 - \sum_{l \in \mathcal{L}_0} w_l p_{l,0,S_t}\right) dt.$

$l^* \in \arg\max_{l' \in \mathcal{L} \setminus \mathcal{L}_0} \left(\frac{N}{s} \int_0^T p_{l',0,S_t} dt + \sum_{n=1}^{N} \frac{p_{l',b_n,S_{t_n}}}{\sum_{l \in \mathcal{L}_0} w_l p_{l,b_n,S_{t_n}}} \right).$ ▷ discover a new location

$\mathcal{L}_0 \leftarrow \mathcal{L}_0 \cup \{l^*\}.$

Update $\boldsymbol{w}$ by running Algorithm 1 with location set $\mathcal{L}_0$.

end while

Output the location set $\mathcal{L}_0$ and its corresponding weight vector $\boldsymbol{w}$.

Figure 4.2. Estimation algorithm with a location-discovery procedure

4.2.2. Finding the Location with the Largest Partial derivative of the Lagrangian

At each iteration of Algorithm 2, a new location is selected by solving

$$\max_{l' \in \mathcal{L} \setminus \mathcal{L}_0} \frac{N}{s} \int_0^T p_{l',0,S_t} dt + \sum_{n=1}^{N} \frac{p_{l',b_n,S_{t_n}}}{\sum_{l \in \mathcal{L}_0} w_l p_{l,b_n,S_{t_n}}}$$

Equation 4.35. partial derivative of the full Lagrangian function

to maximize the partial derivative of the full Lagrangian function $\Theta^{\mathcal{L}}(\mathbf{w}, \mu)$. The terms in the objective function have intuitive interpretations. It finds a location that strikes a balance between explaining riders' leaving without picking up bikes (the first term) and the observed booking sequence (the second term). On the basis of current estimates, if s is small or $\sum_{l \in \mathcal{L}_0} w_l \, p_{l,b_n,S_{t_n}}$ is large, then the first term possesses a heavier weight than the second term. The new location l' tends to improve the explanation of riders' leaving behaviors by uplifting the leaving probability $p_{l',0,S_t}$. Conversely, with a large s or small booking probability $\sum_{l \in \mathcal{L}_0} w_l \, p_{l,b_n,S_{t_n}}$, the new location l' focuses more on explaining the booking behaviors by increasing the value of booking probabilities $p_{l',b_n,S_{t_n}}$.

As we will illustrate later, this objective as a function of the new location coordinates $(x_{l'}, y_{l'})$ is not concave in general and may exist with multiple local optima. Therefore, Equation 4.35, unfortunately, does not possess much structure. We could apply a general nonlinear

programming algorithm, such as gradient-based methods, to approach its local optimum. On the other hand, because only two variables, $x_{l'}$ and $y_{l'}$, need to be optimized, an alternative simple and effective method is to use a *grid search*. The grid search method optimizes the objective function in a full factorial sampling plan, which places a grid of evenly spaced points across the search area. We implement a multi-round variant of the search method to speed up the search process. In each round, the granularity of the search grid becomes finer and finer within a shrinking and more targeted search region. More details are provided in Chapter 5.

Another benefit of running a grid-search is to easily enable the *batch* addition of new locations. To be specific, instead of exploring one location at a time, we can simultaneously discover a batch of locations in each iteration. For example, all local maxima, in addition to the global maximum, can be included in each iteration. In the grid search procedure, we identify local maxima by discovering new locations whose partial derivatives are greater than those of eight neighboring locations in the search grid. As we will show later in Chapter 5, batch addition often speeds up the convergence and improves the accuracy of the estimator.

CHAPTER 5. NUMERICAL EXPERIMENTS

In this chapter, we present numerical experiments that were conducted with our demand model and estimation procedures. These are broadly divided into two parts: experiments based on synthetic data and those based on real-world bike-sharing data. In section 5.1, we demonstrate the performance of the algorithm with synthetically generated data on a square. We benchmark the EM algorithm and give evidence that the location-discovery procedure significantly improves the accuracy of the estimator. In section 5.2, we illustrate our algorithm and its estimation results with a set of real-world dockless bike-sharing data in Seattle. We show that our methods produce high accuracy in predicting ridership with out-of-sample tests. In addition, based on the estimation results, we provide managerial insights regarding bike allocations to increase service levels in the Seattle area. All algorithms were implemented in Python 3.8.10 with NumPy 1.23.3 on a virtual machine with 8 vCPUs using Microsoft Azure.

5.1. Experiments Based on Synthetic Data

We describe the generating process of our synthetic data. We consider an arrival period of length T. For each simulation run, we sample the initial locations of each bike (at $t = 0$) independently and uniformly from a 10×10 square. Specifically, the (x, y) coordinates live in $[-5,5]^2$. We then sample a sequence of arrivals based on a homogeneous Poisson process with rate λ within $[0, T]$ representing the rider arrival times. We use a discrete-event simulation that processes arrival times one by one in chronological order. Each rider arrives at a location $l \in \mathcal{L}$ sampled from a multinoulli distribution with probability w_l^* such that $\sum_{l \in \mathcal{L}} w_l^* = 1$. A rider chooses to book an available bike based on an underlying MNL choice model with model parameters $\beta_{0,l} = 1$ and $\beta_{1,l} = -1$ for all $l \in \mathcal{L}$. We use Euclidean distance (the L_2-norm between two location coordinates) as our measurement of distance. When a rider books a bike, we randomly sample her destination uniformly from the 10×10 square. We also generate her booking duration (in hours) according to a rectified Gaussian distribution $\max\{N(\text{walking time} + \text{traveling time},\ 0.1),\ 0.05\}$. Here walking time is computed as the distance from the rider's arrival location to the bike location divided by 4 km per hour (walking speed) and the traveling time is computed as the distance from the bike location to the rider's destination divided by 18 km per hour (cycling speed). If the rider chooses to leave, we simply move on to the next arrival. Bike patterns are updated accordingly. We repeat this procedure until all rider arrivals are processed.

5.1.1. Performance of the EM Algorithm

We first benchmark the performance of our EM algorithm in maximizing the log-likelihood function by comparing it with an MM algorithm described at the end of section 4.1. This approach iteratively optimizes a concave surrogate function, which minorizes the log-likelihood function. In particular, ignoring the constant, we re-arrange the log-likelihood function into a difference of two concave functions $g(\mathbf{w}) - h(\mathbf{w})$ where $g(\mathbf{w}) = \sum_{n=1}^{N} \log \sum_{l \in \mathcal{L}} w_l \, p_{l,b_n,s_{t_n}}$ and $h(\mathbf{w}) = N \log \int_0^T \left(1 - \sum_{l \in \mathcal{L}} w_l \, p_{l,0,s_{t_n}}\right) dt$. To find a concave surrogate, we replace $h(\cdot)$ with its first-order approximation. Then the problem reduces to iteratively solving the following concave program with a simplex constraint

$$\boldsymbol{w}^{(m)} \in \underset{\boldsymbol{w} \geq 0, \sum_{l \in \mathcal{L}} w_l = 1}{\arg\max} \; g(\boldsymbol{w}) - h\left(\boldsymbol{w}^{(m-1)}\right) - \nabla h\left(\boldsymbol{w}^{(m-1)}\right)^T \left(\boldsymbol{w} - \boldsymbol{w}^{(m-1)}\right)$$

Equation 5.1. Objective function for the MM algorithm

Although Equation is convex, it does not have a closed-form solution. Observing that the feasible region of Equation is a simplex, we use the Frank-Wolfe (FW) algorithm (Frank and Wolfe 1956) to solve this concave maximization problem. At each iteration with current estimate $\boldsymbol{w}'$, FW solves a linear program $\underset{\boldsymbol{w} \in \Delta^L}{max} a^T \boldsymbol{w}$, where $a \in \mathbb{R}^L$ is the gradient of the function $g(\boldsymbol{w}) - \nabla h\left(\boldsymbol{w}^{(m-1)}\right)^T \boldsymbol{w}$ evaluated at $\boldsymbol{w} = \boldsymbol{w}'$. It is easy to check that an optimal solution of this linear program is a standard unit vector e_j for some $j \in \arg\max_l a_l$. Thus, each iteration of FW is reduced to a simple findmax operation. The current $\boldsymbol{w}'$ is then updated to $\boldsymbol{w}' + \alpha\left(e_j - \boldsymbol{w}'\right)$, where $\alpha \in [0,1]$ is a step size. To determine α, we perform a line search in $[0,1]$ that maximizes the objective function $g\left(w' + \alpha\left(e_j - w'\right)\right) - \nabla h\left(w^{(m-1)}\right)^T \left(w' + \alpha\left(e_j - w'\right)\right)$. To compare their performances, we position a 5×5 Cartesian grid in the aforementioned 10×10 square and randomly select certain points from the 25 intersections of the Cartesian grid as the underlying true rider arrival locations. Their corresponding weights are independently sampled from a symmetric Dirichlet distribution with parameter 1. We consider the number of ground-truth rider arrival locations and bikes $(L^*, B) \in \{(2,10), (5,20), (10,40)\}$. The length of the arrival period T is set to be 100, and the underlying rider arrival rate is set to be $\lambda = 10$. We consider all 25 intersections on the 5×5 Cartesian grid as the set of potential rider arrival locations $\mathcal{L}$. For both the EM and MM algorithms, we terminate the algorithm when the L_1 norm of the difference of $\mathbf{w}$ between two consecutive iterations is less than 0.0001. To evaluate the accuracy of the estimator,

we compute the Wasserstein distance between the predicted location weights and the true location weights. In our model, given the estimated rider locations and their weights $(\hat{\mathcal{L}}, \hat{\mathbf{w}})$ and the underlying truth $(\mathcal{L}^*, \mathbf{w}^*)$, the Wasserstein 2-distance is defined as

$$W\left((\hat{\mathcal{L}}, \hat{\boldsymbol{w}}), (\mathcal{L}^*, \boldsymbol{w}^*)\right) = \inf_{\lambda_{i,j}} \left\{ \sum_{i=1}^{|\hat{\mathcal{L}}|} \sum_{j=1}^{|\mathcal{L}^*|} \lambda_{i,j} \|\hat{l}_i - l_j^*\|_2^2 : \sum_{i=1}^{|\hat{\mathcal{L}}|} \lambda_{i,j} = w_j^*, \sum_{j=1}^{|\mathcal{L}^*|} \lambda_{i,j} = \hat{w}_i, \lambda_{i,j} \geq 0 \right\}^{1/2},$$

where $\| \hat{l}_i - l_j^* \|_2 := \sqrt{\left(\hat{x}_i - x_j^*\right)^2 + \left(\hat{y}_i - y_j^*\right)^2}$ refers to the Euclidean distance between locations $\hat{l}_i$ and l_j^*.

Table 5.1 reports the number of iterations until convergence, Wasserstein distance (WD) between the predicted location weights and the true weights, the log-likelihood values upon convergence (Lkd), and the CPU times in seconds of the algorithms averaged over 10 simulation runs. When computing the log-likelihood value, we exclude the constant term $-N + N\log N$. We observe that both algorithms produce very similar prediction accuracy, and the converging location weights are in close proximity. However, the EM algorithm requires significantly less computation time; it benefits from closed-form updates, which make each iteration much faster than the MM algorithm. On the other hand, the MM algorithm requires fewer iterations, but each iteration is more expensive, as it requires tuning a step size by a line search.

Table 5.1. Performance comparison of the EM and MM algorithms in estimating location weights.

	EM algorithm				MM algorithm			
(L^*, B)	Iterations	WD	Lkd	Time	Iterations	WD	Lkd	Time
2,10	1,236.2	4.10	-465	0.94	202.0	4.06	-465	4.08
5,20	595.0	4.44	-2,411	4.18	294.6	4.44	-2,411	13.40
10,40	350.9	3.45	-4,533	8.09	303.4	3.45	-4,533	21.96

5.1.2. Performance of the Location-Discovery Procedure

We now evaluate the performance of the location discovery procedure. As in section 5.1.1, we randomly generate rider arrival locations and their corresponding weights on Cartesian grids of 10 x 10 in the same square service region. Rider arrival locations can only be sampled from the intersections of the Cartesian grids. For each grid size, we consider two scenarios with

different numbers of ground-truth rider locations and bikes: $(L^*, B) \in \{(5,20), (10,40), (25,100)\}$. We test with different amounts of data by setting the length of the arrival period to $T \in \{100,500\}$. We compare the following methods.

- **All-in algorithm.** We choose all 100 intersections of the 10 x 10 grid as the set of candidate rider locations L, respectively. We then run Algorithm 1 to obtain their location weights.
- **Location-discovery algorithm.** We start with two rider arrival locations sampled uniformly from the service region. We implement two variants: a single mode and a batch mode. In the single mode, we begin our search by initializing a coarse Cartesian grid of dimensions 10×10 onto the square service region. We find the rider location in the grid that maximizes the partial derivative. We then perform a second grid search that is confined to a smaller square region whose boundary is defined by the neighboring locations of the one selected from the first round. We again overlay a 10×10 Cartesian grid onto this smaller square region and identify the location that has the largest partial derivative. In the batch mode, we select all local maxima in the first round, and a second grid search is conducted near all locations selected in the first round. We use the following two stopping criteria for the location-discovery algorithm: (1) we compute the Bayesian Information Criteria (BIC) value in each iteration, and the algorithm stops when the BIC value increases; (2) we use the first 80 percent of the data generated from the arrival period for training and use the remaining 20 percent out-of-sample data for validation. We compute the out-of-sample likelihood in each iteration. In the single mode, the algorithm stops if the likelihood decreases consecutively in two successive iterations. In the batch mode, the algorithm stops as soon as there is any decrease in the likelihood.
- ***K*-means algorithm.** We also test a simple baseline method based on K-means clustering that partitions all bike booking locations into different clusters, and each cluster centroid is considered to be a rider arrival location. Location weights are then computed by normalizing the number of bookings belonging to each cluster so that they sum up to 1. We also give this K-means clustering some advantages by assuming that the number of underlying rider arrival locations L^* is known beforehand. That is, we set the number of clusters to $K = L^*$.

Table 5.2 reports the performances of different algorithms by averaging over 50 simulation runs. We exclude all rider locations with predicted weights smaller than 0.01 to ensure numerical stability. The table includes several metrics: "Locs" refers to the number of predicted locations; "WD" refers to the Wasserstein distance between the predicted location weights and the underlying true location weights; "Lkd" and "BIC" refer to the log-likelihood and BIC values over the entire arrival period, respectively. We also report computation time measured in seconds. We have the following key observations.

- All algorithms significantly outperform the baseline clustering algorithm in terms of the Wasserstein distance and the BIC value.
- The all-in algorithm significantly overestimates the number of rider locations. This leads to worse predictive accuracy in terms of Wasserstein distance and BIC value. Another observation about the all-in algorithm is that the Wasserstein distance does not significantly decrease as one uses more data, which is likely related to stricter identifiability conditions caused by a large set of candidate locations (Theorem 1). In contrast, our location discovery algorithms, especially the one with batch addition, achieve improved results with larger sample sizes, as demonstrated by the significantly lower Wasserstein distance obtained for $T = 100$ in comparison to $T = 500$.
- The location-discovery algorithms have overall the best performance, and the batch mode improves over the single mode as evidenced by the lower Wasserstein distance and similar BIC values. We observe that the batch mode discovers more locations than the single mode under both stopping criteria. The location-discovery procedure stops when the discovered location does not significantly improve the likelihood. The batch mode usually continues longer than the single mode, as it can identify multiple locations simultaneously in each iteration, thereby increasing the chance of discovering good locations.
- Stopping the location-discovery algorithm using validation likelihood leads to a lower Wasserstein distance in comparison to BIC, albeit at the expense of discovering more locations and requiring more CPU time.

Table 5.2. Prediction performance when the set of candidate rider locations is unknown.

T	(L^*, B)	Algorithm	Locs	WD	Lkd	BIC	Time	Algorithm	Locs	WD	Lkd	BIC	Time
100	(5,20)	K-means	5.0	2.87	-2,267	2,282	<0.01	All-in	25.6	3.10	-2,146	2,219	2.38
	(10,40)		10.0	2.37	-4,628	4,659	0.02		27.0	2.48	-4,503	4,589	11.83
	(25,100)		24.9	1.91	-7,438	7,522	0.22		31.0	1.65	-7,334	7,439	43.39
500	(5,20)		5.0	2.89	-12,839	12,857	0.01		27.8	3.05	-12,234	12,335	14.91
	(10,40)		10.0	2.37	-26,726	26,766	0.11		29.7	2.51	-26,100	26,218	54.19
	(25,100)		25.0	1.94	-43,179	43,283	1.21		35.0	1.76	-42,666	42,812	150.84
100	(5,20)	Loc. Disc. (Single)	6.4	3.02	-2,158	2,176	1.09	Loc. Disc. (Batch)	8.5	2.74	-2,150	2,174	1.58
	(10,40)		8.5	2.57	-4,519	4,546	5.29		12.1	2.23	-4,505	4,543	7.76
	(25,100)		11.6	2.15	-7,351	7,390	24.23		15.5	1.76	-7,337	7,389	35.47
500	(5,20)		7.8	2.92	-12,251	12,279	9.19		10.0	2.64	-12,228	12,264	11.86
	(10,40)		10.1	2.35	-26,123	26,163	44.10		14.6	2.02	-26,087	26,145	59.39
	(25,100)		12.9	1.90	-42,676	42,730	219.98		18.4	1.58	-42,643	42,719	249.27
100	(5,20)	Loc. Disc. (Single, Valid Lkd.)	6.4	3.02	-2,158	2,176	1.09	Loc. Disc. (Batch, Valid Lkd.)	10.4	2.59	-2,152	2,182	2.08
	(10,40)		8.5	2.57	-4,519	4,546	5.29		14.9	2.19	-4,515	4,563	10.30
	(25,100)		11.6	2.15	-7,351	7,390	24.23		20.7	1.64	-7,338	7,408	53.08
500	(5,20)		7.8	2.92	-12,251	12,279	9.19		11.8	2.61	-12,235	12,279	12.83
	(10,40)		10.1	2.35	-26,123	26,163	44.10		16.7	1.95	-26,097	26,164	64.00
	(25,100)		12.9	1.90	-42,676	42,730	219.98		22.5	1.43	-42,647	42,741	313.98

In Figure 5.1, we visualize an instance of the predicted locations and their corresponding weights with $L^* = 10,\ B = 40$, and $\lambda = 10$ under the K-means algorithm and the location-discovery algorithm with batch mode when $T = 100$ and $T = 2{,}000$. We observe that as T

increases from 100 to 2,000, the predicted locations in the discovery algorithm approach the true locations more closely, whereas the K-means algorithm does not exhibit any significant improvement.

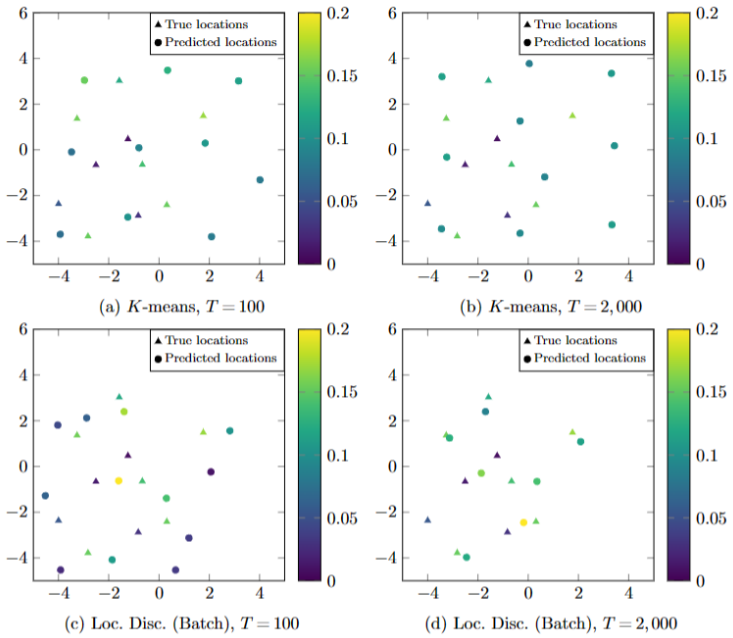


Figure 5.1. Predicted and true rider locations and their corresponding weights. We use triangles to represent the underlying true locations and circles to represent the predicted locations.

Finally, Figure 5.2 displays the objective values of all locations in the square region during the first six iterations of the location discovery algorithm with batch addition, on an instance with $B = 40,\ L^* = 10,\ \lambda = 10$ and $T = 100$. The peak values undergo a sharp slump in the first three iterations and then fluctuate around 1,000. Moreover, iteration by iteration, the function becomes more multi-modal — the first graph depicts only two local maxima whereas the last graph exhibits more than five local maxima.

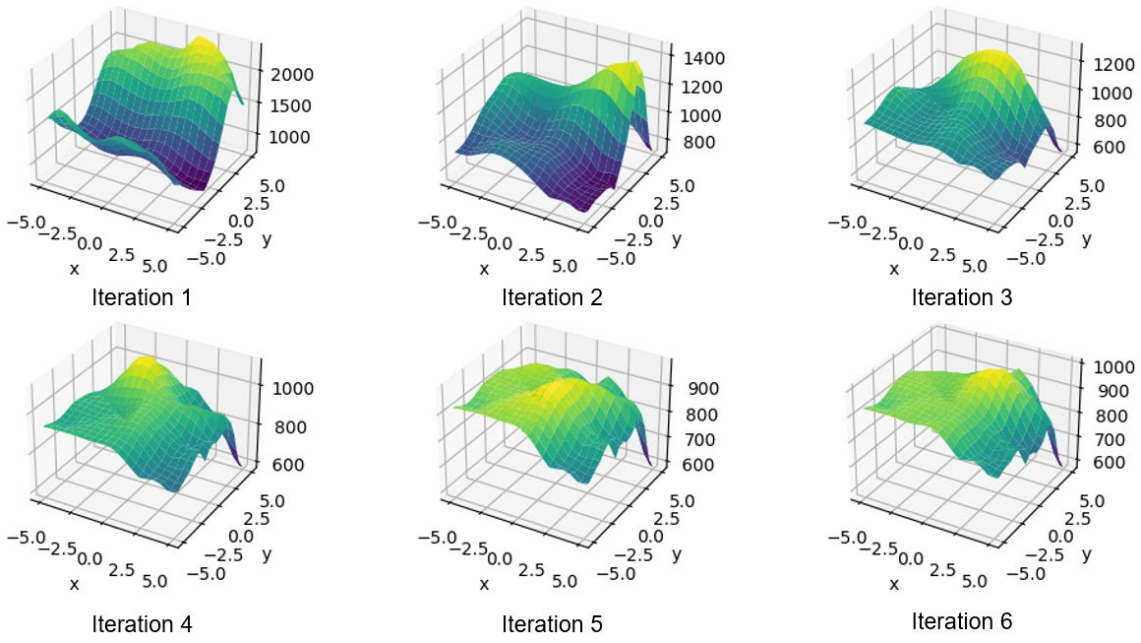


Figure 5.2. Objective values of all locations in the square service region at each iteration of the location discovery algorithm with batch addition.

5.2. Experiments Based on Seattle Bike-Sharing Data

We ran experiments using data from Seattle's dockless bike-sharing system. The data set recorded all bookings and real-time bike locations and statuses from a dockless bike-sharing company in the Seattle region during July and August 2019. The data set and code can be accessed here. We confined our analysis to the downtown Seattle area (see Figure 5.3 for the service region) and looked at a fixed 30-minute period in the morning rush hour (from 8:00 am to 8:30 am) every day. There were 4,024 bookings in total and 2,317 bikes.

We use an MNL model to fit rider choice behaviors. Kabra et al. (2019) estimated a structural demand model using Paris bike-sharing data. Their user choice model was specified as an MNL model with walking distance (in km) as the feature and time and location as fixed effects. They specified the disutility of walking distance by using a piecewise linear structure in which the coefficient of walking distance was estimated to be -2.229 for walking distances of less than 300 meters and -15.445 for walking distances of greater than 300 meters. In our model, we assume that $\beta_{0,l}$ and $\beta_{1,l}$ have the same value across all rider locations l. We set $\beta_{0,l} = 1$, and we estimate $\beta_{1,l}$ with an initial value of -3 in the EM algorithm.

To fit the model, we use a location-discovery algorithm of batch addition with a two-round grid search each with a granularity of 20 x 20 for the first round and 10 x 10 for the second round. We initialize the location-discovery algorithms by randomly generating 20 coordinates on the service region as initial locations. We start estimating β_1 when we have discovered at least 60 locations, as our service region comprised approximately 60 census tracts.

We partition our data into training and testing sets. The training and testing sets consist of bookings ranging from July 1st to July 31st and from August 1st to August 15th, respectively. We further split the training data into two datasets to guide the stopping criterion. Specifically, we use the bookings ranging from July 1st to July 24th to train the model, and the bookings from July 25th to July 31st as a validation set. We terminate the discovery algorithm as long as the validation likelihood decreases. We implement the aforementioned algorithms for the training set and evaluate their performances on the testing set. Based on our predictions of locations $\hat{\mathcal{L}}$ and weights $\widehat{\mathbf{w}}$ using the training set, the arrival rate $\hat{\lambda}$ can be estimated as $\hat{\lambda} = N_{\text{train}} / \int_0^{T_{\text{train}}} \left(1 - \sum_{l \in \hat{\mathcal{L}}} \widehat{w}_l \, \hat{p}_{l,0,S_t}\right) dt$, where N_{train} is the number of bookings observed in the training set and T_{train} is the length of the training period. Let T_{test} be the length of the testing period. Then an estimate of the number of bookings in the testing period $\widehat{N}_{\text{test}}$ can be computed as $\widehat{N}_{\text{test}} = \hat{\lambda} \int_{T_{\text{train}}}^{T_{\text{train}}+T_{\text{test}}} \left(1 - \sum_{l \in \hat{\mathcal{L}}} \widehat{w}_l \, \hat{p}_{l,0,S_t}\right) dt$. We measure the predictive accuracy in the testing set by comparing the actual bookings with the predicted bookings at the census tract level. Let C be the set of census tracts. The predicted number of bookings in census tract $c \in C$ can be expressed as $\widehat{N}_{\text{test},c} = \hat{\lambda} \int_{T_{\text{train}}}^{T_{\text{train}}+T_{\text{test}}} \left(1 - \sum_{l \in \hat{\mathcal{L}}} \sum_{b \in \mathcal{B}_{t,c}} \widehat{w}_l \, \hat{p}_{l,b,S_t}\right) dt$, where $\mathcal{B}_{t,c}$ is the indexed set of available bikes in tract c at time t. We use the weighted mean absolute percentage error (WMAPE) to evaluate the performance, where the weights are given by the booking volume in each test period. Specifically, this quantity can be computed as $WMAPE = \frac{1}{\sum_{c \in C} N_{\text{test},c}} \sum_{c \in C} \left| N_{\text{test},c} - \widehat{N}_{\text{test},c} \right|$.

Table 5.3 reports the performance of the discovery algorithm. The estimated value for β_1 is -4.4. We observe a relatively small WMAPE for both training and testing data. Moreover, our implementation of the discovery algorithm does not require any additional data sources, such as point-of-interest and social-economic data. The nonparametric nature of the location-discovery procedure also makes it robust for modeling misspecification. This makes the discovery algorithm a natural fit for estimating inputs for downstream decision making. For example,

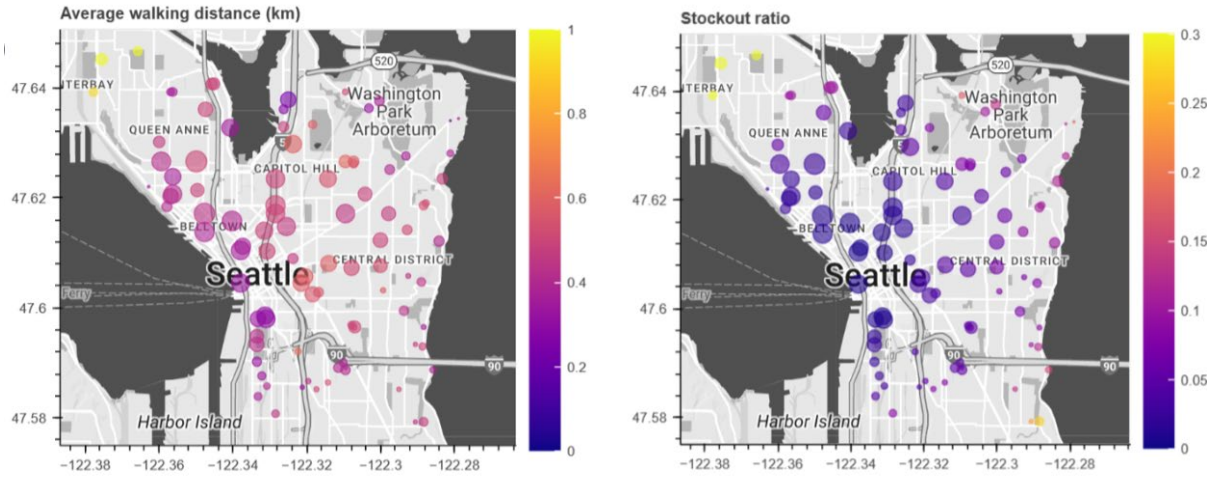
predicted locations could be naturally used in Luo et al. (2022), who discussed free-floating bike allocation and rebalancing problems faced by Mobike, the largest dockless bike-sharing company in China.

Table 5.3. Prediction performance of the location-discovery algorithm with Seattle data.

	Locs	Lkd	WMAPE (%)	Time (min:sec)
Train	115	-47,426	15.93	191:54
Test	-	-24,088	25.35	-

Figure 5.3 shows the predicted locations as well as two important service-level metrics based on the estimation results from the location-discovery algorithm based on all data in July 2019. We note that booking frequencies are comparatively low in the areas of the Washington Park Arboretum and Queen Anne. Figure 5.3(a) shows the average walking distance at each rider's arrival location conditional on riders choosing to pick up bikes. Formally, for each location $l \in \mathcal{L}$, this quantity can be computed as $(1/N) \cdot \sum_{n=1}^{N} \sum_{b=1}^{B} \hat{p}_{l,b,t_n} d_{l,b,t_n} / \left(1 - \hat{p}_{l,0,t_n}\right)$. Figure 5.3(b) depicts the stockout ratio of each location, which is the probability that an arriving rider chooses to leave without picking up a bike. These two metrics do not have to be perfectly (positively) correlated. For example, according to Figure 5.3(b), the arrival locations near Lake Washington (the lake in the rightmost of the figure) have a moderate walking distance but a high stockout ratio. This likely suggests that there is a scarcity of bikes but once there is a bike, it is often close to the rider's location.

Some operational insights can be inferred from these results. The stockout ratio visualization shows the sufficiency of bikes in the downtown area. However, there may exist a shortage of bikes between Queen Anne and Interbay (upper left corner) where the discovery algorithm detects that certain locations are more prone to stockouts.



(a) Avg. dist. for discovery algorithm

(b) Stockout ratio for discovery algorithm

Figure 5.3. The average walking distance and bike stockout ratio for each arrival location under the location-discovery algorithm. Different colors represent different average walking distances (left) or stockout ratios (right). Each circle represents a discovered location. For a clear view, the radius of each circle is computed as a log-linear transformation of the weight in that location (a larger radius corresponds to a larger weight). Specifically, the radius for location l in the plot is $\log(\widehat{w}_l) + 9$.

CHAPTER 6. CONCLUSION

In this report, we present a locational demand model for a bike-sharing system that aims to recover rider arrival locations and their corresponding arrival intensities by using only booking and vehicle availability data. We give conditions under which our model is identifiable and devise an efficient EM algorithm for estimation. This EM algorithm is further complemented with a location-discovery procedure for efficient implementation in large metropolitan areas. It is worth noting that our model is also suitable in other contexts such as ridesharing, car rental or retail where customers originate over space and the locational aspect of demand is pronounced.

Future studies can be directed toward addressing the following gaps. First, as recognized by He et al. (2021) and Kabra et al. (2019), endogenity issues may be introduced by unobservable factors, such as the political importance of each arrival location affecting its nearby bike availability. Features such as walking distance to the closest bike thus can be correlated with these unobserved factors. In addition, our experiments can be enhanced by including additional features, such as battery usage and the maintenance status of the bikes, that help to explain rider choice behaviors. In a docked-based or hybrid system, features such as the number of available bikes in a dock or convenience of the drop-off location (docks are needed for a dock-based system but not for a free-floating system) can also influence rider decisions.

CHAPTER 7. REFERENCES

Abdallah, T., and Vulcano, G. (2020). Demand Estimation under the Multinomial Logit Model from Sales Transaction Data. Manufacturing & Service Operations Management, 23(5), 1196-1216.

Anupindi, R., Dada, M., and Gupta, S. (1998). Estimation of Consumer Demand with Stock-Out Based Substitution: An Application to Vending Machine Products. Marketing Science, 17, 406-423.

AW, v. d. (2000). Asymptotic statistics. Cambridge university press.

Banerjee, S., Freund, D., and Lykouris, T. (2022). Pricing and optimization in shared vehicle systems: An approximation framework. Operations Research, 70(3), 1783-1805.

Bhat, C. (1997). An Endogenous Segmentation Mode Choice Model with an Application to Intercity Travel. Transportation Science, 31(1), 34-48.

Bureau of Transportation Statistics (2023). 2010 Census Tract Seattle - Population Statistics. Retrieved 5 4, 2023, from https://data.bts.gov/stories/s/Bikeshare-and-e-scooters-in-the-U-S-/fwcs-jprj/

Dempster, P.A., Laird, M.N., and Rubin, B.D. (1977). Maximum Likelihood from Incomplete Data via the {EM} Algorithm. Journal of the Royal Statistical Society. Series B (Methodological), 39(1), 1-38.

El-Assi, W., Mahmoud, M. S., and Habib, K. N. (2017). Effects of Built Environment And Weather on Bike Sharing Demand: A Station Level Analysis of Commercial Bike Sharing In Toronto. Transportation, 44(3), 589-613.

Open Mobility Foundation, (2022). Mobility-Data-Specification: A Data Standard to Enable Right-of-Way Regulation and Two-Way Communication between Mobility Companies and Local Governments. Retrieved 11 29, 2022, from https://github.com/openmobilityfoundation/mobility-data-specification/

Frank, M., and Wolfe, P. (1956). An algorithm for quadratic programming. Naval Research Logistics Quarterly, 95-110.

Freund, D., Henderson, S. G., and Shmoys, D. B. (2019). Sharing Economy: Making Supply Meet Demand. Springer.

Gaur, V., and Honhon, D. (2006). Assortment planning and inventory decisions under a locational choice model. Management Science, 52(10), 1528-1543.

GeoData, S. (2020). 2020 Census Tract Seattle. Retrieved 3 17, 2023, from https://data-seattlecitygis.opendata.arcgis.com/datasets/SeattleCityGIS::2020-census-tract-seattle-redistricting-data-1990-2020/

Greene, W. H., and Hensher, D. A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. Transportation Research Part B: Methodological, 27(8), 681-698.

Grun, B., and Leisch, F. (2008). Identifiability of finite mixtures of multinomial logit models with varying and fixed effects. Journal of Classification, 25(2), 225-247.

He, P., Zheng, F., Belavina, E., and Girotra, K. (2021). Customer preference and station network in the {London} bike share system. Management Science, 67(3), 1392-1412.

Hotelling, H. (1929). Stability in competition. The Economic Journal, 39(153), 41-57.

Hunter, D. R., and Lange, K. (2004). A tutorial on MM algorithms. The American Statistician, 58(1), 30-37.

Hunter, D.R., and Lange, K. (2000). Optimization Transfer Using Surrogate Objective Functions: Rejoinder. Journal of Computational and Graphical Statistics, 9(1), 52-59.

Jagabathula, S., Subramanian, L., and Venkataraman, A. (2020). A conditional gradient approach for nonparametric estimation of mixing distributions. Management Science, 66(8), 3635--3656.

Kabra, A., Belavina, E., and Girotra, K. (2019). Bike-share systems: Accessibility and availability. Management Science, 66(9), 3803-3824.

Lancaster, K.J. (1975). Socially Optimal Product Differentiation. American Economic Review, 65(4), 567-585.

Lancaster, K. J. (1966). A new approach to consumer theory. Journal of Political Economy, 74(2), 132-157.

Luo, X., Li, L., Zhao, L., and Lin, J. (2022). Dynamic Intra-Cell Repositioning in Free-Floating Bike-Sharing Systems Using Approximate Dynamic Programming. Transportation Science, 56(4), 799-826.

Mellou, K., and Jaillet, P. (2019). Dynamic resource redistribution and demand estimation: An application to bike sharing systems. Available at SSRN 3336416.

Newman, J., Ferguson, M., Garrow, L., and Jacobs, T. (2014). Estimation of choice-based models using sales data from a single firm. Manufacturing & Service Operations Management, 16(2), 184-197.

O'Mahony, E. (2015). Smarter tools for (Citi) bike sharing. Cornell University.

O'Mahony, E., and Shmoys, D. B. (2015). Data Analysis and Optimization for (Citi) Bike Sharing. (pp. 687-694). Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence.

Rixey, R. A. (2013). Station-level forecasting of bikesharing ridership: Station network effects in three US systems. Transportation research record, 2387(1), 46-55.

Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 461-464.

Seattle Department of Transportation (2022). Scooter Share Data & Permitting. Retrieved 05 14, 2023, from {https://www.seattle.gov/transportation/projects-and-programs/programs/new-mobility-program/scooter-share/

Shapiro, A., Dentcheva, D., and Ruszczynski, A. (2009). Lectures On Stochastic Programming: Modeling and Theory. SIAM.

Singhvi, D., Singhvi, S., Frazier, P. I., Henderson, S. G., O'Mahony, E., Shmoys, D. B., and Woodard, D. B. (2015). Predicting Bike Usage for New York City's Bike Sharing System. AAAI Workshop On Computational Sustainability.

Talluri, K., and van Ryzin, G. (2004). Revenue management under a general discrete choice model of consumer behavior. Management Science, 50(1), 15-33.

Transportation Research Board. (2010). NCHRP Report 672 - Roundabouts: An Informational Guide - 2nd Edition. Washington, D.C.: Transportation Research Board of the National Academies.

van Ryzin, G. and Vulcano, G. (2014). A Market Discovery Algorithm to Estimate a General Class of Nonparametric Choice Models. Management Science, 61(2), 281-300.

Vulcano, G., van Ryzin, G., and Ratliff, R. (2012). Estimating Primary Demand for Substitutable Products from Sales Transaction Data. Operations Research, 60(2), 313-334.

Wu, Jeff C. F.

. (1983). On the Convergence Properties of the EM Algorithm. The Annals of Statistics, 11(1), 95–103. http://www.jstor.org/stable/2240463