EE 621 **Spring 2020-21**

HW 3: Endgame

Assigned: 29/04/17 Due: No need to submit

1 CTMCs and their EMCs

Consider a positive recurrent CTMC over a countable state space S. Assuming the corresponding embedded Markov chain (EMC) is also positive recurrent, relate the stationary distribution π of the CTMC to the stationary distribution $\hat{\pi}$ of the EMC. You should be able to express each of these distributions in terms of the other.

2 M/M/1 response time distribution

In this problem, you will derive the *distribution* of the steady state response time T in an M/M/1 queue under FCFS scheduling.

Recall our notation from class: the job arrival rate is λ, and the average service time of a job equals $1/\mu$, where $\lambda < \mu$. Recall that we have proved that $\mathbb{E}\left[T\right] = \frac{1}{\mu-\lambda}$.

Proceed as follows:

1. It will be useful to first prove the following result: Suppose that $\{X_i\}_{i\geq 1}$ is a sequence of i.i.d. $\mathrm{Exp}(\mu)$ random variables. M is a $\mathrm{Geometric}(p)$ random variable (i.e., $P\left(X = k\right) = (1-p)^{k-1}p$ for $k \in \mathbb{N}$) independent of $\{X_i\}_{i\geq 1}$. Prove that the random sum $\sum_{i=1}^{M} X_i$ is an $\mathrm{Exp}(p\mu)$ random variable.
2. Let N' denote the number of jobs as seen by an arriving job in steady state. What is the distribution of N'? (Hint: PASTA)
3. Express the response time T of our arriving job as a random sum.
4. Specify the distribution of T.

Caution: Be careful in distinguishing between the geometric distribution (with support $\mathbb{N}$) and the time-shifted geometric distribution (with support $\mathbb{Z}_+$).

3 Taxi stand

Consider a taxi stand with room for W taxis. Suppose that people arrive at the taxi stand as per a Poisson process with rate λ. If an arriving person finds one or more taxis in the stand, he/she boards a taxi and leaves. Else, the person waits in queue (this queue can grow unboundedly). Taxis arrive to the stand as per a Poisson process with rate $\mu > \lambda$. If an arriving taxi finds the stand full (i.e., with W taxis waiting), it departs immediately. Otherwise, it picks up a waiting customer if available, or joins the stand of waiting taxis.

(a) What is the long term fraction of time there are k people waiting in line?

(b) What is the steady state average waiting time for a person until he/she boards a taxi?

4 M/M/1 with finite buffer[1]

Consider an M/M/1/n queue. That is, consider a queueing system where jobs arrive according to a Poisson process with rate λ, service times are exponentially distributed with mean $1/\mu$, and there is a finite buffer size of n (i.e., the system can accommodate at most n jobs at a time, including the one in service). If a job arrives when the buffer is full, it gets rejected (lost). The jobs in the buffer are served in a FCFS manner.

(a) Set up the CTMC corresponding to the number of jobs in the system.

(b) Derive the stationary probabilities.

(c) Compute the utilization ρ of the system (i.e., the long-term fraction of time the server is busy).

(d) What is the long run fraction of jobs that get lost? (Don't forget to use "PASTA" in your justification.)

(e) What is the long term rate of lost jobs?

(f) Derive $\mathbb{E}[N]$, the steady state average number of jobs in the system.

(g) Derive $\mathbb{E}[T]$, the steady state average response time, for the jobs that enter the system.

(h) Your goal as system operator is to minimise the fraction of lost jobs. Suppose n = 5 and $\lambda/\mu = 0.4$; you have to choose between two system upgrades – (i) double the buffer size, (ii) double the server speed μ. What should you do?

(i) Repeat the above exercise for $n = 5$ and $\lambda/\mu = 0.8$. If your answer the same? Why?

5 To balance or not to balance

Consider the two-server setup depicted in Fig. 1. The job arrival stream is Poisson with rate λ. Each arriving job is sent to Server 1 with probability p and to Server 2 with probability $1 - p$. The service rate of Server i equals μ_i (i.e., a job takes $\text{Exp}(\mu_i)$ service time on Server i), where $\lambda < \mu_1 + \mu_2$. Each server processes the jobs in its queue in a FCFS manner.

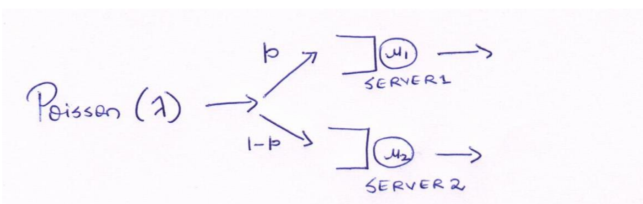

Figure 1: Two-server setup

1. Derive an expression for $\mathbb{E}[T]$, the steady state average response time of a job in this system.
2. Suppose $\mu_1 = \mu_2$. Prove or disprove that $\mathbb{E}[T]$ is minimized by balancing the utilization of the two servers.
3. Suppose $\mu_1 \neq \mu_2$. Prove or disprove that $\mathbb{E}[T]$ is minimized by balancing the utilization of the two servers. Can you justify your solution intuitively?

[1]This is Problem 13.5 from Mor's text.

6 CTMCs: Average time between visits

Consider an irreducible, positive recurrent CTMC with a recurrent EMC.

1. Prove, using a renewal reward argument, that the mean time between visits to state $i \in S$ equals $\frac{1}{\nu_i \pi_i}$ where $\nu_i = \sum_{j \neq i} q_{ij}$ is the total outgoing rate from state i and π_i is the stationary probability of state i.

   Note: This gives positive recurrnce the same meaning as in the DTMC case: in a positive recurrent chain, the mean return time to each state is finite.

2. Prove, also using a renewal reward argument, that for $i, j \in S$, prove that the long run rate of $i \to j$ transitions equals $\pi_i q_{ij}$.

   Note: The stationary equations for a CTMC are:

$$\sum_j \pi_i q_{i,j} = \sum_j \pi_j q_{ji}.$$

   With the above result, you can interpret the stationary equations as balancing the rate of outgoing transitions from state i to the rate of incoming transitions into state i. You should also be able to re-interpret the definition of time-reversibility using this result!