Math 525: Lecture 16

March 8, 2018

1 Markov chains

Consider a sequence $(X_n)_{n\geq 0}$ of random variables (defined on the same probability space) each taking values in some countable set. If we reinterpret the index n as "time", we can think of the random variable X_n "occurring" before X_{n+1}. Similarly, we may interpret the finite sequence $(X_n)_{n\in\{0,\ldots,N\}}$ as indexed by time. Let's give this concept a name:

Definition 1.1. A *discrete stochastic process* is a sequence $(X_n)_{n\in T}$ such that

1. each X_n takes values in some countable set (i.e., there exists a countable set S such that $X_n(\omega) \in S$ for all n and ω) and
2. the index set is either $T = \{0, 1, \ldots\}$ or $T = \{0, \ldots, N\}$.

We call the set S the *state space*.

In practice, many discrete time stochastic processes satisfy a sort of "memoryless" property:

$$\mathbb{P}(X_{n+1} = j \mid X_0 = i_0,\ \ldots,\ X_{n-1} = i_{n-1},\ X_n = i) = \mathbb{P}(X_{n+1} = j \mid X_n = i) \tag{1}$$
$$\text{for all } 0 \leq n < \sup T \text{ and } i_0, \ldots, i_{n-1}, i, j \in S.$$

That is, if we know the value of X at the "current" time n, knowing the value of X at "previous" times $0, \ldots, n-1$ provides no further insight as to the value of X at "future" time $n+1$. In other words, only the present value of the process is important, not its history.

Definition 1.2. (1) is called the *Markov property*. A discrete stochastic process which satisfies the Markov property is called a *Markov chain*.

Technically, (1) is not properly stated since the event $\{X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i\}$ can have probability zero. Therefore, when we say that the Markov property holds, it is understood that we are ignoring (1) when the event $\{X_0 = i_0,\ \ldots,\ X_{n-1} = i_{n-1},\ X_n = i\}$ has probability zero.

Example 1.3. Let $(Y_n)_{n\geq 0}$ be sequence of independent Bernoulli (coin flip) random variables. Let $X_n = Y_1 + \cdots + Y_n$. Since

$$X_{n+1} = X_n + Y_{n+1},$$

it is obvious that $(X_n)_{n\geq 0}$ is a Markov process. In particular,

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \begin{cases} \mathbb{P}(Y_{n+1} = 1) & \text{if } j = i + 1 \\ \mathbb{P}(Y_{n+1} = 0) & \text{if } j = i \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 1.4. *If $(X_n)_{n\in T}$ is a Markov chain, then*

$$\mathbb{P}(X_{n+1} = j \mid (X_0, \ldots, X_{n-1}) \in A,\ X_n = i) = \mathbb{P}(X_{n+1} = j \mid X_n = i)$$

for all $0 \leq n < \sup T$, $i, j \in S$, and $A \subset \mathbb{R}^{n+1}$.

Proof. First, note that

$$\mathbb{P}(X_{n+1} = j \mid (X_0, \ldots, X_{n-1}) \in A,\ X_n = i) = \frac{\mathbb{P}(X_{n+1} = j,\ (X_0, \ldots, X_{n-1}) \in A,\ X_n = i)}{\mathbb{P}((X_0, \ldots, X_n) \in A,\ X_n = i)}.$$

Next,

$$\begin{aligned} &\mathbb{P}(X_{n+1} = j,\ (X_0, \ldots, X_{n-1}) \in A,\ X_n = i) \\ &= \sum_{a\in A} \mathbb{P}(X_{n+1} = j,\ (X_0, \ldots, X_{n-1}) = a,\ X_n = i) \\ &= \sum_{a\in A} \mathbb{P}(X_{n+1} = j \mid (X_0, \ldots, X_{n-1}) = a,\ X_n = i)\mathbb{P}((X_0, \ldots, X_{n-1}) = a,\ X_n = i) \\ &= \mathbb{P}(X_{n+1} = j \mid X_n = i) \sum_{a\in A} \mathbb{P}((X_0, \ldots, X_{n-1}) = a,\ X_n = i) \\ &= \mathbb{P}(X_{n+1} = j \mid X_n = i)\mathbb{P}((X_0, \ldots, X_n) \in A,\ X_n = i). \end{aligned}$$

The desired result follows by combining these equalities. □

So far, we have seen that the future value X_{n+1} of a Markov chain may depend on the present value X_n and the time n. This was indeed the case in Example 1.3. However, what if in Example 1.3, we made the variables Y_n independent? Then, the Markov chain would not depend on the time n. This is a very common scenario, so we give it a name.

Definition 1.5. A Markov chain X is *stationary* if

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \mathbb{P}(X_1 = j \mid X_0 = i)$$

for all states i, j and $0 \leq n < \sup T$. If the state space is finite, we can create a matrix of *transition probabilities* $P = (P_{ij})$ with

$$P_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$$

that wholly describes the evolution of the Markov chain.

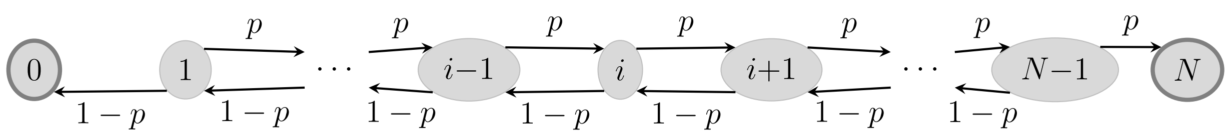
Remark 1.6. Actually, we can create the matrix P even if the state space S is countable. In this case, we have what is called a *denumerable matrix* in $\mathbb{R}^{\mathbb{N}\times\mathbb{N}}$. When we say "transition matrix", we will assume that the state space is finite unless otherwise specified.

Example 1.7. Recall the game of gambler's ruin: a gambler repeatedly plays a game against an opponent in which they receive a dollar with probability p and lose a dollar with probability $1-p$. Both the gambler and opponent start off with initial stakes of x and y dollars, respectively. The game ends when either the gambler or the opponent are broke.

Let X_n be the amount of money the gambler has at time n. Let $N = x + y$ be total wealth in play. It is easy to see that $(X_n)_n$ is a stationary Markov chain, whose transition probabilities are given by the tridiagonal $(N+1)\times(N+1)$ matrix

$$P = \begin{pmatrix} 1 & 0 & & & & & & \\ 1-p & 0 & p & & & & & \\ & 1-p & 0 & p & & & & \\ & & \ddots & \ddots & \ddots & & & \\ & & & 1-p & 0 & p & & \\ & & & & 1-p & 0 & p \\ & & & & & 0 & 1 \end{pmatrix}.$$

Pictorially,



Any Markov chain can always be transformed into a stationary Markov chain by "unrolling" the state space:

Proposition 1.8. *Let $(X_n)_{n\in T}$ be a Markov chain that is not necessarily stationary. Define a new discrete stochastic process $(Y_n)_{n\in T}$ by*

$$Y_n = (X_0, \ldots, X_n).$$

Then, $(Y_n)_n$ is a stationary Markov chain (if the state space of the original Markov chain was S, the state space of the new Markov chain is $\cup_n S^n$).

Let's recall some definitions and results from linear algebra.

Definition 1.9. Let $A \in \mathbb{C}^{m\times m}$. We call $\lambda \in \mathbb{C}$ an *eigenvalue* of A if there exists a nonzero vector $x \in \mathbb{C}^n$ such that

$$Ax = \lambda x.$$

In this case, x is an *eigenvector* associated to the eigenvalue λ. The *spectrum* $\sigma(A)$ of A is the set of all the eigenvalues of A. The *spectral radius* $\rho(A)$ of A is defined as

$$\rho(A) = \max_{\lambda\in\sigma(A)} |\lambda|.$$

We say A is *nonsingular/invertible* if $0 \notin \sigma(A)$.

Definition 1.10. A norm is a function $\|\cdot\| : \mathbb{C}^m \to \mathbb{R}$ satisfying, for all $c \in \mathbb{C}$ and $x, y \in \mathbb{C}^m$

1. $\|cx\| = |c| \, \|x\|$,
2. $\|x + y\| \leq \|x\| + \|y\|$,
3. $\|x\| \geq 0$,
4. $\|x\| = 0$ if and only if $x = 0$.

There is some redundancy in the above definition: points (1) and (2) imply point (3).

Proposition 1.11. *Let $\|\cdot\|$ be a norm on $\mathbb{C}^m$, $A \in \mathbb{C}^{m\times m}$, and define the operator norm of A by*

$$\|A\| = \sup_{\|x\|=1} \|Ax\| \,.$$

Then,

$$\rho(A) \leq \left\|A^k\right\|^{1/k}$$

Proof. Let (λ, x) be an eigenvalue-eigenvector pair. Then,

$$|\lambda|^k \|x\| = \left\|\lambda^k x\right\| = \left\|A^k x\right\| \leq \left\|A^k\right\| \|x\|$$

and hence $|\lambda|^k \leq \|A\|^k$, from which the desired result follows. □

Example 1.12. The infinity norm of a vector is

$$\|x\|_\infty = \max_i |x_i| \,.$$

First, note that for $\|x\|_\infty = 1$,

$$\|Ax\|_\infty = \max_i \left|\sum_j A_{ij} x_i\right| \leq \max_i \sum_j |A_{ij}| \, |x_i| \leq \max_i \sum_j |A_{ij}| \,.$$

Now, let i^* be such that

$$\max_i \sum_j |A_{ij}| = \sum_j |A_{i^* j}|$$

let y be the vector with entries

$$y_j = \begin{cases} +1 & \text{if } A_{i^* j} \geq 0 \\ -1 & \text{otherwise.} \end{cases}$$

Note that $\|y\| = 1$ and

$$\|Ay\|_\infty = \sum_j |A_{i^* j}| \,.$$

Since

$$\|Ay\|_\infty \leq \|A\|_\infty \leq \|Ay\|_\infty \,,$$

it follows that

$$\|A\|_\infty = \max_i \sum_j |A_{ij}| .$$

By Proposition 1.11 and the above,

$$\rho(A) \leq \|A\|_\infty = \max_i \sum_j |A_{ij}| . \tag{2}$$

Proposition 1.13. *Let $P = (P_{ij})$ be a transition matrix associated to a stationary Markov chain. Then,*

1. *$P_{ij} \geq 0$ for all i, j and*
2. *$\sum_j P_{ij} = 1$ for all i.*
3. *$\lambda = 1$ is an eigenvalue of P.*
4. *$I - P$ is singular.*
5. *$\rho(P) \leq 1$.*

Proof. The first two points are trivial consequences of the transition matrix being defined by

$$P_{ij} = \mathbb{P}(X_1 = j \mid X_0 = i).$$

As for the third and fourth points, note that

$$Pe = e \qquad \text{and hence} \qquad (I - P)\, e = 0.$$

where $e = (1, \ldots, 1)$ is the vector of ones. The last point is a consequence of (2). ☐

Definition 1.14. Let $\mu_i = \mathbb{P}(X_0 = i)$. μ is called the *initial distribution* of the Markov chain $(X_n)_{n \in T}$.

Note that the initial distribution is just a vector. If we know the initial distribution of a stationary Markov chain and its transition matrix, we can determine the distribution X_n at any future time $n > 0$.

Proposition 1.15. *Let $(X_n)_{n \in T}$ be a stationary Markov chain with transition matrix $P = (P_{ij})$ and initial distribution μ_i. Then,*

$$\mathbb{P}(X_n = j) \textit{ is the } j\textit{-th entry of the vector } \mu P^n.$$

where it is understood that $P^0 = I$.

Proof. Let $\mu_i^{(n)} = \mathbb{P}(X_n = i)$ so that $\mu = \mu^{(0)}$. Then,

$$\begin{aligned}
\mu_j^{(n+1)} = \mathbb{P}(X_{n+1} = j) &= \sum_i \mathbb{P}(X_{n+1} = j \mid X_n = i)\mathbb{P}(X_n = i) \\
&= \sum_i \mathbb{P}(X_1 = j \mid X_0 = i)\mathbb{P}(X_n = i) \\
&= \sum_i P_{ij}\mu_i^{(n)} = \sum_i \mu_i^{(n)} P_{ij}
\end{aligned}$$

and hence

$$\mu^{(n+1)} = \mu^{(n)} P.$$

The desired result now follows by induction. □

2 Examples of Markov chains

Example 2.1 (Gambler's ruin)**.** Consider the gambler's ruin with $N = 4$ and $p = 1/2$. The transition matrix is

$$P = \begin{pmatrix} 1 & 0 & & & \\ 1/2 & 0 & 1/2 & & \\ & 1/2 & 0 & 1/2 & \\ & & 1/2 & 0 & 1/2 \\ & & & 0 & 1 \end{pmatrix}.$$

Let μ be the initial distribution of the gambler's wealth. For example, if the gambler's initial wealth is 2, we can represent this with the initial distribution vector

$$\mu = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \tag{3}$$

Then,

$$\mu^{(n)} = \mu^{\intercal} P^n$$

is the distribution of the gambler's wealth after n plays. We can diagonalize the matrix P by

$$P = SJS^{-1}$$

where

$$J = \begin{pmatrix} 0 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & -\frac{1}{\sqrt{2}} & \\ & & & & -\frac{1}{\sqrt{2}} \end{pmatrix} \qquad S = \begin{pmatrix} 0 & -3 & 4 & 0 & 0 \\ -1 & -2 & 3 & 1 & 1 \\ 0 & -1 & 2 & -\sqrt{2} & \sqrt{2} \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

Since

$$P^n = \left(SJS^{-1}\right)^n = \underbrace{\left(SJS^{-1}\right)\cdots\left(SJS^{-1}\right)}_{n \text{ times}} = SJ^nS^{-1},$$

it follows that

$$\lim_{n\to\infty} P^n = S\left(\lim_{n\to\infty} J^n\right)S^{-1} = S \begin{pmatrix} 0 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & 0 \end{pmatrix} S^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 3/4 & 0 & 0 & 0 & 1/4 \\ 1/2 & 0 & 0 & 0 & 1/2 \\ 1/4 & 0 & 0 & 0 & 3/4 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Therefore,

$$\mu^{(\infty)} \equiv \lim_{n\to\infty} (\mu^\intercal P^n) = \mu^\intercal \lim_{n\to\infty} P^n.$$

If we plug in 3, we get

$$\mu^{(\infty)} = \begin{pmatrix} 1/2 \\ 0 \\ 0 \\ 0 \\ 1/2 \end{pmatrix}.$$

Unsurprisingly, the probability of ruin is the same as that of victory.

Example 2.2 (Symmetric random walk)**.** Consider once again the gambler's ruin. Suppose now that instead of the game ending when one player accumulates all the wealth, the game never ends, with players allowed to be in debt. This is a stationary Markov chain with the denumerable transition matrix

$$P = \begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ \cdots & 1/2 & 0 & 1/2 & 0 & 0 & \cdots \\ \cdots & 0 & 1/2 & 0 & 1/2 & 0 & \cdots \\ \cdots & 0 & 0 & 1/2 & 0 & 1/2 & \cdots \\ & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Example 2.3 (Success runs)**.** Consider an experiment with probability p of success and $1-p$ of failure. Let

$$Y_{-1} = F$$

and

$$Y_n = \begin{cases} S & \text{if the } n\text{-th experiment was a success} \\ F & \text{otherwise} \end{cases}$$

for $n \geq 0$. Then, the variable

$$X_n = n - \max\{k \leq n \colon Y_k = F\}$$

counts the consecutive number of successes leading up to time n. For example, if

$$SFSSSF\cdots$$

is a sequence of trials, then $X_0 = 1$, $X_1 = 0$, $X_2 = 1$, $X_3 = 2$, $X_4 = 3$, $X_5 = 0$, etc.

This is a stationary Markov chain with

$$\mathbb{P}(X_1 = j \mid X_0 = i) = \begin{cases} p & \text{if } j = i+1 \\ 1-p & \text{if } j = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, this Markov chain admits the denumerable transition matrix

$$P = \begin{pmatrix} 1-p & p & 0 & 0 & \cdots \\ 1-p & 0 & p & 0 & \cdots \\ 1-p & 0 & 0 & p & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$