M/M/s Queuing Systems

Alejandro C. Frery

March 2023

School of Mathematics and Statistics
New Zealand

What is it about?

We will see in more details useful properties of some of the most widely used queuing systems: M/M/s systems.

Series

$$\sum_{k=0}^{\infty} p^k = \frac{1}{1-p}.$$

$$\sum_{k=1}^{\infty} p^k = \frac{p}{1-p}.$$

$$\sum_{k=0}^{n} p^k = \frac{1-p^{n+1}}{1-p}.$$

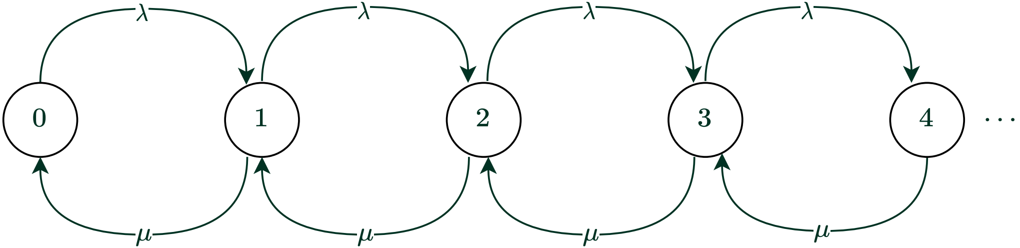
$$\sum_{k=0}^{\infty} kp^k = \sum_{k=1}^{\infty} kp^k = \frac{p}{(1-p)^2}.$$

M/M/1 Queueing Systems

The Simplest M/M/1 System

Recall that M/M/1 stands for M/M/1/GI/∞/∞.

We characterize the behaviour of the simplest M/M/1 system by means of its arrival (birth) λ and departure (death) μ rates.



Steady-State of the Simplest M/M/1 System

We already saw that the steady-state, if it exists, is characterized by the probabilities $\pi_k = \Pr(N = k)$, $k = 0, 1, 2, \ldots$: the long-term probability of finding k customers in the system.

We denoted $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2, \ldots)$, and we found that

$$\pi_0 = \frac{1}{1 + \frac{\lambda}{\mu} + \frac{\lambda}{\mu}\frac{\lambda}{\mu} + \frac{\lambda}{\mu}\frac{\lambda}{\mu}\frac{\lambda}{\mu} + \cdots} = \frac{1}{1 + \sum_{k=1}^{\infty} \rho^k} = \frac{1}{\sum_{k=0}^{\infty} \rho^k} = 1 - \rho$$

if $\rho = \lambda/\mu < 1$, and

$$\pi_k = \rho^k \pi_0 = \pi_0 (1 - \pi_0)^k,$$

So N, the number of customers in the system, follows a Geometric distribution:

$$\Pr(N = k) = \pi_0 (1 - \pi_0)^k = (1 - \rho)\rho^k.$$

Properties

Denoting the Geometric distribution as $N \sim \mathrm{Geom}(p)$, the following properties hold:

$$\begin{aligned}
\mathrm{Support}(N) &= \{0, 1, 2, \dots\} = \mathbb{N}_0, \\
\Pr(N = k) &= (1-p)p^k, \\
F_N(k) &= 1 - (1-p)^{k+1}, \\
\mathrm{E}(N) &= \frac{1-p}{p}, \\
Q_{1/2}(N) &= \left\lceil -\frac{1}{\log_2(1-p)} \right\rceil - 1, \\
\mathrm{Mode}(N) &= 0, \\
\mathrm{Var}(N) &= \frac{1-p}{p^2}.
\end{aligned}$$

How much is the system in use?
The system is in use when it is not idle.

The system is idle with probability $\mathrm{Pr}(N = 0) = \pi_0 = 1 - \rho = 1 - \lambda/\mu$, so the system is in use with probability

$$\mathrm{Pr}(\text{System is busy}) = \frac{\lambda}{\mu}.$$

But remember that the system is stable only if $\lambda < \mu$.

Interesting questions ii

What is the expected number of clients in the system?
You should prove that

$$\mathrm{E}(N) = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}.$$

What is the expected number of clients queueing?
If there is zero or one client in the system, nobody is queueing. The number of clients in the queue is

$$N_q = \begin{cases} 0 & \text{if } N = 0, 1 \\ N-1 & \text{if } N = 2, 3, 4, \ldots. \end{cases}$$

Let us write the probability function, i.e. the pairs $\big(n, \Pr(N_q = n)\big)$:

$$\big((0,\pi_0),(0,\pi_1),(1,\pi_2),(2,\pi_3),(3,\pi_4),\dots,(n,\pi_{n+1}),\dots\big)$$
$$\big((0,1-\rho),(0,(1-\rho)\rho),(1,(1-\rho)\rho^2),(2,(1-\rho)\rho^3),\dots,(n,(1-\rho)\rho^{n+1}),\dots\big)$$

So its expected value is

$$\mathrm{E}(N_q) = 0\big(1-\rho+(1-\rho)\rho\big) + \sum_{n=1}^{\infty} n(1-\rho)\rho^{n+1} = (1-\rho)\rho\sum_{n=1}^{\infty} n\rho^n$$
$$(1-\rho)\rho\frac{\rho}{(1-\rho)^2} = \frac{\rho^2}{1-\rho}.$$

What is the average number of customers being serviced?

The probability function of N_s is

$$\big((0,\pi_0),(1,\pi_1),(1,\pi_2),(1,\pi_3),\ldots,(1,\pi_n),\ldots\big) = \big((0,\pi_0),(1,1-\pi_0)\big),$$

so its expected value is

$$\mathrm{E}(N_s) = 0\pi_0 + 1(1-\pi_0) = 1-\pi_0 = 1-(1-\rho) = \rho = \frac{\lambda}{\mu}.$$

What is the average time it takes to serve a customer?

$$W = \frac{L}{\lambda} = \frac{\rho}{\lambda(1-\rho)} = \frac{1}{\mu - \lambda}$$
$$W_q = \frac{L_q}{\lambda} = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{\rho^2}{\lambda(1-\rho)}$$
$$W_s = \frac{L_s}{\lambda} = \frac{\rho}{\lambda} = \frac{1}{\mu}$$

Recap

The Maclaurin series of e^x is

$$e^x = \sum_{j=0}^{\infty} \frac{x^n}{n!},$$

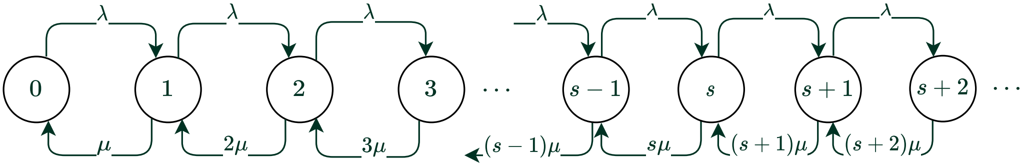
and it converges for all $x \in \mathbb{R}$.

M/M/∞ Queuing Systems

M/M/∞ Queuing Systems

Such systems are also known as IS (Infinite Servers) Queuing Systems.

M/M/∞ Queuing Systems

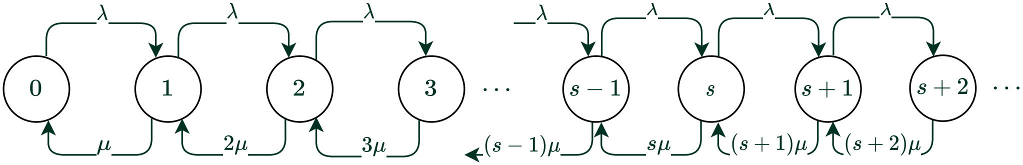


As there is always an available server, there is no queue in such systems.

IS systems can be described by the following rates:

$$\begin{cases} \lambda_i = \lambda & j = 0, 1, 2, \ldots, \\ \mu_0 = 0, \\ \mu_j = j\mu & j = 1, 2, 3, \ldots \end{cases}$$

The variable serving rate



The serving rate now varies: $\mu_j = j\mu$.

Does it mean that the j-th client will be served with rate $j\mu$; in other words, will each arriving client be served faster that the previous ones?

M/M/∞ Queuing Systems i

We will compute the steady-state probabilities starting by π_k and then finding π_0.
Denote $\rho = \lambda/\mu$.

$$
\begin{aligned}
\pi_1 &= \frac{\lambda}{\mu}\pi_0 = \rho\pi_0 \\
\pi_2 &= \frac{\lambda}{2\mu}\pi_1 = \frac{1}{2!}\left(\frac{\lambda}{\mu}\right)^2\pi_0 = \frac{1}{2!}\rho^2\pi_0 \\
&\vdots \\
\pi_{s-1} &= \frac{1}{(s-1)!}\rho^{s-1}\pi_0 \\
\pi_s &= \frac{1}{s!}\rho^s\pi_0 \\
\pi_{s+1} &= \frac{1}{(s+1)!}\rho^{s+1}\pi_0 \\
\pi_{s+2} &= \frac{1}{(s+2)!}\rho^{s+2}\pi_0 \\
&\vdots \\
\pi_{s+k} &= \frac{1}{(s+k)!}\rho^{s+k}\pi_0.
\end{aligned}
$$

M/M/∞ Queuing Systems ii

Use that $\boldsymbol{\pi} = (\pi_0, \pi_1, \dots, \pi_s, \pi_{s+1}, \dots, \pi_{s+k}, \dots)$ is a vector of probabilities, i.e. $\sum_{k=0}^{\infty} \pi_k = 1$, so

$$\begin{aligned}
1 &= \pi_0 + \rho\pi_0 + \frac{1}{2!}\rho^2\pi_0 + \cdots + \frac{1}{s!}\rho^s\pi_0 + \frac{1}{(s+1)!}\rho^{s+1}\pi_0 + \frac{1}{(s+2)!}\rho^{s+2}\pi_0 + \cdots \\
&= \pi_0\left(1 + \rho + \frac{1}{2!}\rho^2 + \cdots + \frac{1}{s!}\rho^s + \frac{1}{(s+1)!}\rho^{s+1} + \frac{1}{(s+2)!}\rho^{s+2} + \cdots\right) \\
&= \pi_0 \sum_{k=0}^{\infty} \frac{\rho^k}{k!} = \pi_0 e^{\rho}.
\end{aligned}$$

Finally,

$$\pi_0 = e^{-\rho}$$

and there is always a steady-state distribution.

The steady-state distribution of an M/M/∞ system with arrival rate λ and serving rate μ is characterized by the following probability distribution function, in which we denote $\rho = \lambda/\mu$:

$$
\begin{aligned}
\pi_0 &= e^{-\rho}, \\
\pi_1 &= \rho\pi_0 = \rho e^{-\rho}, \\
\pi_2 &= \frac{1}{2}\rho\pi_1 = \frac{1}{2}\rho^2 e^{-\rho}, \\
&\vdots \\
\pi_k &= \frac{\rho^k}{k!}e^{-\rho}.
\end{aligned}
$$

M/M/∞ Queuing Systems iv

Therefore, the steady-state distribution of the number of customers in the system follows a Poisson distribution with mean ρ:

$$\Pr(N = k) = \frac{\rho^k}{k!} e^{-\rho}.$$

Interesting question

What can we say about the system performance?

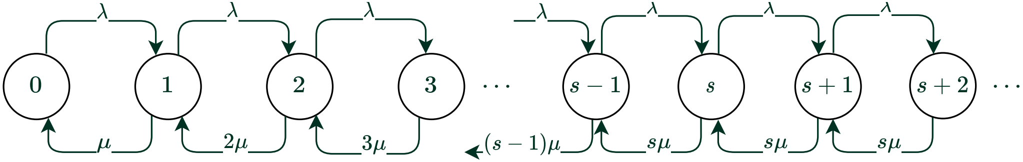
- The average number of customers in the system is $L = L_s = \mathrm{E}(N) = \rho = \lambda/\mu$, and there is no queue ($L_q = 0$).
- The average time in the system is $W = W_s = 1/\mu$, and $W_q = 0$.

M/M/s Queueing Systems

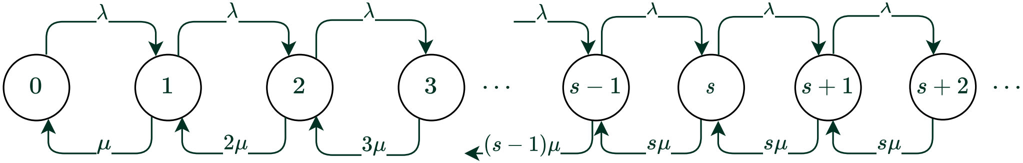
What's the difference?

An M/M/s system has more servers than an M/M/1 system, and less servers than an M/M/∞ system. M/M/s systems should, in principle, have an "intermediate" behaviour.

Customers arrive at the same rate, but they are served at a rate that depends on the number of available servers. At some point, customers have to start queuing.



There are s identical and independent servers in the system. While busy, each of them serves customers according to a Poisson Process with rate μ. If $k \leq s$ servers are busy, then the total output of the system is a Poisson process with rate $k\mu$ because it is a merging of k independent Poisson processes.



The Birth-and-Death process obeys the following rates:

$$\begin{cases} \lambda_i = \lambda & \text{for every } i = 0, 1, 2, \ldots \\ \mu_j = \begin{cases} 0 & \text{if } j = 0, \\ j\mu & \text{if } j = 1, \ldots, s \\ s\mu & \text{if } j = s+1, s+2, \ldots \end{cases} \end{cases}$$

M/M/s Queueing Systems iii

We will compute the steady-state probabilities starting by π_k and then finding π_0.

$$\begin{aligned}
\pi_1 &= \frac{\lambda}{\mu}\pi_0 = \rho\pi_0 \\
\pi_2 &= \frac{\lambda}{2\mu}\pi_1 = \frac{1}{2!}\left(\frac{\lambda}{\mu}\right)^2\pi_0 = \frac{1}{2!}\rho^2\pi_0 \\
\pi_3 &= \frac{\lambda}{3\mu}\pi_1 = \frac{1}{3!}\left(\frac{\lambda}{\mu}\right)^3\pi_0 = \frac{1}{3!}\rho^3\pi_0 \\
&\vdots \\
\pi_s &= \frac{1}{s!}\rho^s\pi_0 \\
\pi_{s+1} &= \frac{1}{s!}\frac{1}{s}\rho^{s+1}\pi_0 \\
\pi_{s+2} &= \frac{1}{s!}\frac{1}{s^2}\rho^{s+2}\pi_0 \\
&\vdots \\
\pi_{s+k} &= \frac{1}{s!}\frac{1}{s^k}\rho^{s+k}\pi_0 = \frac{\rho^s}{s!}\left(\frac{\rho}{s}\right)^k\pi_0.
\end{aligned}$$

M/M/s Queueing Systems iv

Use that $\boldsymbol{\pi} = (\pi_0, \pi_1, \dots, \pi_s, \pi_{s+1}, \dots, \pi_{s+k}, \dots)$ is a vector of probabilities, i.e. $\sum_{k=0}^{\infty} \pi_k = 1$, so

$$
\begin{aligned}
1 &= \pi_0 + \rho\pi_0 + \frac{1}{2!}\rho^2\pi_0 + \cdots + \frac{1}{s!}\rho^s\pi_0 + \frac{1}{s!s}\rho^{s+1}\pi_0 + \frac{1}{s!s^2}\rho^{s+2}\pi_0 + \cdots + \frac{\rho^s}{s!}\left(\frac{\rho}{s}\right)^k\pi_0 + \cdots \\
&= \pi_0\left(\underbrace{1 + \rho + \frac{1}{2!}\rho^2 + \cdots + \frac{1}{(s-1)!}\rho^{s-1}} + \underbrace{\frac{1}{s!}\rho^s + \frac{1}{s!s}\rho^{s+1} + \frac{1}{s!}\frac{1}{s^2}\rho^{s+2} + \cdots + \frac{\rho^s}{s!}\left(\frac{\rho}{s}\right)^k + \cdots}\right) \\
&= \pi_0\left(\sum_{k=0}^{s-1}\frac{\rho^k}{k!} + \frac{\rho^s}{s!}\sum_{k=0}^{\infty}\left(\frac{\rho}{s}\right)^k\right) = \pi_0\left(\sum_{k=0}^{s-1}\frac{\rho^k}{k!} + \frac{\rho^s}{s!}\frac{1}{1-\frac{\rho}{s}}\right)
\end{aligned}
$$

Finally,

$$
\pi_0 = \frac{1}{\sum_{k=0}^{s-1}\frac{\rho^k}{k!} + \frac{\rho^s}{s!}\frac{1}{1-\frac{\rho}{s}}},
$$

and there is a steady-state distribution iff $\rho/s < 1$ or, equivalently, iff $\lambda < s\mu$.

Interesting questions i

What is the average number of customers queuing?

$$L_q = \pi_0 \frac{\frac{\rho^{s+1}}{s!s}}{\left(1 - \frac{\rho}{s}\right)^2}$$

What is the average waiting time?

$$W_q = \frac{L_q}{\lambda}$$

What is the average number of customers being serviced?

$$L_s = \frac{\lambda}{\mu} = \rho$$

Interesting questions ii

And the average number of costumers in the system?

$$L = L_q + L_s = \pi_0 \frac{\frac{\rho^{s+1}}{s!s}}{\left(1 - \frac{\rho}{s}\right)^2} + \rho$$

What is the probability that an arriving customer will have to wait?

$$
\begin{aligned}
\Pr(\text{wait}) = \sum_{n=s}^{\infty} \pi_n &= \sum_{j=0}^{\infty} \frac{\rho^{s+j}}{s!s^j}\pi_0 \\
&= \frac{\rho^s}{s!}\pi_0 \sum_{j=0}^{\infty} \left(\frac{\rho}{s}\right)^j \\
&= \underbrace{\frac{\rho^s}{s!}\pi_0}_{\pi_s} \frac{1}{1-\frac{\rho}{s}} = \frac{\pi_s}{1-\frac{\rho}{s}}
\end{aligned}
$$

References

Hillier, F. S. & Lieberman, G. J. (2001), *Introduction to Operations Research*, 7 edn, McGraw-Hill, New York.

Little, J. D. C. (1961), 'A proof for the queueing formula: $L = \lambda W$', *Operations Research* **9**(3), 383–387.