Week 11

31 Stationary and Limit Distributions

Stationary Distribution

Let X_t be a continuous-time Markov chain with transition matrix $P(t) = (p_{ij}(t))_{i,j \in S}$ and state space $S = \{0, 1, 2, \cdots\}$. A probability distribution π on S, i.e, a vector $\pi = [\pi_0, \pi_1, \pi_2, \cdots]$, where $\pi_i \in [0, 1]$ and

$$\sum_{i \in S} \pi_i = 1,$$

is said to be a **stationary distribution** for X_t if

$$\pi = \pi \, P(t), \qquad \text{for all } t \geq 0.$$

The intuition here is exactly the same as in the case of discrete-time chains. If the probability distribution of X_0 is π, then the distribution of X_t is also given by π, for any $t \geq 0$.

One of the main uses of the generator matrix Q is finding the stationary distribution.

Th11.1. Consider a continuous Markov chain X_t with the state space S and the generator Matrix Q. The probability distribution π on S is a stationary distribution for X_t if and only if it satisfies

$$\pi Q = 0.$$

Proof. We assume S is finite for simplicity, i.e., $\pi = [\pi_0, \pi_1, \cdots, \pi_r]$, for some $r \in \mathbb{N}$. If π is a stationary distribution, then $\pi = \pi P(t)$. Differentiating both sides, we obtain

$$\begin{aligned} 0 &= \frac{d}{dt}[\pi P(t)] \\ &= \pi P'(t) \\ &= \pi Q P(t) \quad \text{(backward equations)} \end{aligned}$$

Now, let $t = 0$ and remember that $P(0) = I$, the identity matrix. We obtain

$$0 = \pi Q P(0) = \pi Q.$$

Next, let π be a probability distribution on S that satisfies $\pi Q = 0$. Then, by backward equations,

$$P'(t) = Q P(t).$$

Multiplying both sides by π, we obtain

$$\pi P'(t) = \pi Q P(t) = 0.$$

Note that $\pi P'(t)$ is the derivative of $\pi P(t)$. Thus, we conclude $\pi P(t)$ does not depend on t. In particular, for any $t \geq 0$, we have

$$\pi P(t) = \pi P(0) = \pi.$$

Therefore, π is a stationary distribution.

Remark: Recall that if $\lambda_i = \lambda, i = 0, 1, 2, ...$, then $Q = \lambda(P - I)$, where P is the transition matrix of the jump chain and

$$I = \begin{bmatrix} 1 & 0 & ... & 0 & ... \\ 0 & 1 & ... & 0 & ... \\ \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & ... & 1 & ... \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

In this case, $\pi Q = 0$ iff $\pi P = \pi$, i.e., the stationary distribution of a continuous MC is the same as the stationary distribution of its jump chain.

Limiting Distributions

The probability distribution $\pi = [\pi_0, \pi_1, \pi_2, \cdots]$ is called the **limiting distribution** of the continuous-time Markov chain X_t if

$$\pi_j = \lim_{t\to\infty} P(X_t = j | X_0 = i) = \lim_{t\to\infty} P(X_t = j)$$

for all $i, j \in S$, and we have

$$\sum_{j\in S} \pi_j = 1.$$

As in discrete case, the limiting distribution, if it exists, is a stationary distribution. Indeed, by recalling

- $p_{ij}(s+t) = \sum_{k\in S} p_{ik}(s)\, p_{kj}(t)$, for all $s, t \geq 0, i, j \in S$;
- $P_j(s+t) = \sum_{k\in S} P_k(s)\, p_{kj}(t)$, for all $s, t \geq 0, j \in S$, where $P_j(s) = P(X_s = j)$,

if the limit distribution exists, then

$$\begin{aligned} \pi_j &= \sum_{k\in S} \pi_k\, p_{kj}(\delta), \quad \text{(for any } \delta > 0) \\ &= \pi_j(1 - \lambda_j \delta) + \sum_{k\neq j} \pi_k\, q_{kj}\, \delta + o(\delta), \end{aligned}$$

indicating

$$\sum_{k\neq j} \pi_k\, q_{kj} = \lambda_j \pi_j, \quad j \in S,$$

or $\pi\, Q = 0$.

However, the inverse is not necessarily true and depends on the structure of the chain.

In general, for "nice" chains, there exists a unique stationary distribution which will be equal to the limiting distribution. In theory, we can find the stationary (and limiting) distribution by solving $\pi P(t) = \pi$, or by finding $\lim_{t\to\infty} P(t)$. However, in practice, finding $P(t)$ itself is usually very difficult.

It is easier if we think in terms of the jump (embedded) chain.

Th11.2 Let $\{X_t, t \geq 0\}$ be a continuous-time Markov chain with an irreducible positive recurrent jump chain. Suppose that the unique stationary distribution of the jump chain (with transition matrix P) is given by

$$\tilde{\pi} = \big[\tilde{\pi}_0, \tilde{\pi}_1, \tilde{\pi}_2, \cdots\big],$$

i.e., $\tilde{\pi} = \tilde{\pi}P$. Further assume that

$$0 < \sum_{k\in S} \frac{\tilde{\pi}_k}{\lambda_k} < \infty.$$

Then,

$$\pi_j = \lim_{t\to\infty} P(X_t = j | X_0 = i) = \frac{\frac{\tilde{\pi}_j}{\lambda_j}}{\sum_{k\in S} \frac{\tilde{\pi}_k}{\lambda_k}}.$$

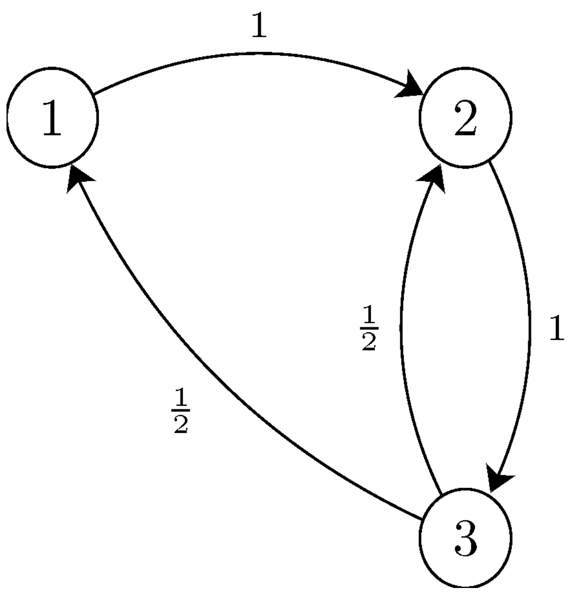
for all $i, j \in S$. That is, $\pi = [\pi_0, \pi_1, \pi_2, \cdots]$ is the limiting distribution of X_t.

Note: In this case, we have $\pi Q = 0$, i.e., the limiting dist and the statioanry dist are the same.

Remark: The following intuitive argument gives us the idea why Theorem 11.2 is true.

Since $\tilde{\pi} = \left[\tilde{\pi}_0, \tilde{\pi}_1, \tilde{\pi}_2, \cdots\right]$ is the limiting distribution of the jump chain, the discrete-time Markov chain associated with the jump chain will spend a fraction $\tilde{\pi}_j$ of time in state j in the long run. Note that, for the corresponding continuous-time Markov chain, any time that the chain visits state j, it spends on average $\frac{1}{\lambda_j}$ time units in that state. Thus, we can obtain the limiting distribution of the continuous-time Markov chain by multiplying each $\tilde{\pi}_j$ by $\frac{1}{\lambda_j}$. We also need to normalize (divide by $\sum \frac{\tilde{\pi}_k}{\lambda_k}$) to get a valid probability distribution.

Example 11.1. Consider a continuous-time Markov chain X_t that has the jump chain shown below. Assume the holding time parameters are given by $\lambda_1 = 2$, $\lambda_2 = 1$, and $\lambda_3 = 3$. Find the limiting distribution of X_t.



Solution: We first note that the jump chain is irreducible. In particular, the transition matrix of the jump chain is given by

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}.$$

The next step is to find the stationary distribution for the jump chain by solving $\tilde{\pi} P = \tilde{\pi}$. We obtain

$$\tilde{\pi} = \frac{1}{5}[1,\ 2,\ 2].$$

Finally, we can obtain the limiting distribution of X_t using

$$\pi_j = \frac{\frac{\tilde{\pi}_j}{\lambda_j}}{\sum_{k \in S} \frac{\tilde{\pi}_k}{\lambda_k}}.$$

We obtain

$$\begin{aligned}\pi_1 &= \frac{\frac{\tilde{\pi}_1}{\lambda_1}}{\frac{\tilde{\pi}_1}{\lambda_1}+\frac{\tilde{\pi}_2}{\lambda_2}+\frac{\tilde{\pi}_3}{\lambda_3}} \\ &= \frac{\frac{1}{2}}{\frac{1}{2}+\frac{2}{1}+\frac{2}{3}} \\ &= \frac{3}{19}.\end{aligned}$$

$$\begin{aligned}\pi_2 &= \frac{\frac{\tilde{\pi}_2}{\lambda_2}}{\frac{\tilde{\pi}_1}{\lambda_1}+\frac{\tilde{\pi}_2}{\lambda_2}+\frac{\tilde{\pi}_3}{\lambda_3}} \\ &= \frac{\frac{2}{1}}{\frac{1}{2}+\frac{2}{1}+\frac{2}{3}} \\ &= \frac{12}{19}.\end{aligned}$$

$$\begin{aligned}\pi_3 &= \frac{\frac{\tilde{\pi}_3}{\lambda_3}}{\frac{\tilde{\pi}_1}{\lambda_1}+\frac{\tilde{\pi}_2}{\lambda_2}+\frac{\tilde{\pi}_3}{\lambda_3}} \\ &= \frac{\frac{2}{3}}{\frac{1}{2}+\frac{2}{1}+\frac{2}{3}} \\ &= \frac{4}{19}.\end{aligned}$$

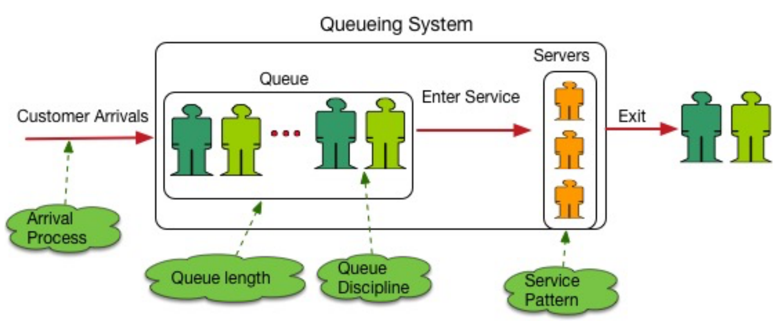
Thus, we conclude that $\pi = \frac{1}{19}[3, 12, 4]$ is the limiting distribution of X_t.

Week 11

Simple Queuing theory (birth and death processes)

The operation of a queueing system is described as follows: given a number of servers, a customer, upon arrival, will receive service immediately if one of the servers is free, otherwise he has to wait by joining a waiting line (a queue) until serviced. The interesting problems in a queue system are

- the length of the queue;
- the waiting time of a customer in the queue;
- the total time spent by a customer in the system;
- others ...



The queue $M/M/k$

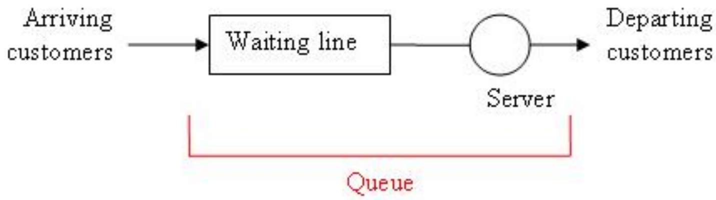
The random components of a queueing system consist of the arriving times and the service times of customers.

- Denote by $T_0 = 0, T_n, n \geq 1$, the successive arrival times of customers.
- Let $E_n, n \geq 1$ be the interarrival times, i.e., $E_n = T_n - T_{n-1}$, and $Y_n, n \geq 1$ be the service time of the n-th customer.
- It is natural to assume that E_n and Y_n are independent and both form sequences of iid random variables with distribution functions F and H, respectively.
- Under the service policy "first come, first serviced", a queueing system is characterized by the distributions F and H, and the number k of servers.
- A queueing system is usually denoted as the triplet $F/H/k$ and queueing systems are classified according to the nature of F, H and k.

$M/M/k$ denotes a queueing system, where both F **and** H **are exponential with parameters** $\lambda > 0$ **and** $\mu > 0$, respectively. (In the case that $k = \infty$, each customer will be served immediately on arrival and there is no queue.)

32 $M/M/1$ queue

This section considers $M/M/1$ queue.



In this system, the arrival process N_t is a Poisson process with rate λ and the service time $Y \sim Exp(\mu)$.

Let L_t denote the number of customers (either waiting or being served) in the $M/M/1$ queue.

- L_t is referred to as **the system length** at the time t.
- $\{L_t, t \geq 0\}$ is a continuous homogenous MC with state space $S = \{0, 1, 2, ...\}$. Explicitly, L_t is a birth and death process with birth rate $b_i = \lambda > 0, i \geq 0$, and death rate $\mu_0 = 0, \mu_i = \mu$, $i \geq 1$, i.e.,

$$
\begin{aligned}
& P(L_{t+h} = k \mid L_t = i) \\
= & \begin{cases} 1 - (\lambda + \mu_i)h + o(h) & \text{for } k = i, \ i \geq 0, \\ \lambda h + o(h) & \text{for } k = i + 1, \ i \geq 0, \\ \mu h + o(h) & \text{for } k = i - 1, \ i \geq 1 \\ o(h) & \text{for } |k - i| \geq 2 \text{ or } i = 0, k = -1, \end{cases}
\end{aligned}
$$

Let B_t denote the number of customers leaving out the system until time t. $\{B_t\}_{t\geq 0}$ is a Poisson process with rate μ.

$$\begin{aligned} P(L_{t+h} = i+1 \mid L_t = i) &= P(\text{one arrival, no departure in } (t, t+h]) \\ &= P(N_{t+h} - N_t = 1, B_{t+h} - B_t = 0) \\ &= \lambda h \, e^{-\lambda h} e^{-\mu h} = \lambda h + o(h), \quad i \geq 0. \end{aligned}$$

$$\begin{aligned} P(L_{t+h} = i-1 \mid L_t = i) &= P(\text{no arrival, one served in } (t, t+h]) \\ &= P(N_{t+h} - N_t = 0, B_{t+h} - B_t = 1) \\ &= e^{-\lambda h} \, \mu h e^{-\mu h} = \mu h + o(h), \quad i \geq 1. \end{aligned}$$

$$\begin{aligned} P(|L_{t+h} - i| \geq 2 \mid L_t = i) &= P(\text{at least 2 arrivals or 2 departures in } (t, t+h]) \\ &= \sum_{k=2}^{\infty} P(N_{t+h} - N_t = k) + \sum_{k=2}^{\infty} P(B_{t+h} - B_t = k) \\ &= \sum_{k=2}^{\infty} \left[(\lambda h)^k \, e^{-\lambda h} + (\mu h)^k e^{-\mu h} \right] = o(h), \end{aligned}$$

$$\begin{aligned} P(L_{t+h} = i \mid L_t = i) &= 1 - \sum_{k=0, k\neq i}^{\infty} P(L_{t+h} = k \mid L_t = i) \\ &= 1 - (\lambda + \mu_i) h + o(h), \quad i \geq 0. \end{aligned}$$

Write $\rho = \lambda/\mu$ (known as *the traffic intensity*) and

$$p_{ij}(t) = P(L_{t+s} = j \mid L_s = i).$$

Th11.3. For all $i, j \geq 0$,

$$\pi_j := \lim_{t\to\infty} P(L_t = j) = \lim_{t\to\infty} p_{ij}(t) = \begin{cases} (1-\rho)\rho^j & \text{if } \rho < 1, \\ 0 & \text{if } \rho \geq 1. \end{cases}$$

Proof. Omitted.

Note: Since $\sum_{j=0}^{\infty} \pi_j = 1$ when $\rho < 1$, the limit distribution of $M/M/1$ queue exists, which is the stationary distribution.

Note that the Q-matrix is given by

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & ... \\ \mu & -(\lambda+\mu) & \lambda & 0 & .. \\ 0 & \mu & -(\lambda+\mu) & \lambda & 0.. \\ . & . & . & & \end{pmatrix}.$$

Now, by using Theorem 11.1, the $\pi = (\pi_0, \pi_1, ...)$ is a stationary distribution iff

$$\pi\, Q = 0,$$

i.e.,

$$\begin{aligned} \lambda\pi_0 &= \mu\pi_1 \\ (\lambda+\mu)\pi_j &= \lambda\pi_{j-1} + \mu\pi_{j+1}, \quad j \geq 1. \end{aligned}$$

This yields that

$$\begin{aligned} \pi_1 &= \rho\,\pi_0 \\ \pi_{j+1} - \pi_j &= \rho(\pi_j - \pi_{j-1}) = ... = \rho^j(\pi_1 - \pi_0), \quad j \geq 1. \end{aligned}$$

Hence, if $\rho \geq 1$, then $\pi_j = 0, j \geq 0$. No stationary distribution exists.

If $\rho < 1$, then

$$\pi_{k+1} = \pi_1 + \sum_{j=1}^{k} \rho^j(\pi_1 - \pi_0) = \rho^{k+1}\pi_0, \quad k \geq 0.$$

Note that $\sum_{k=0}^{\infty} \pi_k = 1$. It follows that $\pi_k = (1-\rho)\rho^k, \quad k \geq 0$.

In this case, stationary distribution is the same as the limit distribution.

Remark: By Th11.3, as expected, if $\lambda < \mu$, then the system length will be stable after a long period of time, whereas if $\lambda > \mu$, with probability one, the system length grows to infinity.

Write $L = L_\infty$. In the long run, the queueing system is governed by the random variable L whose distribution is $\pi = \{\pi(j), j \geq 0\}$, a geometric distribution. Therefore, in the long run, the expected number of customers in the queue $M/M/1$ is

$$EL = \sum_{j=0}^{\infty} j\pi(j) = \lambda/(\mu - \lambda);$$

the expected number of customers waiting in the queue $M/M/1$ is

$$EL^* = \sum_{j=1}^{\infty} (j-1)\pi(j) = \lambda^2/[\mu(\mu - \lambda)],$$

where L^* is the waiting number of customers, i.e.,

$$\begin{aligned} L^* &= \max\{L-1, 0\} = (L-1)I(L \geq 1) \\ EL^* &= EL - P(L \geq 1) = EL - \sum_{k=1}^{\infty} \pi_k. \end{aligned}$$

33 $M/M/1$ queue

Next we assume the queue $M/M/1$ is in a stable state.

Let W and V denote the waiting time and the service time of a customer in the queue $M/M/1$, respectively. Note that

$$W = \begin{cases} 0 & \text{if } L = 0, \\ Y_1^* + Y_2 + ... + Y_L & \text{if } L \geq 1, \end{cases}$$

where Y_1^* is the residual service time of the customer being served and Y_j is the service time for j-th customer in the queue . By the fact that Y_j are iid $Exp(\mu)$ and Y_1^* is also exponential with parameter μ, due to the lack of memory property of the exponential distribution, we easily obtain the following theorem.

Th11.4. For $x \geq 0$,

$$P(W \leq x) = 1 - \rho e^{-\mu(1-\rho)x}, \quad P(W + V \leq x) = 1 - e^{-\mu(1-\rho)x}.$$

Also, we have $E(W) = \rho/[\mu(1-\rho)]$ and $E(W+V) = 1/[\mu(1-\rho)]$.

Note that $W + V$ is the total time spent by a customer in the system. Th11.4 provides solutions of the interesting problems raised in a queueing system.

Proof. Note that L is independent of Y_j and Y_1^*.

$$\begin{aligned}
P(W \le x) &= P(W = 0) + P(0 < W \le x) \\
&= P(L = 0) + \sum_{n=1}^{\infty} P(0 < W \le x \mid L = n) P(L = n) \\
&= 1 - \rho + \sum_{n=1}^{\infty} P(\sum_{k=1}^{n} Y_k \le x)\, P(L = n) \\
&= 1 - \rho + \sum_{n=1}^{\infty} (1-\rho)\rho^n \int_0^x \frac{\mu^n t^{n-1} e^{-\mu t}}{(n-1)!} dt \\
&\quad \text{(see Lecture 23)} \\
&= 1 - \rho + (1-\rho)\rho\mu \int_0^x e^{-(1-\rho)\mu t} dt \\
&= 1 - \rho\, e^{-\mu(1-\rho)x}.
\end{aligned}$$

Also,

$$\begin{aligned}
P(W + V \le x) &= \int_0^x P(W \le x - y) \mu e^{-\mu y} dy \\
&= \ldots = 1 - e^{-\mu(1-\rho)x}.
\end{aligned}$$

Note: It is easy to check **Little's law**:

$$EL = \lambda\, E(W + V), \qquad EL^* = \lambda\, E(W).$$

That is, in the long run,

> the long-term average number L of customers in a stationary system is equal to the long-term average effective arrival rate λ multiplied by the average time that a customer spends in the system.