Stochastic Processes I: Course Outline

Chris Hayduk

December 17, 2020

1 Markov Chains

1.1 Definitions and Examples

We say that X_n is a discrete time **Markov chain** with **transition matrix** $p(i, j)$ if for any $j, i, i_{n-1}, \dots, i_0$

$$P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = p(i, j)$$

We have restricted our attention to the **temporally homogeneous** case in which the **transition probability**

$$p(i, j) = P(X_{n+1} = j \mid X_n = i)$$

does not depend on the time n.

1.2 Multistep Transition Probabilities

Theorem 1.1. The m step transition probability $P(X_{n+m} = j \mid X_n = i)$ is the mth power of the transition matrix p (i.e. $p^m(i, j)$).

1.3 Classification of States

We begin with some important notation. We are often interested in the behavior of the chain for a fixed initial state, so we will introduce the shorthand

$$P_x(A) = P(A \mid X_0 = x)$$

Later we will have to consider expected values for this probability and we will denote them by $\mathbb{E}_x$.

Let $T_y = \min\{n \geq 1 : X_n = y\}$ be the **time of the first return to** y (i.e. being there at time 0 doesn't count), and let

$$\rho_{yy} = P_y(T_y < \infty)$$

be the probability X_n returns to y when it starts at y.

We say that T is a **stopping time** if the occurrence (or nonoccurrence) of the event "we stop at time n," $\{T = n\}$, can be determined by looking at the values of the process up to that time: $X_0, \dots, X_n$. To see that T_y is a stopping time, note that

$$\{T_y = n\} = \{X_1 \neq y, \dots, X_{n-1} \neq y, X_n = y\}$$

and that the right-hand side can be determined from $X_0, \dots, X_n$.

Since stopping at time n depends only on the values $X_0, \dots, X_n$, and in a Markov chain the distribution of the future only depends on the past through the current state, it should not be hard to believe that the Markov property holds at stopping times. This fact can be stated formally as

Theorem 1.2 (Strong Markov Property). Suppose T is a stopping time. Given that $T = n$ and $X_T = y$, any other information about $X_0, \dots, X_T$ is irrelevant for predicting the future, and $X_{T+k}, k \geq 0$ behaves like the Markov chain with initial state y.

Let $T_y^1 = T_y$ and for $k \geq 2$ let

$$T_y^k = \min\{n > T_y^{k-1} : X_n = y\}$$

be the **time of kth return to** y. The strong Markov property implies that the conditional probability we will return one more time given that we have returned $k-1$ times is ρ_{yy}. This and induction imply that

$$P_y(T_y^k < \infty) = \rho_{yy}^k$$

At this point, there are two possibilities:

1. $\rho_{yy} < 1$: The probability of return k times is $\rho_{yy}^k \to 0$ as $k \to \infty$. Thus, eventually the Markov chain does not find its way back to y. In this case the state y is called **transient**, since after some point it is never visited by the Markov chain

2. $\rho_{yy} = 1$: The probability of returning k times $\rho_{yy}^k = 1$, so the chain returns to y infinitely many times. In this case, the state y is called **recurrent**, it continually recurs in the Markov chain.

Lemma 1.3. Suppose $P_x(T_y \leq k) \geq \alpha > 0$ for all x in the state space S. Then

$$P_x(T_y > nk) \leq (1 - \alpha)^n$$

Definition 1.1. We say that x **communicates with** y and write $x \to y$ if there is a positive probability of reaching y starting from x, that is, the probability

$$\rho_{xy} = P_x(T_y < \infty) > 0$$

Lemma 1.4. If $x \to y$ and $y \to z$, then $x \to z$.

Theorem 1.5. If $\rho_{xy} > 0$ but $\rho_{yx} < 1$, then x is transient.

Lemma 1.6. If x is recurrent and $\rho_{xy} > 0$, then $\rho_{yx} = 1$.

A set A is **closed** if it is impossible to get out, i.e., if $i \in A$ and $j \notin A$ then $p(i,j) = 0$. A set B is called **irreducible** if whenever $i, j \in B$, i communicates with j.

Theorem 1.7. If C is a finite, closed, and irreducible set, then all states in C are recurrent.

Theorem 1.8. If the state space S is finite, then S can be written as a disjoint union $T \cup R_1 \cup \cdots \cup R_k$, where T is a set of transient states and R_i, $1 \le i \le k$ are closed irreducible sets of recurrent states.

Lemma 1.9. If x is recurrent and $x \to y$, then y is recurrent.

Lemma 1.10. In a finite closed set there has to be at least one recurrent state.

Let $N(y)$ be the number of visits to y at $n \ge 1$. Then we can compute $\mathbb{E}\, N(y)$:

Lemma 1.11. $\mathbb{E}_x N(y) = \rho_{xy}/(1 - \rho_{yy})$

Lemma 1.12. $\mathbb{E}_x N(y) = \sum_{n=1}^{\infty} p^n(x,y)$

Theorem 1.13. y is recurrent if and only if

$$\sum_{n=1}^{\infty} p^n(y,y) = \mathop{\mathbb{E}}_{y} N(y) = \infty$$

1.4 Stationary Distributions

If $qp = q$, then q is called a **stationary distribution**, If the distribution at time 0 is the same as the distribution at time 1, then by the Markov property it will be the same distribution at all times $n \ge 1$.

Stationary distributions have a special importance in the theory of Markov chains, so we will use a special letter π to denote solutions of the equation

$$\pi p = \pi$$

1.4.1 Doubly Stochastic Chains

Definition 1.2. A transition matrix p is said to be **doubly stochastic** if its columns um to 1, or in symbols, $\sum_x p(x,y) = 1$.

Theorem 1.14. If p is a doubly stochastic transition probability for a Markov chain with N states, then the uniform distribution, $\pi(x) = 1/N$ for all x, is a stationary distribution.

1.5 Detailed Balance Condition

π is said to satisfy the **detailed balance condition** if

$$\pi(x)p(x,y) = \pi(y)p(y,x)$$

This a stronger condition than $\pi p = \pi$.

1.6 Limit Behavior

The **period** of a state is the largest number that will divide all the $n \geq 1$ for which $p^n(x,x) > 0$. That is, it is the greatest common divisor of $I_x = \{n \geq 1 : p^n(x,x) > 0\}$.

Lemma 1.17. If $\rho_{xy} > 0$ and $\rho_{yx} > 0$, then x and y have the same period.

Lemma 1.18. If $p(x,x) > 0$, then x has period 1.

We now come to the main results of the chapter. We first list the assumptions. All of these results hold when S is finite or infinite.

- I: p is irreducible
- A: aperiodic, all states have period 1
- R: all states are recurrent
- S: there is a stationary distribution π

Theorem 1.19 (Convergence Theorem). Suppose I, A, S. Then as $n \to \infty$, $p^n(x,y) \to \pi(y)$

Theorem 1.20 (Asymptotic Frequency). Suppose I and R. If $N_n(y)$ is the number of visits to y up to time n, then

$$\frac{N_n(y)}{n} \to \frac{1}{\mathbb{E}_y T_y}$$

Theorem 1.21. If I and S hold, then

$$\pi(y) = 1/\mathop{\mathbb{E}}_y T_y$$

and hence the stationary distribution is unique.

Theorem 1.22. Suppose I, S, and $\sum_x |f(x)|\pi(x) < \infty$. Then,

$$\frac{1}{n}\sum_{m=1}^{n} f(X_m) \to \sum_x f(x)\pi(x)$$

Theorem 1.23. Suppose I, S.

$$\frac{1}{n}\sum_{m=1}^{n} p^m(x,y) \to \pi(y)$$

Thus while the sequence $p^m(x,y)$ will not converge in the periodic case, the average of the first n values will.

1.7 Returns to a Fixed State

Theorem 1.20. Suppose p is irreducible and recurrent. Let $N_n(y)$ be the number of visits to y at times $\leq n$. As $n \to \infty$,

$$\frac{N_n(y)}{n} \to \frac{1}{\mathbb{E}_y T_y}$$

Theorem 1.21. If p is irreducible and has stationary distribution π, then

$$\pi(y) = 1/\mathop{\mathbb{E}}_{y} T_y$$

.

Theorem 1.24. Suppose p is irreducible and recurrent. Let $x \in S$ and let $T_x = \inf\{n \geq 1 : X_n = x\}$.

$$\mu_x(y) = \sum_{n=0}^{\infty} P_x(X_n = y, T_x > n)$$

defines a stationary measure with $0 < \mu_x(y) < \infty$ for all y.

Theorem 1.22. Suppose p is irreducible and has stationary distribution π, and $\sum_x |f(x)|\pi(x) < \infty$ then

$$\frac{1}{n}\sum_{m=1}^{n} f(X_m) \to \sum_x f(x)\pi(x)$$

The key idea here is that by breaking the path at the return times to x we get a sequence of random variables to which we can apply the law of large numbers.

1.9 Exit Distributions

Theorem 1.28. Consider a Markov chain with state space S. Let A and B be subsets of S, so that $C = S - (A \cup B)$ is finite. Suppose $h(a) = 1$ for $a \in A$, $h(b) = 0$ for $b \in B$ and that for $x \in C$, we have

$$h(x) = \sum_y p(x,y)h(y)$$

If $P_x(V_A \wedge V_B < \infty) > 0$ for all $x \in C$, then $h(x) = P_x(V_a < V_b)$. Note that for a set F, $V_F = \min\{n \geq 0 : X_n \in F\}$ and for a point x, $V_x = \min\{n \geq 0 : X_n = x\}$

1.10 Exit Times

Theorem 1.29. Let $V_A = \inf\{n \geq 0 : X_n \in A\}$. Suppose $C = S - A$ is finite and that $P_x(V_A < \infty) > 0$ for any $x \in C$. If $g(a) = 0$ for all $a \in A$, and for $x \in C$ we have

$$g(x) = 1 + \sum_y p(x, y)g(y)$$

Then $g(x) = \mathbb{E}_x(V_A)$.

1.11 Infinite State Spaces

In this section we consider chains with an infinite state space. The major new complication is that recurrence is not enough to guarantee the existence of a stationary distribution.

Note: x is said to be **positive recurrent** if $\mathbb{E}_x T_x < \infty$. If a state is recurrent but not positive recurrent, i.e., $P_x(T_x < \infty) = 1$ but $\mathbb{E}_x T_x = \infty$, then we say that x is **null recurrent**.

Theorem 1.30. For an irreducible chain the following are equivalent:

1. Some state is positive recurrent.
2. There is a stationary distribution π
3. All states are positive recurrent

2 Poisson Processes

2.1 Exponential Distribution

A random variable T is said to have **an exponential distribution with rate** λ, or $T =$ exponential(λ) if

$$P(T \leq t) = 1 - e^{-\lambda t} \text{ for all } t \geq 0$$

Here we have described the distribution b giving the **distribution function** $F(t) = P(T \leq t)$. We can also write the definition in terms of the **density function** $f_T(t)$ which is the derivative of the distribution function.

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

We have $\mathbb{E}\, T = 1/\lambda$ and $\mathrm{var}(T) = 1/\lambda^2$

Lack of Memory Property It is traditional to formulate this property in terms of waiting for an unreliable bus driver. In words "if we've been waiting for t units of time then the probability we must wait s more units of time is the same as if we haven't waited at all." In symbols

$$P(T > t + s \mid T > t) = P(T > s)$$

Theorem 2.1. Let $V = \min(T_1, \ldots, T_n)$ and I be the (random) index of the T_i that is smallest.

$$P(V > t) = \exp(-(\lambda_1 + \cdots + \lambda_n)t)$$
$$P(I = i) = \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_n}$$

I and $V = \min\{T_1, ldots, T_n\}$ are independent.

Theorem 2.2. Let $\tau_1, \tau_2, \ldots$ be independent exponential(λ). The sum $T_n = \tau_1 + \cdots + \tau_n$ has a gamma(n, λ) distribution. That is, the density function of T_n is given by

$$f_{T_n}(t) = \lambda e^{-\lambda t} \cdot \frac{(\lambda t)^{n-1}}{(n-1)!} \text{ for } t \geq 0$$

and 0 otherwise.

2.2 Defining the Poisson Process

Definition. We say that X has a **Poisson distribution** with mean λ, or $X =$ Poisson(λ), for short, if

$$P(X = n) = e^{-\lambda} \frac{\lambda^n}{n!} \text{ for } n = 0, 1, 2, \ldots$$

Theorem 2.3. For any $k \geq 1$,

$$\mathbb{E}\, X(X-1)\cdots(X-k+1) = \lambda^k$$

and hence $\mathrm{var}(X) = \lambda$.

Theorem 2.4. If X_i are independent Poisson(λ_i), then

$$X_1 + \cdots + X_k = \text{Poisson}(\lambda_1 + \cdots \lambda_n)$$

Let $N(s)$ be the number of arrivals in $[0, s]$

Definition. $\{N(s), s \geq 0\}$ is a Poisson process if

1. $N(0) = 0$
2. $N(t+s) - N(s) = \text{Poisson}(\lambda t)$, and
3. $N(t)$ has **independent increments**, i.e., if $t_0 < t_1 < \cdots < t_n$, then

$$N(t_1) - N(t_0), \ldots, N(t_n) - N(t_{n-1}) \text{ are independent}$$

Theorem 2.5. If n is large, the binomial($n, \lambda/n$) distribution is approximately Poisson(λ).

2.2.1 Constructing the Poisson Process

Definition. Let $\tau_1, \tau_2, \ldots$ be independent exponential(λ) random variables. Let $T_n = \tau_1 + \cdots + \tau_n$ for $n \geq 1$, $T_0 = 0$ and define $N(s) = \max\{n : T_n \leq s\}$.

We think of the τ_n as times between arrivals of customers at the ATM, so $T_n = \tau_1 + \cdots + \tau_n$ is the arrival time of the nth customer and $N(s)$ is the number of arrivals by time s.

Lemma 2.6. $N(s)$ has a Poisson distribution with mean λs.

Lemma 2.7. $N(t+s) - N(s), t \geq 0$ is a rate λ Poisson process and independent of $N(r), 0 \leq r \leq s$

Lemma 2.8. $N(t)$ has independent increments.

2.2.2 More Realistic Models

Theorem 2.9. Let $X_{n,m}, 1 \leq m \leq n$ be independent random variables with $P(X_m = 1) = p_m$ and $P(X_m = 0) = 1 - p_m$. Let

$$S_n = X_1 + \cdots + X_n,\ \lambda_n = \mathbb{E}\, S_n = p_1 + \cdots p_n$$

and $Z_n = \text{Poisson}(\lambda_n)$. Then for any set A,

$$|P(S_n \in A) - P(Z_n \in A)| \leq \sum_{m=1}^{n} p_m^2$$

Nonhomogenous Poisson Processes We say that $\{N(s), s \geq 0\}$ is a Poisson process with rate $\lambda(r)$ if

1. $N(0) = 0$
2. $N(t)$ has independent increments
3. $N(t) - N(s)$ is Poisson with mean $\int_s^t \lambda(r)dr$

In this case, the interarrival times are not exponential and they are not independent.

2.3 Compound Poisson Process

In this section we will embellish our Poisson process by associating an independent and identically distributed (i.i.d) random variable Y_i with each arrival. By independent, we mean that the Y_i are independent of each other and of the Poisson process of arrivals.

Theorem 2.10. Let $Y_1, Y_2, \cdots$ be independent and identically distributed, let N be an independent nonnegative integer valued random variable, and let $S = Y_1 + \cdots + Y_N$ with $S = 0$ when $N = 0$.

1. If $\mathbb{E}\,|Y_i|, \mathbb{E}\,N < \infty$, then $\mathbb{E}\,S = \mathbb{E}\,N \cdot \mathbb{E}\,Y_i$
2. If $\mathbb{E}\,Y_i^s, \mathbb{E}\,N^2 < \infty$, then $\text{var}(S) = \mathbb{E}\,N \text{var}(Y_i) + \text{var}(N)(\mathbb{E}\,Y_i)^2$
3. If N is Poisson(λ), then $\text{var}(S) = \lambda\,\mathbb{E}\,Y_i^2$

2.4 Transformations

2.4.1 Thinning

In the previous section, we added up the Y_i's associated with the arrivals in our Poisson process to see how many customers, etc., we had accumulated by time t. In this section, we will use the Y_i to split the Poisson process into several. Let $N_j(t)$ be the number $i \leq N(t)$ with $Y_i = j$. In Example 2.3, where Y_i is the number of people in the ith car, $N_j(t)$ will be the number of cars that have arrived by time t with exactly j people. The somewhat remarkable fact is

Theorem 2.11. $N_j(t)$ are independent rate $\lambda P(Y_i = j)$ Poisson processes.

Why is this remarkable? There are two "surprises" here: the resulting processes are Poisson and they are independent.

Theorem 2.12. Suppose that in a Poisson process with rate λ, we keep a point that lands at s with probability $p(s)$. Then the result is a nonhomogeneous Poisson process with rate $\lambda p(s)$.

Theorem 2.13. In the long run the number of calls in the system will be Poisson with mean

$$\lambda \int_{r=0}^{\infty} (1 - G(r))dr = \lambda\mu$$

2.4.2 Superposition

Taking one Poisson process and splitting it into two or more by using an i.i.d. sequence Y_i is called **thinning**. Going in the other direction and adding up a lot of independent processes is called **superposition**. Since a Poisson process can be split into independent Poisson processes, it should not be too surprising that when the independent Poisson processes are put together, the sum is Poisson with a rate equal to the sum of the rate.

Theorem 2.14. Suppose $N_1(t), \dots, N_k(t)$ are independent Poisson processes with rates $\lambda_1, \dots, \lambda_k$, then $N_1(t) + \cdots + N_k(t)$ is a Poisson process with rate $\lambda_1 + \cdots \lambda_k$.

2.4.3 Conditioning

Let $T_1, T_2, T_3, \dots$ be the arrival times of a Poisson process with rate λ, let $U_1, U_2, \dots, U_n$ be independent and uniformly distributed on $[0, t]$ and let $V_1 < \cdots < V_n$ be the U_i rearranged into increasing order. This section is devoted to the proof of the following remarkable fact.

Theorem 2.15. If we condition on $N(t) = n<$ then the vector $(T_1, T_2, \dots T_n)$ has the same distribution as $(V_1, V_2, \dots, V_n)$ and hence the set of arrival times $\{T_1, T_2, \dots, T_n\}$ has the same distribution as $\{U_1, U_2, \dots, U_n\}$

Theorem 2.16. If $s < t$ and $0 \leq m \leq n$, then

$$P(N(s) = m \mid N(t) = n) = \binom{n}{m} \left(\frac{s}{t}\right)^m \left(1 - \frac{s}{t}\right)^{n-m}$$

That is, the conditional distribution of $N(s)$ given $N(t) = n$ is binomial$(n, s/t)$

3 Renewal Processes

3.1 Law of Large Numbers

Theorem 3.1. Let $\mu = \mathbb{E}\, t_i$ be mean interarrival time. If $P(t_i > 0) > 0$, then with probability one,

$$N(t)/t \to 1/\mu \text{ as } t \to \infty$$

In words, if our light bulb lasts μ years on the average then in t years we will use up about t/μ light bulbs. Since the interarrival times in a Poisson process are exponential with mean $1/\lambda$, Theorem 3.1 implies that if $N(t)$ is the number of arrivals up to time t in a Poisson process, then

$$N(t)/t \to \lambda \text{ as } t \to \infty$$

Theorem 3.2 (Strong Law of Large Numbers). Let $x_1, x_2, x_3, \dots$ be i.i.d. with $\mathbb{E}\, x_i = \mu$,, and let $S_n = x_1 + \cdots + x_n$. Then with probability one,

$$S_n/n \to \mu \text{ as } n \to \infty$$

Now suppose at the time of the ith renewal we earn a reward r_i. Let

$$R(t) = \sum_{i=1}^{N(t)} r_i$$

be the total amount of rewards earned by time t. The main result about renewal reward processes is the following strong law of large numbers:

Theorem 3.3. With probability one,

$$\frac{R(t)}{t} \to \frac{\mathbb{E}\, r_i}{\mathbb{E}\, t_i}$$

Now we can move onto an extension of the renewal reward process: **alternating renewal reward processes**.

Example 3.4 (Alternating Renewal Processes). Let $s_1, s_2, \dots$ be independent with a distribution F that has mean μ_F, and let $u_1, u_2, \dots$ be independent with distribution G that has mean μ_G. For a concrete example consider the machine in Example 3.2 that works for amount of time s_i before needing a repair that takes u_i units of time. However, to talk about things in general we will say that the alternating renewal process spends an amount of time s_i in state 1, an amount of time u_i in state 2, and then repeats the cycle again.

Theorem 3.4. In an alternating renewal process, the limiting fraction of time in state 1 is

$$\frac{\mu_F}{\mu_F + \mu_G}$$

3.2 Applications to Queuing Theory

In this section we will use the ideas of renewal theory to prove results for queuing systems with general service times. In the first part of this section we will consider general arrival times. In the second we will specialize to Poisson arrivals.

3.2.1 GI/G/1 Queue

Here the GI stands for general input. That is, we suppose that the times t_i between successive arrivals are independent and have a distribution F with mean $1/\lambda$. We make this somewhat unusual choice of notation for mean so that if $N(t)$ is the number of arrivals by time t, then Theorem 3.1 implies that the long-run arrival rate is

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mathbb{E}\, t_i} = \lambda$$

The second G stands for general service times. That is, we assume that the ith customer requires an amount of service s_i, where the s_i are independent and have a distribution G with mean $1/\mu$. Again, the notations for the mean is chosen so that the service rate is μ. The final 1 indicates there is one server. Our final result states that the queue is stable if the arrival rate is smaller than the long-run service rate.

Theorem 3.5. Suppose $\lambda < \mu$. If the queue starts with some finite number $k \geq 1$ customers who need service, then it will empty out with probability one. Furthermore, the limiting fraction of time the server is busy is $\leq \lambda/\mu$.

3.2.2 Cost Equations

Let X_s be the number of customers in the system at time s. Let L be the long-run average number of customers in the system:

$$L = \lim_{t\to\infty} \frac{1}{t} \int_0^t X_s ds$$

Let W be the long-run average amount of time a customer spends in the system:

$$W = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} W_m$$

where W_m is the amount of time the mth arriving customer spends in the system. Finally, let λ_a be the long-run average rate at which arriving customers join the system, that is,

$$\lambda_a = \lim_{t\to\infty} N_a(t)/t$$

where $N_a(t)$ is the number of customers who arrive before time t and enter the system. Ignoring the problem of proving these limits, we can assert that these quantities are related by,

Theorem 3.6 (Little's Formula) $L = \lambda_a W$

3.2.3 M/G/1 Queue

Here the M stands for Markovian input and indicates that we are considering the special case of the $GI/G/1$ queue in which the inputs are a rate λ Poisson process. The rest of the setup is as before: there is one server and the ith customer requires an amount of service s_i, where the s_i are independent and have a distribution G with mean $1/\mu$.

Let X_n be the number of customers in the queue when the nh customer enters service. To be precise, when $X_0 = x$, the chain starts with x people waiting in line and customer) just beginning her service. To understand the definition, the following picture is useful:

To begin to define our Markov chain X_n, let

$$a_k = \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^k}{k!} dG(t)$$

be the probability that k customers arrive during a service time.

$$\sum_k k a_k = \int_0^\infty \lambda t dG(t) = \lambda \mathbb{E}\, s_i = \lambda/\mu$$

To construct the chain now let $\xi_1, \xi_2, \ldots$ be i.i.d. with $P(\xi_i = k) = a_k$. We think of ξ_i as the number of customers to arrive during the ith service time. If $X_n > 0$ then

$$X_{n+1} = X_n + \xi_n - 1$$

In words, ξ_n customers are added to the queue and then it shrinks by one when the nth customer departs. If $X_n = 0$ and $\xi_n > 0$, then the same logic applies but if $X_n = 0$ and $\xi_n = 0$ then $X_{n+1} = 0$. To fix this we write

$$X_{n+1} = (X_n + \xi_n - 1)^+$$

where $z^+ = \max\{z, 0\}$ is the positive part.

From this formula, we see that the transition probability is

$$\begin{array}{c|c} & \begin{array}{cccccccc} 0 & 1 & 2 & 3 & 4 & 5 & \cdots \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} a_0 + a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & \\ a_0 & a_1 & a_2 & a_3 & a_4 & a_5 & \\ 0 & a_0 & a_1 & a_2 & a_3 & a_4 & \\ 0 & 0 & a_0 & a_1 & a_2 & a_3 & \\ 0 & 0 & 0 & a_0 & a_1 & a_2 & \end{pmatrix} \end{array}$$

To explain the value of $p(0,0)$, we note that if $X_n = 0$ and $\xi_n = 0$ or 1 then $X_{n+1} = 0$. In all other cases, if $X_{n+1} = X_n = \xi_n - 1$, so, for example, $p(2,4) = P(\xi_n = 3)$.

Our result for the $GI/G/I$ queue implies $1/\mathbb{E}_0 T_0 = \pi(0) = 1 - \lambda/\mu$ so we have proved the first part of the following:

Theorem 3.7.

1. If $\lambda < \mu$, then X_n is positive recurrent and $\mathbb{E}_0 T_0 = \mu/(\mu - \lambda)$
2. If $\lambda = \mu$, then X_n is null recurrent
3. If $\lambda > \mu$, then X_n is transient

Busy Periods We learned in Theorem 3.7 that if $\lambda < \mu$ then an $M/G/1$ queue will repeatedly return to an empty state. Thus the server experiences alternating busy periods with duration B_n and idle periods with duration I_n. The lack of memory property implies that I_n has an exponential distribution with rate λ. Combining this observation with our result for alternating renewal processes, we see that the limiting fraction of time the server is idle is

$$\frac{1/\lambda}{1/\lambda + \mathbb{E}\, B_n} = \pi(0)$$

After some simplification, we have,

$$EB_n = \frac{1}{\mu - \lambda}$$

PASTA These initials stand for "Poisson arrivals see time averages." To be precise, if $\pi(n)$ is the limiting fraction of time that there are n individuals in the queue and α_n is the limiting fraction of arriving customers that see a queue of size n, then

Theorem 3.8. $\alpha_n = \pi(n)$

Theorem 3.9 (Pollaczek-Khintchine Formula). The long run average waiting time in the queue:

$$W_Q = \frac{\lambda\, \mathbb{E}(s_i^2/2)}{1 - \lambda\, \mathbb{E}\, s_i}$$

4 Continuous Time Markov Chains

4.1 Definitions and Examples

In continuous time, it is technically difficult to define the conditional probability given all of the X_r for $r \leq s$, so we instead say that X_t,, $t \geq 0$ is a Markov chain if for any $0 \leq s_0 < s_1 < \cdots < s_n < s$ and possible states $i_0, \ldots, i_n, i, j$, we have,

$$P(X_{t+s} = j \mid X_s = i, X_{s_n} = i_n, \ldots, X_{s_0} = i_0) = P(X_t = j \mid X_0 = i)$$

In words, given the present state, the rest of the past is irrelevant for predicting the future. Note that built into the definition is the fact that the probability of going from i at time s to j at time $s+t$ only depends on t, the difference in times.

In continuous time there is no first time $t > 0$, so we introduce for each $t > 0$ a **transition probability:**

$$p_t(i,j) = P(X_t = j \mid X_0 = j)$$

In continuous time, as in discrete time, the transition probability satisfies,

Theorem 4.1 (Chapman-Kolmogorov Equation).

$$\sum_k p_s(i,k) p_t(k,j) = p_{s+t}(i,j)$$

Formal Construction. Suppose, for simplicity, that $\lambda_i > 0$ for all i. Let Y_n be a Markov chain with transition probability $r(i,j)$. The discrete time chain Y_n gives the road map that the continuous time process will follow. To determine how long the process should stay in each state let $\tau_0, \tau_1, \tau_2, \ldots$ be independent exponentials with rate 1.

At time 0, the process is in state Y_0 and should stay there for an amount of time that is expressed with rate $\lambda(Y_0)$, so we let the time the process stay in state Y_0 be $t_1 = \tau_0/\lambda(Y_0)$.

At time $T_1 = t_1$ the process jumps to Y_1, where it should stay for an exponential amount of time with rate $\lambda(Y_1)$, so we let the time time process stays in state Y_1 be $t_2 = \tau_1/\lambda(Y_1)$.

At time $T_2 = t_1 + t_2$ the process jumps to Y_2, where it should stay for an exponential amount of time with rate $\lambda(Y_2)$, so we let the time the process stays in state Y_2 be $t_3 = \tau_2/\lambda(Y_2)$.

Continuing in the obvious way, we can let the amount of time the process stays in Y_n be $t_{n+1} = \tau_n/\lambda(Y_n)$, so that the process jumps to Y_{n+1} at time

$$T_{n+1} = t_1 + \cdots + t_{n+1}$$

In symbols, if we let $T_0 = 0$, then for $n \geq 0$, we have

$$X(t) = Y_n \text{ for } T_n \leq t \leq T_{n+1}$$

4.2 Computing the Transition Probability

In the last section we saw that given jump rates $q(i,j)$ we can construct a Markov chain that has these jump rates. This chain, of course, has a transition probability

$$p_t(i,j) = P(X_t = j \mid X_0 = i)$$

Our next question is: How do you compute the transition probability p_t from the jump rates 1? By the following,

$$p_t'(i,j) = \sum_{k \neq i} q(i,k) p_t(k,j) - \lambda_i p_t(i,j)$$

To simplify the last expression we introduce a new matrix:

$$Q(i,j) = \begin{cases} q(i,j) & j \neq i \\ -\lambda_i & j = i \end{cases}$$

For future computations note that the off-diagonal elements $q(i,j)$ with $i \neq j$ are non-negative while the diagonal entry is a negative number chosen to make the row sum equal to 0. Using matrix notation, we can rewrite the equation for $p_t'(i,j)$ as

$$p_t' = Q p_t$$

This is **Kolmogorov's backward equation**. If Q were a number instead of a matrix, the last equation would be easy to solve. We would set $p_t = e^{Qt}$ and check by differentiating that the equation held. Inspired by this observation, we define the matrix

$$e^{Qt} - \sum_{n=0}^{\infty} \frac{(Qt)^n}{n!} = \sum_{n=0}^{\infty} Q^n \cdot \frac{t^n}{n!}$$

Kolmogorov's Forward Equation. This time we split $[0, t+h]$ into $[0,t]$ and $[0, t+h]$ rather than into $[0,h]$ and $[h, t+h]$.

$$\begin{aligned} p_{t+h}(i,j) - p_t(i,j) &= \left(\sum_k p_t(i,k) p_h(k,j) \right) - p_t(i,j) \\ &= \left(\sum_{k \neq j} p_t(i,k) p_h(k,j) \right) + [p_h(j,j) - 1] p_t(i,j) \end{aligned}$$

Computing as before we arrive at

$$p_t'(i,j) = \sum_{k \neq j} p_t(i,k) q(k,j) - p_t(i,j) \lambda_j$$

Introducing matrix notation again, we can write

$$p_t' = p_t Q$$

So we see that $Q p_t = p_t' = p_t Q$.

4.3 Limiting Behavior

The Markov chain X_t is irreducible if for any two states i and j it is possible to get from i to j in a finite number of steps. To be precise, there is a sequence of states $k_0 = i, k_1, \ldots, k_n = j$ so that $q(k_{m-1}, k_m) > 0$ for $1 \leq m \leq n$.

Lemma 4.6. If X_t is irreducible and $t > 0$, then $p_t(i, j) > 0$ for all i, j.

Lemma 4.7. π is a stationary distribution if and only if $\pi Q = 0$.

Why is this true? Filling in the definition of Q and rearranging, the condition $\pi Q = 0$ becomes

$$\sum_{k \neq j} \pi(k) q(k, j) = \pi(j) \lambda_j$$

If we think of $\pi(k)$ as the amount of sand at k, the right-hand side represents the rate at which sand leaves j, while the left gives the rate at which sand arrives at j. Thus, π will be a stationary distribution if for each j the flow of sand into j is equal to the flow out of j.

Lemma 4.6 implies that for any $h > 0$, p_h is irreducible and aperiodic, so by Theorem 1.19

$$\lim_{n \to infty} p_{nh}(i, j) = \pi(j)$$

From this, it is intuitively clear that

Theorem 4.8. If a continuous time Markov chain X_t is irreducible and has a stationary distribution π, them

$$\lim_{t \to \infty} p_t(i, j) = \pi(j)$$

4.3.1 Detailed Balance

Generalizing from discrete time we can formulate this condition as:

$$\pi(k) q(k, j) = \pi(j) q(j, k) \text{ for all } j \neq k$$

The reason for interest in this concept is

Theorem 4.9. If the above equation holds, then π is a stationary distribution.