CARLETON UNIVERSITY

SCHOOL OF
MATHEMATICS AND STATISTICS

HONOURS PROJECT

TITLE: Estimating the sample size of Geo/Geo/1

AUTHOR: Yue Xu

SUPERVISOR: Dr. Yiqiang Zhao

DATE: April 7th, 2020

Abstract

The statistical inference of queuing system is very important, because we can predict or calculate different queuing models. The determination of sample size is also a very important subject in statistics. However, for inference on queues, studies on sample size are very limited. In this project,we consider two simpler queuing models, M/M/1 and Geo/Geo/1, with infinite capacity. Some papers have given an approach, Bayesian, to estimate the sample size of M/M/1. Hence, motivated by the study of sample size on the M/M/1 queueing system, we apply the same method to a discrete-time queueing system, Geo/Geo/1, for estimating the sample size. Thus, this project will summary the approach and verify that the same approach can find out the sample size of Geo/Geo/1.

1 Background

It is widely known that Statistics is related to data, for example, according to Romeijn (2014), Statistics is a mathematical and conceptual discipline dedicated to the relationship between data and assumptions. It is not hard to see that Statistics is useful and significant as data exists everywhere. But, in the beginning, Desrosières (2013) points out those Statistical institutions in the 19th and early 20th centuries were usually small, and the main job was the population census (p.11). Nowadays, Statistics are not limited to population censuses, but can also include data collection on different subjects such as life satisfaction. In addition, Statistics includes not only data collection but also many other operations on data, such as data visualization and data analysis. Hence, in other fields, like medical, mathematical and economic, Statistics can also play an important role in improvement, recommendations, or otherwise. To be specific, people can get facts from the data, and then analyze, calculate and carry out other operations to determine whether they have reached the expected state, or predict the probability of future events. Secondly, statistics is just the general name of discipline. Then, it has many more specific research directions, like stochastic process, data mining, time series, and so on. And different directions also have more specific branches. For instance, take the survey sampling related courses I took earlier as an example, there are different sample selection methods such as simple random sampling, stratified sampling, cluster sampling, and among others, and different situations will have different related better selection methods. This project will be in the area of the stochastic processes.

According to Fajardo (1985), the stochastic process is one of the main topics of probability theory research (p.174). We know that we can use the given data to calculate probabilities, so the stochastic processes are important in Statistics.

I choose this research area based on what I have learned. Because I learned the introduction to the stochastic process before, I know a little knowledge from the stochastic process. Hence, compare to the other research areas like data mining, I already have a little basis. And so, I finally choose the stochastic process as my paper's topic to get a little faster start. Next, I hope I can learn more from the stochastic process because a certain amount of research is required in the process of writing a paper. And based on my own situation, I am good at analysis and quite weak at computational. So, following the suggestion from the supervisor, queues became my specific topic for the paper.

Furthermore, it is widely known that sample size plays an important role in a statistical test. If the sample size in a study is too small, then the study finally may produce uncertain results. And if the sample size is too large, the study will waste resources. Therefore, the determination of appropriate sample size is also very important in the queue. On the other hand, only limited studies for determining the sample size of queuing systems are available (such as Quinino, R.C., & Cruz, F. R. B. (2016), Dai, Tianyi. (2018), Choudhury, Amit.,& Basak, Arpita. (2017)...), Among them, the larger part is about M/M/1 and the parameters play an extremely important role. As a result, M/M/1 is available and some variants of M/M/1 may also available. Then, the rest of this article is organized as follows. Section 2 will explain the queuing theory and queuing system. And section 3 will introduce some preliminaries of a selected statistical method first and then show the process of that statistical method for estimating the sample size of the model M/M/1 and discuss the model Geo/Geo/1. Finally, section 4 is a simple analysis and a conclusion about whether the statistical method used in M/M/1 can also estimate the sample size of the model Geo/Geo/1.

2 Background on queue

The queues are important as they are very familiar in our daily life, we can see queues that are in hospitals, banks, post office, and so on. Then it is necessary to know what are queuing theory and queuing systems first. Lim (2018) gives an explanation such that when the number of service objects (customers) is larger than the number of service facilities (servers), which provide service, and then queues will be formed. And queuing theory is the mathematical study of queuing and has customers and servers 2 sides (paras.1). In addition, customers and servers not only have the superficial meaning of words but also can become pronouns. By giving an example, when a machine is broken, then the machine is a “customer” and maintenance is a "server".

Armero and Bayarri (2015) give the summary, for example, that queuing systems are used to explain congestion and they are also simplified mathematical models. Under normal circumstances, queuing systems happen when customers require servers to provide services (p.784). In other words, when customers who find that all servers are busy may join the queues in front of servers, then it is a name queuing system. Also, Armero and Bayarri (2015) point out that queuing systems have three components: the arrival pattern, the service mechanism and the queue discipline (p.786). And according to Almeida and Cruz (2018), it mentions some queuing system features such as the **arrival rate** (λ), the **service rate** (μ) and the **traffic intensity** (ρ, the ratio between λ and μ). And these features are important to calculate the average number of customers in the system (L) and in the queue (L_q) (p.2578). In addition, an equilibrium which represents that traffic intensity is less than 1 is a general assumption in queuing theory. Hence, the steps in section 3 for the models will around the parameter ρ.

Consequently, queuing theory and queuing systems have very close ties. Sometimes, people will use both two to solve problems. According to Azam et al. (2017), it lists 8 queues applications such as applying queues to Transportation, Telephone Call Centres, and Health Care among others (p.7). More specifically, using health care as an example, the hospital is busy every day. It includes different types of outpatient, emergency rooms, ambulances and pharmacies, and all of them may involve queuing as the number of doctors, nurses and other staff is limited. C and Appa Iyer (2013) mention that, as queuing models demand fewer data, queuing theory becomes a powerful tool and it helps to improve server utilization and reduce patient waiting times (p.26). The queuing model can be viewed as a queuing system because sometimes both two can be exchanged. Hence, queuing theory and queuing systems help to reduce congestion issues and improve efficiency.

In addition to applications, Azam et al. (2017) also list models. From the listed models, we can see different types of models like M/M/1, M/D/1, M/G/1 and so on (p.2-5). We need to know what the symbols mean, so we need to know a little about Kendall first. C and Appa Iyer (2013) mentions that Kendall which consists of symbols separated by slashes introduces a useful stenography notation for simple queuing systems. And it is written in the form of A/B/C, which A represents the interarrival times between arrivals (customers), B describes the time it takes for a customer to be serviced and C refers to the number of servers (p.786).

2.1 The M/M/1 queue

M represents the exponential (Markovian), and one represents one server. In addition, we also introduce the waiting line (queueing system) capacity, queueing discipline, and independent assumptions to the M/M/1 model. The M/M/1 queue is a short hand notation for the M/M/1/∞/∞/FIFO, where the first "∞" represents an infinite queue capacity and the second "∞" represents an infinite population size, and FIFO represents first in first out service. Among them, the waiting line capacity in M/M/1 model refers to the first ∞ (“Introduction to queueing theory," n.d., para.2). Using another example of M/M/3/20/∞/FIFO, we have waiting line capacity of 20 with 3 in service and 17 in waiting. Moreover, queueing discipline refers to the rule when the server completes the service of the current customer and the next customer from the queue (if any) is selected by the server. And in M/M/1, queueing discipline refers to FIFO. And independent assumptions mean that all customers are assumed to be independent, customers arrive independently and service times for customers are independent.

According to Almeida and Cruz (2018), it explains that the M/M/1 model assumes that the number of arrivals follows a Poisson process with the rate of arrivals λ, and an exponential distribution of service time with the service rate μ (p.2578).
(1) So, interarrival times follow an exponential distribution, then:

$$P(x) = \lambda e^{-\lambda x}, x > 0.$$

(2) And the probability density of service times is

$$P(x) = \mu e^{-\mu x}, x > 0.$$

This is the simplest situation.

Next, Almeida and Cruz (2018) mention that the queuing system eventually will reach a steady-state (equilibrium) after a long period, which the traffic intensity $\rho < 1$ (p.2579). Then, the geometric probability distribution of the number of customers (N) at departure time in the system follows:

$$P(N=n)=\begin{cases}(1-\rho)\rho^n & \text{for} \quad n=0,1,2...\\ 0 & \text{for} \quad otherwise\end{cases} \tag{1}$$

So, from Eq. (1), $P(N=0)=1-\rho$, $L=\frac{\rho}{1-\rho}$, $L_q=\frac{\rho^2}{1-\rho}$.

The above probability distribution shows that the probability of a departing customer leaving behind an empty system is 1-ρ. And the probability of leaving behind a non-empty system is ρ. Thus, each departing customer can be seen as a Bernoulli variable.

Besides, M/M/1 is the simplest and one of the most suitable queuing models in many practical applications and it is continuous.

2.2 The Geo/Geo/1 queue

Geo represents geometric, and then it is a model that refers to the system with one server has different geometric distributions of both interarrival times and service times. And it is in discrete time.

According to Li and Tian (2016), it analyzes the discrete-time Geo/Geo/1 queue under *one working vacation* that when the service is completed, once there are no customers, the system will randomly start working vacation. Then a customer will be serviced at a lower rate if the customer arrivals at vacation period. Until vacation ends, the service rate changes to normal (p.78).

Further, Li and Tian (2016) give different geometric distributions for different specific situations. We can use this as a reference to infer the distribution of the standard Geo/Geo/1 model, which has no working vacation. Suppose that a potential arrival occurs in the interval (n^-, n), and a potential departure occurs in the interval (n, n^+):

(1) A customer arrives at the end of the slot $t = n^-$, n= 0,1, 2…, where n^- represents the moment before n. So, if we label the slot t=1, t=2…the customer will arrive before t=1, t=2…And then interarrival times are independent and follow a geometric distribution with parameter λ:

$$P(T = k) = \lambda\overline{\lambda}^{k-1}, k \geq 1, 0 < \lambda < 1.$$

where $\overline{\lambda} = 1 - \lambda$, and each arrival customer can be seen as a geometric variable.

(2) A service from the beginning to ending occurs at the slot t=n, which means that observing a server begins at t=1, t=2...The service times, S_b, with the parameter μ_b in a regular busy period are independent and identically distributed random variables following geometric distribution:

$$P(S_b = k) = \mu_b\overline{\mu_b}^{k-1}.k \geq 1, 0 < \mu_b < 1.$$

where $\overline{\mu_b} = 1 - \mu_b$. and each server can be seen as Geometric variable.

The distribution of service time S_v with parameter μ_v in a working vacation is:

$$P(S_v = k) = \mu_v\overline{\mu_v}^{k-1}.k \geq 1, 0 < \mu_v < 1.$$

where $\overline{\mu_v} = 1 - \mu_v$.

(3) A server begins a working vacation and the queue becomes empty, then the distribution for vacation period is:

$$P(V = k) = \theta\overline{\theta}^{k-1}, k \geq 1, 0 < \theta < 1.$$

where $\overline{\theta} = 1 - \theta$ (Li andTian, 2016, pp.78-79).

Hence, suppose there is no working vacation now. Time (t) is slotted, and suppose customers will arrive at the beginning of time slots and depart at the end of time slots. For instance, a customer arrives at the beginning of t=1, and leaves at the end of t=1. At the same time, there is at most one arrival or one departure per time slot ("H 10," n.d., p.3), then:
(1) Arrivals are followed Bernoulli distribution. And interarrival times (T) are independent and following a geometric distribution with parameter λ:

$$P(T = k) = \lambda\overline{\lambda}^{k-1}, k \geq 1, 0 < \lambda < 1.$$

where $\overline{\lambda} = 1 - \lambda$, and each arrival customer can be seen as a geometric variable.

(2) Service times (S) are also independent and following a geometric distribution with parameter μ:

$$P(S = k) = \mu\overline{\mu}^{k-1}, k \geq 1, 0 < \mu < 1.$$

where $\overline{\mu} = 1 - \mu$.

3 Bayesian Method

In addition to the models and applications mentioned above, Azam et al. (2017) also list 9 Statistical Paradigms such as Bayesian, Maximum Entropy, Non-parametric and so on (p.6). Bayesian inference is a well-known method of statistical inference and it also can be applied to many models in different areas. As some studies have used the Bayesian method to find the sample size of M/M/1, this paper will use Bayesian to estimate the sample size of different models.

Next, we need to know how Bayesian works to find out the sample size. For example, according to Sadia and Hossain (2014), it proposes an unknown parameter vector (θ) that is derived to be estimated. Then, assuming that the sample size n needs to be determined, a random sample X= (X_1, X_2, ...,X_n) will be used to estimate θ. If $f(\theta)$ is the **prior distribution** for the parameter and the likelihood function given the data x = (x_1, x_2, $\ldots$, x_n) is $L(\theta; x)$, and $L(\theta; x) \propto f(x|\theta)$. The preposterior marginal distribution of x is thus given by:

$$f(x) = \int_{\Theta} f(x|\theta) f(\theta)\, d\theta \tag{2}$$

Now, the **posterior distribution** of θ given data x with sample size n is:

$$\begin{aligned} f(\theta|x,n) &= \frac{f(x|\theta)f(\theta)}{\int_{\Theta} f(x|\theta)f(\theta)\, d\theta} \\ &= \frac{f(x|\theta)f(\theta)}{f(x)} \propto L(\theta;x)f(\theta) \end{aligned} \tag{3}$$

Additionally, if the prior distribution f(θ) and the posterior distribution f(θ|x) are in the same distribution family, the prior is the conjugate prior for the likelihood function.

Further, using **Highest Posterior Density** (HPD) Interval approach with given fixed internal, by finding n which gives the maximum coverage of Eq.(3), we can get the sample size n which is most suitable to estimate θ. And the **average coverage criterion** (ACC) could find the smallest n that meets the following condition:

$$\int_{\chi} \left(\int_{a(x,n)}^{a(x,n)+l} f(\theta|x,n)\, d\theta \right) f(x)\, dx \geq 1 - \alpha$$

where f(x) and f(θ|x, n) are given in Eq.(2) and Eq.(3). And *l* is the HPD credible set of length, a(x,n) is the lower limit of *l* for posterior density $f(\theta|x,n)$. (pp.421-422). Hence, **ACC** will find the minimum sample size n so that the aveage coverage probability of the given fixed HPD interval length *l* is at least $(1-\alpha)$

Moreover, as a region to which the parameter of interest belongs with high probability is meaningful, we use **average length criteria** (ALC) first determine the length of the credible interval for the fixed coverage $1-\alpha$ as follow:

$$\int_{a(x,n)}^{a(x,n)+l'(x,n)} f(\theta|x,n)\ d\theta = 1-\alpha$$

where $l'(x,n)$ is the length of the $(1-\alpha)100\%$ posterior credible interval for *x*.

Second, we can find the smallest n such that:

$$\int_{\chi} l'(x,n)f(x)\ dx \leq l,$$

where χ is the data space for *x*, and *l* is the desired pre-specified average length. This **ALC** is used to find sample size n which would determine the coverage probability $(1-\alpha)$ of θ's HPD credible set (Sadia and Hossainp, 2014, p.422).

If we prefer a conservative sample size intead of averaging over χ to guarantee the desired coverage and interval length, we could use worst outcome criterion (**WOC**) to get the minimum sample size n such that:

$$\inf_{x\in\chi}\left[\int_{a(x,n)}^{a(x,n)+l(x,n)} f(\theta|x,n)\ d\theta\right] \geq 1-\alpha$$

where both α and *l* are fixed (Sadia and Hossainp, 2014, p.422).

Exponential family

We also need to know a little about the exponential family, because we may use it to deduce the conjugate prior more easily. The exponential family is defined as a family of the probability density functions or probability mass functions which can be expressed as:

$$f_\theta(x) = h(x)c(\theta)\exp(\sum_{j=1}^{k} w_j(\theta)t_j(x)) \tag{4}$$

in which h(x)≥0, c(θ)≥0 and $t_1(x)\ldots t_k(x)$ are functions of x. f(x|θ) does not depend on θ.

Example 1: $x_1\ldots x_n \sim$ N(μ, σ^2), in which μ is unknown and σ^2 is known. The density of *X* is:

$$f_\mu(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}$$

Then it can be rewritten as:

$$\begin{aligned} f_\mu(x) &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{x^2}{2\sigma^2}+\frac{\mu x}{\sigma^2}-\frac{\mu^2}{2\sigma^2}\} \\ &= \frac{\exp\{-\frac{\mu^2}{2\sigma^2}\}}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{x^2}{2\sigma^2}\}\exp\{\frac{\mu x}{\sigma^2}\} \end{aligned}$$

where

$$c(\mu) = \frac{\exp\{-\frac{\mu^2}{2\sigma^2}\}}{\sqrt{2\pi\sigma^2}}, h(x) = \exp\{-\frac{x^2}{2\sigma^2}\}, w(\mu) = \frac{\mu}{\sigma^2}, t(x) = x.$$

Example 2: $x_1\ldots x_n \sim$ N(μ, σ^2), in which μ is known and σ^2 is unknown. The density of *X* is:

$$f_{\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}$$

Put θ = σ^2, then $f_{\sigma^2}(x)$ can be replaced as:

$$f_\theta(x) = \frac{1}{\sqrt{2\pi\theta}}\exp\{-\frac{(x-\mu)^2}{2\theta}\}$$

where

$$c(\theta) = \frac{1}{\sqrt{2\pi\theta}}, h(x) = 1, w(\theta) = -\frac{1}{2\theta}, t(x) = (x-\mu)^2.$$

By the way, both exponential and geometric distribution are belonged to exponential family distribution.

3.1 Sample size determination for the M/M/1 queue

Now, considering the model M/M/1, Quinino and Cruz (2016) will use $\Pi(.)$ instead of f(.). It mentions that it is necessary to observe the system at departure time. And as mentioned above, the number of customers in the system at the departure epoch is given by Eq.(1). Then, we randomly select a sample of the number of customers n left after a departing customer. Therefore, Quinino and Cruz (2016) assume that the number of customers left is given by x_i and that $x = (x_1, x_2, \ldots, x_n)$ composes our sample size n (p.997). Next, we can estimate the traffic intensity ρ and get likelihood function from Eq.(1) with parameter ρ as follow:

$$L(x|\rho) = \rho^y (1-\rho)^n \tag{5}$$

where y=$\sum_{i=1}^{n} x_i$ and n is the sample size. Also, Almeida and Cruz(2018) explain that considering the ergodic property of Markov chain, the above data generation process can ensure the independence of sample observations if there is enough space between them (p.2579).

As Eq.(1) belongs to the exponential family, we can infer the prior for ρ directly:

$$\Pi(\rho) \propto \rho^{a-1}(1-\rho)^{b-1}, 0 < \rho < 1, a > 0, b > 0 \tag{6}$$

where $\Pi(\rho)$ gives Beta distribution with parameter a and b.

Immediately after, the posterior distribution corresponding to the prior from Eq.(6)

would be given by:

$$\Pi(\rho|data) = \begin{cases} \frac{1}{Beta(a+y,n+b)}\rho^{y+a-1}(1-\rho)^{n+b-1} & 0<\rho<1 \\ 0 & otherwise \end{cases} \tag{7}$$

in which Beta(a+y, n+b) follows a beta function with parameters y+a and n + b.

Besides,

$$\Pi(\rho|data) = \frac{1}{Beta(a+y,n+b)} \exp\{(a+y-1)\log(\rho)(n+b-1)\log(1-\rho)\}$$

where

$$c(data) = \frac{1}{Beta(a+y,n+b)}, h(\rho) = 1.$$

$w_1(data) = a+y-1, w_2(data) = n+b-1, t_1(\rho) = \log(\rho), t_2(\rho) = \log(1-\rho).$

So, we can say that $\Pi(\rho|data)$ also belongs to the exponential family.

And then, the Bayes point estimator, which is also the mean of the posterior beta distribution, for $\hat{\rho}$ would be:

$$\hat{\rho} = \frac{y+a}{y+a+n+b}$$

we can also deduce the mean system size $\hat{L}$ and mean queue size $\hat{L_q}$, but we ignore here.

Secondly, as mentioned earlier that we state n customers stayed after one customer left, and then Quinino and Cruz (2016) think that it is also possible to estimate the credible region (p.997). In general, the closer the estimated value to the actual value, the more accurate the probability, but the greater the sample size requirement. Also, Quinino and Cruz (2016) mention that people are looking for the minimal n such that the coverage probability of the credible region of width

(w) is at least 1-α (p.997). For example, letting d = w/2, then it might find the minimal n such that:

$$\int_{\hat{\rho}-d}^{\hat{\rho}+d} \Pi(\rho|data)d\rho \geq 1-\alpha \tag{8}$$

where $\hat{\rho}$ is the posterior mean and the credible region is [$\hat{\rho}$-d, $\hat{\rho}$+d]. And the esimator $\hat{\rho}$ and the coverage probability depend on x=(x_1, x_2..., x_n).

Next, Quinino and Cruz (2016) mention that we can choose n such that the expected posterior coverage probability, m(x), is at least 1-α given by:

$$m(x) = \int_0^1 L(x|\rho)\Pi(\rho)d\rho \tag{9}$$

in which the expected value exceeds the marginal distribution of *x* induced by the prior distribution (p.997).

Therefore, it is recommended to seek a minimum n by using average coverage criterion (ACC) and satisfies:

$$\sum_{\forall x \in \chi} \left[\int_{\hat{\rho}-d}^{\hat{\rho}+d} \Pi(\rho|data)d\rho\right] m(x) \geq 1-\alpha \tag{10}$$

where m(x) gives weights (p.998).

Besides, if we need a conservative sample size and to satisfy the required coverage and interval length on any possible sample x, we could use **WOC** rather than averaging to find the minimum n as follow:

$$\inf_{x \in \chi} \left[\int_{\hat{\rho}-d}^{\hat{\rho}+d} \Pi(\rho|data)d\rho\right] \geq 1-\alpha$$

3.2 Sample size determination for the Geo/Geo/1 queue

By using a similar process to Geo/Geo/1 model. Firstly, we may try to get a probability distribution. Based on the information we described above:

(1) Interarrival times (T) are independent and following a geometric distribution with parameter λ:

$$P(T = k) = \lambda\overline{\lambda}^{k-1}, k \geq 1, 0 < \lambda < 1.$$

where $\overline{\lambda} = 1 - \lambda$, and each arrival customer can be seen as geometric variable.

(2) Service times (S) are also independent and following a geometric distribution with parameter μ:

$$P(S = k) = \mu\overline{\mu}^{k-1}, k \geq 1, 0 < \mu < 1.$$

where $\overline{\mu} = 1 - \mu$.

Also, we assume arrival and service processes are mutually independent and services are provided on a first-in-first-out rule. Now, let Q be the steady-state number of customers in the queue. And let the size of the queue at time t be Q(t) with the state space {0, 1, 2...}, we may see that Q(t) is a discrete-time Markov chain. Also, the transition diagram can be seen as a birth-death process, which is similar to the model M/M/1. Next, let b=birth probability and d=death probability, hence, we have a transition matrix as follow:

$$P = \begin{bmatrix} P_{0,0} = 1-\lambda & P_{0,1} = \lambda & P_{0,2} = 0 & P_{0,3} = 0 & \cdots & \\ P_{1,0} = d & P_{1,1} = r & P_{1,2} = b & P_{1,3} = 0 & \cdots & \\ P_{2,0} = 0 & P_{2,1} = d & P_{2,2} = r & P_{2,3} = b & P_{2,4} = 0 & \cdots \\ \vdots & \ddots & & & & \vdots \\ P_{i-2,0} = 0 & \cdots & P_{i-2,j-3} = d & P_{i-2,j-2} = r & P_{i-2,j-1} = b & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

where r=1-b-d, b=λ (1-μ) and d=μ (1-λ) ("H 10," n.d., p.4).

Thereafter, we try to get the equilibrium distribution by solving π= πP: $\pi_0 = (1-\lambda)\pi_0 + d\pi \Rightarrow \pi_0 = \frac{d}{\lambda}\pi_1 = \frac{\mu(1-\lambda)}{\lambda}\pi_1$

$\pi_1 = \lambda\pi_0 + r\pi_1 + d\pi_2 \Leftrightarrow b\pi_1 = d\pi_2$

$\vdots$

$\pi_i = b\pi_{i-1} + r\pi_r + d\pi_{i+1}, i \geq 2 \Leftrightarrow b\pi_i = d\pi_{i+1}$

Finally, we can get a probability distribution as follow:

$$\pi_i = (\frac{b}{d})^{i-1}\pi_1 \tag{11}$$

And since $\sum_{i\geq 0}\pi_i = 1$, we get

$$\frac{\mu(1-\lambda)}{\lambda}\pi_1 + \pi_1 + \frac{b}{d}\pi_1 + (\frac{b}{d})^2\pi_1 + \cdots + (\frac{b}{d})^{i-1}\pi_1 = 1$$

Let ρ=b/d, we will get

$$\frac{\mu(1-\lambda)}{\lambda}\pi_1 + \pi_1 + \rho\pi_1 + \rho^2\pi_1 + \cdots + \rho^{i-1}\pi_1 = 1$$

At last, we could get $\pi_1 = \frac{\lambda(\mu-\lambda)}{\mu^2(1-\lambda)} = \frac{\lambda(1-\rho)}{\mu}$

Here, λ and μ can be estimated, respectively, and the posterior for ρ can be conditioned on λ and μ. And next, we will focus on ρ.

So, continuous to Eq.(11) and assume ρ (=$\frac{b}{d}$) < 1 for stability which is equivalent to λ < μ, then we can get a probability distribution of the number of customers (Q):

$$\begin{aligned} P(Q=i) &= \pi_1\rho^{i-1}, i \geq 1. \\ &= \frac{\lambda}{\mu}(\frac{1-\rho}{\rho})\rho^i \qquad (12) \\ &= \frac{\lambda}{\mu}(1-\rho)\rho^{i-1} \qquad (13) \end{aligned}$$

So, P(Q=0) = $\pi_1 \frac{\mu(1-\lambda)}{\lambda}$

Followed by Eq.(13), let assume the number of customer in queue at time t is defined x_i, and that x = (x_1, x_2 ... x_n) composes the sample size n. So, the likelihood function for parameter ρ will be:

$$\begin{aligned} L(\rho; x) &= \prod_{x=1}^{n} \pi_1 \rho^{x-1} \\ &= \prod_{x=1}^{n} \frac{\lambda(1-\rho)}{\mu} \rho^{x-1} \\ &= \frac{\lambda^n (1-\rho)^n}{\mu^n} \rho^{\sum_{x=1}^{n} x_i - n} \\ &= \frac{\lambda^n (1-\rho)^n}{\mu^n} \rho^{y-n} \end{aligned} \tag{14}$$

where $y = \sum_{i=1}^{n} x_i$.

Continuous to Eq. (12), we try to deduce a prior distribution for ρ. As the Eq. (12) can be rewritten that follows the form of Eq.(4) as:

$$P(x|\rho) = \frac{\lambda}{\mu}(\frac{1-\rho}{\rho}) \exp\left[xIn(\rho)\right] \tag{15}$$

where h(x) = 1, c(ρ) = $\frac{\lambda}{\mu}(\frac{1-\rho}{\rho})$, and w($\rho$) = In($\rho$) and t(x) = x.

Hence, Eq.(12) also belongs to the exponential family, and then we could try a prior distribution, which is similar to the model M/M/1, for ρ as follow:

$$\Pi(\rho) \propto \rho^{a-1}(1-\rho)^{b-1}, 0 < \rho < 1, a > 0, b > 0 \tag{16}$$

where $\Pi(\rho)$ gives Beta distribution with parameter a and b.

Next, the posterior distribution corresponding to Eq.(15) can be obtained by multiplying L(ρ; x) and $\Pi(\rho)$:

$$\begin{aligned}\Pi(\rho|data) &\propto \frac{\lambda^n(1-\rho)^n}{\mu^n}\rho^{y-n}\rho^{a-1}(1-\rho)^{b-1} \\ &= (\frac{\lambda}{\mu})^n\rho^{y+a-n-1}(1-\rho)^{n+b-1} \\ &\propto \rho^{y+a-n-1}(1-\rho)^{n+b-1} \qquad (17)\end{aligned}$$

where y=$\sum_{i=1}^{n} x_i$. And it is the posterior distribution of Beta (y+a-n, b+n).

Hence, the Bayes point estimator, which is also the mean of the posterior beta distribution, for $\hat{\rho}$ would be:

$$\hat{\rho} = \frac{y+a-n}{y+a+b}$$

And then, the next steps are almost the same as the model M/M/1 from Eq.(8) to Eq.(10). We look for the minimal n such that the coverage probability of the credible region of width (w) is at least 1-α. For example, letting d = w/2, then it might find the minimal n such that:

$$\int_{\hat{\rho}-d}^{\hat{\rho}+d} \Pi(\rho|data)d\rho \geq 1-\alpha \qquad (18)$$

where $\hat{\rho}$ is the posterior mean and the credible region is [$\hat{\rho}$-d, $\hat{\rho}$+d]. And the esimator $\hat{\rho}$ and the coverage probability depend on x=(x_1, x_2..., x_n).

Further, n can be selected to make the expected posterior coverage probability be at least 1-α.

$$m(x) = \int_0^1 L(x|\rho)\Pi(\rho)d\rho \qquad (19)$$

in which the expected value exceeds the marginal distribution of *x* induced by the prior distribution.

Therefore, it is recommended to seek a minimum n by using the average coverage criterion (ACC) and satisfies:

$$\sum_{\forall x \in \chi} \left[\int_{\hat{\rho}-d}^{\hat{\rho}+d} \Pi(\rho|data) d\rho \right] m(x) \geq 1 - \alpha \tag{20}$$

where m(x) gives weights.

And last, if we need a conservative sample size and to satisfy the required coverage and interval length on any possible sample x, we could use **WOC** rather than averaging to find the minimum n as follow:

$$\inf_{x \in \chi} \left[\int_{\hat{\rho}-d}^{\hat{\rho}+d} \Pi(\rho|data) d\rho \right] \geq 1 - \alpha$$

4 Conclusion

In conclusion, the Bayesian approach seems to be well suited to deal with inference goals in queuing systems. According to the data output from Almeida and Cruz (2018), it can conclude that using the Bayesian approach to estimate the traffic intensity of the M/M/1 model has a good prediction. The Bayesian factor may help test the prior distribution of the data. And it is possible to infer the traffic intensity and calculate the probabilities of the number of customers in the system. In addition, Quinino and Cruz (2016) also conclude that it is possible to determine the sample size within a certain range of accuracy by ACC (p.1001). In the model Geo/Geo/1, it is very similar to the M/M/1, and the biggest challenge is to find the probability distribution of the number of customers in the system and the second challenge is to find a suitable posterior distribution. We can see that the process of estimating the sample size of Geo/ Geo/ 1 is so similar to that of model M/M/1. As a result, the Bayesian method may also be used to estimate the sample size of the Geo/Geo/1 model.

Bibliography

Asanjarani, A., Nazarathy, Y., & Pollett, P. K. (2017, January 17). *Parameter and State Estimation in Queues and Related Stochastic Models: A Bibliography*. Researchgate. https://www.researchgate.net/publication/313097665

Armero, C., & Bayarri, S. (2015). Queues. In *International Encyclopedia of the Social & Behavioral Sciences.* (2nd ed., pp. 784-789). Retrieved from https://doi.org/10.1016/B978-0-08-097086-8.42164-1

Almeida, M.A.C., & Cruz, F. R. B. (2018). A note on Bayesian estimation of traffic intensity in single-server Markovian queues. *Communications in Statistics - Simulation and Computation, 47*(9), 2577-2586. https://doi.org/10.1080/03610918.2017.1353614

Choudhury, Amit., & Basak, Arpita. (2017). Statistical inference on traffic intensity in an M / M / 1 queueing system. *International Journal of Management Science and Engineering Management, 13*(4), 274-279. https://doi.org/10.1080/17509653.2018.1436010

C, L., & Appa Iyer, S. (2013). Application of queueing theory in health care: A literature review. *Operations Research for Health Care, 2*(1-2), 25-39. https://doi.org/10.1016/j.orhc.2013.03.002

Desrosières, A. (2013). The History of Statistics as a Genre: Styles of Writing and Social Uses. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, 119*(1), 8–23. https://doi.org/10.1177/0759106313486398

Fajardo, S. (1985). Completeness theorems for the general theory of stochastic processes. *In Methods in Mathematical Logic*, 174-194. Springer, Berlin, Heidelberg. https://link.springer.com/content/pdf/10.1007/BFb0075311

H 10. (n.d.). Retrieved from https://web.stanford.edu/class/ee384x/EE384X//handouts/H10.pdf

Introduction to queueing theory. (n. d). Retrieved from http://www.mathcs.emory.edu/~cheung/Courses/558/Syllabus/00/queueing/queueing.html

Lim, A. (2018, July 09). *An Introduction to Queuing Theory. The mathematical study of waiting in line*. ThoughtGo, https://www.thoughtco.com/queuing-theory-4171870

Li, Ji-homh., & Tian, Nai-shuo. (2016). Analysis of the Discrete Time Geo/Geo/1 Queue with Single Working Vacation. *Quality Technology & Quantitative Management*, *5*(1), 77-89. https://doi.org/10.1080/16843703.2008.11673177

Quinino, R.C., & Cruz, F. R. B. (2016). Bayesian sample sizes in an M/M/1 queueing systems. *The International Journal of Advanced Manufacturing Technology*, 88, 995-1002. https://link.springer.com/article/10.1007/s00170-016-8855-2

Romeijn, J. W. (2014, August 19). *Philosophy of Statistics*. The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/archives/spr2017/entries/statistics/

Sadia, F., & Hossain, S. S. (2014). Contrast of Bayesian and Classical Sample Size Determination. Journal of Modern Applied Statistical Methods, 13(2), 420-431. https://digitalcommons.wayne.edu/jmasm/vol13/iss2/23/