



NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

**ONLINE LEARNING OF MARKOVIAN SYSTEMS
WITH CENSORED POISSON ARRIVALS**

by

Cedric G. Gibbons Mac-Lean

June 2021

Thesis Advisor:

Roberto Szechtman

Co-Advisor:

James Grant,

Lancaster University

Second Reader:

Javier Salmeron-Medrano

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			<i>Form Approved OMB No. 0704-0188</i>	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC, 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE June 2021	3. REPORT TYPE AND DATES COVERED Master's thesis	
4. TITLE AND SUBTITLE ONLINE LEARNING OF MARKOVIAN SYSTEMS WITH CENSORED POISSON ARRIVALS			5. FUNDING NUMBERS	
6. AUTHOR(S) Cedric G. Gibbons Mac-Lean				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.			12b. DISTRIBUTION CODE A	
13. ABSTRACT (maximum 200 words) This thesis deals with online optimization of discrete performance measures in Markovian models with incomplete information. We consider a setting where a physical realization of the model is sequentially obtained over a number of periods. The information gathered to date is used in order to efficiently run the model in future days. The information is incomplete in two ways: (i) model parameters are initially unknown (the demand rates in our case), but can be estimated from the physical realizations; and (ii), the demands are censored when the system is in some boundary states. The method of Sample Average Approximation is used to solve the optimization problem. More precisely, in each period, sample paths are generated from the distributions estimated to date, and the best model configuration is determined with respect to these sample paths. Sequential observation of the system's behavior allows for information to be gathered and a more informed decision to be made in each future round. The method developed in this thesis can be applied in a variety of contexts, where no information is known about the system beforehand, but can be observed at least partially in a sequential manner, such as assigning assets for surveillance of remote geographical regions for illicit activity. The motivating setting of this work is the operation of a bike-sharing system with fixed capacity stations, where an initial number of bikes must be set each day to minimize unsatisfied customers.				
14. SUBJECT TERMS bike-sharing system, online learning, machine learning			15. NUMBER OF PAGES 79	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UU	

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

**ONLINE LEARNING OF MARKOVIAN SYSTEMS WITH CENSORED
POISSON ARRIVALS**

Cedric G. Gibbons Mac-Lean
Lieutenant, Chilean Navy
BSEE, Académia Politécnica Naval, 2011

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

**NAVAL POSTGRADUATE SCHOOL
June 2021**

Approved by: Roberto Szechtman
Advisor

James Grant
Co-Advisor

Javier Salmeron-Medrano
Second Reader

W. Matthew Carlyle
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This thesis deals with online optimization of discrete performance measures in Markovian models with incomplete information. We consider a setting where a physical realization of the model is sequentially obtained over a number of periods. The information gathered to date is used in order to efficiently run the model in future days. The information is incomplete in two ways: (i) model parameters are initially unknown (the demand rates in our case), but can be estimated from the physical realizations; and (ii), the demands are censored when the system is in some boundary states. The method of Sample Average Approximation is used to solve the optimization problem. More precisely, in each period, sample paths are generated from the distributions estimated to date, and the best model configuration is determined with respect to these sample paths. Sequential observation of the system's behavior allows for information to be gathered and a more informed decision to be made in each future round.

The method developed in this thesis can be applied in a variety of contexts, where no information is known about the system beforehand, but can be observed at least partially in a sequential manner, such as assigning assets for surveillance of remote geographical regions for illicit activity. The motivating setting of this work is the operation of a bike-sharing system with fixed capacity stations, where an initial number of bikes must be set each day to minimize unsatisfied customers.

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

1	Introduction	1
1.1	Motivation	1
1.2	Scope, Limitations, and Assumptions	2
1.3	Research Questions	3
1.4	Thesis Outline	3
2	Literature Review and Background	5
2.1	Background on Bike Sharing Systems	5
2.2	Literature Review	5
2.3	Contribution to Existing Literature	11
3	Analytic Models	13
3.1	Introduction	13
3.2	Rates Estimation	15
3.3	Convergence to the Optimal Solution	22
3.4	Analyses Left for Future Work	25
4	Analysis and Simulation Results	27
4.1	Introduction	27
4.2	Fixed Bins and Piece-Wise Constant Rates	30
4.3	Fixed Bins and Smooth Rates	39
4.4	Adaptive Bins and Smooth Rates	40
4.5	Reuse of Samples	44
5	Conclusion and Recommendations	49
5.1	Conclusion.	49
5.2	Future Work	50

Appendix:	53
A.1 Fixed Time Bin Algorithm	53
A.2 Rate Estimation Fixed Bin.	55
A.3 Adaptive Bin Size Algorithm	56
A.4 Rate Estimation Using CIF	57
A.5 Simulate Customer Pick-Up and Drop-Off Flow	58
 List of References	 59
 Initial Distribution List	 61

List of Figures

Figure 1	Average cumulative regret observed as a result for variants of the proposed model.	xiii
Figure 2.1	Graphical representation of the Markov chain model used by Raviv and Kolka (2013) to represent the bike station and demands. . . .	7
Figure 4.1	Customer demand rate	28
Figure 4.2	Expected number of unsatisfied customers for each initial number of bikes assuming demand rates in equation 4.1.	29
Figure 4.3	Piecewise constant customer demand rates approximating baseline scenario defined by equation 4.1.	30
Figure 4.4	Rent (Blue) and Return (Red) Demand rate estimate by days 1, 10, 100, and 1000	34
Figure 4.5	$\log_{10}(RMSE)$ of the estimation of demand rates.	35
Figure 4.6	Effect of censoring on demand rate estimation, shaded areas are the corresponding probabilities of rack being empty (blue) or full (red).	36
Figure 4.7	Distribution of computational budget in planned time horizon.	37
Figure 4.8	Cumulative regret for various budget distributions.	37
Figure 4.9	Cumulative regret for different total computational budgets	38
Figure 4.10	Cumulative regret for different time bin sizes	40
Figure 4.11	Cumulative regret using a range of exponents for the rate of decay of the bin size.	43
Figure 4.12	Measure of dissimilarity of the rates in a 364 days of operation of the station. Red line represents the selected threshold Θ	46
Figure 4.13	Cumulative regret plots for different selections of α and Θ	47
Figure 4.14	Cumulative regret plots for Method 1 ($\alpha = 1$ and $\Theta = 1$)	48

THIS PAGE INTENTIONALLY LEFT BLANK

Executive Summary

Across a wide variety of applications, we find situations where a decision about the allocation of resources must be made for the optimal exploitation of a system, under conditions where little to no information is available to support such decision. In this thesis we propose a model that addresses the lack of information by observing, at least partially, the system's behavior, and proposes a stochastic approximation aided by simulation, to sequentially converge to an optimal solution that minimizes the expected regret. The problem addressed in this research can be seen as an online optimization of discrete performance measures in Markovian models with incomplete information. We consider a setting where a physical realization of the model is sequentially obtained over a number of periods (days in our case), and the analyst can use the information gathered to-date in order to efficiently run the model in future days. The information is incomplete in two ways: (i) model parameters are initially unknown (the demand rates in our case), but can be estimated from the physical realizations; and (ii), the demands are censored when the system is in some boundary states.

Consider, for example, a naval force is assigned the task of enforcing maritime law in a determined geographical region and has to decide the number of assets to assign for patrolling it. This specific region is known to have illicit activity occurring in it, but the rates at which they happen are not available to the decision makers. Each time an illicit activity is carried out and is undetected, there is a penalty or a cost; on the other, hand there is also a cost associated with assigning assets and a penalty if they patrol the area in times no illicit activities are carried out. This kind of cost could be associated with the operation, and perhaps the lost opportunity of using it in areas where it could have potentially been more useful.

In order to constrain the problem at hand to a situation widely discussed in the available literature and provide a scenario suitable for the analysis, we will discuss an analogous situation in the context of a bike-sharing system (BSS). These stations have a fixed capacity, where some optimal initial number of bikes must be allocated at the beginning of each day of operation. The method of Sample Average Approximation is used to solve the optimization problem. More precisely, in each period of a day, sample paths are generated from the customer rates estimated to-date, and the best model configuration is determined

with respect to these sample paths. Sequential observation of the system's behavior enables better estimates of the arrival rates which in turn allow for an informed decision to be made in each future round. In the instances when the rack is full there is no visibility of customers wishing to return bikes, and in the case when it is empty there is no visibility of customers who wish to rent bikes. We wish to minimize the expected number of these unsatisfied customers whose demand could not be immediately met, by selection of the initial number of bikes on the station at the beginning of the day. The model we propose in this thesis was developed in a base scenario where the arrival of customers was modeled in hourly constant rates. Then we see the effect of approximating to constant values using smooth functions. We then develop two variants of this model. The first one uses sequentially smaller time periods, which we refer to as bins, starting from hour-size bins, and the second one relies on reusing of past samples. We study the problem in the following sequence:

1. True rates are piecewise constant with a value equal to the mean of the corresponding fixed-size interval.
2. True rates are the true continuous functions, and rates are assumed constant for the corresponding fixed- size interval.
3. True rates are the true continuous functions, and rates are assumed constant for the corresponding interval that decreases in size as time progresses and more information about the rates is captured.
4. We reuse sample paths obtained previously as weighted averages that consider recency and similarity of the rates used for the estimation of the objective functions.

We are able to provide a detailed analytical proof for the convergence of the model, and empirically test that the adaptive sizing of the time bins did not improve our results. The reusing of past samples significantly accelerated the convergence towards the optimal decision, reducing the variability of this estimate, and average cumulative regret as show in Figure 1. We will refer to the regret as the difference in the number of expected unsatisfied customers of the selected initial state of the station for the following day, and the optimal number.

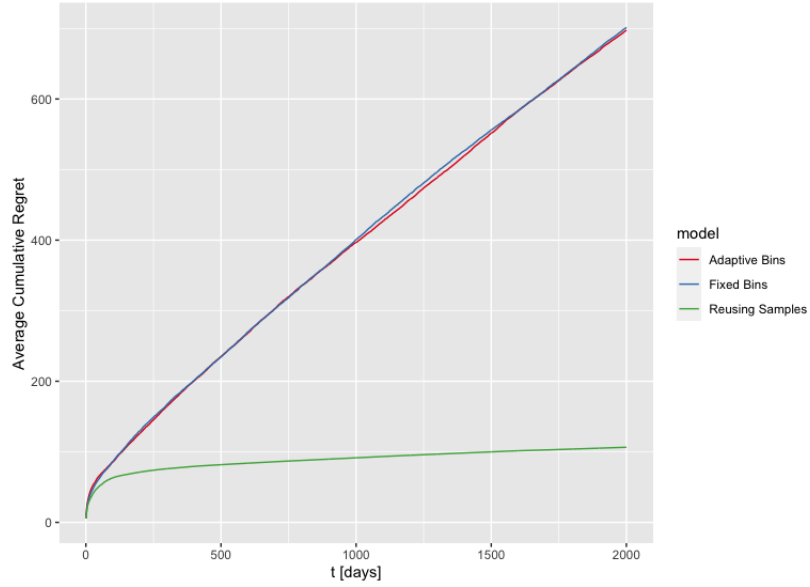


Figure 1. Average cumulative regret observed as a result for variants of the proposed model.

Notwithstanding the fact that the model we propose has proven to work analytically for the base case and performs at least as good for the other cases, we believe that further research is required for the selection of model parameters that could optimize the performance of the model.

THIS PAGE INTENTIONALLY LEFT BLANK

Acknowledgments

I would like to start by thanking my best friend and wife Javi for her never-ending love, kindness and patience. To my boy Matt, thanks for brightening our days with that joyful smile, even in the gloomiest of times.

A special expression of both admiration and gratitude to my parents, who have always been an clear example of love, humbleness and hard work.

Finally, I would like to acknowledge my gratitude to my advisors Dr. Roberto Szechtman and Dr. James Grant. To Dr. Szechtman, I would like to express my deepest appreciation for the time and dedication he has invested in guiding me through this research endeavor, along with his always present kind and caring words for me and my family. I cannot thank Dr. Grant enough for his accurate guidance and orientation, along with encouraging words, even at times where my contribution felt trifling.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 1:

Introduction

1.1 Motivation

In practical applications where the decision that yields the most beneficial return depends on a large or unknown number of parameters, defining an analytical model that characterizes the behavior of that system, is not always possible. If the knowledge of the system's parameters is limited, deciding correctly, in most applications, becomes impossible. Under this setting, we can use the ideas proposed by Hazan (2016), in that optimization can be approached as a process where the observation of the environment allows more informed decisions to be taken as time goes on.

Hazan (2016) states, "The growing literature of machine learning, statistics, decision science and mathematical optimization blur the classical distinctions between deterministic modeling, stochastic modeling and optimization methodology." Based on techniques inspired by these ideas, this thesis proposes an algorithm that sequentially learns from observing a system that presents a behaviour that is partially concealed or not visible, and uses it as the input for the optimization process. Conventional optimization techniques rely on being able to observe the outcome of each decision; however, in the setting we intend to tackle here, the outcome of each decision will be concealed.

The framework proposed in this thesis is applicable to scenarios that can be classified as birth and death processes, in which the knowledge of the birth and death rates are initially unknown. The censoring of the process when the bounding states have been reached is considered as part of the model. A canonical example of the problem we are modelling would apply to a finite capacity queue with non-homogeneous Poisson Process arrivals and departures, where balking is not observable and server rates are not captured during instances the queue is empty.

Applications of the proposed model could arise in a wide variety of situations, for instance a war hospital with a fixed number of beds where a single physician was available, upon call, in order to attend the wounded warriors. If the hospital beds fill up it would imply that

no more wounded soldiers could be attended to, which leads to a determined cost in lives. On the other extreme case where the doctor was called to attend the site and no patients were there to be attended, every potential wounded warrior that could have been attended to elsewhere will represent a cost.

Another practical application where this proposed framework would provide useful insight is the assignment of patrolling resources, say Coast Guard ships in a defined region where the occurrence of illegal fishing activity would pass unnoticed if no means are assigned at the right time and geographical space. Since the operation of these assets generates an expense, assigning an excessive number will reduce the number of illicit activities that can take place, but at a cost that might make it unsustainable or limit the amount of time it can take place.

As a motivating example with which we will develop the model, we will aid the operator of a bicycle-sharing system (BSS) in determining the number of bicycles that a fixed size station must have at the beginning of the day, where no prior information is known about the demands. As with the situations described previously, the number of initial bikes on the station (decided by the operator), will impact the number of customers who walk away from the station empty-handed. They wish to pick up or rent a bike for their own use, or have to ride somewhere else in order to drop off their bicycles in some other station after they have finished using it. So the only source of information we will have available for our decision-making process will be the number of bicycles in the station throughout the day, and the times this number changes as customers return and rent bikes during a day's normal operation.

We will rely solely on observable events to develop a model that sequentially improves the decision, eventually converging to an optimal solution. The techniques used in this model are based on simulating the scenario based on the information that has been available for us to observe.

1.2 Scope, Limitations, and Assumptions

The research presented in this thesis applies online learning techniques to a problem constrained to the assumptions listed below. The results do not pretend to directly provide

answers to related questions that could arise from the operation of the BSS, such as the rebalancing of the bikes between stations or the capacity of stations. For this thesis we make the following assumptions:

- The demand for arriving and departing customers is characterized by Non-Homogeneous Poisson Processes, that are identical for each day. There are no stationary or exogenous effects that modify the rates that are followed by demand.
- There is no prior knowledge of the rates of customers arriving to (drop off) and departing off (pick up) to the rack, but they can be estimated from successive observation of the flow of customers.
- Restocking of the station occurs instantaneously
- There is a limited “computational budget” that must be met for the computation of our best estimate.
- No rebalancing of bikes occurs between stations.
- Penalties associated to the two kinds of unsatisfied demand will be considered equal.

1.3 Research Questions

The primary objective of our research is to answer the following questions:

- What is the target level of bikes to fill a station at the beginning of each day?
- Does this dynamic learning methodology reduce regret as more information is collected from the observation of the system?
- What military application decisions can be informed with the use of the algorithm proposed in this thesis?

1.4 Thesis Outline

Chapter 2 provides a review of the literature that has addressed the modeling of BSS as well as an examination on publications that describe the mathematical techniques that will be used to tackle the problem. Chapter 3 presents the analytical proofs for the base scenario with fixed-size time bins. Chapter 4 implements the models detailed in Chapter 3 in a fictitious, yet realistic scenario, and compares the results of variants of the fixed-size bin model, addressing the selection of tuning parameters that are believed to be case dependent.

Finally, Chapter 5 presents summary observations and concludes the thesis.

CHAPTER 2:

Literature Review and Background

2.1 Background on Bike Sharing Systems

Great interest has arisen in the research community to study Bike Sharing Systems given the growing demand for such transport modalities in cities that are both highly congested and also densely populated. DeMaio (2009) provides a historical narrative for the growth in popularity of such transportation alternatives throughout the world, providing a categorization for the different systems that have been developed. According to the National Association of City Transportation Officials (NATCO 2020), in 2019, people took 136 million trips on shared bikes and scooters, which represented a 60% growth in demand with respect to 2018. This contrast is even more dramatic if we go back to 2010, back when the total sum of trips carried out in shared systems only reached 321,000 trips. This same report states that most of these trips are shorter than 2 miles, but so are 35% of the automobile trips carried out in the United States. This notoriously growing trend and the interest in the operators of these systems to efficiently exploit them, has been the source of many research efforts by the Operations Research community.

2.2 Literature Review

In this section we will present previous work, related to the operation of BSS as well as the supporting theoretical research relevant to the techniques that were employed for the development of our framework.

2.2.1 Bike Sharing System

Perhaps the most comprehensive publication on the operation of BSS, was conducted in Freund et al. (2019). In this work the authors present a methodology to find an optimal allocation of docks to stations, and include the use of incentive programs that motivate users to contribute in the redistribution of bikes, in the form of crowdsourcing. The authors relied on the knowledge of the rates at which customers rented bikes generating empty docks on

the station and customers who dropped off their bicycles in the station after their use. The work done by Raviv and Kolka (2013) is fundamental to the BSS topic, including Freund et al. (2019). In it the authors define the User Dissatisfaction Function (UDF) that will be used as the objective function of the model we will propose in Chapter 3. This function maps the number of expected unsatisfied customers to the number of initial bikes on each station at the beginning of each day. In Raviv and Kolka (2013) the authors define the UDF , $F(I_0)$ given an initial number of bikes on the rack, I_0 , according to the following equation:

$$F(I_0) = \mathbb{E}_R \left\{ \sum_{i=1}^{|E^R|} \left(p(-I_{E_{i-1}^R} - d_{E_i^R}^R)^+ + h(I_{E_{i-1}^R} - d_{E_i^R}^R - C)^+ \right) \right\} \quad (2.1)$$

where E^R corresponds to the set of epochs where a realization of a demand for bikes or docks occurs. The demand for bikes $d_{E_i^R}^R$ takes a value 1 when a demand for bike occurs and -1 when a demand for a dock occurs, and 0 in all other times. The number of customers that walk away without being able to satisfy their demand for bikes at time epoch E_i^R is $(-I_{E_{i-1}^R} - d_{E_i^R}^R)^+$ and the number of customers who cannot satisfy the demand for empty slots is given by $(I_{E_{i-1}^R} - d_{E_i^R}^R - C)^+$. The $(\cdot)^+$ operator is defined as $\max\{0, \cdot\}$. In this setting the satisfaction of the demand is determined by the inventory level restricted to be between 0 (empty) and C (rack at full capacity), defined by:

$$I_{E_i^R} = \begin{cases} 0 & I_{E_{i-1}^R} - d_{E_i^R}^R < 0 \\ C & I_{E_{i-1}^R} - d_{E_i^R}^R > C \\ I_{E_{i-1}^R} - d_{E_i^R}^R & otherwise \end{cases}$$

p and h are the corresponding penalties associated with unsatisfied demand of customers picking up and dropping off bikes, which for the purposes of our study will be considered as a constant value of equal weight that considers both unsatisfied demands with equal penalty.

In summary this function returns the expected demand events that were not satisfied, given an initial I_0 number of bikes on the station at time 0. The proof of the convexity provided by Raviv and Kolka (2013) is fundamental in determining the optimality of our result. The authors provide an approach for the determination of the UDF assuming piece-wise

constant approximation of demand rates. Of course, these approximations rely on the prior knowledge of the rates, unlike what we will confront in the assumptions of this thesis. In order to obtain these results the station and corresponding demand rates were modeled as a non-time-homogenous birth-and-death process depicted as a Markov chain, where the limiting probabilities of each piece-wise constant time bin are employed to obtain the function. The same modeling approach of individual stations is carried out by Schuijbroek et al. (2017) but also focus their study in the system-wide performance, providing a heuristic approach to in dealing with the rebalancing problem.

A different strategy is used by George and Xia (2011), who model vehicle availability at each rental station in a closed queuing network to derive the asymptotic behavior of vehicle availability using Continuous Time Markov chains that are graphically depicted in Figure 2.1.

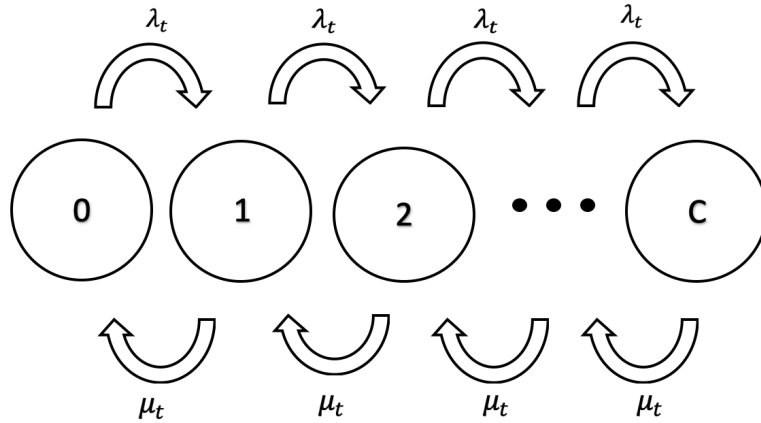


Figure 2.1. Graphical representation of the Markov chain model used by Raviv and Kolka (2013) to represent the bike station and demands.

An interesting approach to defining the problem is performed by Leurent (2012), in which he models a single station, as a customers arrive at the station as a stochastic, as a bi-sided waiting Markovian process with rates that are homogeneous. In the cases where the rack is full dropping off customers wait until there is an available dock on the station. Conversely, when it is empty, customers picking up a bicycle wait in a queue until a dropping off customers make bikes available. The work performed by Leurent (2012) differs from our setting in two main aspects: We will consider that customers will balk after their demand

has not been met, and that rates are both known and homogeneous.

The approach we adopted for in this research for defining the *UDF* is based on the simulation of the flow of customers demanding bikes and available docks on the station for a large number of days. We will describe more details about the method employed for obtaining the *UDF* in Chapter 3.

2.2.2 Poisson Processes

The results we obtained through this research rely on the overarching assumption that the arrival of customers who wish to rent or return bikes at the station are both independent *non-homogeneous Poisson Processes (NHPP)* with corresponding rates $\lambda(t)$ and $\mu(t)$. This case corresponds to a generalization of the *homogeneous Poisson Process*, in which the rates are constant. According to Ross (1997) a NHPP with rate $\lambda(t)$ is a counting process $\{N(t), t \geq 0\}$ that is defined by the following properties:

1. $N(0) = 0$
2. $\{N(t), t \geq 0\}$ has independent increments
3. $P(N(t+h) - N(t) \geq 2) = o(h)$
4. $P(N(t+h) - N(t) = 1) = \lambda(t)h + o(h)$

Poisson Processes Simulation

To generate the flow of customers, we employed the “Thinning” technique defined by Lewis and Shedler (1979) for the simulation of a NHPP. In a few words, this technique consists of simulating a flow of customers picking up and dropping off bicycles, according to their corresponding maximum rate of the time frame to be simulated. Then we randomly select to keep arrivals with probability equal to the ratio of the rate at the instant of arrival and the maximum rate used to generate each customer arrival.

2.2.3 Rate Estimation

As is stated in Grant (2020), a widely used method of estimating the underlying rate function that a NHPP has is the use non-parametric inference, where the rate $\lambda(t)$ is modeled as a combination of piecewise combination of simple functions, which in our case will be constant. Much of our work relies on being able to estimate the rates at which demand is

met for customers renting bikes and returning them. To determine the rates, we relied on the definitions of the NHPP (Ross 1997), to say that for an interval of duration δ the number of arrivals in an interval from time t to $t + \delta$, $E(N(t + \delta) - N(t))$, over a large sample of replications with rate $\lambda(t)$ converges to $\int_t^{t+\delta} \lambda(s) ds \cdot \delta^{-1}$. Henderson (2003) generalizes this definition to:

$$\tilde{\lambda}_n(t) = \frac{1}{n\delta} \sum_{i=1}^n N(\ell(t), \ell(t) + \delta)$$

where $\ell(t) = \delta \lfloor \frac{t}{\delta} \rfloor$ so t belongs to subinterval $[\ell(t), \ell(t) + \delta)$

These estimators correspond to the maximum likelihood estimators for Poisson variables. Nonetheless, these observations do not capture the demand that was not met that we will define as censoring, that if not dealt with properly, will generate a bias in the estimator as described in Nair and Miller-Hooks (2011).

An alternative method for estimating the rates is proposed by Henderson (2003) by using aggregated data, by using the linear approximation of the piecewise constant cumulative intensity function:

$$\tilde{\Lambda}(t) = \frac{1}{n} \sum_{i=1}^n N_i(0, \ell(t)) + \frac{t - \ell(t)}{\delta} \sum_{i=1}^n N_i(\ell(t), \ell(t) + \delta).$$

This formulation will prove to be useful for conducting experiments with a time-dependant reduction of the size of the time bins. Henderson (2003) also defines the rate at which the size of the piecewise constant intervals as the number of samples n increases. The size of the recommended time bin is defined as:

$$\delta^* = \left(\frac{\lambda(h)}{2\lambda'(h)^2} \right)^{\frac{1}{3}} * h^{-\frac{1}{3}}.$$

This choice of δ^* minimizes the mean-squared error of $\tilde{\lambda}_n(h)$. A caveat is to be considered in the selection of a time h to calculate the coefficient. Since no clear guidance is provided in the selection of the time, and is limited to indicating a fixed value, we selected the time

that had the highest probability of being censored, because they are the times in the day of most interest.

2.2.4 On-line Convex Optimization Models

Hazan (2016) states that in on-line convex optimization models (OCO), a decision is made in an iterative or sequential fashion, and the outcome of the decision becomes unknown to the decision maker. In the OCO framework, it is possible to have an adversarial selection of the losses associated with the alternative selected by the decision maker, under certain conditions. However they will not apply to our setting since the loss only depends on what alternative is selected. The OCO framework considers the possible decision set as a convex in a Euclidean space denoted by $\kappa \subseteq \mathbb{R}^n$ where the costs are bounded convex functions over κ . For the problem defined for this thesis, the set κ represents the possible number of bicycles we can select to allocate on the station at the beginning of the day, and the function that models the cost would correspond to the UDF (Raviv and Kolka 2013).

2.2.5 Sample Average Approximation

Most of the techniques described in Hazan (2016) rely on the knowledge of the objective function that we wish to learn by observing the function's response to the selected decisions. Now, in a setting where there was no knowledge of the underlying objective function and the consequence of each decision is not observable, we would not be able to apply these on-line convex optimization techniques as described by Hazan (2016). This is where the idea of Sample Average Approximation proposed by Kleywegt et al. (2001) and described in Kim et al. (2015) to compute real valued functions through estimated stochastic simulation. As explained in Kim et al. (2015) an outline of Sample Average Approximation (SAA) is that a function f has no way of being calculated or observed but it is known that $f(x) = E[Y(x, \xi)]$, where ξ is a random element with a distribution that is independent from x , and $Y(\cdot, \cdot)$ is a deterministic real valued function. In order to obtain an approximation to the real valued function f samples $\xi_1, \xi_2, \xi_3, \dots, \xi_n$ are selected with the same distribution as ξ , and define:

$$f_n(x) = \frac{1}{n} \sum_{i=1}^n Y(x, \xi_i).$$

The resulting $f_n(\cdot)$ is deterministic and can be used to solve optimization problems of the form:

$$\min_{x \in \Theta} f_n(x).$$

This minimization problem is quite simple to solve if $f(\cdot)$ is known to be convex. For our problem $f(\cdot)$ corresponds to the UDF proposed in Raviv and Kolka (2013), who show this function's convexity.

2.2.6 Green Simulation

In the definition of the algorithm we established a computational budget that was distributed along the simulation horizon. This computational budget defines the number of SAA samples employed for the simulation. Inspired by the ideas of Feng and Staum (2017) that propose reusing samples of past simulations to compliment the results of the newest simulation. The *green simulation* concept originates from the *reusing* of the past samples in favor of computational efficiency. We adapted the selection of weights and filtering in a way that empirically proved to improve the results dramatically with respect to using only the last simulated run.

2.3 Contribution to Existing Literature

Even though a large number of research efforts have analyzed methods for optimally defining the number of bicycles to allocate in a single station, they all have done so taking into account an absolute knowledge of the expected rates that drive customer behavior. This allows them, as a consequence, to develop optimization techniques based on more classical methods, given that the objective function is known. Our research adds an extra layer of complexity to this problem by ignoring previous knowledge of the rates that are estimated exclusively through the observation of customer behaviour, that considers censored periods where observability of customer behavior is not possible.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 3:

Analytic Models

3.1 Introduction

In this chapter we will describe the analytical model used to solve the problem addressed in this thesis, in the context of a BSS. We will present the mathematical proof for the convergence of the rate estimators in the situation where these are in reality piece-wise constant. In Section 3.3 we will present the proof for the convergence of the solution to the optimal decision. We will not provide a proof for the convergence of the rates to their piece-wise constant mean, in the case where the true rates are smooth functions, however we will provide empirical evidence in Chapter 4 that these rates approach their true mean, and that the optimal solution converges to the true objective function minimizer.

Consider customers who arrive to a bike rack of size B to drop off or pick up bikes, where the drop offs (or arrivals) and pick ups (or departures) are Poisson processes. Customers who attempt to pick up a bike when the rack is empty or to drop off a bike when the rack leave unsatisfied, so the operator's goal is to minimize the expected number of unsatisfied customers. We model a situation where at the end of each period (e.g., a day), the rack operator employs the arrival and departure observations to-date to restock the rack at some desired level. If the arrival and departure process parameters were known, this could be done using, for instance, sample-average approximation (SAA) methods. In reality, however, these parameters can be estimated on a daily basis, albeit in the presence of censored arrivals, when the rack is full, or censored departures, when the rack is empty.

For simplicity, we assume both arrival and departure processes are independent of each other and i.i.d. across days, with unknown rates $\lambda(h)$ for the arrivals and $\mu(h)$ for the departures, for $h \in [0, 24]$ hours. The counting process of bike drop offs is $(X_\lambda(h) : h \in [0, 24])$, and the one corresponding to the pickups is $(Y_\mu(h) : h \in [0, 24])$.

For a realization of (X_λ, Y_μ) over a 24 hour period, the function $f(b; X_\lambda, Y_\mu)$ measures the (random) number of unsatisfied customers over a 24-hour period when the initial number

of bikes in the rack is b . Its expected value is

$$F(b; \lambda, \mu) := E[f(b; X_\lambda, Y_\mu)],$$

where the expectation is with respect to X_λ and Y_μ . As mentioned earlier, if the rates λ and μ were known, the rack operator would attempt to minimize F over the possible values of $b \in [B] := \{0, 1, \dots, B\}$, and obtain an optimal daily restock level $b^* \in \arg \min_{b \in [B]} F(b; \lambda, \mu)$. Since F is convex (Raviv and Kolka 2013), there exist at most two maximands of F .

In this work we assume the rates are initially unknown, but can be estimated. By day t , there are rate estimates λ_t and μ_t (whose exact form is specified later), leading to the approximate problem

$$\min_{b \in [B]} F(b; \lambda_t, \mu_t),$$

with (possibly non-unique) solution b_t . The function F is not known but can be estimated by Monte Carlo sampling, leading to the estimator

$$F_{n_t, t}(b; \lambda_t, \mu_t) := \frac{1}{n_t} \sum_{i=1}^{n_t} f(b; X_{\lambda_t, i}, Y_{\mu_t, i}),$$

where $(X_{\lambda_t, i}, Y_{\mu_t, i})$, $i = 1, \dots, n_t$, are i.i.d. copies drawn from the distribution of $(X_{\lambda_t}, Y_{\mu_t})$ (which is available to the analyst), and the sample size n_t (which may depend on t). A key issue for the analyst is to judiciously choose n_t subject to a constraint on the computational budget over all rounds $1, \dots, T$.

In summary, the operator starts day t with b_t bikes and observes censored versions of (X_λ, Y_μ) . These censored observations, along with similar ones collected in earlier days, are used to form estimators λ_t and μ_t . The operator then uses SAA to determine the value of b_{t+1} , and the process repeats itself in day $t + 1$. The decisions taken by the operator are two: the number of bikes in each day, $b_1, b_2, \dots$, and the number of samples used for estimation, $n_1, n_2, \dots$. The main modeling assumption, in force throughout this work, is that $\lambda(h) > 0$ and $\mu(h) > 0$ for all $h \in [0, 24]$.

When the context is clear, we write $F(b)$, $F_t(b)$, and $\bar{F}_t(b)$ in lieu of $F(b; \lambda, \mu)$, $F(b; \lambda_t, \mu_t)$, and $F_{n_t, t}(b; \lambda_t, \mu_t)$, respectively.

3.2 Rates Estimation

In this section we consider the problem of estimating the Poisson rate functions for the drop-off and pick-up processes. We assume that both λ and μ are constant over each of the 24-hour time bins.

3.2.1 Piece-Wise Constant Rates

We develop and study estimators for λ and μ under the assumption that both intensity functions are piece-wise constant over each hourly interval. While this assumption is unrealistic from a physical standpoint, it is commonly used by BSS operators to simplify their models. We write λ_h and μ_h for the Poisson rates in period $h \in [h] := \{1, \dots, 24\}$.

For each time period $h \in [24]$ and each day $s \in [t]$, let $C_h(s)$ and $D_h(s)$ be the amount of time drop offs and pick ups were observable. Let $X_h(s)$ be the number of observed drop-offs and $Y_h(s)$ be the number of observed pick-ups in period h of day s . The natural estimators of the drop-off and pick-up rates based on these observations are then $X_h(s)/C_h(s)$ and $Y_h(s)/D_h(s)$. At time t we aggregate these estimates to generate estimators

$$\lambda_{t,h} = \frac{\sum_{s=1}^t X_h(s)}{\sum_{s=1}^t C_h(s)} \mathbb{I} \left(\sum_{s=1}^t C_h(s) > 0 \right) \text{ and } \mu_{t,h} = \frac{\sum_{s=1}^t Y_h(s)}{\sum_{s=1}^t D_h(s)} \mathbb{I} \left(\sum_{s=1}^t D_h(s) > 0 \right), \forall h \in [H].$$

We emphasize the fact that these estimators are based off real, physical, realizations. For the purpose of testing the model in Chapter 4 we simulated these arrivals. As a clarifying note, we refer to the $\mathbb{I}(\cdot)$ as the indicator function.

The expectation of $\lambda_{t,h}$ is computed by recognising that $X_h(s)|C_h(s)$ has a Poisson distribution with parameter $\lambda_h C_h(s)$. We have, for each $t = 2, 3, \dots$, and $h \in [H]$,

$$\begin{aligned} E(\lambda_{t,h}) &= E \left(\frac{\sum_{s=1}^t X_h(s)}{\sum_{s=1}^t C_h(s)} \mathbb{I} \left(\sum_{s=1}^t C_h(s) > 0 \right) \right) = \\ &= E \left(E \left(\frac{\sum_{s=1}^t X_h(s)}{\sum_{s=1}^t C_h(s)} \mathbb{I} \left(\sum_{s=1}^t C_h(s) > 0 \right) \middle| C_h(1), \dots, C_h(t) \right) \right) \\ &= \lambda_h P \left(\sum_{s=1}^t C_h(s) > 0 \right). \end{aligned}$$

The $\mu_h > 0$ assumption leads to,

$$\begin{aligned}
P\left(\sum_{s=1}^t C_h(s) = 0\right) &\leq P(C_h(1) = 0) \\
&\leq P(\text{no departures during period } h \text{ on day } 1) \\
&= \exp(-\mu_h) \\
&< 1,
\end{aligned}$$

We conclude that $P(\sum_{s=1}^t C_h(s) = 0) \rightarrow 0$ as $t \rightarrow \infty$, so $\lambda_{t,h}$ is asymptotically unbiased.

Next we develop a central limit theorem (CLT) for $\lambda_{t,h}$ under the assumption that $b_1, \dots, b_t$ are constant. Since $X_h(s)$ are i.i.d. with $\text{Poisson}(\lambda_h E[C_h(1)])$ distribution, as stated in Proposition 4.10.1 of Resnick (1992), and $C_h(1), \dots, C_h(t)$ are i.i.d. (with a distribution that depends only on b_1), the CLT yields

$$\sum_{s=1}^t (X_h(s) - \lambda_h C_h(s)) + \sqrt{t}(\sigma_h N(0, 1) + o_p(1)) = 0, \quad (3.1)$$

where the variance constant is

$$\begin{aligned}
\sigma_h^2 &= \text{Var}(X_h(1) - \lambda_h C_h(1)) \\
&= E[\text{Var}(X_h(1)|C_h(1))] + \text{Var}(E[X_h(1)|C_h(1)]) \\
&= \lambda_h E[C_h(1)] + \lambda_h^2 \text{Var}(C_h(1)).
\end{aligned} \quad (3.2)$$

Dividing by $\sum_{s=1}^t C_h(s)$ on the event $\{\sum_{s=1}^t C_h(s) > 0\}$ yields

$$\left(\lambda_{t,h} - \lambda_h + \sqrt{t} \frac{\sigma_h N(0, 1) + o_p(1)}{\sum_{s=1}^t C_h(s)} \right) \mathbb{I}\left(\sum_{s=1}^t C_h(s) > 0\right) = 0.$$

The Strong Law of Large Numbers (SLLN) results in $t^{-1} \sum_{s=1}^t C_h(s) = E[C_h(1)] + o(1)$ a.s. Also $P(\sum_{s=1}^t C_h(s) > 0) \rightarrow 1$. Hence, Slutsky's theorem leads to

$$\sqrt{t}(\lambda_{t,h} - \lambda_h) \Rightarrow \frac{\sigma_h}{E[C_h(1)]} N(0, 1),$$

as $t \rightarrow \infty$.

To relax the assumption that $b_1, \dots, b_t$ are constant we argue as follows. First, under any non-anticipating policy π there exists a set $\mathcal{B}_\pi \subseteq [B]$ containing elements that get selected infinitely often. Let $T_{t,b}$ is the set of random days that $b \in \mathcal{B}_\pi$ gets selected by round t . To simplify what follows, assume that $|T_{t,b}|/t \rightarrow k_b > 0$ a.s.

Therefore, Equation (3.1) applies to each element $b \in \mathcal{B}_\pi$ via the random index CLT:

$$\sum_{s \in T_{t,b}} (X_h(s) - \lambda_h C_h(s)) + k_b \sqrt{t} (\sigma_{h,b} N(0, 1) + o_p(1)) = 0, \quad (3.3)$$

Adding (3.3) for all $b \in \mathcal{B}_\pi$ leads to

$$\sum_{s \in T_{t,b}, b \in \mathcal{B}_\pi} (X_h(s) - \lambda_h C_h(s)) + \sqrt{t} \left(\left(\sum_{b \in \mathcal{B}_\pi} k_b^2 \sigma_{h,b}^2 \right)^{1/2} N(0, 1) + o_p(1) \right) = 0,$$

This can be justified by the independence of the X_h , C_h , and Gaussian terms.

Adding and subtracting the contributions of the elements not in $\mathcal{B}_\pi$ results in

$$\sum_{s=1}^t (X_h(s) - \lambda_h C_h(s)) + \sqrt{t} \left(\left(\sum_{b \in \mathcal{B}_\pi} k_b^2 \sigma_{h,b}^2 \right)^{1/2} N(0, 1) + o_p(1) \right) = \sum_{s \in [t]: b(s) \notin \mathcal{B}_\pi} (X_h(s) - \lambda_h C_h(s))$$

Dividing by $\sum_{s=1}^t C_h(s)$ on the event $\{\sum_{s=1}^t C_h(s) > 0\}$,

$$\begin{aligned} & \left(\lambda_{t,h} - \lambda_h + \sqrt{t} \frac{(\sum_{b \in \mathcal{B}_\pi} k_b^2 \sigma_{h,b}^2)^{1/2} N(0, 1) + o_p(1)}{\sum_{s=1}^t C_h(s)} \right) \mathbb{I} \left(\sum_{s=1}^t C_h(s) > 0 \right) \\ &= \frac{\sum_{s \in [t]: b(s) \notin \mathcal{B}_\pi} (X_h(s) - \lambda_h C_h(s))}{\sum_{s=1}^t C_h(s)} I \left(\sum_{s=1}^t C_h(s) > 0 \right). \end{aligned} \quad (3.4)$$

The assumption that $|T_{t,b}|/t \rightarrow k_b > 0$ a.s. and the SLLN lead to,

$$\sum_{s=1}^t C_h(s) = \sum_{b \in B_\pi} \sum_{s \in T_{t,b}} C_h(s) + \sum_{s: b(s) \notin B_\pi} C_h(s) \quad (3.5)$$

$$= t \sum_{b \in B_\pi} k_b E[C_{h,b}(1)] + \sum_{s: b(s) \notin B_\pi} C_h(s) + o(t), \quad (3.6)$$

a.s.

Since the number of summands in $\sum_{s: b(s) \notin B_\pi} C_h(s)$ is a.s. sublinear, with each summand in $[0, 1]$, we get $t^{-1} \sum_{s: b(s) \notin B_\pi} C_h(s) = o(1)$. It follows that

$$\frac{1}{t} \sum_{s=1}^t C_h(s) \rightarrow \sum_{b \in B_\pi} k_b E[C_{h,b}(1)] + o(1),$$

a.s. Hence, in (3.4) a converging together argument leads to

$$\sqrt{t}(\lambda_{t,h} - \lambda_h) \Rightarrow \frac{\left(\sum_{b \in B_\pi} k_b^2 \sigma_{h,b}^2 \right)^{1/2}}{\sum_{b \in B_\pi} k_b E[C_{h,b}(1)]} N(0, 1),$$

as $t \rightarrow \infty$.

The analogue CLT applies to the estimator of μ_h ; namely,

$$\sqrt{t}(\mu_{t,h} - \mu_h) \Rightarrow \frac{\left(\sum_{b \in B_\pi} k_b^2 \sigma_{h,b,\mu}^2 \right)^{1/2}}{\sum_{b \in B_\pi} k_b E[D_{h,b}(1)]} N(0, 1),$$

as $t \rightarrow \infty$, where the standard deviation is given by $\sigma_{h,b,\mu} = (\mu_h E[D_h(1)] + \mu_h^2 \text{Var}(D_h(1)))^{1/2}$.

3.2.2 General Rate Functions with Fixed Number of Bins

In this subsection we relax the assumption of the Poisson intensity functions being piecewise constant, while preserving the premise that the rates are estimated over hourly buckets.

Consider some hour of the day h , and let $Z_{\text{drop}}(u) = \mathbb{I}(\text{the number of bikes is } < m \text{ at time } u)$. The model construction ensures that $(Z_{\text{drop}}(u) : u \in [0, H])$ is a continuous time Markov

chain, and that $\int_h P(Z_{\text{drop}}(u) = 1)du > 0$ for all $h \in [H]$. To simplify the developments we initially assume that $b_1, b_2, \dots, b_t$ is constant and equal to b , and omit the dependence on b

Proposition 4.10.1 of Resnick (1992) implies

$$X_h \sim \text{Poisson} \left(\int_h \lambda(u) P(Z_{\text{drop}}(u) = 1) du \right),$$

where X_h is the number of detected arrivals during period h .

Letting

$$\hat{\lambda}_h = \frac{\int_h \lambda(u) P(Z_{\text{drop}}(u) = 1) du}{\int_h P(Z_{\text{drop}}(u) = 1) du}, \quad (3.7)$$

we have $X_h \sim \text{Poisson}(\hat{\lambda}_h \int_h P(Z_{\text{drop}}(u) = 1) du)$. We interpret $\hat{\lambda}_h$ as the average uncensored arrival rate during period h . In case $\lambda(u) = \lambda_h$ constant for all u in period h , we have $\hat{\lambda}_h = \lambda_h$; that is, we recover the estimator of the last subsection.

The estimator of $\hat{\lambda}_h$ is

$$\hat{\lambda}_{t,h} = \frac{\sum_{s=1}^t X_h(s)}{\sum_{s=1}^t \int_h Z_{\text{drop}}(u, s) du} \mathbb{I}(\sum_{s=1}^t \int_h Z_{\text{drop}}(u, s) du > 0).$$

where $\int_h Z_{\text{drop}}(u, s) du$ is the amount of time that arrivals are uncensored during period h on day s . Conditioning then leads to,

$$\begin{aligned} E(\hat{\lambda}_{t,h}) &= E \left(\frac{\sum_{s=1}^t X_h(s)}{\sum_{s=1}^t \int_h Z_{\text{drop}}(u, s) du} \mathbb{I}(\sum_{s=1}^t \int_h Z_{\text{drop}}(u, s) du > 0) \right) = \\ &= E \left(E \left(\frac{\sum_{s=1}^t X_h(s)}{\sum_{s=1}^t \int_h Z_{\text{drop}}(u, s) du} \mathbb{I}(\sum_{s=1}^t \int_h Z_{\text{drop}}(u, s) du > 0) \mid Z_{\text{drop}}(u, 1), \dots, Z_{\text{drop}}(u, t), \forall u \in [h] \right) \right) = \\ &= E \left(\frac{\sum_{s=1}^t \int_h \lambda(u) Z_{\text{drop}}(u, s) du}{\sum_{s=1}^t \int_h Z_{\text{drop}}(u, s) du} \mathbb{I}(\sum_{s=1}^t \int_h Z_{\text{drop}}(u, s) du > 0) \right), \end{aligned}$$

for each $t \in \{2, \dots, T\}$ and $h \in [H]$.

The SLLN implies

$$\frac{1}{t} \sum_{s=1}^t \int_h \lambda(u) Z_{\text{drop}}(u, s) du \rightarrow \int_h \lambda(u) P(Z_{\text{drop}}(u) = 1) du \text{ a.s.} \quad (3.8)$$

and

$$\frac{1}{t} \sum_{s=1}^t \int_h Z_{\text{drop}}(u, s) du \rightarrow \int_h P(Z_{\text{drop}}(u) = 1) du \text{ a.s.} \quad (3.9)$$

Moving the expectation inside the integral in (3.8) and (3.9) can be justified by Fubini's theorem. Clearly, $P(Z_{\text{drop}}(u) = 1)$ depends on b_s . Relaxing the constant b_s assumption can be argued as in (3.5–3.6).

The positive intensity assumption implies $P(\int_h Z_{\text{drop}}(u, s) du > 0) > \epsilon$ for some $\epsilon > 0$. This and the Borel-Cantelli lemma leads to $\mathbb{I}(\sum_{s=1}^t \int_h Z_{\text{drop}}(u, s) du > 0) \rightarrow 1$ a.s. By continuous mapping we then have

$$\frac{\sum_{s=1}^t \int_h \lambda(u) Z_{\text{drop}}(u, s) du}{\sum_{s=1}^t \int_h Z_{\text{drop}}(u, s) du} \mathbb{I}(\sum_{s=1}^t \int_h Z_{\text{drop}}(u, s) du > 0) \rightarrow \frac{\int_h \lambda(u) P(Z_{\text{drop}}(u) = 1) du}{\int_h P(Z_{\text{drop}}(u) = 1) du} \text{ a.s.}$$

In addition, $0 \leq \int_h \lambda(u) Z_{\text{drop}}(u, s) du \leq \int_h \lambda(u) du < \infty$, and $\int_h Z_{\text{drop}}(u, s) du > \epsilon$ for some $\epsilon > 0$, by the positive intensity assumption. Hence, by bounded convergence,

$$E(\hat{\lambda}_{t,h}) \rightarrow \hat{\lambda}_h, \quad (3.10)$$

as $t \rightarrow \infty$. That is, $\hat{\lambda}_{t,h}$ is asymptotically unbiased for $\hat{\lambda}_h$.

Relaxing the assumption that the bike levels, $b_1, b_2, \dots$, are constant can be argued as in the preceding subsection. That is, we assume there exists a set $\mathcal{B}_\pi \subseteq [B]$ of elements that get selected with positive probability. The details are left for future work.

The CLT follows along the same lines of the constant rate case since the random variables $X_h(s) - \hat{\lambda}_t \int_h Z_{\text{drop}}(u, s) du$, for $s = 1, 2, \dots$, are i.i.d. with mean zero. In this setting we obtain,

$$\sqrt{t}(\hat{\lambda}_{t,h} - \hat{\lambda}_h) \Rightarrow \frac{\left(\sum_{b \in \mathcal{B}_\pi} k_b^2 \hat{\sigma}_{h,b}^2\right)^{1/2}}{\sum_{b \in \mathcal{B}_\pi} k_b \int_h P(Z_{\text{drop}}(u, 1) = 1) du} N(0, 1),$$

as $t \rightarrow \infty$, where the variance constant is given by

$$\begin{aligned}\hat{\sigma}_{h,b}^2 &= \text{Var}\left(X_h(1) - \hat{\lambda}_h \int_h Z_{\text{drop}}(u, 1) du\right) \\ &= E[\text{Var}(X_h(1) | \int_h Z_{\text{drop}}(u, 1) du)] + \text{Var}(E[X_h(1) | \int_h Z_{\text{drop}}(u, 1) du]) \\ &= \hat{\lambda}_h \int_h P(Z_{\text{drop}}(u, 1) = 1) du + \hat{\lambda}_h^2 \text{Var}\left(\int_h \mathbb{I}(Z_{\text{drop}}(u, 1) = 1) du\right).\end{aligned}$$

Two comments are in order: First, we omitted the dependence on b in the notation of the variance derivation; and second, we recover the same variance expression as in (3.2) in case the rates are constant within buckets.

To close this subsection, we mention the matching estimator of

$$\hat{\mu}_h = \frac{\int_h \mu(u) P(Z_{\text{pick}}(u) = 1) du}{\int_h P(Z_{\text{pick}}(u) = 1) du}, \quad (3.11)$$

by day t ,

$$\hat{\mu}_{t,h} = \frac{\sum_{s=1}^t Y_h(s)}{\sum_{s=1}^t \int_h Z_{\text{pick}}(u, s) du} \mathbb{I}\left(\sum_{s=1}^t \int_h Z_{\text{pick}}(u, s) du > 0\right),$$

where $Z_{\text{pick}}(u, s) = \mathbb{I}(\text{the number of bikes is } > 0 \text{ at time } u \text{ on day } s)$, and $\int_h Z_{\text{pick}}(u, s) du$ is the amount of time that departures are uncensored during period h on day s .

The corresponding CLT is,

$$\sqrt{t}(\hat{\mu}_{t,h} - \hat{\mu}_h) \Rightarrow \frac{\left(\sum_{b \in B_\pi} k_b^2 \hat{\sigma}_{h,b,\mu}^2\right)^{1/2}}{\sum_{b \in B_\pi} k_b \int_h P(Z_{\text{pick}}(u, 1) = 1) du} N(0, 1),$$

as $t \rightarrow \infty$, where

$$\hat{\sigma}_{h,b,\mu}^2 = \hat{\mu}_h \int_h P(Z_{\text{pick}}(u, 1) = 1) du + \hat{\mu}_h^2 \text{Var}\left(\int_h \mathbb{I}(Z_{\text{pick}}(u, 1) = 1) du\right)$$

is the variance constant.

3.3 Convergence to the Optimal Solution

In this section we contend that the solution by day t ,

$$b_t = \operatorname{argmin}_{b \in [B]} \bar{F}_t(b),$$

approaches

$$b^* \in \arg \min_{b \in [B]} F(b),$$

as t gets large. We do so for the case where λ and μ are piece-wise constant in each hourly period; the general case, where λ and μ are not constant, will be treated elsewhere.

We will argue that if $n_t \rightarrow \infty$ as $t \rightarrow \infty$ and $\arg \min_{b \in [B]} F(b)$ is a singleton (meaning that b^* is unique), the decision-making scheme described previously (selecting b_t as above) is asymptotically consistent, in the sense that,

$$\lim_{t \rightarrow \infty} P(b_t = b^*) = 1. \quad (3.12)$$

To wit, the uniqueness of b^* implies the existence of a constant Δ such that,

$$\Delta := \min_{b \in [B] \setminus b^*} F(b) - F(b^*) > 0.$$

From the definition of b_t , a sufficient condition for the selection of the optimal action is that the approximation error is less than $\Delta/2$, i.e.,

$$|\bar{F}_t(b) - F(b)| < \Delta/2, \quad \forall b \in [B].$$

It follows that, if $\lim_{t \rightarrow \infty} \mathbb{P}(|\bar{F}_t(b) - F(b)| \geq \Delta/2) = 0$ for all $b \in [B]$, then (3.12) holds. To prove such a result, we decompose the approximation error for any $b \in [B]$, and $t \in \mathbb{N}$, as follows,

$$|\bar{F}_t(b) - F(b)| \leq |F_t(b) - F(b)| + |\bar{F}_t(b) - F_t(b)|, \quad (3.13)$$

and consider the right hand side terms in isolation.

The first term considers the error induced by the approximation of the parameters λ, μ with

λ_t, μ_t in the latent estimator F_t . The second term considers the error induced by the SAA approximation to the function F_t . To prove the convergence of F_t to F we first demonstrate that $F(b)$ is continuously differentiable with respect to λ and μ , with bounded gradients.

We recall the definition

$$F(b, \lambda, \mu) = \int_0^H \lambda(s)P(N(s) = B|b_0 = b) + \mu(s)P(N(s) = 0|b_0 = b)ds,$$

where $N(s)$ is the number of bikes in the station at time s , and consider the matrix-exponential version of the finite-time transition probabilities. By the Kolmogorov ordinary differential equations for a CTMC we have

$$P(N(s) = B|b_0 = b) = \left[\exp \int_0^s Q(h)ds \right]_{bB} = \left[\sum_{n=0}^{\infty} \frac{\left(\int_0^s Q(h)dh \right)^n}{n!} \right]_{bB} \quad (3.14)$$

where $\{Q(s); s \geq 0\}$ are the transition matrices of the CTMC, and we have

$$\int_0^s Q(s)dh = \begin{pmatrix} -\Lambda(s) & \Lambda(s) & 0 & 0 & 0 & \dots & 0 \\ M(s) & -\Lambda(s) - M(s) & \Lambda(s) & 0 & 0 & \dots & 0 \\ 0 & M(s) & -\Lambda(s) - M(s) & \Lambda(s) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & M(s) & -\Lambda(s) - M(s) & \Lambda(h) \\ 0 & \dots & 0 & 0 & 0 & M(s) & -M(s) \end{pmatrix},$$

where $\Lambda(s) = \int_0^s \lambda(s)ds$ and $M(s) = \int_0^s \mu(s)ds$.

The elements of the sum in (3.14) are such that there exist real coefficients $\{\{a_{m,n}\}_{m=0}^n\}_{n=0}^{\infty}$, such that we may write

$$\begin{aligned} P(N(s) = B|b_0 = b) &= \sum_{n=0}^{\infty} \frac{\left[\left(\int_0^s Q(h)dh \right)^n \right]_{bB}}{n!} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{m=0}^n a_{m,n} \Lambda(s)^m M(s)^{n-m}. \end{aligned}$$

The upshot of this is that $P(N(s) = B|b_0 = b)$ and similarly $P(N(s) = 0|b_0 = b)$ are $\Omega(\Lambda(s))$ and $\Omega(M(s))$, so small estimation errors will not have an unbounded effect. Formally, if for some $0 < \epsilon < 1/s$, we have Lipschitz-like bounds $|\Lambda(s) - \Lambda'(s)| \leq \epsilon s$ and $|M(s) - M'(s)| \leq \epsilon s$ then

$$P_{\Lambda, M}(N(s) = B|b_0 = b) - P_{\Lambda', M'}(N(s) = B|b_0 = b) = o(\epsilon s).$$

Applying this result to the particular estimators λ_t and μ_t , it follows that there exist finite positive constants k_1, k_2 such that,

$$\begin{aligned} |F_t(b) - F(b)| &= |F(b; \lambda_t, \mu_t) - F(b; \lambda, \mu)| \\ &= \left| \int_0^H \lambda(s) P_{\lambda, \mu}(N(s) = B|b_0 = b) - \lambda_t(s) P_{\lambda_t, \mu_t}(N(s) = B|b_0 = b) \right. \\ &\quad \left. + \mu(s) P(N(s) = 0|b_0 = b) - \mu_t(s) P_{\lambda_t, \mu_t}(N(s) = 0|b_0 = b) ds \right| \\ &\leq \int_0^H \lambda(s) |P_{\lambda, \mu}(N(s) = B|b_0 = b) - P_{\lambda_t, \mu_t}(N(s) = B|b_0 = b)| \\ &\quad + |\lambda(s) - \lambda_t(s)| P_{\lambda_t, \mu_t}(N(s) = B|b_0 = b) ds \\ &\quad + \int_0^H \mu(s) |P_{\lambda, \mu}(N(s) = 0|b_0 = b) - P_{\lambda_t, \mu_t}(N(s) = 0|b_0 = b)| \\ &\quad + |\mu(s) - \mu_t(s)| P_{\lambda_t, \mu_t}(N(s) = 0|b_0 = b) ds \\ &\leq \int_0^H \lambda(s) k_1 |\Lambda(s) - \Lambda_t(s)| + |\lambda(s) - \lambda_t(s)| ds \\ &\quad + \int_0^H \mu(s) k_2 |M(s) - M_t(s)| + |\mu(s) - \mu_t(s)| ds \\ &\leq \lambda_{max} k_1 (|\Lambda - \Lambda_t| + |\lambda - \lambda_t|) + \mu_{max} k_2 (|M - M_t| + |\mu - \mu_t|) \end{aligned}$$

The SLLN and continuous mapping lead to $|\lambda_t - \lambda| + |\mu - \mu_t| \rightarrow 0$ with probability 1. Fatou's lemma completes the proof that $P(|F_t(b) - F(b)| > \Delta/2) \rightarrow 0$.

For the $|\bar{F}_t(b) - F_t(b)|$ term in (3.13), the SLLN applies conditioned on the (random) rates. Then $E[P(|\bar{F}_t(b) - F_t(b)| > \Delta/2 | \lambda_t, \mu_t)] \rightarrow 0$ follows by bounded convergence.

Since

$$P(|\bar{F}_t(b) - F(b)| > \Delta/2) \leq \underbrace{P(|F_t(b) - F(b)| > \Delta/2)}_{\rightarrow 0} + \underbrace{P(|\bar{F}_t(b) - F_t(b)| > \Delta/2)}_{\rightarrow 0},$$

we conclude that $P(b_t = b^*) \rightarrow 1$ as the number of days t observed gets large if n_t grows with t .

3.4 Analyses Left for Future Work

Due to time considerations we could not complete the analysis of several ideas implemented in Chapter 4. In this section we briefly comment on these ideas.

Optimal sampling rates: While we argued that having n_t grow with t leads to the optimal number of bikes b^* being selected (cf., Eq. (3.12)), we left open the issue of how to set $n_1, n_2, \dots, n_t$ given an overall computer budget by time t . Looking at this problem within a regret minimization framework suggests that n_t should be of order $t^{1/10}$. This and other rates are compared in Chapter 4.

Adaptive number of buckets: In case the Poisson intensities λ and μ are not constant within buckets, the estimators $\hat{\lambda}_{t,h}$ and $\hat{\mu}_{t,h}$ have an expectation that approach $\hat{\lambda}_h$ and $\hat{\mu}_h$ (cf., Eq. (3.10)). However, (3.7) and (3.11) suggest that $\hat{\lambda}_h$ and $\hat{\mu}_h$ are not necessarily equal to the average of λ and μ over hour h (namely, $\int_h \lambda(u)du$ and $\int_h \mu(u)du$). In other words, $\hat{\lambda}_{t,h}$ and $\hat{\mu}_{t,h}$ will be biased. This issue can be ameliorated by splitting the 24-hour interval into increasingly more buckets, each of smaller duration, as t gets larger. Results from the literature (Henderson 2003) suggest that the optimal growth of the number of buckets is of order $t^{1/3}$. These ideas are explored in Section 4.3.

Reuse of older samples: Following the notion that the use of a greater number of samples will lead to lower variance in the estimate of F and the idea that estimates generated using rates that are alike will lead to similar results, we tested reusing past results with different methods for developing weighted averages that considered both similarity of the rates and the recency of the results. We empirically tested these ideas in Section 4.5.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 4:

Analysis and Simulation Results

4.1 Introduction

In this chapter we will analyze the implementation of the models described in Chapter 3. We will begin by describing the setup for the experiment and then we will present three variants of the model. For each one we will present the algorithm and provide a detailed description of the process along with the design features that must be considered in the implementation of each one.

The first model we will present is characterized by time bins that are fixed throughout the planning horizon. The initial discussion will be carried out in the context where the arrival rates of customers are piecewise constant as shown in Figure 4.3. The design features that we will discuss in this section are the computational budget distribution along the planned time horizon and the number of bins to use throughout the day. We will present some analysis on the effect of censoring in the estimation of the rates; and then will present the results when true arrival rates are smooth.

Following this, we will present a variation of the model where the size of the bins decrease with the progression of days, hence the number of bins increases as the number of sampled days and the amount of information available for the estimation of the rates allows us to reduce their size. In this implementation of the algorithm, we also modified the method for estimating the rates, which will now rely on the accumulated observation rather than isolating the estimation for each bin. The motivation for this method of estimating the rates was to use a modified version of what is presented in Henderson (2003), where censoring is taken into account.

Finally, we will present a second variant for the fixed bin size algorithm inspired in the ideas proposed in Feng and Staum (2017) and Polyak and Juditsky (1992), where we reuse past samples. We tested two methods for assigning weights for the averaging of past samples, one relying on the similarity of the rates used to obtain the estimate and a second relying on the recency that assigned a larger weight to estimates obtained closer in time.

4.1.1 Experimental Setup

Based on the piecewise constant demand rates presented in the example used in Raviv and Kolka (2013), we will use smooth function to define the rates defined by the following Cauchy functions, where we will again use λ as the rate of customers returning bikes and μ as the customers renting bicycles at time s :

$$\lambda(s) = 5 + \frac{180}{\pi} \frac{1}{(s - 7)^2 + 4} \quad \mu(s) = 6 + \frac{200}{\pi} \frac{1}{(s - 17)^2 + 4} \quad (4.1)$$

where $s \in [0, 24]$, as shown in Figure 4.1.

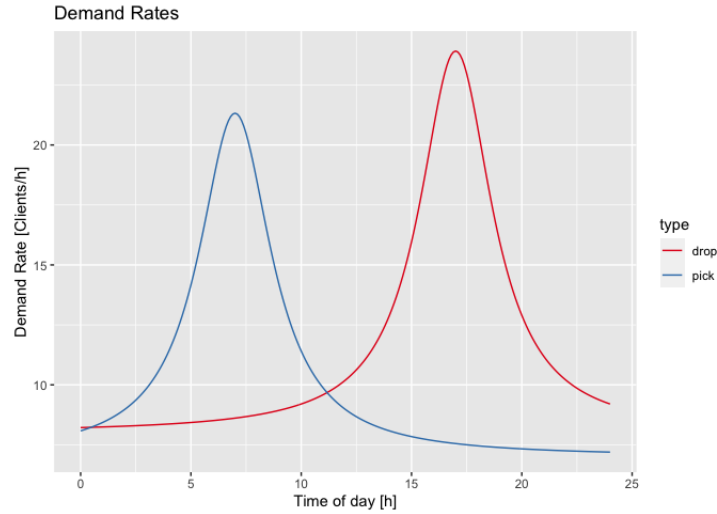


Figure 4.1. Customer demand rate

In order to verify the behaviour of the model in a large enough time horizon, we observed the operation of the station in the simulated experiments for 2,000 days, in the different implementations of the model. All cumulative regret plots shown in this chapter represent the average of 20 repetitions of the experiment.

4.1.2 User Dissatisfaction Function

As with all optimization problems, an objective function will provide the reference upon which the decision maker must select from the possible alternatives to generate the response that best meets his interests. In the scenario we are facing, the operator of the station must

decide on the initial level to which he must set the station for next day's customers. Even though the operator cannot see the effect of his decisions and does not know the true optimal value, there exists a underlying function that traces each possible $b \in \{0, \dots, B\}$ to the corresponding expected number of unsatisfied customers. We will refer to this function as the User Dissatisfaction function, F , throughout this research.

Since there is no closed form for F , and it cannot be computed analytically, we will estimate it using Monte Carlo Sampling using the notion of SAA. The method for estimating the true expected number of unsatisfied customers, given the initial number of bikes on the station, consists of simulating an arbitrarily large number of days of customer arrivals, which in our case was 100,000, using the true rates displayed in Figure 4.1, and averaging the results for each initial number of bikes, as shown in Figure 4.2. As observed in Theorem 1 of Raviv and Kolka (2013), F is discretely convex, and for our example has its minimum at $b_0 = 62$ bicycles.

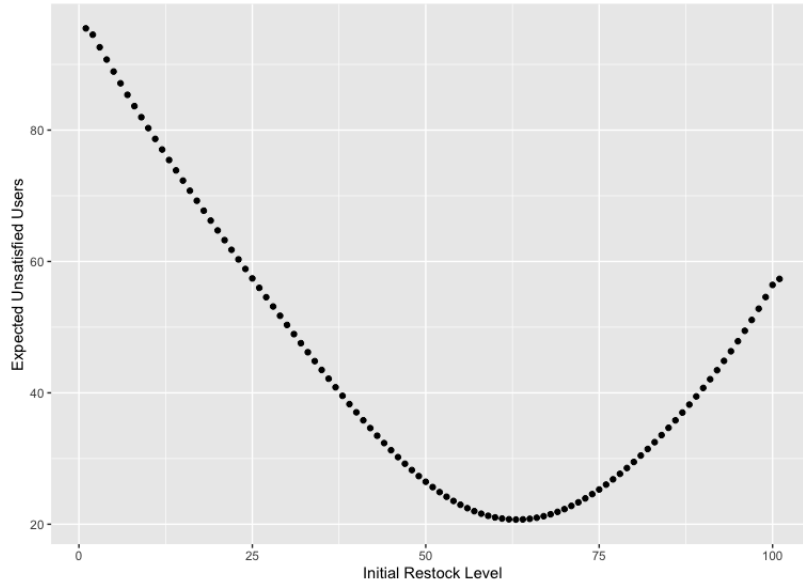


Figure 4.2. Expected number of unsatisfied customers for each initial number of bikes assuming demand rates in equation 4.1.

4.2 Fixed Bins and Piece-Wise Constant Rates

In the initial reference scenario, true rates where the customers demand for bikes to rent and return were modeled as piecewise constant within each 1 hour bin, with respect to the rates presented in Figure 4.1, as shown in Figure 4.3. The piecewise constant values for each time bin h was defined by

$$\lambda_h = \int_{h-1}^{h-1+\delta} \lambda(s) ds$$

where $h = \delta \left\lfloor \frac{s}{\delta} \right\rfloor$ and δ is the size of the time bin. In the context of time during the day, we will use h to represent discrete times and $\mathcal{H}$ the set containing all the breaking points between the time bins. We will use s to represent continuous intervals. We will reserve the commonly used t notation to identify the days. We will arbitrarily select δ as 1 hour, but will verify the effect of using other values in the implementation of the model.

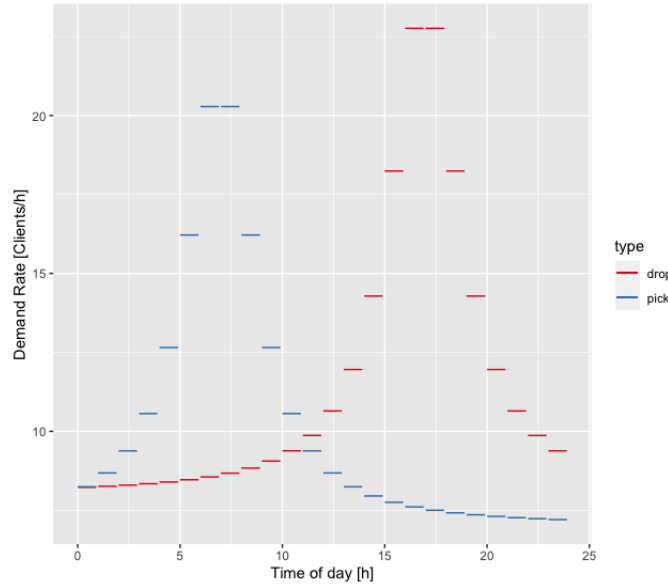


Figure 4.3. Piecewise constant customer demand rates approximating base-line scenario defined by equation 4.1.

As part of the assumptions in the development of the model, we defined that a fixed computational budget will set a constraint on the total number of SAA samples throughout the 2,000-day time horizon. In order to verify the effect of different budget allocations through-

out the time horizon, we assign a total computational budget of 100,000 SAA samples. We allocate this computational budget using five different distributions, graphically displayed in Figure 4.7. The allocation that generates the lowest cumulative regret will be used for the rest of the experiments. In order to gain insight into the effect of using a different total computational budget (with the same distribution) we will repeat the experiment using a total budget that is equal to double of that used as a reference (200,000 samples) and half (50,000 samples).

4.2.1 Algorithm

Algorithm 1 provides an outline of the implementation of the fixed bin model presented in Chapter 3. We will provide a description and brief narration of the steps involved in this implementation in order to clarify the notation that was used as part of the pseudo-code. The primary objective is to inform the operator of the system what the initial number of bikes on the rack should be on the following day in order to minimize the number of unhappy customers.

Definitions

A few prior definitions must be performed before the initial step, which consider δ , the duration of each bin, and a planning horizon in days defines the total number of days that will be observed. We begin by introducing some additional notation. The budget $\mathcal{B}(t)$ is the function that determines the number of SAA samples that will be performed in each day t . $X, Y \in \mathbb{N}^{|H|}$ will accumulate the total number of customers returning bikes and renting bikes correspondingly during each time bin, and $C, D \in \mathbb{R}^{|H|}$ will accumulate the proportion of time each time bin is observable. C_h will accumulate the sum of the proportions each time bin h was not full from day 0 to day t ; likewise, D_h will be the cumulative proportion time bin h was not full from day 0 to t .

Initialization

For all days t a policy must be defined for selecting the initial bikes for the following day; thus, we will use the minimizer of F estimate. This decision for day 1 must be made somewhat arbitrarily since no prior estimate of the objective function F is available. Possible policies for day $t = 1$ would be to fill up to half of the station capacity or select

a random number of bikes. This could be repeated for a certain amount of initial days as an exploration phase in which a range of initial states for bike allocation could be tested. Nonetheless, this is not described as part of the algorithm. This would allow for better estimates of the demand rates prior to the SAA portion of the algorithm (lines 10 to 20 of Algorithm 1 in appendix A.1) where an estimate of the F is generated and can be used to set the initial state of the station on the next iteration (day $t + 1$). The initial estimate of the rates $\bar{\lambda}$ and $\bar{\mu}$ are calculated according to what is described in Algorithm 3 in appendix A.2. The initial estimate is computed by observing the flow of customers for day $t = 1$ and counting the number of customers (pick-ups and drop-offs) during each time bin, as well as the amount of time each bin is censored by being either full or empty. The initial estimate of $\bar{\lambda}_h$ is obtained from the ratio between the number of returning customers and the total time bin h , likewise for $\bar{\mu}_h$, using the renting customers observed and the time bin h is not full, as defined in Chapter 3.

SAA

The main loop corresponding to each day between day 2 and the defined planning horizon $\mathcal{T}$ begins by setting the parameter M_t to the corresponding budget for day t or $\mathcal{B}(t)$ and resetting $U_b \in \mathbb{N}, b \in \{0, 1, \dots, B\}$ and will identify the estimated unsatisfied customers given b bikes on the station. The main loop can be broken down into two phases: the SAA step used to obtain F and then the observation of the station flow of customers and update of demand rates. On each day t , M_t sequences of customer arrivals are generated, for which the rates corresponding to each bin h on every iteration m are drawn from a random sample with a normal distribution of mean equal to the latest estimate of the corresponding rate ($\bar{\lambda}_{t,h}$ for returning and $\bar{\mu}_{t,h}$ for renters) and variance equivalent to $\bar{\lambda}_h$ divided by C_h (sum of the proportions until day t that bin h has been not full) and $\bar{\mu}_h$ divided by D_h (sum of the proportions until day t that bin h has been not empty). The remaining steps in the SAA portion are described in algorithm 2 and consist in repeating each generated sequence of customers with all possible initial number of bikes, ranging from starting with an empty station to starting with the station full, counting the number of unsatisfied customers observed for each initial number. U_b will be used to accumulate the unsatisfied customers over the M SAA samples using b initial bicycles. The policy to select the following day's station initial state can be defined by the operator of the station. In our implementation we always selected $\text{argmin } F$ bikes, but other policies such as selecting the most frequent value

of $\text{argmin } F$ could be explored. Once this is completed, we observe the next day's operation with the corresponding flow of customers and observability within each time bin, and use that information to update the estimation of $\bar{\lambda}_{t,h}$ and $\bar{\mu}_{t,h}$, as described in Algorithm 3.

Rate Estimation

We refer her to Algorithm 3 of appendix A.2. For each day t and time bin h we update the values of $\bar{\lambda}_{t,h}$ and $\bar{\mu}_{t,h}$, using the accumulated data stored in variables X, Y, C and D and observe the physical realization of the operation of the bicycle station (that in our case was simulated). This data is then used to update the rates according to what is described in line 8 of the algorithm.

The observation of customers' behavior at the station in successive days is what improves the estimation of the demand rates, as more information is gathered from observing the flow of customers and the effect this has on the number of bikes docked in the bike station, that will be bounded between empty $b = 0$ and full $b = C$.

We can observe from Figure 4.4 how the estimated rates (represented by a circle) after a single day of observation are not very accurate, and the 95% confidence interval is large, however as, by day 10 our estimation to the real values is much better, and by day 1,000, there seems to be minimal error, represented by the vertical distance between the estimate, and the solid lines of corresponding color that represents the true values.

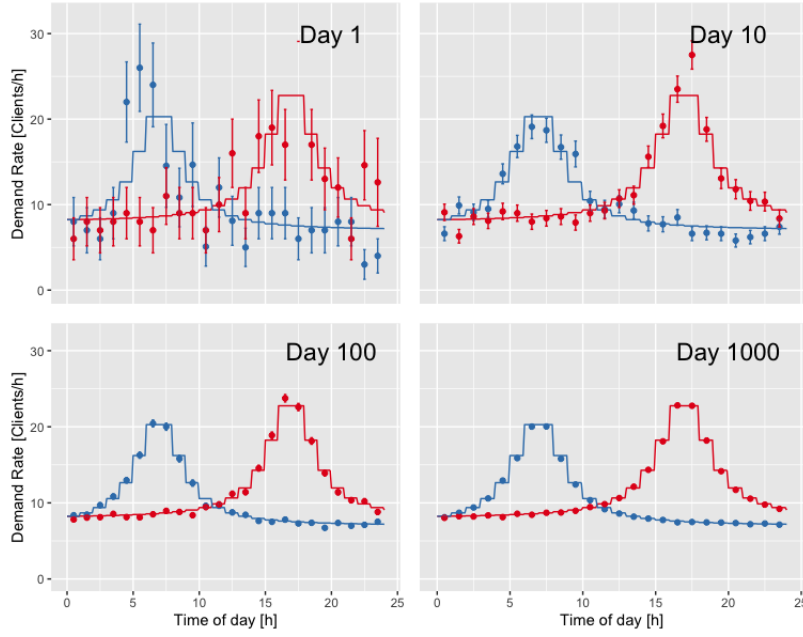


Figure 4.4. Rent (Blue) and Return (Red) Demand rate estimate by days 1, 10, 100, and 1000

To quantify the error we performed 20 repetitions of 1,000 days, and averaged the squared sum of the differences between the estimated values and the true values for each rate on day t , graphically depicted in Figure 4.5, in which the rapid decrease in the error of the rate estimation, is more informative if presented in a $\log_{10}(\cdot)$ scale.

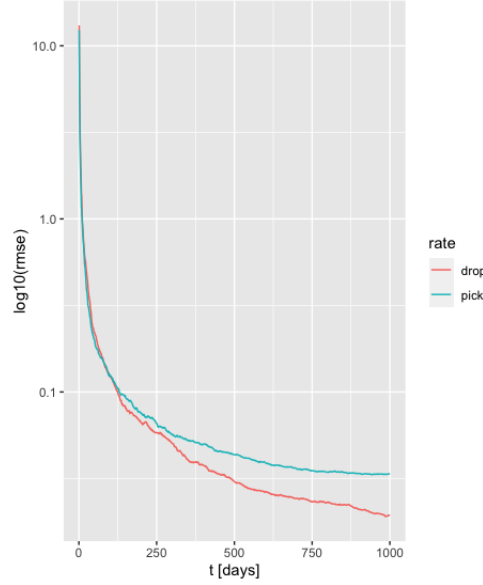


Figure 4.5. $\log_{10}(RMSE)$ of the estimation of demand rates.

The rate estimation presented previously included the effect correction required to account for censoring. Not considering this effect in the estimator would lead to a bias proportional to the probability that the specific time bin presented the corresponding form of censoring. That is, when the station was full in the drop-off demand rate and when all docks in the station were empty for the demand rate of customers picking up bikes. If we wrongly define the estimator of the demand rates for time bin i as:

$$\tilde{\lambda}_i = \frac{X_i}{\delta} \quad \tilde{\mu}_i = \frac{Y_i}{\delta}$$

we observe an underestimation of the rates occurs with some relation to the probability of observing censoring in that time bin. Figure 4.6 shows the bias generated in the estimation of the rate as the probability of censoring increases. The solid blue line represents the true rate, and the point represents the estimate of the rate in a time bin. In the regions where no censoring is observed each point lies on, or close to, the solid line, whereas in regions time bins that present censoring (shaded regions) we can observe a bias between the true value and the estimate that looks somewhat proportional to the probability it is censored or equivalently the proportion of time the corresponding demand was not observable.

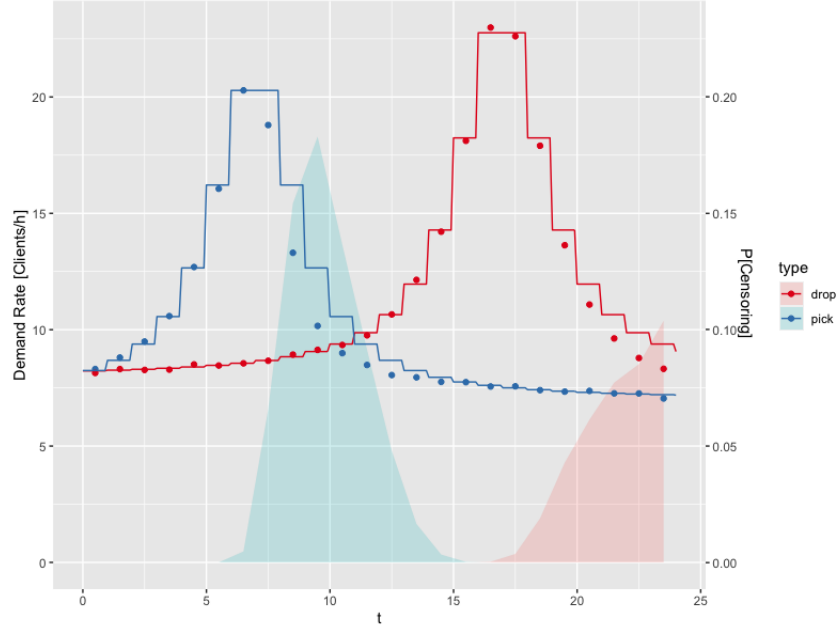


Figure 4.6. Effect of censoring on demand rate estimation, shaded areas are the corresponding probabilities of rack being empty (blue) or full (red).

4.2.2 Experimental Performance

Effect of Budget Distribution

One of the design decisions that were required for the implementation of the model was a distribution of the total computational budget from day 1 until the last day of the planning horizon, which in the experiment was set to 2,000 days.

We tested budget functions scaling like $o(\log(t))$, $o(1)$, $o(t^{1/2})$, $o(t^{1/4})$, $o(t^{1/8})$, $o(t^{1/10})$, and $o(t^{1/15})$, normalised as described above to use the same overall budget. Figure 4.7 plots these $\mathcal{B}(t)$ functions, and Figure 4.8 plots the cumulative regret of the algorithms using each choice of $\mathcal{B}(t)$.

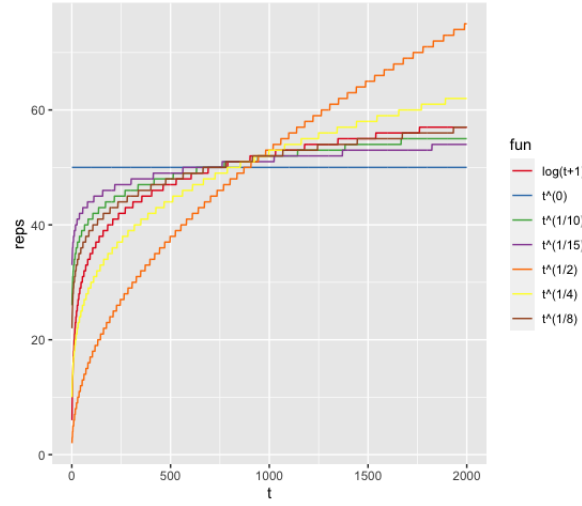


Figure 4.7. Distribution of computational budget in planned time horizon.

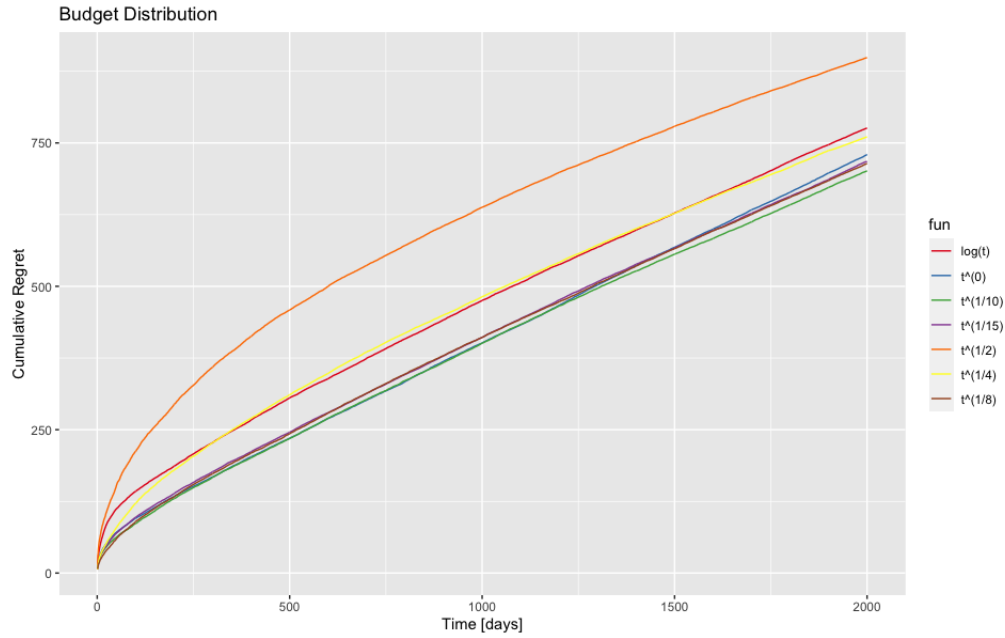


Figure 4.8. Cumulative regret for various budget distributions.

The resulting cumulative regret comparison plot for the different distributions of the budget throughout the planning horizon shows that there is significant effect of the budget allocation in the performance of the algorithm. The case where the distribution growth was proportional to $\sqrt{t}$ showed the worst cumulative regret. The explanation for this behavior

lies in variance generated the low number of samples for the initial days. On the extreme case when a constant number of samples is used for all days t , there is a good initial tendency in the curve, but a worse behavior towards the end of the experiment. Between these two distributions, the budget allocation proportional to $t^{1/10}$ seemed to yield the overall best cumulative regret behavior, and remain fixed throughout the subsequent experiments we demonstrate.

Effect of Total Budget

The selection of the overall budget size of 100,000 SAA samples was selected arbitrarily. Nonetheless we wished to see how using a larger or smaller total budget would impact the performance of the algorithm on the cumulative regret plots. This experiment was not performed with the objective of defining a tuning parameter for the algorithm, but rather with the purpose of empirically testing how a reduction in the error induced by the inequality presented in equation 3.13.

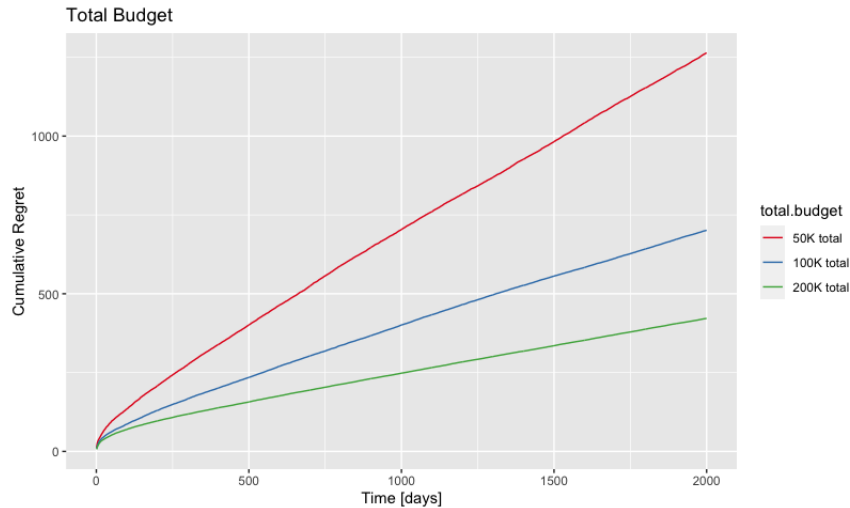


Figure 4.9. Cumulative regret for different total computational budgets

We observe that the total number of SAA samples used for the estimation of the F that informs our bike allocation decision has an important impact in the regret. This conclusion is valid when the policy for initial bike selection is done solely on this resulting minimizing argument of the function. An empirical estimate on the reducing effect of the budget allocation shows that the cumulative regret by day 2,000 decreases proportional to $o(\frac{1}{n})$.

4.3 Fixed Bins and Smooth Rates

We now consider the setting where the bins continue to be fixed but the rates are the smoothly varying functions in Figure 4.1. We will not specify the algorithm in this section since it is identical to the description provided in the previous section. The focus is to evaluate the effect of approximating the smooth function to a piecewise constant approximation. A side-by-side comparison between results of this implementation and the case modeled in the previous cases is not valid, since the underlying F is not exactly the same, because the customer rates are not exactly the same. Notwithstanding the above, their model minimizers match. In order to implement this algorithm, a decision has to be made regarding the number of fixed-size bins that will be used. Until this point, this had not been an issue since the 1 hour. bins happened to coincide with the true constant behaviour of customers for each hour, as shown in Figure 4.1.

4.3.1 Effect of Number of Bins

The next experiment, as for all subsequent ones, will use the continuous functions as presented in Figure 4.1 to characterize the true arrival of customers. The next numerical experiment was conducted with the purpose of comparing the effect on the cumulative regret of using different bin sizes for obtaining the piecewise constant estimate of each bin that, that is then used to generate the SAA samples. Much in the manner in which the comparison between different budget allocations was conducted, we wish to analyze the effect of approximating the continuous function shown in Figure 4.1 to intervals. Time bins of 0.25, 0.5, 1, 2, 4 and 6 hours were tested.

Our initial intuition, perhaps led us to think that smaller size bins that captured the shape of the curve with higher resolution, leading to an improvement in the performance of the algorithm, however the validity of this supposition was limited. We can derive from this plot that using smaller bin sizes does not imply an improvement in the cumulative regret we obtain throughout the planning horizon. The most favorable result was seen when the bin size was set to 30 minutes. The conclusion drawn from this experiment is highly case dependent, since the selection might depend on factors such as the characteristics of the curve that describes the behavior and the capacity of the station. When bins that resulted being too small were used, larger errors in the initial estimates of the rates were generated, leading to the sub-optimal regret plots we see in Figure 4.10 when we used 0.25-hour long

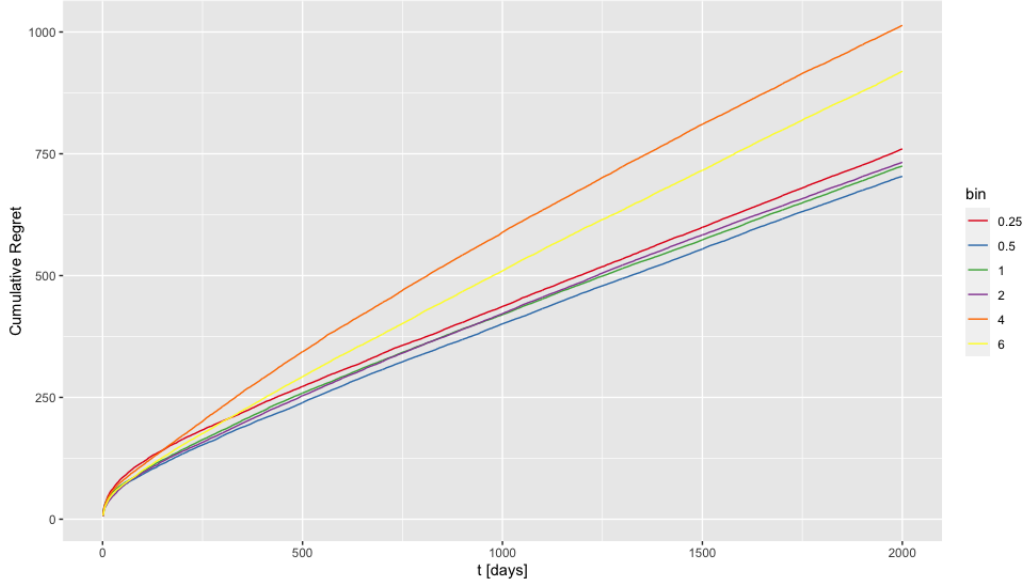


Figure 4.10. Cumulative regret for different time bin sizes

time bins. The observation of this behaviour was the motivator for the next algorithm that will use larger bins in the initial stages, when few observation of days are available for the estimation of the rates, and that gradually decrease in size in an effort to capture the shape of the demand curves with a higher resolution.

4.4 Adaptive Bins and Smooth Rates

We will now review the variant from the fixed-bin algorithm (Algorithm 1) in which the size of time bins gradually decrease in size. The pseudo-code implementation of this algorithm is found in Algorithm 4 in appendix A.3.

4.4.1 Algorithm

The previous section presented a model that was empirically shown to converge to an optimal solution using bins of fixed size. We will now present a modification where we gradually increase the number of time bins into which we divide the 24-hour day, splitting it up into equal size bins of smaller duration. We will employ 1-hour bins as the starting case, but will decrease the size as days progress. The function used to characterize the rate at which the size of the bins get smaller as a function of days t , was defined according to

what is described by Henderson (2003), where the mean-squared error in the estimation of the rate λ is minimized by defining the size of the time step δ as

$$\delta_t^* \sim \left(\frac{\lambda(h)}{2\lambda'(h)^2} \right)^{\frac{1}{3}} \cdot t^{-\frac{1}{3}} \quad (4.2)$$

Since h is assumed to be a fixed point in the day, we selected the moment of day that presented the highest probability of being censored. This decision could be changed to some other criteria, but no other guidance is provided in the literature so this was thought to be at least an initial point for further investigation. λ and μ presented censoring in different moments in the day, as shown in Figure 4.6, hence, the value of h will be selected differently for both parameters, resulting in different coefficients that will imply two separate bin sizes for each demand rate, as is detailed in Algorithm 4.

The overall structure of this algorithm follows the same sequence as Algorithm 1, with a few changes in the method we use to record the data, in order to be able to get the rate estimates with the increasing number of time bins of smaller size.

With this information, the operation of the station can begin to take place, where the next step would be to define the time bin size, for each one of the rate approximations. A few comments must be made for the proper implementation of this step:

1. The coefficients $\Phi_{\text{pick-up}}$ and $\Phi_{\text{drop-off}}$ must be selected by the operator at this moment since no clear guidance has been developed for the implementation. These coefficients define the decay rate in the size of the time bins.
2. For ease in the implementation, we approximated the size of the bin to multiples of $24 \cdot \delta_{\min}^{-1}$. We selected $\delta_{\min}$ as 1 minute, therefore in order to allow our implementation, the possible values of $\delta_{\text{drop-off}}$ and $\delta_{\text{pick-up}}$ were limited to : 1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 15, 16, 18, 20, 24, 30, 32, 36, 40, 45, 48 or 60.
3. A design feature that could be implemented if desired, is a fixed number of days before which $\delta_{\text{drop-off}}$ and $\delta_{\text{pick-up}}$ remain fixed. After that day, the values can be updated according to the corresponding rates.

The remaining sequence of steps correspond to the SAA implementation that is equivalent to the one performed in Algorithm 1, except for the fact that the generation of customer

sample flows of returning and renting bikes must be performed in separate steps of the sequences, due to the fact that they do not always have an equal number of time bins. Once the F has been generated in the SAA step, we can set the next day's initial state, according to the policy, that for our implementation was $\text{argmin } \bar{F}$. The following day's operation is observed and the rates are updated.

Initialization

The initial step consists of setting the initial size of the bins $\delta_{\text{drop-off}}$ and $\delta_{\text{pick-up}}$. In our implementation we set both to an initial value of 1 hour. Perhaps the most significant difference with respect to the fixed bin size algorithm is that we will set a parameter $\delta_{\min}$ that will be the smallest time interval that is possible and variables X, Y, C and D will be of dimension $\mathcal{R}^{\frac{24}{\delta_{\min}}}$. In our implementation, we defined this parameter to be 1 minute, and will be referred to as interval τ . Therefore $\delta_{\text{drop-off}}$ and $\delta_{\text{pick-up}}$ will be themselves composed of $\frac{\delta_{\text{drop-off}}}{\delta_{\min}}$ and $\frac{\delta_{\text{pick-up}}}{\delta_{\min}}$ sub intervals which will decrease in duration as a function of days t . X_h and $Y_h \in X$ and Y correspondingly will hold the cumulative total number of customers returning bikes, and customers renting bikes from $\tau = 0$ until the end of time bin $h \cdot \tau$. Likewise C_h and D_h will contain the total time the station has been full since the beginning of the day until the end of time bin $h \cdot \tau$. In the same manner that was used in Algorithm 1, we will adopt the convention that the capital letter variables will hold the cumulative values from day 0 until day t . Therefore each one of these variables will accumulate their corresponding values from the start of the day until the end of sub interval τ , from day 0 until day t .

SAA

The SAA portion of the algorithm is carried out in the same way it was done in the Fixed Bin case (Algorithm 1).

Rate Estimation

We refer here to Algorithm 5 in appendix A.4. In line 2 of Algorithm 4 where we set the size of the time bin $\delta_{\text{drop-off}}$ and $\delta_{\text{pick-up}}$ for day $t = 1$ we defined 1 hour intervals as a starting point, hence in order to obtain the initial estimate in Algorithm 5 we calculate the rates by obtaining the difference between the elements of the vector X and Y that are at 1

hr intervals, which are the number of customers observed during each $\delta_{\text{drop-off}}$ and $\delta_{\text{pick-up}}$ interval, and divide them by the total time the rack was observable, for those same intervals as described in detail in lines 16 and 20 Algorithm 5. Following estimations will use an updated size for $\delta_{\text{drop-off}}$ and $\delta_{\text{pick-up}}$ which will not necessarily match, since the functions that define their size are obtained have different coefficients ($\Phi_{\text{pick-up}}$ and $\Phi_{\text{drop-off}}$).

Bin Updating

The size of the bins was modeled to Equation 4.2. We limited the possible values to integer multiples of the 1440 minutes (δ_{min}) of the day. This constraint could be relaxed if dealt with appropriately.

4.4.2 Experimental Performance

Effect of Rebinning Rate

We will then repeat the experiments by implementing the time-varying bin size version of the algorithm, where, much like we did previously, we will verify the effect of different exponents in the decay of the bin size.

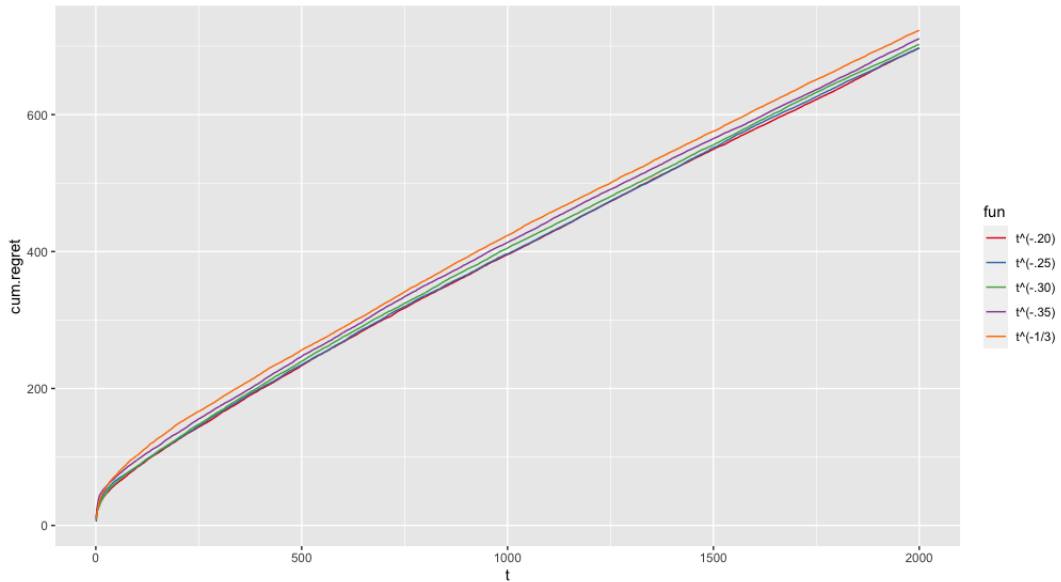


Figure 4.11. Cumulative regret using a range of exponents for the rate of decay of the bin size.

We observe that the best cumulative regret was obtained when the rate of decay was proportional to $t^{-0.25}$ which is different to the value suggested by Henderson (2003) of $t^{-\frac{1}{3}}$. We will not further explore the implication of this selection in this thesis, but it would certainly would be an interesting subject to analyze in further investigations.

4.5 Reuse of Samples

Inspired on the idea of Polyak Averaging, proposed by Polyak and Juditsky (1992), and the Green Simulation concept of Feng and Staum (2017), we wish to get a more efficient use of the limited computational budget by using past estimates of F obtained from rates that are in some measure similar to the most updated pair of rates $\bar{\lambda}$ and $\bar{\mu}$. We implemented two methods through which these past estimates can be averaged with the latest estimates of F that in practice improved our selection of the initial state for the following day's operation significantly.

Implementation required saving all past estimates of $\bar{\lambda}$ and $\bar{\mu}$ and estimates of F . We defined matrix L_t and Q_t to hold all past estimates of the rates. Each row contains the H time bin estimates of the rates for days 1 to t . We define matrix U that will store past estimates of the F in each row corresponding to a day's estimate of the number of expected unhappy customers for each initial number of bikes from 0 to B , e.g. element $u_{i,j}$ would be the estimated number of expected unsatisfied customers obtained on day i starting the day with j bikes on the station.

$$L_t = \begin{pmatrix} \bar{\lambda}_{1,1} & \bar{\lambda}_{1,2} & \cdots & \bar{\lambda}_{1,H} \\ \bar{\lambda}_{2,1} & \bar{\lambda}_{2,2} & \cdots & \bar{\lambda}_{2,H} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\lambda}_{t,1} & \bar{\lambda}_{t,2} & \cdots & \bar{\lambda}_{t,H} \end{pmatrix} \quad Q_t = \begin{pmatrix} \bar{\mu}_{1,1} & \bar{\mu}_{1,2} & \cdots & \bar{\mu}_{1,H} \\ \bar{\mu}_{2,1} & \bar{\mu}_{2,2} & \cdots & \bar{\mu}_{2,H} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\mu}_{t,1} & \bar{\mu}_{t,2} & \cdots & \bar{\mu}_{t,H} \end{pmatrix}$$

$$U_t = \begin{pmatrix} u_{1,0} & u_{1,1} & \cdots & u_{1,B} \\ u_{2,0} & u_{2,1} & \cdots & u_{2,B} \\ \vdots & \vdots & \ddots & \vdots \\ u_{t,0} & u_{t,1} & \cdots & u_{t,B} \end{pmatrix}$$

As we stated previously, we are interested in averaging the latest estimate of the F with past estimates that were generated in the SAA part of the algorithm using estimates of the demand rates that are in "some" measure similar to the most recent update. This similarity between the rates of day t_i and the latest rates for day t_j were saved as the elements of a vector V where each component i was defined as:

$$V_i = \sum_{h=1}^H (\bar{\lambda}_{i,h} - \bar{\lambda}_{j,h})^2 + (\bar{\mu}_{i,h} - \bar{\mu}_{j,h})^2.$$

We believe that there must be an optimal method for implementing the use of past estimates. Nonetheless, we will constrain the research of this idea to testing a number of options. The main issues associated with the fact of using past samples is that their underlying distribution is not identical, since it is defined by the $F_t(b, \bar{\lambda}_t, \bar{\mu}_t)$ and the computational budget $\mathcal{B}(t)$ that defines the number of samples. Another issue that should be accounted for when using past samples is the fact that the variance of the sample rates $\hat{\lambda}$ and $\hat{\mu}$ depend on the total sample size, that in this case corresponds to the total time each specific time bin has been visible. The purpose of presenting these issues is stating the fact that accounting for all of them in a strict analytically correct method can be a challenging task on its own and deviates from the topic of this thesis. Notwithstanding the previous fact, in an effort to try to capture all these issues, we implemented two different methods in a way that is at least intuitively reasonable.

4.5.1 Recency-Based Approach

This method consists on selecting past samples that were generated with rates that were in some measure similar to the latest estimate of λ and μ and then multiplying them by some constant that we will refer to as α raised to the exponent given by the number of days before the latest update of the rates. The selection of the days from which we will use past estimates relies on computing the measure of dissimilarity for each day i , which we will refer to as V_i , and setting a threshold Θ that will be used to select past samples that will be used for the estimation. Therefore we will have past estimates of the form:

$$w_i = \alpha^{t-i} \cdot \mathbf{I}(V_i < \Theta).$$

To select Θ we were interested in evaluating on a single run of the simulation, which would be the measure of dissimilarity after a 365 days of operation. Figure 4.12 presents a plot of value of the dissimilarity of rates for past days. We see from this plot that if we select a threshold $\Theta = 1$ by day $t = 365$, all estimates of F after day $t = 200$ will be used for averaging.

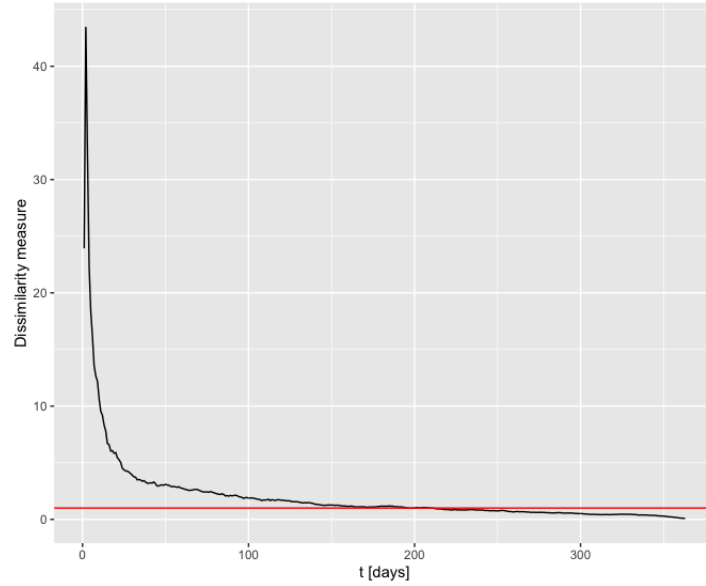


Figure 4.12. Measure of dissimilarity of the rates in a 364 days of operation of the station. Red line represents the selected threshold Θ .

The selection for a best value of Θ will depend on the scale of the rates, and no closed form method for the selection is provided as part of this thesis.

For the selection of the parameter α we tested a range of parameters between 0.9 and 1. Figure 4.13 presents the cumulative regret obtained by using selected values for α .

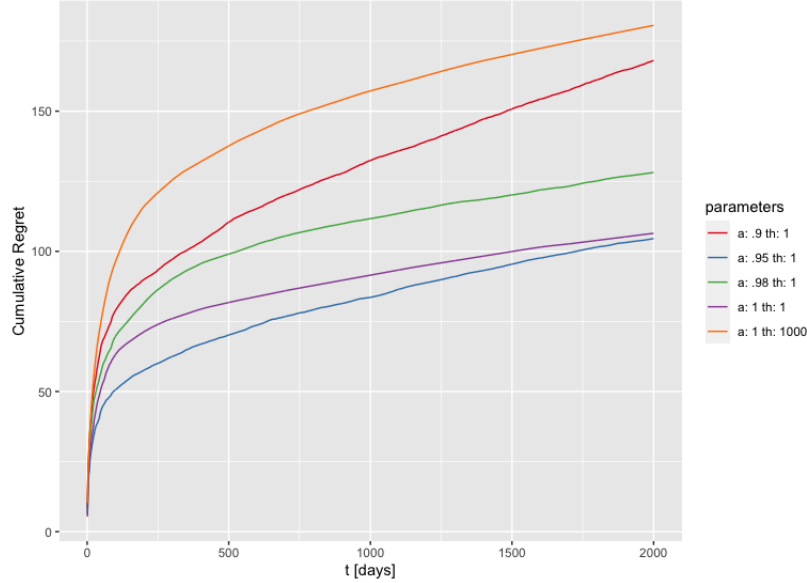


Figure 4.13. Cumulative regret plots for different selections of α and Θ

The value in our implementation that provided the best performance was $\alpha = 0.95$. Again, we state the fact that an optimal selection of this parameter was not covered as part of this thesis.

4.5.2 Similarity-Based Approach

This implementation is somewhat simpler since no parameters are required. It consists of defining a vector of weights by which we multiply past estimates of the F . The weights for each day i were defined by:

$$w_i = \left(1 - \frac{V_i - \min(V)}{\max(V)}\right) \frac{\mathcal{B}(i)}{\max(\mathcal{B}(1), \dots, \mathcal{B}(i), \dots, \mathcal{B}(t))}$$

The weights are defined by the product of the measure of similarity and the proportion of budget with which it was obtained, which is a proxy for the recency with which it was obtained, except if the computational budget is selected as constant.

4.5.3 Experimental Performance

We can observe in Figure 4.14 that the recency-based method (Method 1) outperforms the similarity-based approach (Method 2), for the selection of parameters we have used here.

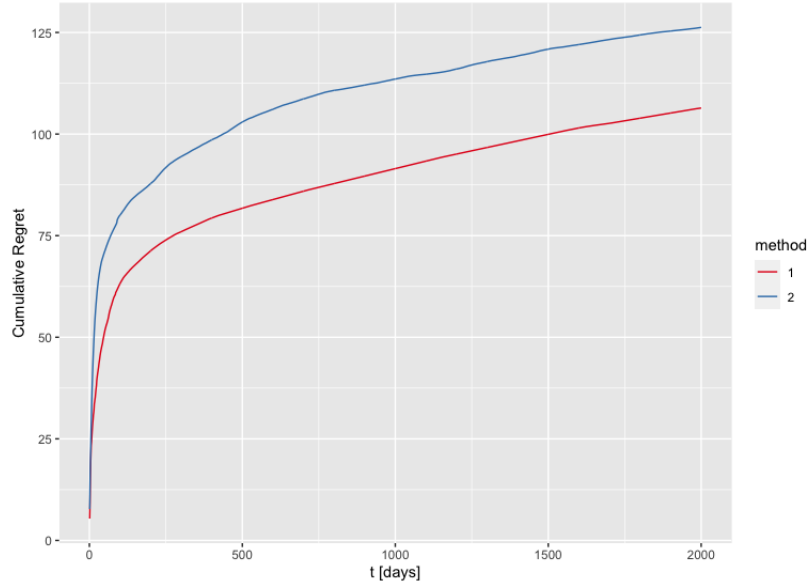


Figure 4.14. Cumulative regret plots for Method 1 ($\alpha = 1$ and $\Theta = 1$)

These two methods are heuristics and are simply a way of providing insight into the fact that the proper use of past samples can improve the performance of the algorithm. The optimal or most efficient use of these samples are not explored in this thesis but could certainly provide an interesting topic for further exploration.

CHAPTER 5:

Conclusion and Recommendations

This chapter provides a summary of the findings this research thesis has developed along with some concluding comments, accompanied by further exploration that can be developed using this work as a starting point.

5.1 Conclusion

This thesis has developed a framework that can be applied to a wide range of problems, where the sequential recollection of information improves the ability of the decision maker to select from the alternatives, even in situations where no feedback is available from observing the outcome of his choice. We have worked throughout Chapters 3 and 4 with a motivating example in the context of a bike-sharing system, since it provides a scenario that could be considered restrictive, in the sense that it is bounded on both extremes and no information on system behavior is available as an initial condition. Through the use of Sample Average Approximation, considering the fact that early on in the operation of the Station, the operator was able to make decisions that were gradually more and more informed, leading to decisions that converged to the optimal solution, in a situation where the use of classical analytical methods would not have been available. The variants of the model we developed in this research proved in theory and in practice to converge to an optimal solution. Nonetheless, the use of reusing of past samples seemed to really improve the performance of the algorithm, without an additional burden on the computational processing budget and overall running time. In the numerical implementation we were able to select parameters that seemed to be the best out of the selected range of values. Perhaps optimal values do exist for the specific scenario we have defined as an example. However, it was not the objective of this work to find a closed-form expression that facilitates the selection, and we believe that they would be somewhat case-specific.

5.2 Future Work

The ideas described in this thesis presented a model that proved to work when tested in a specific application, bounded by the assumptions and limitations described in Chapters 1 and 3. Hence, we propose a few extensions of the work we have developed, expanding theory, relaxing assumptions, optimization of parameters, and applications of the model.

5.2.1 Develop Theory for Re-Using Samples

In Chapter 4 we implemented a model based on the intuition that past samples empirically improved our estimates by increasing the number of samples with which we obtained $\bar{F}$ without a burden on the computational budget. Even though relevant research has been done (Polyak and Juditsky 1992) and (Feng and Staum 2017), it does not apply directly to our setting, therefore finding a mathematically optimal method for averaging past samples is necessary.

5.2.2 Relaxing Assumptions

Assumptions to develop the theory allowed us to develop a model that is analytically correct, which we then tested in a scenario that met all of these constraints. We understand that real world scenarios do not always behave according to the assumptions and limitations necessary for the development of this model. Nonetheless, it would make an interesting line of research to explore ways in which certain assumptions, such as the customer behaviour not being i.i.d., that could reflect certain seasonality in customer rates, or well the effect of customers leaning that in certain hours, bikes are likely to not be available. An even greater challenge would be including our model as part of a network of stations where the redistribution of bikes between stations occurs, in line with the research pursued by Freund et al. (2019).

5.2.3 Optimize Parameter Selection

In the development of Chapter 4 we presented three models that required a fine-tuning of parameters, such as computational budget allocation, time bin sizes and decreasing of bin sizes in the case of the adaptive model. Our approach to finding a good value for each one of these parameters consisted of testing each one from a limited range and selecting the one that seemed to perform best. We are aware that neither this method was exhaustive nor

found the optimal value for each one of the parameters, which we believe to be highly case dependant. Exploring methods for defining these parameters would constitute a natural complement to our research.

5.2.4 Implement Model in Other Applications

The motivating setting of bike-sharing systems we have used throughout this thesis does not limit the model to this application. We believe that many situations where an optimal decision exists and can be informed by sequentially learning from its behavior. A few ideas are proposed in Chapter 1 in which the model would be adequate. One of these are in the context of a Combat Support Hospital, where wounded warriors arrive and a physician is only available on call, leaving him unavailable for other hospitals. Another example is the assignment of surveillance assets to monitor geographical regions known for being the focus of illegal activity, such as fishing inside the Exclusive Economic Zone by Chinese fishing fleets (Quiñones Tarazona 2021). Here, naval assets need to be assigned, generating a cost if they do not patrol efficiently for the prevention of illegal activity that generates some penalty in case it is not undetected.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX:

A.1 Fixed Time Bin Algorithm

Algorithm 1 Fixed Time Bin Algorithm

Output: UDF Minimizing number of bikes to allocate on station.

- 2: Define δ interval size
Define $\mathcal{H} = \{1, 2, \dots, \frac{24}{\delta}\}$
 - 4: Set $\mathcal{T}$ planning horizon in days.
Define $\mathcal{B}(t)$ {Computational budget for day t }
 - 6: Set b to a number of bikes on day $t = 0$ {A policy such as half of rack capacity would be an example}
Set $X, Y, C, D \leftarrow 0$
 - 8: Initiate $\bar{\lambda}$ and $\bar{\mu}$, see Algorithm 3
for day $t = 1, \dots, \mathcal{T}$ **do**
 - 10: $M \leftarrow \mathcal{B}(t)$
 $U \leftarrow 0$
 - 12: **for** $m = 1, \dots, M$ **do**
 for all Time bin h in $\mathcal{H}$ **do**
 - 14: Sample $\dot{\lambda}_{m,h} \sim \mathcal{N}(\bar{\lambda}_h, \frac{\bar{\lambda}_h}{C_h})^+$ $\dot{\mu}_{m,h} \sim \mathcal{N}(\bar{\mu}_h, \frac{\bar{\mu}_h}{D_h})^+$ { $(\cdot)^+$ is $\max(\cdot, 0)$ }
 end for
 - 16: $\mathcal{E} \leftarrow$ Simulated arrival sequence of customers (rentals and returns) using $\dot{\lambda}_m$ and $\dot{\mu}_m$ See Algorithm 6
 for b in 0 to B **do**
 - 18: Do Algorithm 2
 end for
 - 20: **end for**
 $\bar{\mathcal{F}} \leftarrow \frac{U}{M}$
 - 22: Set $\text{argmin}(\bar{\mathcal{F}})$ to a number of bikes on day t
 Observe day t and Update $\bar{\lambda}$ and $\bar{\mu}$, see Algorithm 3
 - 24: **end for**
-

Algorithm 2 Generation of F

```
 $u_b \leftarrow 0$  {unhappy customers}  
2: for all customer arrival  $e$  in  $\mathcal{E}$  do  
    if  $e$  is drop-off AND  $b < \mathcal{B}$  then  
4:      $b \leftarrow b + 1$   
    else if  $e$  is drop-off AND  $b = \mathcal{B}$  then  
6:      $u_b \leftarrow u_b + 1$   
    else if  $e$  is pick-up AND  $b > 0$  then  
8:      $b \leftarrow b - 1$   
    else  
10:     $u_b \leftarrow u_b + 1$   
    end if  
12: end for  
 $U_b = U_b + u_b$ 
```

A.2 Rate Estimation Fixed Bin

Algorithm 3 Rate Estimation Fixed Bin

Input: X, Y, C and D
 2: **Output:** $\bar{\lambda}$ and $\bar{\mu}$
 Observe day t (Simulated with thinning)
 4: **for all** Time bin h in $\mathcal{H}$ **do**
 $x_h \leftarrow$ Count bike drop-off $y_h \leftarrow$ Count bike pick-ups
 6: $c_h \leftarrow \frac{\delta - \text{time full}}{\delta}$ $d_h \leftarrow \frac{\delta - \text{time empty}}{\delta}$
 $X_h \leftarrow X_h + x_h$ $Y_h \leftarrow Y_h + y_h$ $C_h \leftarrow C_h + c_h$ $D_h \leftarrow D_h + d_h$ {Capital letter
 variables will hold cumulative values}
 8: Estimate rates for time bin h

$$\bar{\lambda}_h = \frac{X_h}{C_h * \delta} \mathbb{I}(C_h > 0) \quad \bar{\mu}_h = \frac{Y_h}{D_h * \delta} \mathbb{I}(D_h > 0)$$

end for

A.3 Adaptive Bin Size Algorithm

Algorithm 4 Adaptive Bin Size Algorithm

Output: UDF ($\mathcal{F}$) Minimizing number of bikes to allocate on station next day.

- 2: Define $\delta_{\text{drop-off}}$ and $\delta_{\text{pick-up}}$ interval size.
Set $\mathcal{T}$ planning horizon in days.
- 4: Define $\mathcal{B}(t)$ {Computational budget for day t from 1 to $\mathcal{T}$ }
Set b to a number of bikes on day $t = 0$ {A policy such as half of rack capacity would be an example}
- 6: Set $X, Y, C, D \leftarrow 0$
Initiate $\bar{\lambda}$ and $\bar{\mu}$, see Algorithm 5
- 8: **for** day $t = 1, \dots, \mathcal{T}$ **do**
 $\delta_{\text{drop-off}} \leftarrow \Phi_{\text{drop-off}} * t^{-q}$
- 10: $\delta_{\text{pick-up}} \leftarrow \Phi_{\text{pick-up}} * t^{-q}$ {**note:** The selection of $\Phi_{\text{drop-off}}, \Phi_{\text{pick-up}}$ is somewhat arbitrary at the moment and q was tested empirically to be 25^{-1} }
 $M \leftarrow \mathcal{B}(t)$
- 12: $U \leftarrow 0$
for $m = 1, \dots, M$ **do**
- 14: **for all** Time bin h in $1, \dots, 24 * \delta_{\text{drop-off}}^{-1}$ **do**
 Sample $\dot{\lambda}_{m,h} \sim \mathcal{N}(\bar{\lambda}_h, \frac{\bar{\lambda}_h}{C_h})^+$
- 16: **end for**
 for all Time bin h in $1, \dots, 24 * \delta_{\text{pick-up}}^{-1}$ **do**
- 18: Sample $\dot{\mu}_{m,h} \sim \mathcal{N}(\bar{\mu}_h, \frac{\bar{\mu}_h}{D_h})^+ \{(\cdot)^+ \text{ is } \max(\cdot, 0)\}$
 end for
- 20: $\mathcal{E} \leftarrow$ Simulated customer arrival using $\dot{\lambda}_m$ and $\dot{\mu}_m$ See Algorithm 6
 for b in 0 to B **do**
- 22: Do Algorithm 2
 end for
- 24: **end for**
 $\bar{\mathcal{F}} \leftarrow \frac{U}{M}$
- 26: Set $\text{argmin}(\bar{\mathcal{F}})$ to a number of bikes on day t
 Observe day t and update $\bar{\lambda}$ and $\bar{\mu}$, see Algorithm 5
- 28: **end for**

A.4 Rate Estimation Using CIF

Algorithm 5 Rate Estimation Using CIF

```

1: Input:  $X, Y, C, D, \delta_{pick-up}$  and  $\delta_{drop-off}$ 
2: Output:  $\bar{\lambda}$  and  $\bar{\mu}$ 
3: Define  $\delta_{min}$ 
4: Observe day  $t$  (Simulated with thinning)
5: for all  $\tau$  in  $1, \dots, 24 * \delta_{min}^{-1}$  do
6:    $x_\tau \leftarrow x_{\tau-1} + N_{drop-off}((\tau - 1), \tau)$ 
7:    $y_\tau \leftarrow y_{\tau-1} + N_{pick-up}((\tau - 1), \tau)$ 
8:    $T_f \leftarrow$  time full between  $(\tau - 1)$  and  $\tau$ 
9:    $T_e \leftarrow$  time empty between  $(\tau - 1)$  and  $\tau$ 
10:   $c_\tau \leftarrow c_{\tau-1} + \frac{\delta_{min} - T_f}{\delta_{min}}$ 
11:   $d_\tau \leftarrow d_{\tau-1} + \frac{\delta_{min} - T_e}{\delta_{min}}$ 
12:   $X_\tau \leftarrow X_\tau + x_\tau$     $Y_\tau \leftarrow Y_\tau + y_\tau$     $C_\tau \leftarrow C_\tau + c_\tau$     $D_\tau \leftarrow D_\tau + d_\tau$ 
13: end for
14:  $\kappa \leftarrow \delta_{drop-off} \delta_{min}^{-1}$ 
15: for all  $h$  in  $1, \dots, 24 * \delta_{drop-off}^{-1}$  do
16:   $\bar{\lambda}_h \leftarrow \frac{X_{(h)\kappa} - X_{(h-1)\kappa}}{C_{(h)\kappa} - C_{(h-1)\kappa}}$ 
17: end for
18:  $\kappa \leftarrow \delta_{pick-up} * \delta_{min}^{-1}$ 
19: for all  $h$  in  $1, \dots, 24 * \delta_{pick-up}^{-1}$  do
20:   $\bar{\mu}_h \leftarrow \frac{Y_{(h)\kappa} - Y_{(h-1)\kappa}}{D_{(h)*\kappa} - D_{(h-1)*\kappa}}$ 
21: end for

```

A.5 Simulate Customer Pick-Up and Drop-Off Flow

Algorithm 6 Simulate Customer Pick-Up and Drop-off flow

Input: $\dot{\lambda}$, $\delta_{\text{drop-off}}$, $\dot{\mu}$ and $\delta_{\text{pick-up}}$

2: **Output:** $\mathcal{E}$ simulated flow of customers

for all j in $1, \dots, 24 * \delta_{\text{drop-off}}^{-1}$ **do**

4: Generate first drop-off in time bin j $e_0 \leftarrow (j - 1) + EXP(\dot{\lambda}_1)$

$\kappa \leftarrow 1$

6: **while** $e_\kappa \leq j$ **do**

$e_\kappa \leftarrow e_{\kappa-1} + EXP(\dot{\lambda}_1)$

8: $\kappa \leftarrow \kappa + 1$

end while

10: $\mathcal{E}_{\text{drop-off}} \leftarrow \{\mathcal{E}_{\text{drop-off}}, e_0, \dots, e_\kappa\}$

end for

12: **for all** i in $1, \dots, 24 * \delta_{\text{pick-up}}^{-1}$ **do**

 Generate first pick-up in time bin i $e_0 \leftarrow (i - 1) + EXP(\dot{\mu}_1)$

14: $\kappa \leftarrow 1$

while $e_\kappa \leq i$ **do**

16: $e_\kappa \leftarrow e_{\kappa-1} + EXP(\dot{\mu}_1)$

$\kappa \leftarrow \kappa + 1$

18: **end while**

$\mathcal{E}_{\text{pick-up}} \leftarrow \{\mathcal{E}_{\text{pick-up}}, e_0, \dots, e_\kappa\}$

20: **end for**

$\mathcal{E} \leftarrow \text{Sorted}\{\mathcal{E}_{\text{drop-off}}, \mathcal{E}_{\text{pick-up}}\}$

List of References

- DeMaio P (2009) Bike-sharing: History, impacts, models of provision, and future. *Journal of Public Transportation* 12, URL <http://dx.doi.org/10.5038/2375-0901.12.4.3>.
- Feng M, Staum J (2017) Green simulation: Reusing the output of repeated experiments. *ACM Transactions on Modeling and Computer Simulation* 27:1–28, URL <http://dx.doi.org/10.1145/3129130>.
- Freund D, Henderson SG, O’Mahony E, Shmoys DB (2019) Analytics and bikes: Riding tandem with motivate to improve mobility. *INFORMS Journal on Applied Analytics* 49(5):310–323, URL <http://dx.doi.org/10.1287/inte.2019.1005>.
- George DK, Xia CH (2011) Fleet-sizing and service availability for a vehicle rental system via closed queueing networks. *European Journal of Operational Research* 211(1):198–207, URL <http://dx.doi.org/https://doi.org/10.1016/j.ejor.2010.12.015>.
- Grant J (2020) *Poisson process bandits: Sequential models and algorithms for maximising the detection of point data*. Ph.D. thesis, Lancaster University, URL <http://dx.doi.org/10.17635/lancaster/thesis/871>.
- Hazan E (2016) Introduction to online convex optimization. *Foundations and Trends in Optimization* 2:157–325, URL <http://dx.doi.org/10.1561/24000000013>.
- Henderson S (2003) Estimation for nonhomogeneous poisson processes from aggregated data. *Operations Research Letters* 31:375–382, URL [http://dx.doi.org/10.1016/S0167-6377\(03\)00027-0](http://dx.doi.org/10.1016/S0167-6377(03)00027-0).
- Kim S, Pasupathy R, Henderson S (2015) *A Guide to Sample Average Approximation*, volume 216, 207–243 (Springer New York), URL http://dx.doi.org/10.1007/978-1-4939-1384-8_8.
- Kleywegt A, Shapiro A, Homem-De-Mello T (2001) The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization* 12:479 – 483, URL <http://dx.doi.org/10.1137/S1052623499363220>.
- Leurent F (2012) Modelling a vehicle-sharing station as a dual waiting system: stochastic framework and stationary analysis. *Working paper. LVMT - Laboratoire Ville, Mobilité, Transport* URL <https://hal.archives-ouvertes.fr/hal-00757228>, 19 pages.
- Lewis P, Shedler G (1979) Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly* 26:403–413, URL <http://dx.doi.org/10.1002/nav.3800260304>.

- Nair R, Miller-Hooks E (2011) Fleet management for vehicle sharing operations. *Transportation Science* 45(4):524–540, URL <http://dx.doi.org/10.1287/trsc.1100.0347>.
- NATCO (2020) Bike share snapshot 2020-bike share and shared micromobility initiative. NACTO Website , August 27,2020. [Online], URL <https://nacto.org/2020/08/27/136-million-trips-taken-on-shared-bikes-and-scooters-across-the-u-s-in-2019/>.
- Polyak B, Juditsky A (1992) Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization* 30:838–855, URL <http://dx.doi.org/10.1137/0330046>.
- Quiñones Tarazona P (2021) Pesca ilegal China, la gran amenaza para los océanos latinoamericanos [illegal Chinese fishing, the great threat to latinamerican oceans]. *Revista de Marina (Chile)* 1(981), URL <https://revistamarina.cl/es/articulo/pesca-ilegal-china-la-gran-amenaza-para-los-oceanos-latinoamericanos>.
- Raviv T, Kolka O (2013) Optimal inventory management of a bike-sharing station. *IIE Transactions* 45:1077–1093, URL <http://dx.doi.org/10.1080/0740817X.2013.770186>.
- Resnick SI (1992) *Adventures in stochastic processes* (Boston: Birkhaeuser).
- Ross SM (1997) *Introduction to Probability Models* (San Diego, CA, USA: Academic Press), 6th edition.
- Schuijbroek J, Hampshire R, van Hoeve WJ (2017) Inventory rebalancing and vehicle routing in bike sharing systems. *European Journal of Operational Research* 257(3):992–1004, ISSN 0377-2217, URL <http://dx.doi.org/https://doi.org/10.1016/j.ejor.2016.08.029>.

Initial Distribution List

1. Defense Technical Information Center
Ft. Belvoir, Virginia
2. Dudley Knox Library
Naval Postgraduate School
Monterey, California