Planning Decisions Under Uncertainty: A Data-driven Approach

Natarajan Gautam

Professor, Electrical Engineering & Computer Science, Syracuse University
Amazon Scholar, Amazon Inc.

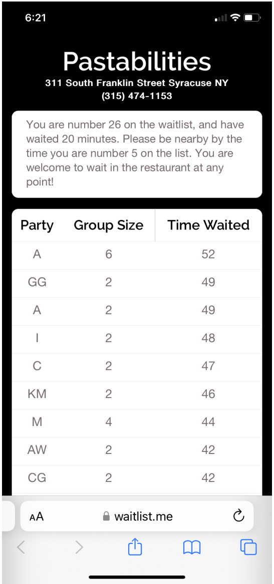
Acknowledgments for Research

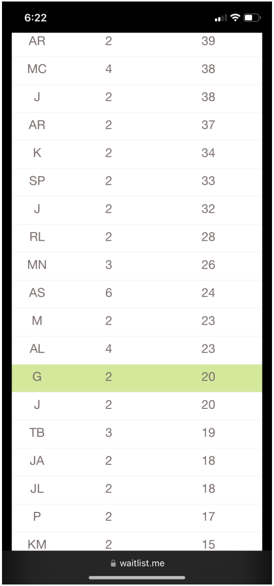
- Grateful to Prof. George Shanthikumar for his insights and ideas that shaped the analysis
- This presentation is predominantly based on a paper that is to appear in *INFORMS Journal on Data Science*
- The presentation also has results from research with former student Preet Karia; a big thanks to him
- DISCLAIMER: This work is completely independent of the research done at Amazon

Motivation: Going to a Restaurant

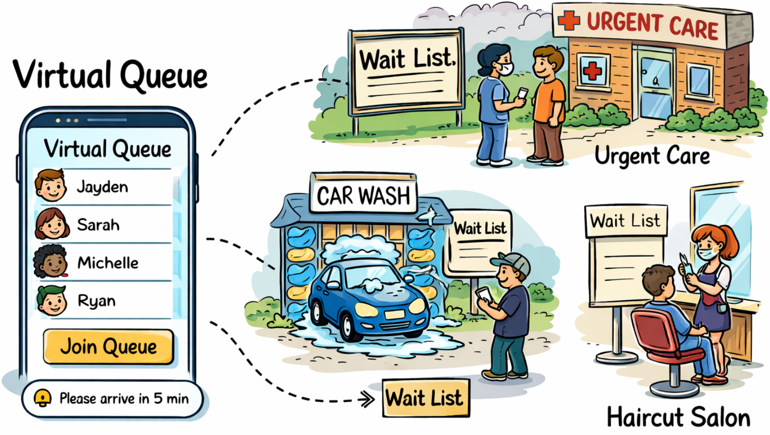
- Pastabilities is a popular restaurant in Syracuse
- Pastabilities does not take reservations
- But you can get on a wait list (through `waitlist.me`), i.e., virtual queue
- When you check your phone, it will tell you how much time each person ahead of you has waited
- Say that you live 30 minutes away and it is necessary to show up when your turn happens
- When should you leave home? Arriving early or late is a problem

Motivation: After 20 Minutes of Waiting

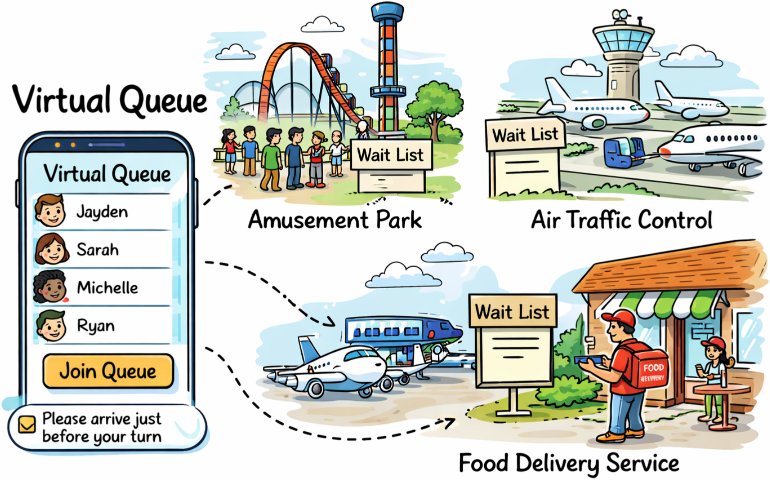




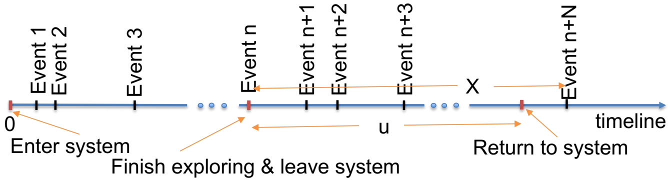
Examples of Virtual Queues



More Examples of Virtual Queues

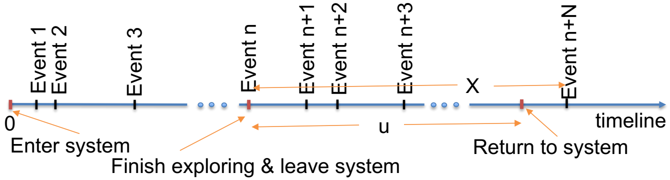


Generic Stylized Setting



- Get on a virtual queue and you are $N + n$ in line at time 0
- Events corresponding to change in your position occur according to a renewal process
- Inter-event times according to gamma distribution with (unknown) shape parameter k and scale parameter θ
- *Explore* by observing the time between events for the first n events
- *Exploit* to return at time u from now when the N^{th} event from now occurs, i.e., the $N + n^{th}$ event since time 0

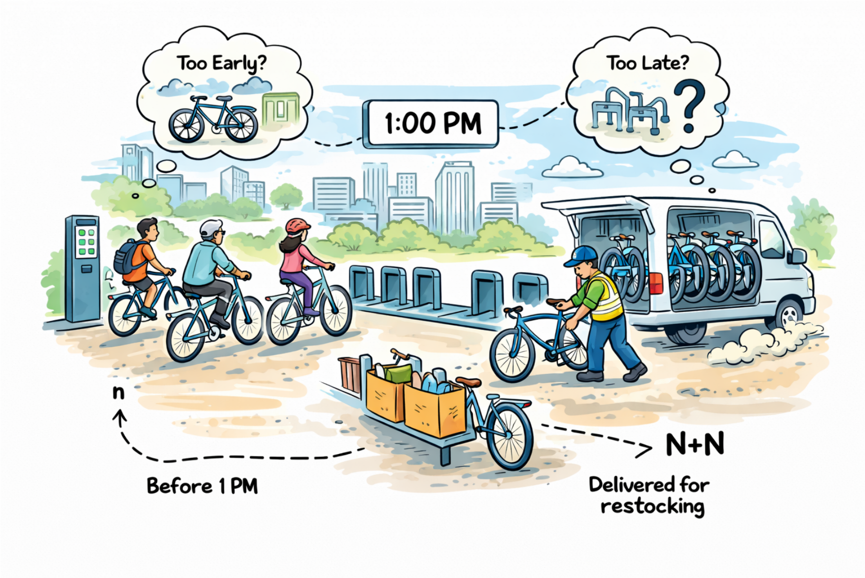
Problem Statement in Stylized Setting



- Let X (random variable) denote the time between the n^{th} and $N+n^{th}$ event
- Decision variable u is when to arrive back to the system after exploring
- Let h be the holding cost per unit time spent waiting for the $N+n^{th}$ event to occur (when $u < X$)
- Let p be the penalty cost per unit time late in reaching the system (when $u > X$)
- The data-driven online optimization problem under uncertainty is: *What should the arrival time u be so that the expected cost is minimized?*
- Notice that between the n^{th} and $N+n^{th}$ event we do not observe the system (due to travel time, but also others as we will see)

Customized Fulfilment of $N + n$ Items: Storage Cost vs. QoS

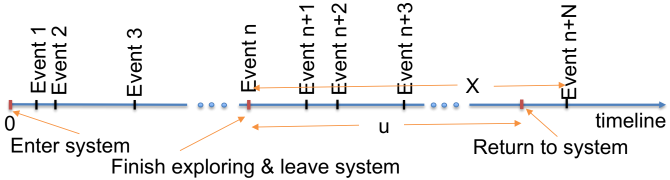
Restocking $N + n$ Bikes: Wait vs. Availability



New Fashion Item Reorder $N + n$: Inventory Holding vs. Opportunity

- A retail store has $N + n$ units of a new fashion accessory
- The demand process for the fashion accessory is uncertain as this is a new item
- So, the store samples the first n customers to understand the demand rate
- Then the store needs to decide when to place an order for the next shipment of the fashion accessory so that with sufficient lead time the order would arrive at the right time
- If the orders arrive too early, then the store would incur an inventory cost as well as any insurance-related holding cost
- If the order would arrive after the store runs out of the fashion accessory, then the store loses opportunities to make sales of the fashion accessory

Reiterating Notation for Stylized Setting



- *Explore* by observing the time between events for the first n events, then *Exploit* to return at time u
- Inter-event times according to gamma distribution with (unknown) shape parameter k and scale parameter θ
- Let X (random variable) denote the time between the n^{th} and $N + n^{th}$ event
- Then X is according to a Gamma distribution with unknown shape parameter Nk and scale parameter θ
- The mean of X is $Nk\theta$ and variance is $Nk\theta^2$
- A holding cost h per unit time or a penalty cost p per unit time is incurred upon coming early or late

Analysis by An Oracle: Full Knowledge of k and θ

- Let $C(k, \theta, u)$ be the expected cost if the time to return to the system to observe the $n + N^{th}$ event is u
- The CDF $F_X(x)$ is

$$F_X(x) = \mathbb{P}(X \le x) = \int_0^x \frac{1}{\theta} e^{-y/\theta} \frac{(y/\theta)^{Nk-1}}{\Gamma(Nk)} dy$$

- Then,

$$\begin{aligned} C(k, \theta, u) &= \int_0^u (u - x) p dF_X(x) + \int_u^\infty (x - u) h dF_X(x) \\ &= puF_X(u) - p \int_0^u x dF_X(x) + h \int_u^\infty x dF_X(x) - hu(1 - F_X(u)) \end{aligned}$$

- This is a standard news vendor problem

News Vendor Solution

- The value u^* that minimizes $C(k, \theta, u)$ is

$$u^* = F_X^{-1}\left(\frac{h}{h+p}\right).$$

- Gaussian approximation: When N is large, using Central Limit Theorem, X is approximately normally distributed with mean $Nk\theta$ and variance $Nk\theta^2$
- Let α be the score where the standard normal random variable has a CDF $\Phi(\alpha) = h/(h+p)$
- The α does not depend on N, k, or θ,

$$u^* = (Nk + \alpha\sqrt{Nk})\theta$$

Modeling and Analysis for Unknown θ

- We use Operational Data Analytics (ODA) to derive results for the case assuming k is known but θ is unknown
- Feng and Shanthikumar (2023) also obtain ODA results for different systems using gamma distribution where either k is known or is estimated and used directly
- It will also allow us to study the exponential distribution inter-event times which is a special case with $k = 1$
- When θ is unknown and k is known, we can estimate θ using the exploration phase where we collect n sample realizations of $D_1, \ldots, D_n$, which we denote as $d_1, \ldots, d_n$
- Both method of moments and maximum likelihood estimator give us

$$\hat{\theta} = \sum_{i=1}^{n} \frac{d_i}{kn}$$

since the average of the n realizations has a mean of $k\theta$
- It is important to note that n is small (we do not have the luxury to explore forever)
- So there could be a significant difference in $\hat{\theta}$ values, depending on the realizations

Predict and Then Optimize (PTO)

- Predict-then-optimize (PTO) is to make decisions in two stages
- In the first stage one makes predictions of the unknown θ
- Then in the second stage one makes the optimal decision assuming the prediction or estimate is accurate
- In PTO, we predict θ as $\hat{\theta}$ first, then we optimize to reach the system at time u^*_{PTO} given by
$$u^*_{PTO} = (kN + \alpha\sqrt{kN})\hat{\theta}$$
- If the realized value of X is x, the cost is
$$\phi(u^*_{PTO}, x) = p\max(u^*_{PTO} - x, 0) + h\max(x - u^*_{PTO}, 0)$$
which we test as an average across multiple such sets of realizations
- How different the PTO-based costs would be compared to
$$\phi(u^*, x) = p\max(u^* - x, 0) + h\max(x - u^*, 0)$$
where $u^* = (Nk + \alpha\sqrt{Nk})\theta$
- Would the average across multiple sets of realizations differ between using θ and $\hat{\theta}$?
- How different u^* and u^*_{PTO} are on average? In addition, we will test them against a method using ODA described next

ODA Preliminaries

- On one hand, the only thing we have is $\hat{\theta}$, the best estimate of θ
- On the other hand, we also want to pick a more optimal u than u^*_{PTO}
- We ask if there is a z so that we use a biased estimate $z\hat{\theta}$ for θ
- To obtain such a z, we use ODA for which it is useful to see the connection between $C(k, \theta, u)$ and

$$\phi(u, x) = p \max(u - x, 0) + h \max(x - u, 0)$$

- Based on that,

$$C(k, \theta, u) = \int_0^\infty \phi(u, x) dF_X(x) \ = \mathbb{E}\left[\phi(u, X) | k, \theta\right]$$

  where the expression in the middle has k and θ embedded in $F_X(\cdot)$
- Our goal is to find a z that minimizes the expected value of $C(k, \theta, u)$ when θ is unknown but only an estimate $\hat{\theta}$ can be obtained
- For ODA to work (i.e., to find a z that is independent of θ), $\phi(u, x)$ needs to satisfy a homogeneous property
- A sufficient condition is to show that for any c_0 we have

$$\phi(c_0 u, c_0 x) = c_0 \phi(u, x)$$

  which is satisfied, hence homogeneous

ODA Preliminaries (Continued)

- All we have at the end of exploration are the realizations $d_1, \ldots, d_n$ with which we compute the estimate $\hat{\theta}$
- Using that estimate we compute the optimal value of u where instead of using
$$u^*_{PTO} = (Nk + \alpha\sqrt{Nk})\hat{\theta}$$
we use
$$u_z = z(Nk + \alpha\sqrt{Nk})\hat{\theta}$$
which is nothing but zu^*_{PTO}
- The value of z must be such that:
(a) there exists an optimal z^* so that among all values of z, the one that produces the lowest cost on average is z^*;
(b) the optimal value z^* is independent of the unknown θ which is assured through the homogeneous property;
(c) when $n \to \infty$, $z^* \to 1$ so that the resulting z^* is asymptotically exact

Optimal z using ODA

- To find a z that minimizes the expected cost, we start with the n observations D_1 to D_n
- D_i is a gamma random variable with shape parameter k and scale parameter θ
- $D_1 + \ldots + D_n$ is according to a gamma distribution with parameters nk and $1/\theta$
- Let random variable $\hat{\Theta}$ be

$$\hat{\Theta} = \frac{1}{kn} \sum_{i=1}^{n} D_i$$

  differentiating with $\hat{\theta}$ which is based on the realizations $d_1, \ldots, d_n$
- Then we write random variable U_z as

$$U_z = (Nk + \alpha\sqrt{Nk})\hat{\Theta}z = z(Nk + \alpha\sqrt{Nk}) \sum_{i=1}^{n} \frac{D_i}{kn}$$

  for some real-valued z

Optimal z using ODA (Continued)

- The random variable version of the expected cost $C(k, \theta, u)$ is now described as $C(k, \theta, U_z)$
- Let $\eta(z)$ be the expectation of this cost

$$\eta(z) = \mathbb{E}\left[C(k, \theta, U_z)\right] = \mathbb{E}\left[C\left(k, \theta, \frac{z(Nk + \alpha\sqrt{Nk})}{kn}\sum_{i=1}^{n} D_i\right)\right]$$

- As $n \to \infty$, since $\frac{1}{n}\sum_{i=1}^{n} D_i \to k\theta$, we have $\eta(z) \to C(k, \theta, zu^*)$ which attains a minimum at $z = 1$, hence $z^* \to 1$ as $n \to \infty$
- Since $\mathbb{E}[C(k, \theta, U_1)|D_1, \ldots, D_n]$ equals the PTO expected cost which can be written as $\mathbb{E}[\mathbb{E}[\phi(U_1, X)|k, \theta, D_1, \ldots, D_n]]$ and due to Jensen's inequality is larger than or equal to $\mathbb{E}[\phi(\mathbb{E}[U_1], X)|k, \theta, D_1, \ldots, D_n]] = C(k, \theta, u^*)$ as $\phi(u, x)$ is a convex function of u
- By selecting a z different from 1, we try to get as close to $C(k, \theta, u^*)$ as possible

Computing the Optimal z

- Let $\beta = (kN + \alpha\sqrt{kN})/(nk)$ and $Y = D_1 + \ldots + D_n$ which is a gamma random variable with mean $nk\theta$ and variance $nk\theta^2$
- Hence we have

$$U_z = z\beta Y \quad \text{and}$$
$$\eta(z) = \mathbb{E}[C(k, \theta, z\beta Y)]$$

which can be written as

$$\eta(z) = \int_0^\infty C(k, \theta, z\beta y) f_Y(y) dy$$

where

$$f_Y(y) = \frac{1}{\theta} \frac{e^{-y/\theta}}{\Gamma(nk)} \left(\frac{y}{\theta}\right)^{nk-1}$$

is the PDF of Y which is a gamma random variable with shape parameter nk and scale parameter θ

- Substituting for $C(k, \theta, z\beta y)$ and state as the CDF of X, $F_X(x)$ as

$$\begin{aligned} \eta(z) \quad = \quad & (h+p) \int_0^\infty z\beta y \int_0^{z\beta y} dF_X(x) f_Y(y) dy \\ & -(h+p) \int_0^\infty \int_0^{z\beta y} x dF_X(x) f_Y(y) dy + hk\theta N - hz\beta nk\theta \end{aligned}$$

Computing the Optimal z (Continued)

- The derivative of $\eta(z)$ with respect to z is given by

$$\eta'(z) = (h+p)\int_0^\infty \beta y \int_0^{z\beta y} dF_X(x) f_Y(y) dy - h\beta nk\theta$$

Rewriting that as

$$\eta'(z) = (h+p)\int_0^\infty \beta y F_X(z\beta y) f_Y(y) dy - h\beta nk\theta$$

we get a form that would use useful in the proof of the next theorem

- *Theorem*: There exists a unique solution to $\eta'(z^*) = 0$ that results in an optimal z^* that is independent of θ given by the solution to

$$(h+p)\int_0^\infty \beta w F_{X_1}(z\beta w) f_{Y_1}(w) dw = h\beta nk$$

where X_1 and Y_1 are special cases of X and Y respectively with scale parameters of 1 (i.e., $\theta = 1$)

- Using ODA, the optimal time to reach the system is

$$u_z^* = z^*(Nk + \alpha\sqrt{Nk})\hat{\theta}$$

which uses the same $\hat{\theta}$ in PTO

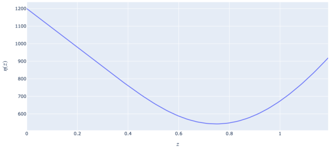
- Then, the cost corresponding to the realization of X and $D_1, \ldots, D_n$ is

$$\phi(u_z^*, x) = p\max(u_z^* - x, 0) + h\max(x - u_z^*, 0)$$

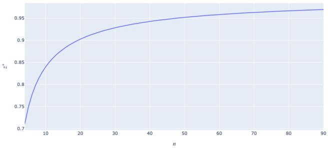
Special Case of $k = 1$: Exponential Distribution

- Nice closed-form expressions exist for some of the calculations for the exponential distribution case
- The optimal value of z that minimizes $\eta(z)$ is the unique solution to

$$
\begin{aligned}
p\beta n - \frac{(h+p)Nn\beta}{(z\beta+1)^{n+1}} + \\
(h+p)\sum_{i=1}^{N}(N-i)\binom{i+n-1}{i}\left(\frac{z\beta}{z\beta+1}\right)^{i}\left(\frac{1}{z\beta+1}\right)^{n+1}\left(\frac{i}{z}-n\beta\right) &= 0
\end{aligned}
$$



(a) $\eta(z)$ versus z



(b) z^* versus n

Extending to Unknown k and θ

- Here both k and θ of a gamma inter-event time are unknown
- We have a single sample of n independent realizations of the gamma random variable with shape parameter k and scale parameter θ
- Once k is "estimated" we simply use in the ODA results, and hence that is easy
- We have are n sample realizations d_1, d_2, ..., d_n that are taken in the exploratory phase
- Define the sample mean $\hat{\mu} = (d_1 + d_2 + \ldots + d_n)/n$ and the sample variance $\hat{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \hat{\mu})^2$
- The sample mean and sample variance are unbiased estimators of $\mathbb{E}[D_i]$ and $\mathbb{V}[D_i]$ respectively

Extending to Unknown k and θ (Continued)

- Using Method of Moments,

$$\hat{\theta}_{MOM} = \frac{\hat{\sigma^2}}{\hat{\mu}} \quad \text{and}$$

$$\hat{k}_{MOM} = \frac{\hat{\mu}}{\hat{\theta}_{MOM}}$$

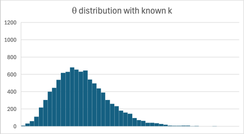
- Using Maximum Likelihood Estimators

$$\hat{\theta}_{MLE} = \frac{3 - \hat{w} + \sqrt{(3 - \hat{w})^2 + 24\hat{w}}}{12\hat{w}} \quad \text{where}$$

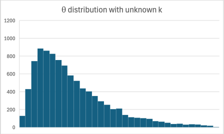
$$\hat{w} = \ln(\hat{\mu}) - \frac{1}{n}\sum_{i=1}^{n} \ln(d_i) \quad \text{and}$$

$$\hat{k}_{MLE} = \frac{\hat{\mu}}{\hat{\theta}_{MLE}}$$

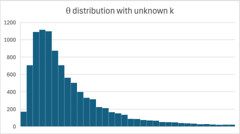
Histogram of $\hat{\theta}$ When k is Unknown



(a) $\hat{\theta}$ with k known



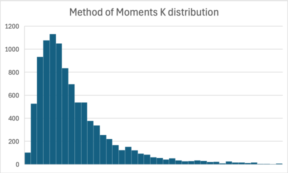
(b) $\hat{\theta}_{MOM}$



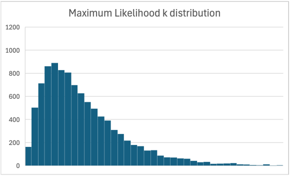
(c) $\hat{\theta}_{MLE}$

Histogram of estimate of θ (0 to 18.3) over 10000 experiments with $k = 1.33$, $\theta = 6$, and $n = 6$

Histogram of $\hat{k}$



(a) $\hat{k}_{MOM}$



(b) $\hat{k}_{MLE}$

Histogram of estimate of k (0 to 10) over 10000 experiments with $k = 1.33$, $\theta = 6$, and $n = 6$

Simulated Data Using Gamma Inter-Event Times

- We performed 5000 experiments with fixed parameters $N = 50$, $n = 7$, $h = 6$, $p = 10$, $k = 1.67$, and $\theta = 6$
- A different set of fixed parameters or varying parameters similar yielded qualitatively similar outcomes
- Each of the 5000 experiments were generated by creating $N + n$ realizations of the gamma random variable with shape parameter k and scale parameter θ
- Although represented as $\overline{z^*}$, it is actually simply z^*
- Aggregate results for gamma when k is known and all parameters fixed

$\overline{\hat{\theta}}$	$\overline{z^*}$	$\overline{u^*}$	$\overline{u^*_{PTO}}$	$\overline{u^*_z}$
6.01	0.89	483.53	484.20	429.50
$\overline{u^*_{bayes}}$	$\overline{\phi(u^*, x)}$	$\overline{\phi(u^*_{PTO}, x)}$	$\overline{\phi(u^*_z, x)}$	$\overline{\phi(u^*_{bayes}, x)}$
467.21	329.87	938.04	865.73	354.72

Simulated Data Using Gamma Inter-Event Times: Unknown k

- We performed 5000 experiments with fixed parameters $N = 50$, $n = 7$, $h = 6$, $p = 10$, $k = 1.67$, and $\theta = 6$
- These results for unknown k are performed alongside the case when k is known, so the gamma random variables realized are exactly the same for both cases
- The average estimate of k and θ are significantly different from the actual
- But the results corresponding to the various approaches (especially PTO and ODA represented as z^*) of both the time to arrive $\overline{u^*_{\cdot,\cdot}}$ as well as the cost $\overline{\phi(u^*_{\cdot,\cdot}, x)}$ averaged over the experiments are relatively similar

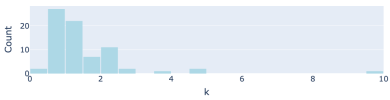
$\overline{\hat{\theta}_{MOM}}$ 5.51	$\overline{\hat{\theta}_{MLE}}$ 5.20	$\overline{\hat{k}_{MOM}}$ 2.61	$\overline{\hat{k}_{MLE}}$ 2.73	$\overline{z^*_{MOM}}$ 0.90	$\overline{z^*_{MLE}}$ 0.90
$\overline{u^*_{PTO,MOM}}$ 485.61	$\overline{u^*_{PTO,MLE}}$ 485.99	$\overline{u^*_{z,MOM}}$ 436.20	$\overline{u^*_{z,MLE}}$ 438.98	$\overline{u^*_{saa}}$ 482.98	$\overline{u^*_{bayes}}$ 484.68
$\overline{\phi(u^*_{PTO,MOM}, x)}$ 941.84	$\overline{\phi(u^*_{PTO,MLE}, x)}$ 943.07	$\overline{\phi(u^*_{z,MOM}, x)}$ 883.50	$\overline{\phi(u^*_{z,MLE}, x)}$ 886.96	$\overline{\phi(u^*_{saa}, x)}$ 944.36	$\overline{\phi(u^*_{bayes}, x)}$ 1300.66

Using Real Bike Share Data for Stock Replenishment

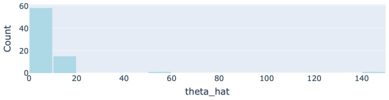
- An automated bicycle rental stand with $n+N$ electronic docks, each initially holding one bike
- The stand opens at 12pm, and customers begin arriving one by one, swiping credit cards and taking bikes, which are returned at other locations
- The arrival times of the first n customers, all before 1pm are observed
- Let T denote the (unknown) time when the $(n+N)^{\text{th}}$ customer arrives and takes the last available bike, leaving all docks empty
- Need to decide at 1pm what time in the evening to return with $n+N$ bikes to restock the stand
- Arriving before T, means a holding cost of h dollars per minute while waiting for a dock to become available
- Arriving after T means the system experiences a stock-out, incurring a penalty of p dollars per minute for the duration it remains empty
- Data from the Metro Bike Share system in Los Angeles County, available at `https://bikeshare.metro.net/about/data/`

A Highly Heterogeneous System

- Although there are 90 days in the first quarter of 2025 between January 1 and March 31, only 74 days were suitable for analysis
- As 16 days had fewer than $n = 4$ between 12pm and 1pm
- For the remaining 74 days, using n as the number of arrivals between 12pm and 1pm we obtain $d_1, \ldots, d_n$ as the time between arrivals and computed their mean and standard deviation for each day
- Every day was different, the distribution was unknown, it was not necessarily IID or homogeneous
- Using method of moments we compute $\hat{k}$ and $\hat{\theta}$ (histogram below) assuming the inter-arrival time is according to a gamma distribution
- The values of n and N were different each day, ranging from 4 to 17 for n and 17 to 107 for N
- We used the same $p = 10$ and $h = 5$ for all days



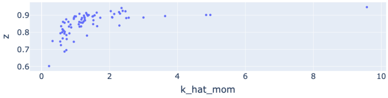
(a) $\hat{k}_{MOM}$



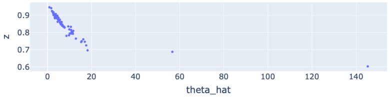
(b) $\hat{\theta}_{MOM}$

Obtaining z^* Values

- Assuming a gamma distribution for the time between customer arrivals each day, albeit with different parameters
- We obtain z^* values for each day by solving for z in the expression in the theorem
- A graph of the resulting z^* is plotted against $\hat{k}_{MOM}$ and $\hat{\theta}_{MOM}$ below
- It is important to note that although z^* does not depend on θ, it is dependent of k whose estimates vary from day to day
- Also, n and N also vary from day to day
- The general trend is that as $\hat{k}_{MOM}$ increases, z^* also increases
- However, as $\hat{\theta}_{MOM}$ increases, z^* decreases
- Notice that in all these examples z^* is less than 1.



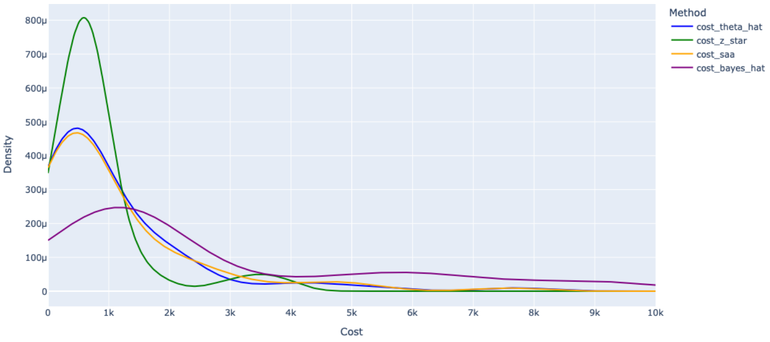
(a) z^* vs. $\hat{k}_{MOM}$



(b) z^* vs. $\hat{\theta}_{MOM}$

Results: Averaged Over 74 Days

$\overline{\hat{\theta}_{MOM}}$ 9.05	$\overline{\hat{k}_{MOM}}$ 1.49	$\overline{z^*_{MOM}}$ 0.85	$\overline{X}$ 355.40
$\overline{u^*_{PTO}}$ 388.16	$\overline{u^*_{z}}$ 328.08	$\overline{u^*_{saa}}$ 393.80	$\overline{u^*_{bayes}}$ 487.95
$\overline{\phi(u^*_{PTO}, x)}$ 1048.40	$\overline{\phi(u^*_{z}, x)}$ 807.46	$\overline{\phi(u^*_{saa}, x)}$ 1078.80	$\overline{\phi(u^*_{bayes}, x)}$ 2646.04

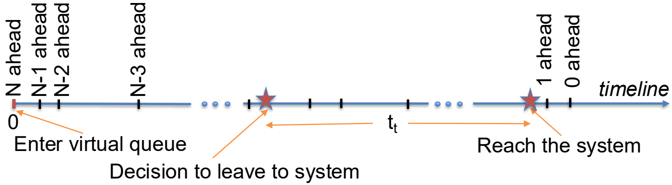


Concluding Remarks

- Impact on stochastic modeling when we assume distributions
- First predicting then optimizing is not as effective
- Even though we do not capture the parameters accurately, but their effects are not contaminated by that
- The MOM and MLE ways to estimate k and θ yield in the average sense almost identical results
- Across solution methods the differences are more pronounced
- The key contribution of this work is in deriving an expression to compute z^* for the stylized setting
- The setting itself can be used in a variety of applications including waiting in virtual queues, inventory reorders, and order promises which are fairly common in the service industry
- We provide description of a bike-rental scenario and use real data where it is not even known whether they are sampled from a gamma distribution
- Nonetheless the results there and in other simulated data instances show that the ODA approach provides the minimal average cost compared to PTO as well as other techniques such as a bootstrapped sample average approximation (SAA) and Bayesian methods

Further Work

- As an immediate next step, we considered a variant of this problem where the penalty cost is a fixed lump sum cost c and does not depend on how much late we are
- Unfortunately, we cannot use ODA for that as the condition for ODA in terms of homogeneous property is not satisfied and we use a reinforcement learning approach (includes a hybrid ODA and a machine learning aspect)
- Further, in this work we fixed n, but that is also a decision variable; for how long should we explore?



Analysis (Known Parameter Case)

- Let h be the holding cost per unit time and c the penalty cost for arriving late
- Let X_N be a random variable denoting the time from now that our turn would come to enter the restaurant
- The cost incurred is $h \max\{X_N - u_N, 0\} + c\mathbb{I}(u_N - X_N > 0)$
- In terms of the CDF $F_N(x)$ and the PDF $f_N(x)$, the value of u_N that minimizes the expected value of the total cost is

$$\frac{f_N(u_N^*)}{1 - F_N(u_N^*)} = \frac{h}{c}$$

  provided X_N is an increasing failure rate (IFR) random variable
- In the dynamic case, let k be the number of customers ahead of us
- Assuming X_k is IFR (which will be true for Gamma if $k > 1$), then u_k^* is the solution to

$$\frac{f_k(u_k^*)}{1 - F_k(u_k^*)} = \frac{h}{c}$$

- Approach: for $k = N$ to 1, wait for the next event or u_k^* whichever happens first, then leave

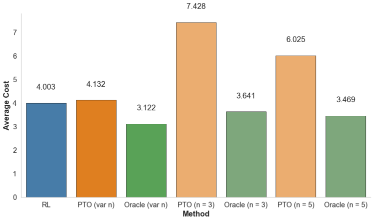
Analysis (Unknown Parameter Case)

- PTO with fixed n (equals 3 and 5)
- PTO under dynamic case (simular to the known parameter case, but with estimates)
- ODA directly does not work as the condition

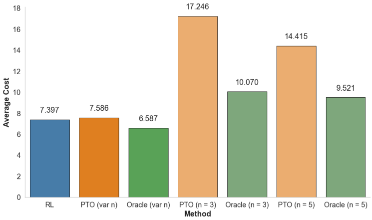
$$\phi(c_0 u, c_0 x) = c_0 \phi(u, x)$$

  is not satisfied when $\phi(u, x) = h \max\{x - u, 0\} + c\mathbb{I}(u - X > 0)$
- Reinforcement Learning State Space
  - $\frac{N-n}{N}$, this gives us a fraction of events that are yet to occur which makes the state between 0 and 1 and hence works for any N;
  - $\frac{\hat{u}_n^* - t_t}{\hat{\alpha}_n}$, this gives us a scaled time that is left at our end, so we know when to leave;
  - $\frac{\hat{u}_n^* - t_t}{\hat{\beta}_n}$, this gives us a scaled time that is left at our end, but this time scaled by $\hat{\beta}_n$ instead of $\hat{\alpha}_n$
- Action space is to wait or leave and reward function is $-\phi(u, x)$ at the end of an episode

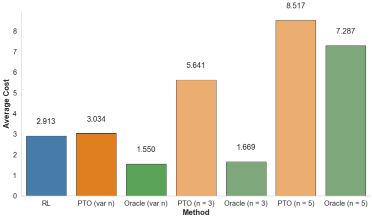
Results: Gamma With Oracle



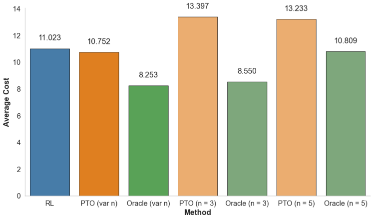
Low-α with Low-t_t



High-α with Low-t_t

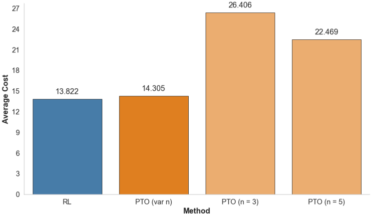


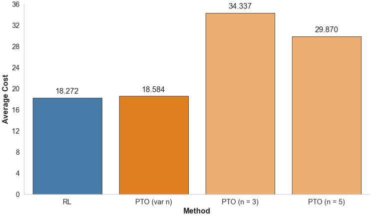
Low-α with High-t_t



High-α with High-t_t

Resuls: Simulated With Different IID and Bike Share





Future Work

- Making the travel time t_t stochastic with estimated parameters that are updated (and routes changed)
- Solving finite state MDPs by learning the P matrices for each action (with application in adversarial radar control)
- Using physics based models for solving machine learning problems (reverse of what was seen here)