Theory of Stochastic Processes
11. Queues

Tomonari Sei
sei@mist.i.u-tokyo.ac.jp

Department of Mathematical Informatics, University of Tokyo

June 29, 2017

http://www.stat.t.u-tokyo.ac.jp/~sei/lec.html

Handouts & Announcements

Handouts:

- Slides (this one)
- Copy of Chapter 8 of PRP
- Copy of §11.1 – 11.3 of PRP

Announcements:

- About the final exam (important!)
  - The final exam is held on July 20 (Thu), 10:25–, 90 min.
  - Contact me if you cannot attend it due to some unavoidable reasons.
  - The exam is open-book and open-note. Write your answer in English.
  - The exam will cover the whole material up to the next lecture (July 6). At least one question will be from the topics before the midterm exam.

Outline today

1. Where are we now?
2. Queues
   - Examples
   - M/M/1
   - M/G/1
3. Recommended problems

Where are we now?

Dependence of topics

$$\begin{tikzcd}
 & \boxed{\text{Stationary processes}} & \\
\boxed{\text{Markov chains}} \arrow[ur] \arrow[r] \arrow[dr] & \boxed{\text{Queues}} & \text{(today)} \\
\boxed{\text{Martingales}} \arrow[r] & \boxed{\text{Diffusion processes}} & \text{(next week)}
\end{tikzcd}$$

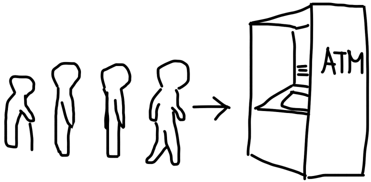
Queues = waiting lines

Today's topic

Example

Consider an ATM which serves customers.

- Customers arrive at ATM according to a stochastic process.
- The service time is also random.



Queueing theory

Other examples of queues:

- security check at Narita airport
- hospitals
- Starbucks coffee
- homework you have to do

The queueing theory concerns a mathematical study of queueing systems.

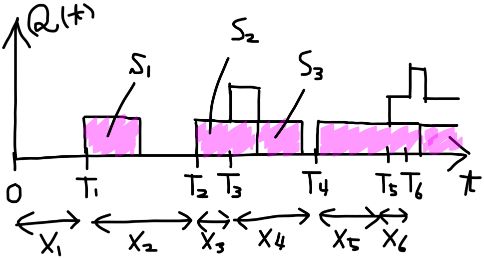
Remark:

- In Japanese, a queue is called "待ち行列".
- Do not confuse it with "行列" (a matrix).

Notation

We only consider single-server queues.

- T_n is the time when the n-th customer arrives.
- $X_n = T_n - T_{n-1}$ is the interarrival time, where $T_0 = 0$.
- S_n is the service time of the n-th customer.
- $Q(t)$ is the number of waiting customers at t, including ones served.



Definition

Definition

A queueing model is a pair of sequences $\{X_n\}$ and $\{S_n\}$ of independent random variables.

Queueing models are classified by the distributions of X_n and S_n, together with the number of servers ($= 1$ here). In particular,

- M/M/1: X_n and S_n are exponential (M stands for Markov).
- M/G/1: X_n is exponential and S_n is any (G stands for general).

Remark:

- Remember that interarrival times of a Poisson process are exponential.
- The notation A/B/s is called Kendall's notation.

An important quantity

Definition

The traffic intensity ρ is defined by

$$\rho = \frac{E[S]}{E[X]} = \frac{\text{(mean service time)}}{\text{(mean interarrival time)}}.$$

We may expect that

- If $\rho < 1$, then $\{Q(t)\}$ is "stable".
- If $\rho > 1$, then $\{Q(t)\}$ is "unstable".

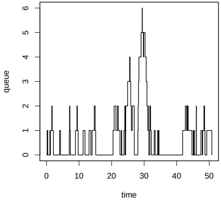
Let us check out this claim for M/M/1 and M/G/1.

M/M/1

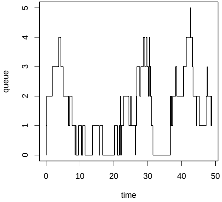
- M(λ)/M(μ)/1: X_n and S_n are exponential with parameters λ and μ, respectively. The pdfs are $f_X(x) = \lambda e^{-\lambda x}$ and $f_S(x) = \mu e^{-\mu x}$.
- The traffic intensity is

$$\rho = \frac{E[S]}{E[X]} = \frac{1/\mu}{1/\lambda} = \frac{\lambda}{\mu}.$$

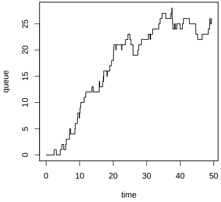
- Sample paths of $Q(t)$:



$\rho = 0.5$



$\rho = 1$



$\rho = 1.5$

An R code is available from the course web site.

Stability of M/M/1

Theorem 11.2.8

Assume the M(λ)/M(μ)/1 model. Then $Q(t)$ is a birth-death process, that is, a continuous-time Markov chain with the generator

$$\mathbf{G} = \begin{pmatrix} -\lambda & \lambda & & 0 \\ \mu & -(\lambda+\mu) & \lambda & \\ & \mu & -(\lambda+\mu) & \ddots \\ 0 & & \ddots & \ddots \end{pmatrix}$$

In particular, the stationary distribution exists if and only if $\rho = \lambda/\mu < 1$.

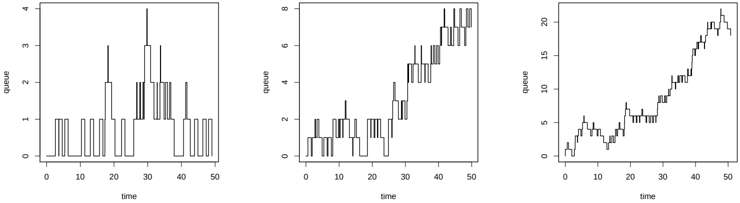
Review exercise

Let $\rho < 1$. Show that the stationary distribution is $\pi_k = (1-\rho)\rho^k$ and the mean number of waiting customers is $\rho/(1-\rho)$.

M/G/1

Now let us go on to the M/G/1 model.

- Sample paths of $Q(t)$ when $S_n \equiv 1$ (deterministic).



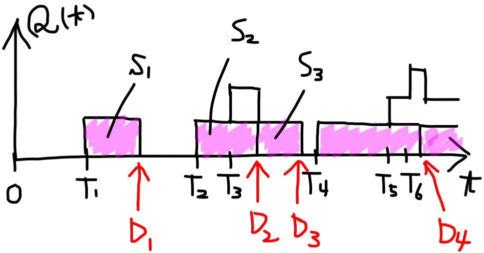
$\rho = 0.5$ $\rho = 1$ $\rho = 1.5$

Exercise

$Q(t)$ is no longer a Markov chain. Why?

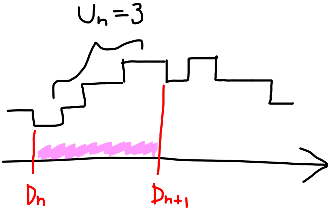
The idea

- Let D_n be the leaving time of the n-th customer. In other words, D_n is the n-th decreasing point of $Q(t)$.
- We will see that $Q(D_n)$ is a discrete-time Markov chain, where $Q(D_n)$ is inerpreted as $Q(D_n+)$.

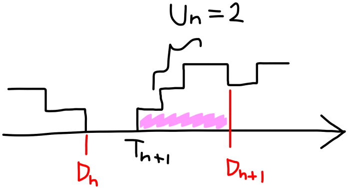


Find a Markov chain

- Let U_n be the number of arrived customers during the service time of the $(n+1)$-th customer.
- U_n is independent of $\{Q(t)\}_{t \leq D_n}$ because of the Markov property of arriving times.



$Q(D_n) > 0$



$Q(D_n) = 0$

Exercise

Check that $Q(D_{n+1}) = U_n + \max(Q(D_n) - 1, 0)$.

Theorem 11.3.4

Assume the M(λ)/G/1 model. Then the discrete-time process $\{Q(D_n)\}$ is a Markov chain with the transition matrix

$$\mathbf{P} = \begin{pmatrix} \delta_0 & \delta_1 & \delta_2 & \cdots \\ \delta_0 & \delta_1 & \delta_2 & \cdots \\ 0 & \delta_0 & \delta_1 & \cdots \\ 0 & 0 & \delta_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad \delta_j = P(U_n = j) = E\left[\frac{(\lambda S_n)^j}{j!} e^{-\lambda S_n}\right].$$

Theorem 11.3.5

The unique stationary distribution $\boldsymbol{\pi}$ exists if and only if $\rho < 1$.
In that case, the generating function of $\boldsymbol{\pi}$ is

$$G(s) = \sum_j \pi_j s^j = (1-\rho)(s-1)\frac{M_S(\lambda(s-1))}{s - M_S(\lambda(s-1))},$$

where $M_S(\theta) = E[e^{\theta S_n}]$.

See §11.3 for proofs and further details.

Recommended problems

Recommended problems:

- §8.4, Problems 1, 2, 3, 4*, 5*.
- §11.2, Problems 3, 7*.
- §11.3, Problem 1*.

The asterisk (*) shows difficulty.