MATH4240 Tutorial 12 Notes

1 Stationary Distribution

For a (continuous-time) Markov process with rate matrix D and transition function P, forward equation implies

$$P' = PD$$

and on the distribution π, like the discrete-time chain we have $\pi = \pi_0 P$ and so

$$\pi' = \pi_0 P' = \pi_0 PD = \pi D$$

This implies that a stationary distribution π must satisfy

$$\pi D = 0$$

(Compare with the discrete-time case where $D = P - I$.)

Furthermore, as the process is aperiodic, we must have convergence to such stationary distribution $\pi(t) \to \pi$, that is the stationary distribution is also a limiting distribution. The many properties and approaches for discrete-time chain (e.g. one-step equations, (null-/positive-) recurrence, mean recurrence time) can be applied similarly to continuous-time Markov processes with obvious modifications. (See lecture for details.)

For example, consider a birth-and-death process with rate matrix

$$D = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \dots \\ \mu_1 & -(\mu_1 + \lambda_1) & \lambda_1 & 0 & \dots \\ 0 & \mu_2 & -(\mu_2 + \lambda_2) & \lambda_2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

(As always, we will assume it is non-explosive.)

As noted in lecture (or just solve it with the same approach of the discrete-time chain), the stationary distribution is

$$\pi(x) = \pi_x / \sum \pi_i \quad \text{where} \quad \pi_x = \prod_{i=0}^{x-1} \frac{\lambda_i}{\mu_{i+1}}$$

assuming that $\sum \pi_i < \infty$, effectively replacing p_i with λ_i and q_i with μ_i in the formula for the discrete-time birth-and-death chain.

2 Computations with Queue

Example 1. Consider the simple case of M/M/1 queue, that is $\lambda_i = \lambda$ and $\mu_i = \mu$ are constants. By the formula, the stationary distribution is given by

$$\pi_x = \rho^x$$
$$\pi(x) = (1 - \rho)\rho^x$$

where $\rho = \lambda/\mu$ is the *offered load* mentioned in the last tutorial session, assuming we have $\lambda < \mu$, that is $\rho < 1$. In particular, we have $\pi(0) = 1 - \rho$ and so the probability that the server is busy is $1 - \pi(0) = \rho$, exactly the same utilization law mentioned in the last session.

Now, what is the average amount of time a customer spent in the queue (before being served)?

From the stationary distribution, we can compute the average length of the *whole system* to be

$$L = \sum n(1-\rho)\rho^n = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}$$

By Little's law, the average time a customer spent in the whole system is $W = L/\lambda = \frac{1}{\mu-\lambda}$.

This implies that the average time of a customer spent in the queue (waiting to be served) is

$$W_q = W - E\,(\text{service time}) = \frac{1}{\mu-\lambda} - \frac{1}{\mu} = \frac{\lambda}{\mu(\mu-\lambda)} = \frac{\rho}{\mu-\lambda}$$

(You can also compute the average *queue* length as $L_q = L - 1 + \pi(0) = L - \rho$ and use Little's law only on the queue, which yields the same result.)

Example 2. Let us consider a M/M/c queue with arrival rate λ and service rate μ. Recall that this is a birth-and-death process with $\mu_x = \min(x,c)\mu$ and $\lambda_x = \lambda$.

You can derive (use the formula, or just verify) that the stationary distribution is

$$\pi(x) = \pi_x / \sum \pi_i \quad \text{where} \quad \pi_n = \begin{cases} E^n/n! & \text{if } n < c \\ \rho^{n-c}E^c/c! = c^c\rho^n/c! & \text{if } n \geq c \end{cases}$$

with *offered traffic intensity* $E = \lambda/\mu$ and *offered load* $\rho = E/c = \frac{\lambda}{c\mu}$, assuming again $\rho < 1$ that so that $\sum \pi_i < \infty$, in which case

$$\sum \pi_i = \sum_{n=0}^{c-1} \frac{E^n}{n!} + \frac{E^c}{c!}\frac{1}{1-\rho} = \sum_{n=0}^{c-1} \frac{E^n}{n!} + \frac{E^c}{c!}\frac{c}{c-E}$$

(On $c \to \infty$, this is just M/M/∞, which the stationary distribution as covered in lecture is of Poisson with parameter $E = \lambda/\mu$.)

Given this, what is the probability that a new customer has to wait in the queue to be served?

A new customer has to wait if at arrival there are at least c people in the system. By PASTA property, this new arrival should see the queue in the stationary distribution. So the probability to wait is

$$P_{\text{wait}} = C(E,c) := \sum_{i=c}^{\infty} \pi(i) = \frac{\frac{E^c}{c!}\frac{c}{c-E}}{\sum_{n=0}^{c-1}\frac{E^n}{n!} + \frac{E^c}{c!}\frac{c}{c-E}} = \frac{c}{c-E}\pi(c)$$

This is commonly referred to as *Erlang-C formula.*

You can also compute the average time a customer has to wait in the queue (before being served). The average queue length is

$$L_q = \sum_{n=c}^{\infty}(n-c)\pi(n) = \frac{E^c}{c!}\sum_{n=0}^{\infty} n\rho^n\pi(0) = \frac{E^c}{c!}\frac{\rho}{(1-\rho)^2}\pi(0) = \frac{E}{c-E}C(E,c)$$

and so by Little's law,

$$W_q = L_q/\lambda = \frac{C(E,c)}{c\mu-\lambda}$$

This also implies the average size of the system is

$$L = \lambda W = \lambda(W_q + E\,(\text{service time})) = L_q + \lambda/\mu = \frac{E}{c-E}C(E,c) + E$$

exactly the same result as computing $\sum n\pi(n)$. (Alternatively, the average number of customer *being served* is $E < c$.)

Example 3. You can also noted that the same formula for stationary distribution holds if the queue has a finite capacity limit. So, for a M/M/c/K queue we have

$$\pi(x) = \left(\sum \pi(n)\right)^{-1} \cdot \begin{cases} E^x/x! & \text{if } x < c \\ \rho^{x-c}E^c/c! & \text{if } x \geq c \end{cases}$$

with

$$\sum \pi(n) = \sum_{n=0}^{c-1} \frac{E^n}{n!} + \frac{E^c}{c!} \cdot \begin{cases} \frac{1-\rho^{K-c+1}}{1-\rho} & \text{if } \rho \neq 1 \\ K-c+1 & \text{if } \rho = 1 \end{cases}$$

In particular, for a M/M/c/c queue ($K = c$) the stationary distribution is just

$$\pi(x) = \frac{E^x/x!}{\sum_{n=0}^{c} E^n/n!}$$

Furthermore, if a new customer is rejected when the queue is full, the probability of rejection is

$$P_{\text{reject}} = B(E, c) \coloneqq \pi(c) = \frac{E^c/c!}{\sum_{n=0}^{c} E^n/n!}$$

which commonly referred to as *Erlang-B formula* or *Erlang's loss formula*. The probability P_{reject} is also called *Grade of Service (GoS)*.

Using Little's law, you can also compute for an M/M/c/K queue e.g. the average waiting time in the queue (if not rejected), but note that as on average the arrival may be rejected (and never enter the system) with probability P_{reject}, the *effective average arrival rate* must be adjusted to be $\lambda_{\text{eff}} = \lambda(1 - P_{\text{reject}}) = \lambda(1 - \pi(K))$. For example, in an M/M/c/c queue, where there is no queue to wait in, Little's law implies

$$L = \lambda_{\text{eff}} W = \lambda(1 - B(E, c))/\mu = E(1 - B(E, c))$$

which is the same as computing $L = \sum n\pi(n)$.

Example 4. Consider the *Engset model*, a variant of M/M/c/c queue, where there is only a finite population of customers $K > c$ and the arrival rate is proportional to the remaining size, that is, $\lambda_x = (K - x)\lambda$. The service rate is still μ, and an arrival is rejected if there is no server available. What is the probability of rejection?

Reusing the formula for stationary distribution, we have for $x \leq c$,

$$\pi_x = \prod_{i=0}^{x-1} \frac{\lambda_i}{\mu_{i+1}} = E^x \prod_{i=0}^{x-1} \frac{K-i}{i+1} = E^x \frac{K!/(K-x)!}{x!} = \binom{K}{x} E^x$$

In the case of infinite population, the rejection probability is just $\pi(c)$. However, we are now considering a model with finite population, so the new customer, at the time of arrival, sees only a system with a population of $K-1$ customers, and so by PASTA should see only the stationary distribution with population $K - 1$. This leads to the *Engset formula*

$$P_{\text{reject}} \coloneqq \pi_{K-1}(c) = \frac{\binom{K-1}{c} E^c}{\sum_{n=0}^{c} \binom{K-1}{n} E^n}$$

In particular, if $K = c + 1$, $P_{\text{reject}} = (\frac{E}{1+E})^c$.[1]

[1] This can be seen as $P_{\text{reject}} = B(E, 1)^c$, that is the probability that the new customer getting rejected by c independent M/M/1/1 queues, although this interpretation does not seem to allow an easy generalization to other cases.