SOC1901 : Parametric Learning of M/M/1 Queue

Kam Ho Cheung
khcheun7 at cse.cuhk.edu.hk

Tsang Kit Cheung
tkcheun7 at cse.cuhk.edu.hk

Siu On Chan*
siuon at cse.cuhk.edu.hk

Abstract

This project tries to perform parametric learning of M/M/1 Queue in general which is a special kind of continuous-time Markov Chain from Queuing Theory in PAC- framework.

1 Introduction

M/M/1 queue is a basic model in queuing theory which is a common model in our daily application, such as the number of users waiting for the ATM banking service, patients waiting at a clinic, a router getting and sending packets in a computer network, and even cars waiting to be serviced by the toll booth. They can all be modeled as an M/M/1 queue.

These cases can imply the M/M/1 queue model to analyze the arrival rate of the customer and the service rate of the server. By taking samples, we can assess the flow of customers the capacity of the service. The model analytic helps with marketing in company business or community study.

M/M/1 queue is a birth-death process. Which the first letter M means that the inter-arrival times follows the exponential distribution with constant parameter $\lambda > 0$, and the second letter M means that the service times(or life-time) follows the exponential distribution with constant parameter $\mu > 0$, and the 1 means the queue only have a single server. And we assume there is no customer in the queue at the beginning. It can be represented in continuous-time Markov chain (Figure 1), with the initial state at state 0

*Project supervisor

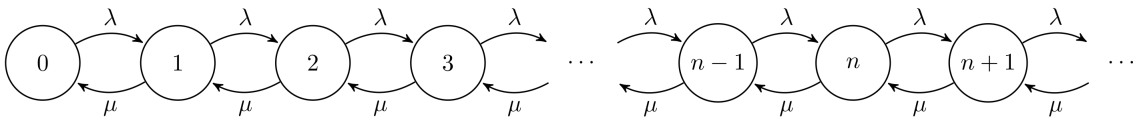

Figure 1: M/M/1 Queue in Markov chain representation(Source from Wikipedia[1])

1.1 Mathematical Characteristic of M/M/1 Queue[1]

In this model, the state space is over an infinite set, $\mathbb{N} = \{0, 1, 2, 3, ...\}$, which the state value represents the number of the customer in the queue. Customers will come in the queue such that,

1. Arrival rate is a constant at every state $k \geq 0$, i.e.

$$\lim_{\Delta t \to 0} \frac{P_{k(k+1)}\left(T \leq \Delta t\right)}{\Delta t} = \lambda$$

   which the inter-arrival time, $T \sim Exp(\lambda)$ i.e.

$$p_{k(k+1)}(T = t) = \lambda e^{-\lambda t}$$

2. Every arrival is independent of each other, i.e.

$$Pr\left(\text{arrival in}[t, t + \Delta t]\right) = Pr\left(\text{arrival in}[0, \Delta t]\right) = P_{k(k+1)}\left(T \leq \Delta t\right)$$

3. Arrivals follows a Poisson process, which in every time intervals $[t, t + \Delta t]$, #arrivals $N \sim Poisson(\lambda t)$, i.e.

$$Pr\left(N = n\right) = \frac{(\lambda \Delta t)^n}{n!} e^{-\lambda \Delta t}$$

While customers will be served and leave the queue such that,

1. Service rate is a constant at every state $k > 0$, i.e.

$$\lim_{\Delta t \to 0} \frac{P_{k(k-1)}\left(T \leq \Delta t\right)}{\Delta t} = \mu$$

   which the service time, $T \sim Exp(\mu)$ i.e.

$$p_{k(k-1)}(T = t) = \mu e^{-\mu t}$$

2. First come first serve with single server, which every service is independent of each other.

Which $P_{ab}(T \leq \Delta t)$ indicates the probability of a state transition, from state a to state b, Δt unit of time after entering state a.

[1]See definition in Queuing systems [22]

1.2 Parametric learning

Parametric learning in distribution learning theory under PAC-framework will be used to apply on M/M/1 Queue in the following context.

Definition 1. ***Parametric learning under PAC-framework.***[2] *The unknown M/M/1 Queue, $Q(\lambda, \mu)$, is successfully learned by the hypothesis M/M/1 Queue, $\hat{Q}(\hat{\lambda}, \hat{\mu})$, for given arbitrary small error $\epsilon > 0$ with high confident level $1 - \delta$, for arbitrary small $\delta > 0$, if and only if,*

$$\mathbb{P}\left[Dist(Q, \hat{Q}) \leq \epsilon\right] \geq 1 - \delta$$

For some distance function indicating the difference between Q and $\hat{Q}$ which we will discuss later.

There are different distance function which each serve different purpose.

Definition 2. ***Absolute error.*** *The most common distance is by the absolute error of parameter For unknown M/M/1 Queue, $Q(\lambda, \mu)$, and the hypothesis M/M/1 Queue, $\hat{Q}(\hat{\lambda}, \hat{\mu})$,*

$$Dist(Q, \hat{Q}) = \max\left(|\lambda - \hat{\lambda}|, |\mu - \hat{\mu}|\right)$$

Which provide accurate error in small scale of parameter.
However, when the parameter is very big, e.g. $\lambda_1 = 100, \lambda_2 = 99, \mu_1 = 1, \mu_2 = 0.1$, although the absolute difference of λ may be larger, in fact λ only differ by 1%, while the number of death process sample may differ by a factor of 10 in the long run.

Definition 3. ***Relative error.*** *Another distance is measured by the relative error of parameter For unknown M/M/1 Queue, $Q(\lambda, \mu)$, and the hypothesis M/M/1 Queue, $\hat{Q}(\hat{\lambda}, \hat{\mu})$,*

$$Dist(Q, \hat{Q}) = \max\left(|\log\frac{\lambda}{\hat{\lambda}}|, |\log\frac{\mu}{\hat{\mu}}|\right)$$

Which provide a relative sense of scale in large scale of parameter.
However, when the parameter is very small, e.g. $\lambda_1 = 1, \lambda_2 = 2, \mu_1 = 0.001, \mu_2 = 0.01$, although the relative difference of μ may be larger, in fact μ only differ by 0.009, while the number of death process sample may differ by 1 or 2 as most of the sample are going to be birth.

Both definition serve the same purpose, as when $Dist(Q, \hat{Q}) = 0, \lambda = \hat{\lambda}, \mu = \hat{\mu}$ which is our goal.

In the following context, we parametric learn unknown M/M/1 queue, $Q(\lambda, \mu)$ with samples drawn from the queue to estimate the birth rate λ and death rate μ under PAC-framework. Achieving arbitrary small error with arbitrary high confidence level with bounded number of time or bounded number of samples. We will introduce two sampling methods, Time-based and Counting-based, in case of different learning limitation.

[2]Similar ideas in P.2-3, Introduction to testing graph properties[8]

2 Related Work

2.1 Concentration inequalities

Concentration inequality is the most used method for estimating the probability that the algorithm successfully learn the M/M/1 Queue.

Theorem 1. ***Folklore Fact.***
There is a good concentration bound for Poisson random variables [3] - namely, sub-exponential. Let $X \sim Poisson(\lambda)$, for $\lambda > 0$. Then for any accuracy parameter $\epsilon > 0$,

$$\mathbb{P}\left[|X - \lambda| \geq \epsilon\right] \leq 2e^{-\frac{\epsilon^2}{2(\lambda+\epsilon)}}$$

Theorem 2. ***Tight tail bounds for sum of N i.i.d. exponential random variables $X_i \sim Exponential(\lambda)$[9]***
For $X = \sum_{i=1}^{N} X_i$, where $X_i \sim Exponential(\lambda)$independently, where $\mathbb{E}[X] = \mu$.
Upper tail bound, for any error parameter $\eta \geq 1$,

$$\mathbb{P}\left[X \geq \eta\mu\right] \leq \eta^{-1}e^{-\lambda\mu(\eta-1-\ln\eta)}$$

Lower tail bound, for any error parameter $\eta \leq 1$,

$$\mathbb{P}\left[X \leq \eta\mu\right] \leq e^{-\lambda\mu(\eta-1-\ln\eta)}$$

In the following context, we consider $\eta = e^{\epsilon}$ in upper tail bound and $\eta = e^{-\epsilon}$ in lower tail bound for $\epsilon > 0$ for simplicity.

Theorem 3. ***Multiplicative Chernoff Bound.****[12] Let $X = \sum_{i=1}^{n} X_i$ where X_i are independent 0-1 random variables. Let $\mu = \mathbb{E}\left[X\right]$, for any $\delta > 0$, upper tail bound*

$$\mathbb{P}\left[X \geq (1+\delta)\mu\right] \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$$

for any $0 < \delta < 1$, lower tail bound

$$\mathbb{P}\left[X \leq (1-\delta)\mu\right] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}$$

Theorem 4. ***The tightness of Chernoff Bound.****[14] Let $X = \sum_{i=1}^{n} X_i$ where X_i are independent 0-1 random variables. Let $\mu = \mathbb{E}\left[X\right]$,*

$$\mathbb{P}\left[X - \mu \geq t\right] \geq \frac{1}{4}e^{-\frac{-2t^2}{\mu}}$$

Remark. *In the following context, we will mention using Chernoff Bound on a Poisson distribution. What we meant is applying different kind of Chernoff style bound on Poisson distribution which we considered as a special case of Binomial Distribution by the Poisson Limit Theorem. For any given Poisson random variable $X \sim Poisson(\lambda)$, there exist $Y \sim Binomial(n, \frac{p}{n})$, which $X = Y$, when $n \to \infty$.*

3 Samples of an unknown M/M/1 Queue

Sampling method: Given an known M/M/1 Queue, we collect all birth time and death time under a time bound T **(Time Based)** or under a Counting bound N**(Counting Based)**.

3.1 Collect samples in bounded time: Time Based

Samples of an unknown M/M/1 Queue will be drawn based on bounded time. **(Time Based)**
This is the most straight forward bound on the number of samples, as customers came at a finite rate, with the finite-state transition in a given time. However, as birth rate λ and death rate μ can be arbitrarily small, which for any given time-bound,
For any time bound $T \in \mathbb{R}, T > 0$. Let a sample set $\Upsilon(T)$ is collected in time interval $[0, T]$ from an unknown M/M/1 Queue $Q(\lambda, \mu)$, including birth time sample multi-set $T^B \in \mathbb{R}^b$, $b \in \mathbb{N}$ and death time sample multi-set $T^D \in \mathbb{R}^d$, $d \in \mathbb{N}$, which indicates the time the customer arrive or leave.

Time Based

$$\begin{aligned}
\Upsilon(T) &= \{T^B, T^D\} \\
&= \{\{t_1^B, t_2^B, ..., t_b^B\}, \{t_1^D, t_2^D, ..., t_d^D\}\}, \forall i : t_i^B \leq t_i^D \\
T^B &= \{t_1^B, t_2^B, ..., t_b^B\}, \forall i : t_i^B \leq t_{i+1}^B \leq T \\
T^D &= \{t_1^D, t_2^D, ..., t_d^D\}, \forall i : t_i^D \leq t_{i+1}^D \leq T
\end{aligned}$$

3.2 Collect samples in bounded number of state transition: Counting Based

Samples of an unknown M/M/1 Queue will be drawn based on the number of state transition in the unknown M/M/1 Queue. **(Counting Based)**
For any counting bound $N \in \mathbb{N}, N > 0$. Let a sample set $\Upsilon(N)$ is collected in N birth or death process from an unknown M/M/1 Queue $Q(\lambda, \mu)$, including birth time sample multi-set $T^B \in \mathbb{R}^b$, $b \in \mathbb{N}$ and death time sample multi-set $T^D \in \mathbb{R}^d$, $d \in \mathbb{N}$, $b + d = N$, which indicate the time the customer arrive or leave.

Counting Based

$$\begin{aligned}
\Upsilon(N) &= \{T^B, T^D\} \\
&= \{\{t_1^B, t_2^B, ..., t_b^B\}, \{t_1^D, t_2^D, ..., t_d^D\}\}, \forall i : t_i^B \leq t_i^D \\
T^B &= \{t_1^B, t_2^B, ..., t_b^B\}, \forall i : t_i^B \leq t_{i+1}^B \\
T^D &= \{t_1^D, t_2^D, ..., t_d^D\}, \forall i : t_i^D \leq t_{i+1}^D, b + d = N
\end{aligned}$$

Which is reasonable to have both sampling method, which will be discuss later.

3.3 Inter-arrival time, service time and idle period

As we will be measuring the inter-arrival time and the service time, for simplicity we define a function for measuring such time interval here. For the k-th customer of the queue,

$$arrival(k) = t_k^B - t_{k-1}^B, t_0^B = 0, 1 \leq k \leq b$$

$$service(k) = t_k^D - \max\left(t_k^B, t_{k-1}^D\right), t_0^D = 0, 1 \leq k \leq d$$

$$idle(k) = \max\left(0, t_k^B - t_{k-1}^D\right), t_0^D = 0, 1 \leq k \leq d+1$$

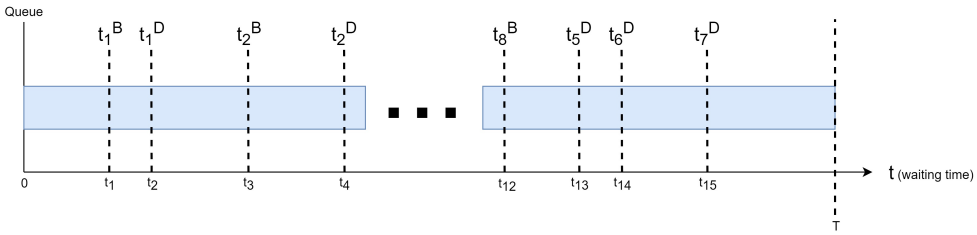

Figure 2: Graph represents the sampling of time of birth and death.

We build a simulator of the M/M/1 Queue sampling that enter the birth rate λ and death rate μ, it will generate a queuing sample set represented by schedule. Y-axis denotes the list of customers and X-axis denotes the timeline the 1 seconds is a unit of time. For each queuing sample, the blue bar represents the waiting time of the customer and the red bar is the service time.

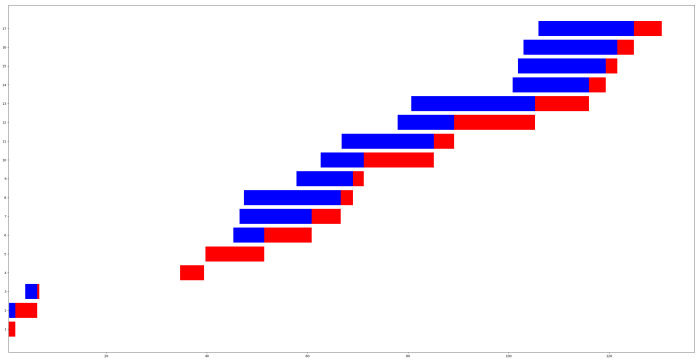

Figure 3: Schedule represents the sampling of time of birth and death.

4 Parametric learning Birth Rate λ, when $\mu = 0$

Consider birth only M/M/1 queue, i.e. $\mu = 0$, for a base case analysis.

4.1 In Time based setting

We are now considering the sample drawn in bounded time, T. [3]

Algorithm 1: parametric learning birth rate λ, when $\mu = 0$

Data: Take samples $\Upsilon(T) = \{T^B, \phi\} \in Q(\lambda, 0)$, for $T = O\left(\frac{\lambda}{\epsilon^2} \log \frac{1}{\delta}\right)$
Result: hypothesis $\hat{\lambda}$ which is ϵ-far from λ
N=number of birth, $|T^B|$
if $N \neq 0$ **then**
  return $\frac{N}{T}$
end
return ϵ

Theorem 5. *In terms of absolute error which, $\epsilon < |\lambda - \hat{\lambda}|$, Algorithm 1 parametric learn birth rate λ in $T = O\left(\frac{\lambda}{\epsilon^2} \log \frac{1}{\delta}\right)$, with high probability $1 - \delta$ for arbitrary small $\epsilon, \delta > 0$*

Proof 1. *As the number of birth $|T^B| \sim Poisson(\lambda T)$, for any birth rate $\lambda > 0$, for arbitrary small $\epsilon, \delta > 0$, by the concentration bound for Poisson Distribution, **Folklore Fact**, when $\mathbb{E}[|T^B|] = \lambda T, \hat{\lambda} = \frac{|T^B|}{T}$*

$$\mathbb{P}\left[||T^B| - \lambda T| \geq \epsilon T\right] \leq 2e^{-\frac{(\epsilon T)^2}{2(\lambda T + \epsilon T)}}$$

$$\mathbb{P}\left[|\hat{\lambda} - \lambda| \geq \epsilon\right] \leq 2e^{-\frac{\epsilon^2 T}{2(\lambda + \epsilon)}} \leq \delta$$

$$T \geq \frac{2\lambda}{\epsilon^2} \ln \frac{2}{\delta} + \frac{2}{\epsilon} \ln \frac{2}{\delta} = O(\frac{\lambda}{\epsilon^2} \ln \frac{1}{\delta})$$

if $|T^B| = 0$, $\hat{\lambda} = 0$ is an improper learning result, as λ should always > 0. Where by letting $\hat{\lambda} = \epsilon$ can fix the problem as $\mathbb{P}\left[|\hat{\lambda} - \lambda| \geq \epsilon\right] = \mathbb{P}\left[\lambda \geq 2\epsilon\right] + \mathbb{P}\left[\lambda \leq 0\right] < \mathbb{P}\left[\lambda \geq \epsilon\right]$ which can have a bound that is not any looser.

Theorem 6. *The time bound $T = O\left(\frac{\lambda}{\epsilon^2} \log \frac{1}{\delta}\right)$ is tight*

Proof 2. *According to **the tightness of Chernoff bound**, for arbitrary $\epsilon, \delta > 0$*

$$\mathbb{P}\left[||T^B| - \lambda T| \geq \epsilon T\right] \geq \frac{1}{4} e^{-\frac{2(\epsilon T)^2}{\lambda T}}$$

$$\mathbb{P}\left[|\hat{\lambda} - \lambda| \geq \epsilon\right] \geq \frac{1}{4} e^{-\frac{2\epsilon^2 T}{\lambda}} \geq \delta$$

$$T \leq \frac{\lambda}{2\epsilon^2} \ln \frac{1}{4\delta} = O(\frac{\lambda}{\epsilon^2} \ln \frac{1}{\delta})$$

[3] the idea of estimating parameter using sample mean is similar to Lee S., Radhika L. Study[19]

Which indicates when $T < O(\frac{\lambda}{\epsilon^2} \ln \frac{1}{\delta})$, probability that the error is larger than ϵ is arbitrary big, i.e. with a arbitrary small confidence level $1-\delta$.

Theorem 7. *In terms of relative error which, $\epsilon < |\log \frac{\lambda}{\hat{\lambda}}|$, Algorithm 1 parametric learn birth rate λ in $T = O\left(\frac{1}{\lambda\epsilon^2} \log \frac{1}{\delta}\right)$ if $|T^B| \neq 0$, with high probability $1-\delta$ for arbitrary small $\epsilon, \delta > 0$*

Proof 3. *As the number of birth $|T^B| \sim Poisson(\lambda T)$, for any birth rate $\lambda > 0$, for arbitrary small $\epsilon, \delta > 0$, by the* ***multiplicative Chernoff bound****, when $\mathbb{E}[|T^B|] = \lambda T, \hat{\lambda} = \frac{|T^B|}{T} \neq 0$*
Performing union bound on both tail, i.e. when

$$\mathbb{P}\left[|\log \frac{\lambda}{\hat{\lambda}}| \geq \epsilon\right] \leq \delta$$

$$\mathbb{P}\left[\log \frac{\lambda}{\hat{\lambda}} \geq \epsilon\right] + \mathbb{P}\left[\log \frac{\hat{\lambda}}{\lambda} \geq \epsilon\right] \leq \delta$$

$$\mathbb{P}\left[\log \frac{\lambda}{\hat{\lambda}} \geq \epsilon\right] \leq \frac{\delta}{2}, \mathbb{P}\left[\log \frac{\hat{\lambda}}{\lambda} \geq \epsilon\right] \leq \frac{\delta}{2}$$

$$\mathbb{P}\left[|T^B| \geq (1+\eta)\lambda T\right] \leq \left(\frac{e^\eta}{(1+\eta)^{1+\eta}}\right)^{\lambda T}$$

$$\mathbb{P}\left[\log \frac{\hat{\lambda}}{\lambda} > \epsilon\right] \leq \left(e^{e^\epsilon - 1 - e^{\epsilon e^\epsilon}}\right)^{\lambda T} \leq \frac{\delta}{2}, 1 + \eta = e^\epsilon$$

$$T \geq \frac{1}{\lambda\left(\epsilon e^\epsilon - e^\epsilon + 1\right)} \log \frac{2}{\delta} = O\left(\frac{1}{\lambda\epsilon^2} \log \frac{1}{\delta}\right)$$

$$\mathbb{P}\left[|T^B| \leq (1-\eta)\lambda T\right] \leq \left(\frac{e^{-\eta}}{(1-\eta)^{1-\eta}}\right)^{\lambda T}$$

$$\mathbb{P}\left[\log \frac{\lambda}{\hat{\lambda}} > \epsilon\right] \leq \left(e^{e^\epsilon - 1 - e^{\epsilon e^\epsilon}}\right)^{\lambda T} \leq \frac{\delta}{2}, 1 - \eta = e^\epsilon$$

$$T \geq \frac{1}{\lambda\left(\epsilon e^\epsilon - e^\epsilon + 1\right)} \log \frac{2}{\delta} = O\left(\frac{1}{\lambda\epsilon^2} \log \frac{1}{\delta}\right)$$

Remark. $\frac{1}{\lambda(\epsilon e^\epsilon - e^\epsilon + 1)} = O\left(\frac{1}{\lambda\epsilon^2}\right)$, *as when $\epsilon \to 0$, by L'Hospital's Rule[11],*

$$\begin{aligned}
\lim_{\epsilon \to 0} \frac{\frac{1}{\lambda(\epsilon e^\epsilon - e^\epsilon + 1)}}{\frac{1}{\lambda\epsilon^2}} &= \lim_{\epsilon \to 0} \frac{\epsilon^2}{\epsilon e^\epsilon - e^\epsilon + 1} \\
&= \lim_{\epsilon \to 0} \frac{2\epsilon}{\epsilon e^\epsilon} \\
&= \lim_{\epsilon \to 0} \frac{2}{(\epsilon + 1)e^\epsilon} \\
&= 2
\end{aligned}$$

Remark. ***The tightness of this bound have not been checked.***
But it should also be guaranteed by the tightness of chernoff bound.

Theorem 8. ***Arbitrary large relative error.*** *In time based setting, Algorithm 1 cannot always guarantee to have any birth measured in time bound T. If the number of birth is 0 in $[0, T]$, the relative error can always be arbitrary large. Which motivate us to use count bound in the following section.*

Proof 4. *For all hypothesis birth rate $\hat{\lambda} > 0$, there exist arbitrary large $\epsilon > 0$, such that $\lambda = \frac{1}{\epsilon}\hat{\lambda} > 0$.*

4.2 In counting based setting

We are now considering the sample drawn in bounded number of state transition (birth process for now), N.

Algorithm 2: parametric learning birth rate λ, when $\mu = 0$

Data: Take samples $\Upsilon(N) = \{T^B, \phi\} \in Q(\lambda, 0)$, for $N = O\left(\frac{\lambda^2}{\epsilon^2}\log\frac{1}{\delta}\right)$
Result: hypothesis $\hat{\lambda}$ which is ϵ-far from λ
T=sum of all inter-arrival time, $\sum_{k=1}^{b} arrival\,(k)$
if $T \neq 0$ **then**
  return $\frac{N-1}{T}$
end
return *error*

Theorem 9. *In terms of absolute error which, $\epsilon < |\lambda - \hat{\lambda}|$, Algorithm 2 parametric learn birth rate λ in $N = O\left(\frac{\lambda^2}{\epsilon^2}\log\frac{1}{\delta}\right)$, with high probability $1 - \delta$ for arbitrary small $\epsilon, \delta > 0$*

Proof 5. *As the inter-arrival time is independent of each other, guaranteed by the definition of Poisson Process, and each of it will follows an Exponential distribution with parameter λ. Each inter-arrival time $T_i \sim Exp(\lambda)$, the time of the last birth $T = \sum_{i=1}^{N} T_i$, follows a Gamma distribution with parameter (N, λ), $T \sim Gamma(N, \lambda)$, while $X = \frac{1}{T}$ follows a Inverse-Gamma distribution[6] with parameter (N, λ), $X \sim IG(N, \lambda)$. For any birth rate $\lambda > 0$, for arbitrary small $\epsilon, \delta > 0$, by the CDF of Inverse-Gamma distribution,*

$$\begin{aligned}
\mathbb{P}\left[X \leq \mathbb{E}\left[X\right] - \eta\right] &= F_{IG(N,\lambda)}(\mathbb{E}\left[X\right] - \eta), \textit{for some } \eta > 0 \\
&= Q\left(N, \frac{\lambda}{\mathbb{E}\left[X\right] - \eta}\right), Q \textit{ is the regularized Gamma function} \\
\mathbb{P}\left[\hat{\lambda} - \lambda \geq \epsilon\right] &= Q\left(N, \frac{\lambda}{\lambda - \epsilon}(N-1)\right), \eta = \frac{\epsilon}{N-1}, \mathbb{E}\left[X\right] = \frac{\lambda}{N-1} \\
&= \mathbb{P}\left[Y \leq (N-1)\right], Y \sim Poisson\left(\frac{\lambda}{\lambda - \epsilon}(N-1)\right) \\
&= \mathbb{P}\left[Y \leq \left(1 - \frac{\epsilon}{\lambda}\right)\mathbb{E}\left[Y\right]\right]
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}\left[X \geq \mathbb{E}\left[X\right] + \eta\right] &= 1 - F_{IG(N,\lambda)}(\mathbb{E}\left[X\right] + \eta), \textit{for some } \eta > 0 \\
&= P\left(N, \frac{\lambda}{\mathbb{E}\left[X\right] + \eta}\right), P \textit{ is the regularized Gamma function} \\
\mathbb{P}\left[\lambda - \hat{\lambda} \geq \epsilon\right] &= P\left(N, \frac{\lambda}{\lambda + \epsilon}(N-1)\right), \eta = \frac{\epsilon}{N-1}, \mathbb{E}\left[X\right] = \frac{\lambda}{N-1} \\
&= \mathbb{P}\left[Y \geq (N-1)\right], Y \sim Poisson\left(\frac{\lambda}{\lambda + \epsilon}(N-1)\right) \\
&= \mathbb{P}\left[Y \geq \left(1 + \frac{\epsilon}{\lambda}\right)\mathbb{E}\left[Y\right]\right]
\end{aligned}$$

Again, with union bound on both tail with multiplicative Chernoff bound as we did for the Time bound with relative error on Poisson distribution in the last section.

$$\begin{aligned}
\mathbb{P}\left[|\hat{\lambda} - \lambda| \geq \epsilon\right] &\leq \delta \\
\mathbb{P}\left[\hat{\lambda} - \lambda \geq \epsilon\right] + \mathbb{P}\left[\lambda - \hat{\lambda} \geq \epsilon\right] &\leq \delta \\
\mathbb{P}\left[\hat{\lambda} - \lambda \geq \epsilon\right] \leq \frac{\delta}{2}, \mathbb{P}\left[\lambda - \hat{\lambda} \geq \epsilon\right] &\leq \frac{\delta}{2}
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}\left[Y \leq \left(1 - \frac{\epsilon}{\lambda}\right)\mathbb{E}\left[Y\right]\right] \leq \left(\frac{e^{-\frac{\epsilon}{\lambda}}}{\left(1 - \frac{\epsilon}{\lambda}\right)^{1-\frac{\epsilon}{\lambda}}}\right)^{\mathbb{E}[Y]} &\leq \frac{\delta}{2}, Y \sim Poisson(\frac{\lambda}{\lambda - \epsilon}(N-1)) \\
\frac{\lambda}{\lambda - \epsilon}(N-1)\left(\left(1 - \frac{\epsilon}{\lambda}\right)\log\left(1 - \frac{\epsilon}{\lambda}\right) + \frac{\epsilon}{\lambda}\right) &\geq \log\frac{2}{\delta}, \mathbb{E}\left[Y\right] = \frac{\lambda}{\lambda - \epsilon}(N-1) \\
N \geq \frac{\lambda - \epsilon}{\lambda}\frac{1}{\left(1 - \frac{\epsilon}{\lambda}\right)\log\left(1 - \frac{\epsilon}{\lambda}\right) + \frac{\epsilon}{\lambda}}\log\frac{2}{\delta} + 1 \\
&= O\left(\frac{\lambda - \epsilon}{\lambda}\frac{\lambda^2}{\epsilon^2}\log\frac{1}{\delta}\right) \\
&= O\left(\frac{\lambda^2}{\epsilon^2}\log\frac{1}{\delta} - \frac{\lambda}{\epsilon}\log\frac{1}{\delta}\right) \\
&= O\left(\frac{\lambda^2}{\epsilon^2}\log\frac{1}{\delta}\right)
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}\left[Y \geq \left(1 + \frac{\epsilon}{\lambda}\right)\mathbb{E}\left[Y\right]\right] \leq \left(\frac{e^{\frac{\epsilon}{\lambda}}}{\left(1 + \frac{\epsilon}{\lambda}\right)^{1+\frac{\epsilon}{\lambda}}}\right)^{\mathbb{E}[Y]} &\leq \frac{\delta}{2}, Y \sim Poisson(\frac{\lambda}{\lambda + \epsilon}(N-1)) \\
\frac{\lambda}{\lambda + \epsilon}(N-1)\left(\left(1 + \frac{\epsilon}{\lambda}\right)\log\left(1 + \frac{\epsilon}{\lambda}\right) - \frac{\epsilon}{\lambda}\right) &\geq \log\frac{2}{\delta}, \mathbb{E}\left[Y\right] = \frac{\lambda}{\lambda + \epsilon}(N-1) \\
N \geq \frac{\lambda + \epsilon}{\lambda}\frac{1}{\left(1 + \frac{\epsilon}{\lambda}\right)\log\left(1 + \frac{\epsilon}{\lambda}\right) - \frac{\epsilon}{\lambda}}\log\frac{2}{\delta} + 1 &= O\left(\frac{\lambda^2}{\epsilon^2}\log\frac{1}{\delta}\right)
\end{aligned}$$

Remark. $\frac{1}{\left(1-\frac{\epsilon}{\lambda}\right)\log\left(1-\frac{\epsilon}{\lambda}\right)+\frac{\epsilon}{\lambda}} = O\left(\frac{\lambda^2}{\epsilon^2}\right)$, *as when* $\epsilon \to 0$, *by L'Hospital's Rule,*

$$\begin{aligned}\lim_{\epsilon\to 0}\frac{\frac{1}{\left(1-\frac{\epsilon}{\lambda}\right)\log\left(1-\frac{\epsilon}{\lambda}\right)+\frac{\epsilon}{\lambda}}}{\frac{\lambda^2}{\epsilon^2}} &= \lim_{\epsilon\to 0}\frac{\epsilon^2}{\lambda^2\left(\left(1-\frac{\epsilon}{\lambda}\right)\log\left(1-\frac{\epsilon}{\lambda}\right)+\frac{\epsilon}{\lambda}\right)}\\ &= \lim_{\epsilon\to 0}\frac{2\epsilon}{\lambda\left(-\log\left(1-\frac{\epsilon}{\lambda}\right)\right)}\\ &= \lim_{\epsilon\to 0}\frac{2}{\frac{1}{\left(1-\frac{\epsilon}{\lambda}\right)}}\\ &= 2\end{aligned}$$

And, $\frac{1}{\left(1+\frac{\epsilon}{\lambda}\right)\log\left(1+\frac{\epsilon}{\lambda}\right)-\frac{\epsilon}{\lambda}} = O\left(\frac{\lambda^2}{\epsilon^2}\right)$*, as when* $\epsilon\to 0$*, by L'Hospital's Rule,*

$$\begin{aligned}\lim_{\epsilon\to 0}\frac{\frac{1}{\left(1+\frac{\epsilon}{\lambda}\right)\log\left(1+\frac{\epsilon}{\lambda}\right)-\frac{\epsilon}{\lambda}}}{\frac{\lambda^2}{\epsilon^2}} &= \lim_{\epsilon\to 0}\frac{\epsilon^2}{\lambda^2\left(\left(1+\frac{\epsilon}{\lambda}\right)\log\left(1+\frac{\epsilon}{\lambda}\right)-\frac{\epsilon}{\lambda}\right)}\\ &= \lim_{\epsilon\to 0}\frac{2\epsilon}{\lambda\left(\log\left(1+\frac{\epsilon}{\lambda}\right)\right)}\\ &= \lim_{\epsilon\to 0}\frac{2}{\frac{1}{\left(1+\frac{\epsilon}{\lambda}\right)}}\\ &= 2\end{aligned}$$

Theorem 10. *In terms of relative error which,* $\epsilon < |\log\frac{\lambda}{\hat{\lambda}}|$*, Algorithm 2 parametric learn birth rate* λ *in* $N = O\left(\frac{1}{\epsilon^2}\log\frac{1}{\delta}\right)$*, with high probability* $1-\delta$ *for arbitrary small* $\epsilon,\delta > 0$*. Which surprisingly independent of* λ*.*

Proof 6. *As for each time interval* $T_i \sim Exp(\lambda)$*, the time of the last birth* $T = \sum_{i=1}^{N} T_i$*, is a sum of independent exponential distribution with same parameter* λ*. For any birth rate* $\lambda > 0$*, for arbitrary small* $\epsilon,\delta > 0$*, by the* ***Tight tail bounds for sum of N i.i.d. exponential random variables****,* $\mathbb{E}[T] = \frac{N}{\lambda}$*, with union bound on both tail,*

$$\begin{aligned}\mathbb{P}[T \le \eta\mu] &\le e^{-\lambda\mu(\eta-1-\ln\eta)}\\ \mathbb{P}\left[\frac{N}{\hat{\lambda}} \le e^{-\epsilon}\frac{N}{\lambda}\right] &\le e^{-N(e^{\epsilon}-1-\epsilon)} \le \frac{\delta}{2}\\ N &\ge \frac{1}{e^{\epsilon}-1-\epsilon}log\frac{2}{\delta} = O\left(\frac{1}{\epsilon^2}\log\frac{1}{\delta}\right)\\ \mathbb{P}[T \ge \eta\mu] &\le \eta^{-1}e^{-\lambda\mu(\eta-1-\ln\eta)}\\ \mathbb{P}\left[\frac{N}{\hat{\lambda}} \ge e^{\epsilon}\frac{N}{\lambda}\right] &\le e^{-\epsilon}e^{-N(e^{\epsilon}-1-\epsilon)} \le \frac{\delta}{2}\\ N &\ge \frac{1}{e^{\epsilon}-1-\epsilon}\left(log\frac{2}{\delta}-\epsilon\right) = O\left(\frac{1}{\epsilon^2}\log\frac{1}{\delta}\right)\end{aligned}$$

The bound is tight guaranteed by the tightness of the tail bound.

Remark. $\frac{1}{e^\epsilon-1-\epsilon} = O\left(\frac{1}{\epsilon^2}\right)$, *as when* $\epsilon \to 0$, *by L'Hospital's Rule,*

$$\begin{aligned}\lim_{\epsilon\to 0} \frac{\frac{1}{e^\epsilon-1-\epsilon}}{\frac{1}{\epsilon^2}} &= \lim_{\epsilon\to 0} \frac{\epsilon^2}{e^\epsilon-1-\epsilon} \\ &= \lim_{\epsilon\to 0} \frac{2\epsilon}{e^\epsilon-1} \\ &= \lim_{\epsilon\to 0} \frac{2}{e^\epsilon} \\ &= 2\end{aligned}$$

5 Parametric learning Birth Rate λ and Death Rate μ in general

In general, we just have to consider applying the same analysis above on both birth and death.

Theorem 11. ***Learning Death Rate.*** *Similar method to learn birth rate* λ *can also be used to learn the death rate,* μ.
By replacing total time T *with busy time* $R = T - I$ *for* $I = \sum_{k=0}^{d+1} idle(k)$ *and number of birth* N *with number of death* M. *Algorithm 1 can be modified to learn death rate for time bound* $R = O\left(\frac{\mu}{\epsilon^2}\log\frac{1}{\delta}\right)$, *in terms of absolute error, and* $R = O\left(\frac{1}{\mu\epsilon^2}\log\frac{1}{\delta}\right)$, *in terms of relative error, with high probability* $1-\delta$ *for arbitrary small* $\epsilon, \delta > 0$.
By replacing sum of inter-arrival time T *with sum of service time* R *and number of birth* N *with number of death* M. *Algorithm 2 can be modified to learn death rate for counting bound* $M = O\left(\frac{\mu^2}{\epsilon^2}\log\frac{1}{\delta}\right)$, *in terms of absolute error, and* $M = O\left(\frac{1}{\epsilon^2}\log\frac{1}{\delta}\right)$. *in terms of relative error, with high probability* $1-\delta$ *for arbitrary small* $\epsilon, \delta > 0$.

Proof. Let R be the busy time interval after removing all time in idle period from the total time, i.e. $R = T - I$ for $I = \sum_{k=0}^{d+1} idle(k)$. After that, it is guaranteed to have customer in the server in R, so that the server is always serving. Number of birth $|T^B| \sim Poisson(\lambda T)$ and so as the number of death $|T^D| \sim Poisson(\mu R)$. The analysis for birth can apply directly on death in the same way.
While for counting based, inverse of the sum of inter-arrival time follows Inverse Gamma distribution and so as the , inverse of the sum of service time. By replacing sum of inter-arrival time T with sum of service time R, number of birth N with number of death M can convert Algorithm 2 to learn death rate. □

5.1 In Time based setting

We are now considering the sample drawn in bounded number of state transition, N.

Algorithm 3: parametric learning birth rate $\lambda, \mu > 0$

Data: Take samples $\Upsilon(T) = \{T^B, T^D\} \in Q(\lambda, \mu)$
Result: hypothesis $\hat{Q}(\hat{\lambda}, \hat{\mu})$ which is ϵ-far from Q
N=number of birth in sample Υ
M=number of death in sample Υ
I=sum of all idle time, $\sum_{k=1}^{d+1} idle\,(k)$
R=T-I
$\hat{\lambda} = \epsilon$
$\hat{\mu} = \epsilon$
if $N \neq 0$ **then**
  $\hat{\lambda} = \frac{N}{T}$
end
if $M \neq 0$ **then**
  $\hat{\mu} = \frac{M}{R}$
end
return $\hat{Q}(\hat{\lambda}, \hat{\mu})$

5.2 In Counting based setting

We are now considering the sample drawn in bounded number of state transition, N.

Algorithm 4: parametric learning birth rate $\lambda, \mu > 0$

Data: Take samples $\Upsilon(S) = \{T^B, T^D\} \in Q(\lambda, 0)$
Result: hypothesis $\hat{Q}(\hat{\lambda}, \hat{\mu})$ which is ϵ-far from Q
N=number of birth in sample Υ
M=$S - N$
B=sum of all inter-arrival time, $\sum_{k=1}^{b} arrival\,(k)$
D=sum of all service time, $\sum_{k=1}^{d} service\,(k)$
$\hat{\lambda} = \frac{N-1}{B}$
$\hat{\mu} = \frac{M-1}{D}$
return $\hat{Q}(\hat{\lambda}, \hat{\mu})$

6 Conclusion

M/M/1 queue is the base case in queuing theory. Learning M/M/1 queue in our project follows PAC-framework from distribution learning theory which differs from many other research papers. Some of them require drawing 1 customer from each n i.i.d M/M/1 Queues which need to assume $\lambda < \mu$.[4] Some may perform MLE learning and confidence interval of M/M/R Queue on some other parameter with similar algorithm.[21]

We define the distance between two M/M/1 queues under two types of distance: Absolute error takes the maximum of the absolute difference of λ and that of μ, which is obviously good for learning the individual value of rate parameters. While, relative error takes the maximum of log-ratio of λ and that of μ, which can serve better in another purpose.

For example, when we are learning the server intensity $\rho = \frac{\lambda}{\mu}$, it is far easier to compare hypothesis $\hat{\rho}$ with concept ρ in terms of log-ratio than the absolute error. Which, when we want to have $|\log\left(\frac{\hat{\rho}}{\rho}\right)| < \epsilon$, we can just perform the same learning algorithm to learn λ and μ with error $< \frac{1}{2}\epsilon$ in terms of relative error, for $\hat{\rho} = \frac{\hat{\lambda}}{\hat{\mu}}$. The algorithm provides the same guaranteed performance as learning rate parameters.

Sample size is bounded by $O\left(\frac{1}{\epsilon^2}\log\frac{1}{\delta}\right)$, for arbitrary small ϵ and δ, with a constant factor in terms of polynomial of arrival rate λ and service rate μ. We think that the counting based bound and time based bound should only differ by a constant factor of the arrival rate λ in the study on the birth only process, as $N = rate \times T$, which should also be true for service rate μ. It means that the counting based bound and time based bound are fundamentally the same.

The arrivals and service time in M/M/1 Queue follow Poisson and exponential distribution. Time-based approach and Counting based approach are our two main ideas to PAC-learn the M/M/1 queue. In the process of PAC-learning, we handle the learning of arrival rate under birth samples, then we handle the learning of service rate under death samples. The reason is that death samples partly depends on the birth rate. Therefore, sample pre-processing allows us to get an accurate value of μ with death samples.

Here we have a simple simulation[4] of our work using Wolfram Language in the link below.
`https://www.wolframcloud.com/obj/kitcheung1480/Published/PACLearnMM1Queue.nb`

[4]The idea of the simulator is based on [Heikki Ruskeep]: `https://demonstrations.wolfram.com/SimulatingTheMM1Queue/`

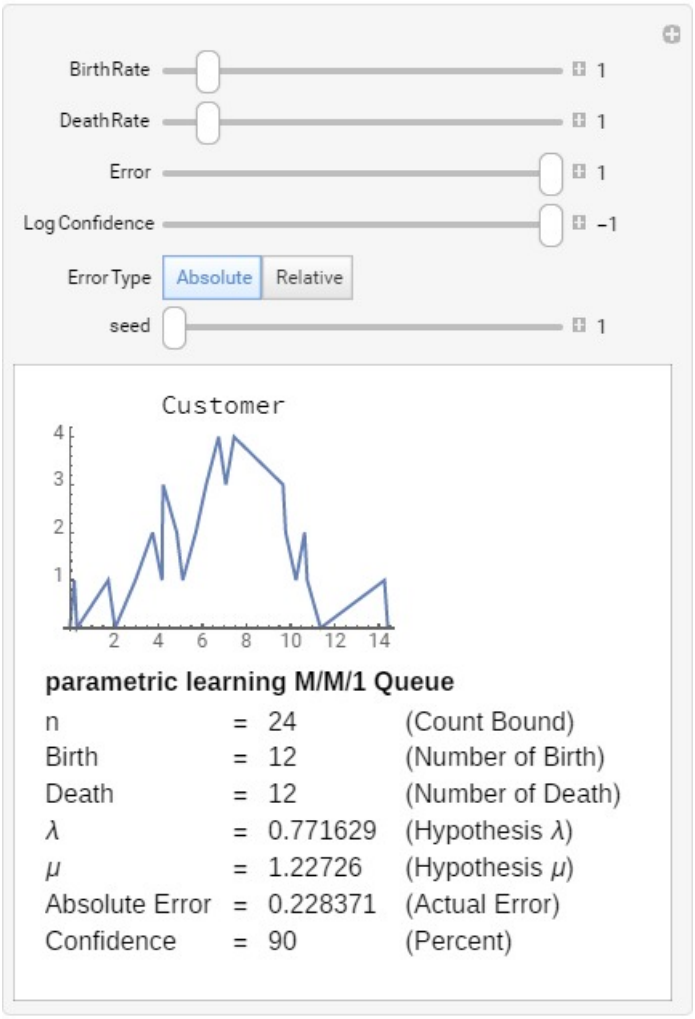

Figure 4: User-Interface showing the relationship between customers and time units

7 Further Work

For further study, we can apply the study on other types of queues such as M/G/1 Queue, G/M/1 Queue, and M/M/c Queue which are more complex than the M/M/1 queue model. For example, M/M/c queue can be further research as there is more than 1 server handing a queue in the real situation. The main idea of estimating the parameter is to calculate the sample mean is ϵ-different to parameters such that the size of the sample is large enough. Therefore, we aim to design the algorithm approach estimator using a small sample size. There are many papers out there that have already mentioned MLE learning and confidence interval of M/M/R Queue, which we may have further study on.

References

[1] Markov chain representation of m/m/1 queue, Nov 2019.

[2] Dimitri P. Bertsekas and John N. Tsitsiklis. *Introduction to probability*. Athena Scientific, 2008.

[3] Clément Canonne. A short note on poisson tail bounds. 2016.

[4] Shovan Chowdhury and S. P. Mukherjee. Estimation of waiting time distribution in an m/m/1 queue. *Opsearch*, 48(4):306–317, 2011.

[5] Hannah Constantin. Markov chains and queuing theory. *Simulating Queuing Systems: A Test of Parameter Change*, 2011.

[6] John D. Cook. Inverse gamma distribution. Oct 2008.

[7] Ilias Diakonikolas. Learning structured distributions - ilias diakonikolas, 2016.

[8] Oded Goldreich. Lecture notes for testing properties of distributions. 2016.

[9] Svante Janson. Tail bounds for sums of geometric and exponential variables. *Statistics & Probability Letters*, 135:1–6, 2018.

[10] Leonard Kleinrock. *Queuing systems*. John Wiley & Sons, 1975.

[11] L'Hospital, Paulian Aime-Henri, John Adams, and John Quincy Adams. *Analyse des infiniment petits*. Chez Didot, le jeune ..., 1768.

[12] Michael Mitzenmacher and Eli Upfal. *Probability and computing: randomized algorithms and probabilistic analysis*. Cambridge university press, 2017.

[13] Chelsea Moore. An introduction to logistic and probit regression models. 2013.

[14] Nima Mousavi. How tight is chernoff bound? *Chernoff-Tightness*, page 2–3, 2012.

[15] Simon Parsons. Introduction to machine learning by ethem alpaydin, mit press, 0-262-01211-1, 400 pp. *The Knowledge Engineering Review*, 20(4):432–433, 2005.

[16] R. C. Quinino and F. R.B. Cruz. Bayesian sample sizes in an m/m/1 queueing systems. *Int J Adv Manuf Technol*, 88(995–1002), 2017.

[17] Philippe Rigollet. 18.s997 high-dimensional statistics: Complete lecture notes, 2015.

[18] Sheldon M. Ross. *Introduction to probability models*. Harcourt/Academic Press, 2000.

[19] Lee Schruben and Radhika Kulkarni. Some consequences of estimating parameters for the m/m/1 queue. *Operations Research Letters*, 1(2):75–78, 1982.

[20] Marco Taboga. Exponential distribution - maximum likelihood estimation.

[21] Kuo-Hsiung Wang, Sheau-Chyi Chen, and Jau-Chuan Ke. Maximum likelihood estimates and confidence intervals of an m/m/r/n queue with balking and heterogeneous servers. *RAIRO - Operations Research*, 38(3):227–241, 2004.

[22] Geoffrey Wolfer and Aryeh Kontorovich. Minimax learning of ergodic markov chains, Mar 2019.

[23] Geoffrey Wolfer and Aryeh Kontorovich. Minimax testing of identity to a reference ergodic markov chain, Sep 2019.

[24] Yining Wang Yuan Zhou Yaonan Jin, Yingkai Li. On asymptotically tight tail bounds for sums of geometric and exponential random variables. 2019.