Math 4232 Stochastic Processes II: Markov chains

Summary

Simple random walk is a random process which takes values on the vertices of a graph. A *graph* G is a set of vertices V and edges E, where each edge is an element of $V \times V$. We assume all graphs are *simple*, which means that $(a, a) \notin E$ and E has no repeated elements, and *undirected*, which means that (a, b) and (b, a) represent the same edge. The *degree* of a vertex of a graph is the number of edges containing that vertex. A random walk is a process in which a walker moves on the vertices of a graph, at each stage moving to the adjacent vertices with probability $1/d$, where d is the degree of the current vertex. Formally, simple random walk is a random process X_n with independent increments on the vertices of the graph, with conditional probabilities

$$(1) \qquad P(X_{n+1} = b | X_n = a) = \begin{cases} \frac{1}{\deg(a)} & \text{if } a \sim b \\ 0 & \text{if } a \nsim b \, . \end{cases}$$

Several of the most interesting questions which come up in connection with this process are

- Let V_1, V_0 be subsets of the vertex set V. Starting from point a, what is the probability that we hit set V_1 before set V_0?
- Does the random walk approach some stable distribution as we let it go forever?
- On an infinite graph, is there a nonzero probability that the random walk will never return to its starting point?

The final question carries a very interesting result. First some notation: For $a \in G$, the expressions E_a and P_a refer respectively to the expectation and probability under the condition that $X_0 = a$ a.s. We will say that X_n is *recurrent* if $P_a(X_n = a \text{ for some } n \geq 1) = 1$, and *transient* if $P_a(X_n = a \text{ for some } n \geq 1) < 1$. We will consider the integer lattice in n dimensions. This is the infinite graph whose vertices are the set of elements of $\mathbb{Z}^n$, with the edge set defined by $\{a_1, \ldots, a_n\} \sim \{b_1, \ldots, b_n\}$ if there exists j such that $|a_j - b_j| = 1$ and $a_i = b_i$ for $i \neq j$. Then we have Pólya's Theorem:

Theorem 1 *Simple random walk is recurrent on* $\mathbb{Z}^1$,$\mathbb{Z}^2$, *and transient on* $\mathbb{Z}^n$ *for* $n \geq 3$.

There are a number of different proofs of this result. Several proceed by letting m be the expected number of times that $X_n = a$, and starting with

$$(2) \qquad m = \sum_{n=1}^{\infty} P_a(X_n = a).$$

It can then also be shown that if $u = P_a(X_n = a \text{ for some } n \geq 1)$, then $m = \frac{1}{1-u}$. Combinatorial arguments can then be used to show that the sum in (2) is ∞ if $n \leq 2$ and finite if $n \geq 3$. Another proof uses the concept of *electric resistance*. Suppose that a graph is taken to represent an electric circuit, where each edge has unit resistance. We imagine that we attach one pole of a battery to a vertex z_0, and the other pole to another vertex z_1, so that z_0 is at voltage 0 and z_1 is at voltage 1. All other points z receive a voltage $V(z)$, which can be calculated using Ohm's Law and Kirchhoff's Current Law.

Ohm's Law: Voltage is equal to current times resistance.

$$V = IR \tag{3}$$

Kirchhoff's Current Law: The sum of the currents entering and leaving any point other than v_0 and v_1 is 0.

As a consequence of Ohm's and Kirchhoff's Laws, the voltage function on G is *harmonic.* That is, for $v \neq v_0, v_1$, We have

$$V(z) = \frac{1}{\deg(z)} \sum_{x \sim z} V(x) \tag{4}$$

This is connected to simple random walk by the following observation. Let $g(z) = P_z(v_1 \text{before} v_0)$ denote the probability that a random walk, started at z, strikes v_1 before hitting v_0. Random walk has no memory, so

$$P_z(v_1 \text{ before } v_0) = \frac{1}{d} P_{x_1}(z_1 \text{ before } z_0) + \ldots + \frac{1}{d} P_{x_d}(z_1 \text{ before } z_0) \tag{5}$$

where $x_1, \ldots, x_d$ are the points adjacent to z. This is the same definition as before, so $g(z)$ is a harmonic function as well on $G - \{z_0, z_1\}$. Given boundary values and a finite graph, there is exactly one possible harmonic function, so we see that $g(z)$ is equal to $V(z)$. Due to Ohm's Law, $V = IR$, the amount of current flowing between two adjacent vertices is given by the difference in voltage. We can therefore measure the amount of current flowing from z_0 to z_1 by summing the voltages of vertices adjacent to z_0. The reciprocal of the amount of current flowing is called the *effective resistance*, and is a metric on the graph. An important consequence of the definitions is

Rayleigh's monotonicity law: If the resistances of one or more edges in a graph are increased, the resistance in the new graph between any pair of points must be at least the resistance between the pair in the old graph.

Of course, this also implies the reverse, that if the resistances of one or more edges in a graph are decreased, the resistance in the new graph between any pair of points must be at most the resistance between the pair in the old graph. Using these concepts, we can determine whether a random walk on an infinite graph is recurrent or transient by calculating the resistance from any point "to infinity". That is, we calculate the resistance between the point and a distant set F_n, then let F_n go to infinity. If this resistance remains finite, the walk is transient. If the resistance becomes infinite, the walk is recurrent. This method can be used to give a neat proof of Pólya's Theorem.

Markov chains are generalizations of simple random walk. Let I be a countable set, with $|I| = n \in (0, \infty]$. For matrices we will write $\{p_{ij}\}$ for the more standard $\{p_{ij}\}_{i,j=1}^{n}$. We let $P = \{p_{i,j}\}$ be a stochastic matrix, that is, an $n \times n$ matrix such that $\sum_{j=1}^{n} p_{ij} = 1$ for all i. We also let λ be a distribution on the elements of I, that is, a non-negative function in i such that $\sum_{i \in I} \lambda_i = 1$. Then a sequence of random variables taking values on I is a *Markov chain* with *initial distribution* λ and *transition matrix* P if

(i) $X_0 \sim \lambda$, and

(ii) Conditioned on $X_n = i$, X_{n+1} has distribution $(p_{ij} : j \in I)$ and is independent of $X_0, X_1, \ldots, X_n$. Put another way,

$$P(X_{n+1} = j | X_0 = i_0, \ldots, X_n = i_n) = P(X_{n+1} = j | X_n = i_n) = p_{i_n j}. \tag{6}$$

We will write $X_n \sim Markov(\lambda, P)$ in this case. Note that n-step transition probabilities are given by $P^n = \{p_{ij}^{(n)}\}$.

Proposition 1 *$X_n \sim Markov(\lambda, P)$ if, and only if,*

$$P(X_0 = i_0, \ldots, X_n = i_n) = \lambda_{i_0} p_{i_0 i_1} \ldots p_{i_{n-1} i_n} \tag{7}$$

We will say i *leads to* j and write $i \longrightarrow j$ if $P_i(X_n = j \text{ for some } n \geq 0) > 0$. If $i \longrightarrow j$ and $j \longrightarrow i$ we will say i and j *communicate* and write $i \longleftrightarrow j$. The relation $\longleftrightarrow$ is clearly an equivalence relation (that is, $i \longleftrightarrow i$, if $i \longleftrightarrow j$ then $j \longleftrightarrow i$, and if $i \longleftrightarrow j$ and $j \longleftrightarrow k$ then $i \longleftrightarrow k$), so it partitions I into equivalence classes, which we call *communicating classes*. If $i \longleftrightarrow j$ for all $i, j \in I$, we say X_n is *irreducible*. If C is a communicating class such that, if $i \in C$, then $i \longrightarrow j$ implies $j \in C$, then we say that C is a *closed class*. If a singleton class $\{i\}$ is closed, we say that i is *absorbing*.

If $i \in I$ and $A \subseteq I$, we can ask for $h_i^A = P_i(X_n \in A \text{ for some } n \geq 0)$.

Theorem 2 *As a function of i, h_i^A is the minimal non-negative solution to*

(8) $$h_i^A = \begin{cases} 1 & \text{if } i \in A \\ \sum_{j \in I} p_{ij} h_j^A & \text{if } i \notin A\ . \end{cases}$$

Let us define things formally. A *probability space* is a triple $(\Omega, \mathcal{F}, P)$: Ω is any set whatsoever; $\mathcal{F}$ is a σ-field, that is, a collection of subsets of Ω, containing Ω itself, and which is closed under complements, finite intersections, and countable unions; P is a probability measure, that is, a measure such that $P(\Omega) = 1$. To say that X is a *random variable* on a probability space $(\Omega, \mathcal{F}, P)$ means that $X = X(\omega)$ is a function on Ω, with $X^{-1}(A) \in \mathcal{F}$ for all Borel sets A. We will also say X is $\mathcal{F}$ measurable. A *discrete time stochastic process* is a sequence of r.v.'s $X_1, X_2, X_3, \ldots$ and an increasing collection of σ-fields $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_3 \ldots$ so that each X_n is $\mathcal{F}_n$-measurable. The collection $\{\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3, \ldots\}$ is called the *filtration* of the process, and in some sense represents the information available to an observer at any time. Generally, we take $\mathcal{F}_n$ to be the σ-field generated by $X_1, \ldots, X_n$, which we denote $\sigma(\{X_i, 1 \le i \le n\})$. A *stopping time* τ is a r.v. taking values in the nonnegative integers and ∞ such that $\{\tau \le n\} \in \mathcal{F}_n$; or, equivalently, $\{\tau = n\} \in \mathcal{F}_n$. Stopping times are very important in the subject of Markov chains. With τ a stopping time, let $\mathcal{F}_\tau = \{A \in \mathcal{F}_\infty : A \cap \{\tau = n\} \in \mathcal{F}_n \text{ for all } n\}$. $\mathcal{F}_\tau$ is essentially all knowledge of the process up to time τ. One reason for the great importance of stopping times is the *strong Markov property* for Markov chains.

Theorem 3 (Strong Markov property of Markov chains) *Suppose $X_n \sim Markov(\lambda, P)$, and let τ be a stopping time for X_n. Then, conditioned on $\tau < \infty$ and $X_\tau = i$, $Y_n = X_{\tau+n}$ is $Markov(\delta_i, P)$, where δ_i is the distribution putting mass 1 on the point i. Furthermore, Y_n is independent of $\mathcal{F}_\tau$.*

Adapting the definitions from simple random walk, we will say that X_n is *recurrent* if $P_a(X_n = a$ for infinitely many $n) = 1$, and *transient* if $P_a(X_n = a$ for infinitely many $n) = 0$.

Theorem 4 *For each $i \in I$, i is either recurrent or transient. In particular, if we let $T_i = \inf\{n \ge 1 : X_n = i\}$, then*

- *if $P_i(T_i < \infty) = 1$, then i is recurrent and $\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$.*
- *if $P_i(T_i < \infty) < 1$, then i is transient, $\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty$, and the expected number of visits the chain makes to i when started at i is $(1 - P_i(T_i < \infty))^{-1}$.*

Proposition 2 *If $i \longleftrightarrow j$, then i and j are either both recurrent or both transient.*

Thus, for irreducible chains, we can talk about whether the chain is recurrent or transient, since all states will have the same behavior in this regard.

Proposition 3 *If X_n is irreducible and recurrent, then $P_i(T_j < \infty) = 1$ for all $i, j \in I$.*

An n-component vector λ is called *invariant* if $\lambda P = \lambda$. If λ itself is a distribution on I and is invariant, we will say λ is an *invariant distribution* or *stationary distribution*. A Markov chain may have no invariant distribution, or it may have many. However, if we set $\gamma_i^k = E_k[\sum_{n=0}^{T_k - 1} 1_{X_n = i}]$, we have

Theorem 5 *Let X_n be irreducible and recurrent. Then*

(i) $\gamma_k^k = 1$.

(ii) $0 < \gamma_i^k < \infty$ *for all* i, k.

(iii) $\gamma^k P = \gamma^k$.

It turns out that when $\sum_{i \in I} \gamma_i^k < \infty$, we can normalize it to form a unique invariant distribution. This happens precisely when $E_i[T_i] < \infty$, and we have

Theorem 6 *Let X_n be irreducible and recurrent. Then the following are equivalent*

(i) $E_i[T_i] < \infty$ *for all* $i \in I$.

(ii) $E_i[T_i] < \infty$ *for some* $i \in I$.

(iii) X_n *has an invariant distribution* π.

If these hold, then in fact π is the unique invariant distribution, and $\pi_i = \frac{1}{E_i[T_i]}$.

In light of this result, we introduce as a definition that an irreducible Markov chain X_n is *positive recurrent* if X_n is recurrent and $E_i[T_i] < \infty$ for some, and hence every, $i \in I$. We see that X_n is positive recurrent if, and only if, X_n has a unique invariant distribution. We say that X_n is *aperiodic* if, for every $i \in I$, there is an N such that $p_{ii}^{(n)} > 0$ for all $n > N$. We have the following important result, which is one of the most important in the theory.

Theorem 7 *Let $X_n \sim Markov(\lambda, P)$ be irreducible, aperiodic, and positive recurrent, with stationary distribution π. Then $P(X_n = i) \longrightarrow \pi_i$ as $n \longrightarrow \infty$ for all $i \in I$.*

Let X be a $\mathcal{F}$-measurable random variable on a space Ω, and let $\mathcal{G}$ be a σ-field on Ω with $\mathcal{G} \subseteq \mathcal{F}$, so that X is not necessarily $\mathcal{G}$-measurable. There is a $\mathcal{G}$-measurable random variable, denoted $E[X|\mathcal{G}]$ and referred to as the *conditional expectation of X with respect to $\mathcal{G}$*, such that $E[X 1_A] = E[E[X|\mathcal{G}] 1_A]$ for all $A \in \mathcal{G}$. There are a few rules for this:

- $E[aX + bY|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}]$.

- If $\mathcal{G} = \{\emptyset, \Omega\}$, then $E[X|\mathcal{G}] = E[X]$.
- If X is $\mathcal{G}$-measurable, then $E[X|\mathcal{G}] = X$, and more generally $E[XY|\mathcal{G}] = XE[Y|\mathcal{G}]$
- If $\mathcal{G}_1 \subseteq \mathcal{G}_2$, then $E[E[X|\mathcal{G}_2]|\mathcal{G}_1] = E[X|\mathcal{G}_1]$.
- If $\sigma(X)$ and $\mathcal{G}$ are independent, then $E[X|\mathcal{G}] = E[X]$.

Discrete time continuous processes have already been defined; a *continuous time stochastic process* is a family of random variables X_t, indexed by $t \in [0, \infty)$, together with an increasing family of σ-fields $\mathcal{F}_t$(the filtration), also indexed by $t \in [0, \infty)$, such that X_t is $\mathcal{F}_t$ measurable for all t. Real-valued stochastic processes for which $E[X_t|\mathcal{F}_s] = X_s$ whenever $t > s$ are called martingales, in either discrete or continuous time, and are very important. Note that $\mathcal{F}_s$ is usually $\sigma(X_r, 0 \leq r \leq s)$, the σ-field generated by all X_r's up to time s. A *stopping time* τ is a r.v. taking values in the nonnegative integers and ∞ such that

$$\{\tau \leq t\} \in \mathcal{F}_t \tag{9}$$

Stopping times are very important in the theory of martingales, in large part due to the following theorem, which has many variants.

Theorem 8 *If M_t is a martingale and τ is a stopping time such that either*

- *τ is bounded, or*
- *$\sup_{t \leq \tau} |M_t| \leq C < \infty$ a.s., for a constant C,*

then $E[M_\tau] = E[M_0]$.

Martingales are closely related to discrete time Markov chains by the following two theorems.

Theorem 9 *Suppose $X_n \sim Markov(\lambda, P)$. Then, for any bounded function $f : I \longrightarrow \mathbb{R}$,*

$$M_n^f = f(X_n) - f(X_0) - \sum_{m=0}^{n-1} (P - I) f(X_m) \tag{10}$$

is a martingale.

Theorem 10 *Suppose $X_n \sim Markov(\lambda, P)$. Suppose further that $f : \mathbb{N} \times I \longrightarrow \mathbb{R}$ satisfies both*

(i) $E|f(n, X_n)| < \infty$, and

(ii) $Pf(n+1, i) = \sum_{j \in I} p_{ij} f(n+1, j) = f(n, i)$.

Then $M_n = f(n, X_n)$ is a martingale.

We now turn to the topic of continuous time Markov chains. The *exponential distribution* $Exp(\lambda)$ is a continuous probability distribution with support on the positive reals, c.d.f. $1 - e^{-\lambda t}$, and p.d.f. $\lambda e^{-\lambda t}$, both for $x \geq 0$. The exponential distribution is generally used for arrival times, since

Proposition 4 *Let T be a positive random variable with a continuous distribution. Then $T \sim Exp(\lambda)$ if, and only if, $P(T > t + s | T > s) = P(T > t)$ for all $t > s$.*

That is, if you have waited time s without an arrival, you are equally likely to get an arrival in the next t time units as you were to get one in the first t time units. The *Poisson process* is the process which results from counting independent, exponentially distributed arrivals. To be precise, let $\tau_1, \tau_2, \ldots \sim Exp(\lambda)$ be independent, and let $T_n = \sum_{i=1}^{n} \tau_i$. Then $N_t = \sup\{n : T_n \leq t\}$ is the Poisson process of rate λ. The following proposition contains some of the most important properties of the Poisson process. Note that the *Poisson distribution* is the discrete distribution with support on the non-negative integers such that, if $X \sim Pois(\mu)$, then $P(X = k) = \frac{e^{-\mu}\mu^k}{k!}$.

Theorem 11 *Let N_t be the Poisson process of rate λ, with associated filtration $\mathcal{F}_t$. Then*

(i) If $t > s$, then $N_t - N_s$ is independent of $\mathcal{F}_s$.

(ii) If $t > s$, then $N_t - N_s$ has a Poisson distribution with parameter $\lambda(t - s)$. In particular, $N_t \sim Pois(\lambda t)$.

(iii) The paths of $N_t, t \geq 0$ are increasing functions of t and change only by jumps of size 1, a.s.

(iv) The following processes are martingales:

- $N_t - \lambda t$.
- $(N_t - \lambda t)^2 - \lambda t$.
- $e^{(\ln(1-u))N_t + u\lambda t}$ *for any $u \in (0, 1)$.*

(vi) The Strong Markov Property holds(see Theorem 13 below for general continuous time Markov chains).

A Q-matrix is an $n \times n$ matrix $\{q_{ij}\}$ such that

(i) $0 \leq q_{ii} < \infty$

(ii) $q_{ij} \geq 0$ for $i \neq j$.

(iii) $\sum_{j\in I} q_{ij} = 0$.

For any matrix M, we define $e^M = \sum_{n=0}^{\infty} \frac{M^n}{n!}$. It was shown in class that this sum always converges. In fact, we have the following theorem.

Theorem 12 *$P_t = e^{tQ}$ is a stochastic matrix if, and only if, Q is a Q-matrix. In this case, we have*

(i) $P_{s+t} = P_s P_t$.

(ii) *P_t is the unique solution to $\frac{d}{dt}P_t = P_t Q$, and $\frac{d}{dt}P_t = QP_t$, with initial condition $P_0 = I$.*

(iii) $(\frac{d}{dt})^k P_t\Big|_{t=0} = Q^k$.

A *right-continous* stochastic process on I is one for which, for a.s. ω, if we choose any $t \geq 0$ then $X_t(\omega) = X_s(\omega)$ for $t \leq s \leq t+\varepsilon$ for some $\varepsilon > 0$. A *continuous time Markov chain* with *initial distribution* λ and *generator matrix* Q, written $X_t \sim Markov(\lambda, Q)$, is a right-continous stochastic process on I with $X_t \sim \lambda$ and transitions governed by Q. To see how the transitions work, let $q_i = \sum_{j\in I, j\neq i} q_{ij}$, and define the *jump matrix* M by

$$m_{ij} = \begin{cases} \frac{q_{ij}}{q_i} & \text{if } q_i \neq 0 \\ 0 & \text{if } q_i = 0 \end{cases} \qquad \text{for } i \neq j \tag{11}$$

$$m_{ii} = \begin{cases} 0 & \text{if } q_i \neq 0 \\ 1 & \text{if } q_i = 0 \end{cases} \tag{12}$$

If we let $J_0, J_1, J_2, \ldots$ be the *jump times* of X_t, that is, $J_0 = 0, J_n = \inf\{t \geq J_{n-1} : X_t \neq X_{J_{n-1}}\}$, then the *jump chain* $Y_n = X_{J_n}$ is a discrete time Markov chain with transition matrix M, and if $S_n = J_n - J_{n-1}$ are the *holding times* then conditional on $Y_0, \ldots, Y_{n-1}$, $S_1, \ldots S_n$ are independent exponential random variables with parameters $q(Y_0), \ldots, q(Y_{n-1})$. In fact, if we set $P_t = e^{tQ}$ as before then the distribution of X_t is λP_t, giving the reason for the importance of the semigroup property contained in Theorem 12. We say that X_t *doesn't explode* if $J_n \nearrow \infty$ a.s. as $n \longrightarrow \infty$. The following tells us when this happens.

Proposition 5 *Let $X_t \sim Markov(\lambda, Q)$. Then X_t doesn't explode if any of the following hold.*

(i) *I is finite.*

(ii) $\sup_{i\in I} q_i < \infty$.

(iii) *$X_0 = i$, and i is recurrent for the jump chain Y_n.*

Continuous time Markov chains possess many of the properties of discrete time Markov chains, or close analogues. In particular, we have the following core results.

Theorem 13 (Strong Markov property) *Suppose $X_t \sim Markov(\lambda, Q)$, and let τ be a stopping time for X_t. Then, conditioned on $\tau < \infty$ and $X_\tau = i$, $Y_t = X_{\tau+t}$ is $Markov(\delta_i, Q)$, where δ_i is the distribution putting mass 1 on the point i. Furthermore, Y_t is independent of $\mathcal{F}_\tau$.*

We define a point $i \in I$ to be *recurrent* if $P_i(\{t : X_t = i\}$ is unbounded$) = 1$, and *transient* if $P_i(\{t : X_t = i\}$ is unbounded$) = 0$. Then

Theorem 14 *For each $i \in I$, i is either recurrent or transient. In particular, if we let $T_i = \inf\{t \geq J_1 : X_t = i\}$, then*

- *if $q_i = 0$ or $P_i(T_i < \infty) = 1$, then i is recurrent and $\int_0^\infty p_{ii}(t)dt = \infty$.*
- *if $q_i > 0$ and $P_i(T_i < \infty) < 1$, then i is transient and $\int_0^\infty p_{ii}(t)dt < \infty$.*

A distribution λ is *invariant* for X_t if $\lambda Q = 0$.

Proposition 6 *The following are equivalent.*

(i) λ is invariant.

(ii) $\mu M = \mu$, where M is the jump matrix of X_t, and $\mu_i = \lambda_i q_i$.

(iii) $\lambda P_t = \lambda$ for all $t \geq 0$, where $P_t = e^{tQ}$.

In light of Theorem 14, we say that $i \in I$ is *positive recurrent* if $E_i[T_i] < \infty$, and *null recurrent* if i is recurrent and $E_i[T_i] = \infty$. We then have

Theorem 15 *Let X_t be an irreducible continuous time Markov chain. Then the following are equivalent.*

(i) Every $i \in I$ is positive recurrent.

(ii) Some $i \in I$ is positive recurrent.

(iii) X_t is non-explosive and has a unique stationary distribution given by $\pi_i = \frac{1}{q_i E_i[T_i]}$.

In this case, $P(X_t = i) \longrightarrow \pi_i$ as $t \longrightarrow \infty$, regardless of the distribution of X_0.