Note on Queueing Theory

Zepeng CHEN
The HK PolyU

Date: January 12, 2023

1 Queueing Theory

Definition 1.1 (The label of queueing system)

Normally a queueing system is noted with $A/S/c$, where A is the property of arrival process, S is the property of service process, c is the number of servers.

	A: arrival process	S: service process
M: memoryless	Poisson arrival	iid exponential service times
G: general	iid interarrival times	iid service times
D: deterministic	fixed interarrival times	fixed service times

2 M/M/k

Lemma 2.1 (P_i **for** $M/M/1$)

Assume $\lambda_n = \lambda, \mu_n = \mu, \lambda/\mu < 1$, *then*

$$P_n = \frac{(\lambda/\mu)^n}{1+\sum_{n'=1}^{\infty}(\lambda/\mu)^{n'}} = \left(\frac{\lambda}{\mu}\right)^n \left(1-\frac{\lambda}{\mu}\right), \quad n \geq 0$$

Proof By Birth and Death process. ∎

Lemma 2.2 (Independence of present and past)

In an ergodic $M/M/1$ queue in steady state,

1. *the number of customers presently in the system is independent of the sequence of past departure times,*
2. *the waiting time spent in the system (waiting in queue plus service time) by a customer is independent of the departure process prior to his departure.*

Lemma 2.3 (Truncated M/M/1 queue)

Consider an M/M/1 queue in which arrivals finding N in the system do not enter but rather are lost. This finite capacity $M/M/1$ system can be regarded as a truncated version of *the $M/M/1$ and so it is time reversible with limiting probabilities given by*

$$P_j = \frac{(\lambda/\mu)^j}{\sum_{i=0}^{N}(\lambda/\mu)^i}, \quad 0 \le j \le N$$

Proof Follow Lemma **??**. ■

Lemma 2.4 (Output of $M/M/s$ queue)

Consider an $M/M/s$ queue in which customers arrive in accordance with a Poisson process having rate λ and are served by any one of s servers – each having an exponentially distributed service time with rate μ. If $\lambda < s\mu$, then the output process of customers departing is, in steady state, a Poisson process with rate λ.

Lemma 2.5 ($M/M/s$ **queue and Birth and Death Process)**

Suppose arrival rate λ and service rate μ, if we let $X(t)$ denote the number in the system at time t, then $\{X(t), t \ge 0\}$ is a birth and death process with

$$\mu_n = \begin{cases} n\mu & 1 \le n \le s \\ s\mu & n > s \end{cases} \quad \text{and} \quad \lambda_n = \lambda, n \ge 0$$

3 M/G/k

When the arrival process with rate λ, and say that a cycle correspond to the start of a busy period, then the time of a cycle inclue (i) the busy period of the cycle and (ii) the time from the departure of the last customer in the cycle to the next arrival. According to the memoryless property of Poisson arrivals, the latter is exponentially distributed with mean $1/\lambda$:

$$E[\text{ time of a cycle }] = E[\text{ time of a busy period }] + 1/\lambda$$

Lemma 3.1 (The # of customers that have completed service by time t and not (Song, 2020, Lec. 2))

The # of customers $N_1(t)$ that have completed service by time t is Poisson with mean

$$E[N_1(t)] = \lambda \int_0^t G(t-s)ds = \lambda \int_0^t G(y)dy$$

and the # of customers $N_2(t)$ being served at time t is Poisson with mean

$$E[N_2(t)] = \lambda \int_0^t \bar{G}(t-s)ds = \lambda \int_0^t \bar{G}(y)dy$$

Further, $N_1(t)$ and $N_2(t)$ are independent.

4 G/M/k

5 G/G/k and Little's Law

Lemma 5.1 (Arrival only when Free server (Song, 2020, PS. 2))

Assume that an arrival only enters the bank if the server is free when he or she arrives, then the events of entering the bank by time t constitute a renewal process, while the events of leaving the bank by time t does not constitute a renewal process. But if the arrival is exponential, then the events of leaving constitutes a (delayed) renewal process.

Proof For entering, both the future arrivals and service times are independent of the history, that is, the process begins anew. For leaving, the next arrival depends on the time of the last arrival before or at t. ■

Customers arrive at a single-server service station according to a renewal process. Upon arrival, she is immediately served if the server is idle, and waits if the server is busy. The service time is i.i.d, and is independent of the arrival stream. Suppose that the first customer arrives at time 0. Let X_i denote the time between the ith and $(i+1)$st arrival, and Y_i denote the ith service time. And assume that $E[Y_i] < E[X_i] < \infty$, which ensures the finiteness conditions for the theorem for renewal reward process and the Wald's equation.

Definition 5.1

- $n(t)$:= *the number of customers in the system at time t*
- $L = \lim_{t\to\infty} \int_0^t n(s)ds/t$:= *long-run average number of customers in the system*
- $n(s)$:= *a reward is earned at time s*
- *L represents the long-run average reward*

$$L = \lim_{t\to\infty} \int_0^t n(s)ds/t = \frac{E[\text{reward during a cycle}]}{E[\text{time of a cycle}]} = \frac{E[\int_0^T n(s)ds]}{E[T]} \quad \textit{Theorem of renewal reward proce}$$

Define a discrete-time renewal reward process

Definition 5.2

- W_i:= *the amount of time the ith customer spends in the system, and we say we received a reward W_i at the arrival of ith customer*
- $W := \lim_{n\to\infty} \frac{W_1+\ldots+W_n}{n}$
- N:= *the number of customers served in a cycle*

$$W = \frac{E[\text{reward during a cycle}]}{E[\text{time of a cycle}]} = \frac{E[\sum_{i=1}^N W_i]}{E[N]}$$

Remark A new cycle begins each time when an arrival finds the system empty, and obviously the process restarts itself each cycle.

Theorem 5.1 (Little's Law)

Let $\lambda = 1/E[X_i]$ denote the arrival rate, then $L = \lambda W$.

Remark Acutally, the little's law holds for more general queueing system.

Proof Note that T is the length of a cycle, and N is the number of customers served in a cycle. Obviously, $T = \sum_{i=1}^{N} X_i$. Note that the event $\{N = n\}$ is equivalent to: (i) for any $k = 1, ..., n-1$, the $(k+1)$st arrival time is before the kth departure time, and (ii) the $(n+1)$st arrival time is after the nth departure time. Therefore, $\{N = n\}$ is independent of $X_{n+1}...$ and is a stopping time for $X_1, ...$

$$X_1 + ... + X_k < Y_1 + ... + Y_k, \quad k = 1, .., n-1$$

$$X_1 + ... + X_n > Y_1 + ... + Y_n$$

Hence by the Wald's equation, we have

$$E[T] = E[N]E[X] = E[N]/\lambda$$

$$L = \lambda \frac{E\left[\int_0^T n(s)ds\right]}{E[N]} = \lambda W \frac{E\left[\int_0^T n(s)ds\right]}{E\left[\sum_{i=1}^{N} W_i\right]}$$

By imaging that each customer pay at a rate of 1 per unit time when in the system, we see

$$\int_0^T n(s)ds = \sum_{i=1}^{N} W_i = \text{ total paid during a cycle}$$

Reduce it can prove the theorem. ■

Lemma 5.2 (Server Utilization)

1. *For a $G/G/1$ queue, let U_t denote whether the server is busy at time t.*

$$\textit{Server Utilization } = E\left[Y_i\right]/E\left[X_i\right]$$

2. *For a $G/G/k$ queue, let U_t denote whether the server is busy at time t.*

$$\textit{Server Utilization } = E\left[Y_i\right]/\left(kE\left[X_i\right]\right)$$

Proof

$$\begin{aligned}
\text{G/G/1's Server Utilization } &= \lim_{t\to\infty} \frac{\int_0^t U_s ds}{t} \\
&= \lim_{t\to\infty} \frac{\int_0^t (\text{ number in service at } s)ds}{t}
\end{aligned}$$

$$= \text{ average number in service}$$

$$= \lambda E\left[Y_i\right] \quad \text{by Little's law}$$

$$= E\left[Y_i\right]/E\left[X_i\right]$$

$$
\begin{aligned}
\text{G/G/k's Server Utilization} &= \lim_{t\to\infty} \frac{\int_0^t U_s ds}{t} \\
&= \lim_{t\to\infty} \frac{\int_0^t (\text{ number served by server } i \text{ at } s) ds}{t} \\
&= \text{average number served by server} i \\
&= \frac{\text{average number in service}}{k} \text{as all servers are identical} \\
&= \lambda E\left[Y_i\right]/k \quad \text{by Little's law} \\
&= E\left[Y_i\right]/\left(kE\left[X_i\right]\right)
\end{aligned} \tag{1}
$$

■

6 Tandem Queue

Theorem 6.1

For the ergodic tandem queue in steady state,

1. *the numbers of customers presently at server 1 and at server 2 are independent, and*

$$
P\left\{\begin{array}{l} n \text{ at server } 1, \\ m \text{ at server } 2 \end{array}\right\} = \left(\frac{\lambda}{\mu_1}\right)^n \left(1 - \frac{\lambda}{\mu_1}\right) \left(\frac{\lambda}{\mu_2}\right)^m \left(1 - \frac{\lambda}{\mu_2}\right)
$$

2. *the waiting time of a customer at server 1 is independent of its waiting time at server 2.*

Proof Follow Lemma 2. ■

7 Jackson Network

Bibliography

Song, Miao (2020). *LGT6202 Stochastic Models and Decision under Uncertainty*.