Lecture 5. Markov chains

Mathematical Statistics and Discrete Mathematics

November 16th, 2015

Stochastic processes

- What we learned so far was about random variables interpreted e.g. as measurements of some random systems.
- Most (physical) systems *evolve in time* and one wants to be able to analyze such systems.
- This can be done by, say, repeated measurements indexed by the time of the measurement.
- This is modelled by objects called stochastic processes.

Main motivating question: How to analyze random systems evolving in time?

Let S be a sample space with a probability measure $\mathbf{P}$ on the events contained in S.

A *stochastic process* is a collection of random variables $\{X_t\}_{t \in T}$ defined on S indexed by time $t \in T$. If T is countable, then we say that the process is a *discrete-time* stochastic process. Otherwise we say that it is a *continuous-time* stochastic process.

Markov chains

We will use the notation $\{X_n\} = \{X_n\}_{n=0}^{\infty} = X_0, X_1, X_2, \ldots$ for a sequence of random variables. We will consider random variables taking values in countable sets (not necessarily subsets of the real numbers).

By $\mathcal{S}$ we will denote the *state space*, that is, a countable set that our random variables will take values in.

A sequence of random variables $\{X_n\}$ defined on the sample space S and taking values in a state space $\mathcal{S}$ is called a *Markov chain* if it satisfies the *Markov property*: for all n and all $s_0, s_1, \ldots, s_n \in \mathcal{S}$,

$$\mathbf{P}(X_n = s_n | X_{n-1} = s_{n-1}, X_{n-2} = s_{n-2}, \ldots, X_0 = s_0) = \mathbf{P}(X_n = s_n | X_{n-1} = s_{n-1}).$$

One can put the Markov property into words in the following way: the future state of a Markov chain is influenced by its past states only through its present state.

Markov chains

Note that a Markov chain is a discrete-time stochastic process.

A Markov chain is called *stationary*, or *time-homogeneous*, if for all n and all $s, s' \in \mathcal{S}$,

$$\mathbf{P}(X_n = s' | X_{n-1} = s) = \mathbf{P}(X_{n+1} = s' | X_n = s).$$

The above probability is called the *transition probability* from state s to state s'.

- ✓ A game of tennis between two players can be modelled by a Markov chain X_n which keeps track of the current score. The transition probabilities depend on the skills of players, and probably also on the score itself.
- ✓ A game of chess can be modelled by a Markov chain X_n whose state space is the space of all possible chessboard configurations.

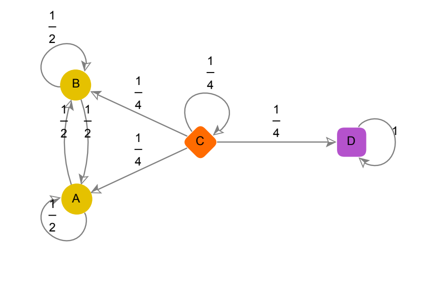
Markov chains

Let $\{X_n\}$ be a sequence of independent random variables. Then,

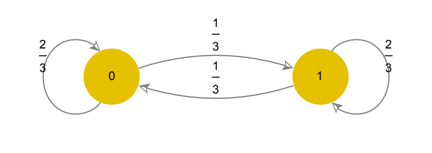
- ✓ $\{X_n\}$ is a Markov chain (The Markov property holds here *trivially* since the past does not influence the future *at all*). It is stationary if and only if the variables have the same distribution .
- ✓ $\{Y_n^{(1)} = \max\{X_1, X_2, \ldots, X_n\}\}$ is a Markov chain.
- ✓ $\{Y_n^{(2)} = \min\{X_1, X_2, \ldots, X_n\}\}$ is a Markov chain.
- ✓ $\{Y_n^{(1)} + Y_n^{(2)}\}$ is *not* a Markov chain.
- ✓ $(Y_n^{(1)}, Y_n^{(2)})$ is a Markov chain.
- ✓ $\{Y_n = X_1 + X_2 + \ldots + X_n\}$ is a Markov chain. If X_i takes values $\{-1, 1\}$, it is called a random walk.

One can define Markov chains by using state diagrams:

- ✓ The following state diagram describes a Markov chain on the state space $\mathcal{S} = \{A, B, C, D\}$.



- ✓ The following diagram describes a Markov chain formed from a repeated execution of **Experiment 2** from the previous lecture: the next coin toss is always biased towards the outcome of the current toss.



Transition matrix

We will only deal with stationary Markov chains on finite state spaces and we will drop the word stationary in the statements.

Let $\{X_n\}_{n=1}^{\infty}$ be a Markov chain on a finite state space $\mathcal{S} = \{s_1, s_2, \ldots, s_k\}$. The *transition matrix* P is a $k \times k$ matrix

$$P = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,k} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k,1} & p_{k,2} & \cdots & p_{k,k} \end{pmatrix}$$

where

$$p_{i,j} = \mathbf{P}(X_n = s_j | X_{n-1} = s_i) \qquad \text{for all} \quad i, j = 1, 2, \ldots, k.$$

A matrix P is a transition matrix of some Markov chain if and only if all its entries are non-negative and they sum up to 1 in each row, that is, for every i, $\sum_{j=1}^{k} p_{i,j} = 1$.

Transition matrix

✓ The matrices

$$P_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix},$$

define deterministic (cellular automaton) Markov chains on a four-element set.

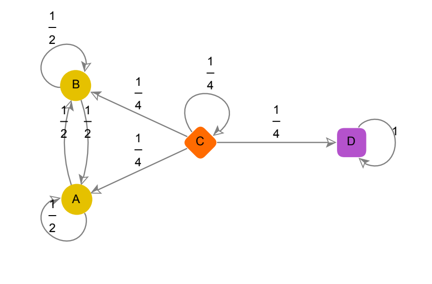
✓ The matrices

$$P_1 = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}, \quad P_2 \begin{pmatrix} 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 \end{pmatrix},$$

define a symmetric and an asymmetric random walk on a four-element cyclic set.

Transition matrix

✓ Let $\mathcal{S} = \{s_1, s_2, s_3, s_4\} = \{A, B, C, D\}$ and consider a Markov chain be given by the following diagram.

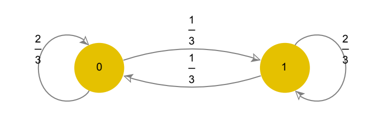


Then,

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Transition matrix

- ✓ Let $\mathcal{S} = \{s_1, s_2\} = \{0, 1\}$ and consider a Markov chain given by the following diagram



Then,

$$P = \begin{pmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{pmatrix}.$$

- ✓ Let $\{X_n\}$ be a sequence of independent coin tosses and let $\{Y_n = \max\{X_1, X_2, \ldots, X_n\}\}$. Then,

$$P = \begin{pmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 2/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 3/6 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 4/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 0 & 5/6 & 1/6 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Distribution vector

One can represent a probability distribution of a random variable X taking values in a finite state space $\mathcal{S} = \{s_1, s_2, \ldots, s_k\}$ by the *distribution vector* $\vec{u}$ given by

$$\vec{u} = (u_1, u_2, \ldots, u_k), \qquad \text{where} \qquad u_i = \mathbf{P}(X = s_i) \quad \text{for } i = 1, 2, \ldots, k.$$

We choose $\vec{u}$ to be a row-vector.

Let $\{X_n\}$ be a Markov chain on $\mathcal{S} = \{s_1, s_2, \ldots, s_k\}$, and let

$$\vec{u}^{(n)} = (u_1^{(n)}, u_2^{(n)}, \ldots, u_k^{(n)})$$

be the distribution vector of X_n. Then,

$$\vec{u}^{(n)} = \vec{u}^{(n-1)} \cdot P = \vec{u}^{(n-2)} \cdot P^2 = \cdots = \vec{u}^{(0)} \cdot P^n,$$

where $\cdot$ denotes matrix multiplication. We call $\vec{u}^{(0)}$ the *initial distribution*.

Distribution vector

Proof. Note that the events $\{X_{n-1} = s_1\}, \{X_{n-1} = s_2\}, \ldots, \{X_{n-1} = s_k\}$ are pairwise disjoint and their union is the full sample space S. We can hence use the total probability formula for the event $\{X_n = s_j\}$. Fix a state s_j. We can write

$$\begin{aligned}
\mathbf{P}(X_n = s_j) &= \sum_{i=1}^{k} \mathbf{P}(X_n = s_j | X_{n-1} = s_i)\mathbf{P}(X_{n-1} = s_i) \\
&= \sum_{i=1}^{k} p_{i,j}\mathbf{P}(X_{n-1} = s_i) \\
&= \sum_{i=1}^{k} p_{i,j}u_i^{(n-1)} \\
&= u^{(n-1)} \cdot c_j, \qquad \text{(dot product of } u^{(n-1)} \text{ and } c_j\text{)}.
\end{aligned}$$

where c_j is the j-th colum vector of P. This agrees with the definition of matrix multiplication.

We say that the chain $\{X_n\}$ starts in state s_i if the initial distribution satisfies $u_i^{(0)} = 1$.

Distribution vector

The i, j entry of the matrix P^n is

$$p_{i,j}^{(n)} = \mathbf{P}(X_n = s_j | X_0 = s_i).$$

✓ Consider a Markov chain $\{X_n\}$ given by the matrix

$$P = \begin{pmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{pmatrix}$$

and started at s_1. What is the distribution vector of X_2, X_4, X_8 and X_{16}? We have $u^{(0)} = (1, 0)$, and

$$\begin{aligned} \vec{u}^{(2)} &= u^{(0)} \cdot P^2 = (5/9, 4/9), \\ \vec{u}^{(4)} &= u^{(0)} \cdot P^4 = (41/81, 40/81), \\ \vec{u}^{(8)} &= u^{(0)} \cdot P^8 = (3281/6561, 3280/6561), \\ \vec{u}^{(16)} &= u^{(0)} \cdot P^{16} = (21523361/43046721, 21523360/43046721). \end{aligned}$$

Absorbing Markov chains

A state $s_i \in \mathcal{S}$ of a stationary Markov chain is called

- *absorbing* if $p_{i,i} = 1$,
- *transient* if $p_{i,i} < 1$.

A Markov chain is called *absorbing* if from any transient state, one can reach an absorbing state.

Absorbing Markov chains

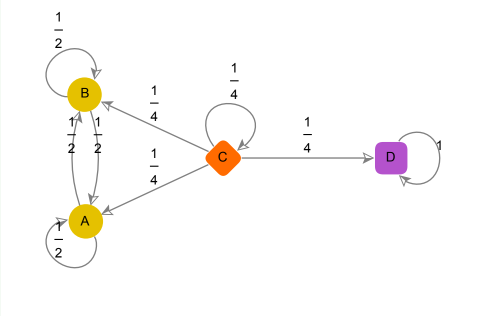
✓ Let a Markov chain be given by the following transition matrix

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

This is an *absorbing* Markov chain. This chain can be interpreted as a mouse jumping on a table of length $1m$ and making jumps of $0.5m$ to the left or to the right with equal probability. The absorbing state is when the mouse falls off the table (either on the left-hand side or the right-hand side).

Absorbing Markov chains

✓ Consider a Markov chain given by the following diagram:



This Markov chain is *not absorbing* since the only absorbing state D cannot be reached from the transient states A and B.

Absorbing Markov chains

The probability that an absorbing Markov chain will eventually end up in one of its absorbing states is 1.

Proof. From each nonabsorbing state s_j it is possible to reach an absorbing state. Let m_j be the minimum number of steps required to reach an absorbing state, starting from s_j. Let p_j be the probability that, starting from s_j , the process will not reach an absorbing state in m_j steps. Then $p_j < 1$. Let m be the largest of the m_j and let p be the largest of p_j. The probability of not being absorbed in m steps is less than or equal to p, in $2m$ steps less than or equal to p^2, etc. Since $p < 1$, these probabilities tend to 0.

For the interested: this argument is very similar to the one for the *infinite monkey* theorem which states that a monkey hitting keys at random on a typewriter keyboard for an infinite amount of time will with probability 1 type a given text, such as the complete works of William Shakespeare.

Canonical form of the transition matrix

Let $\{X_n\}$ be an absorbing Markov chain with l transient states and r absorbing states. Let $\{s_1, s_2, \dots, s_{l+r}\}$ be a numbering of the state space where all the absorbing states come at the end. Let P be the transition matrix for this particular numbering. Then P is in the *canonical form*:

$$P = \begin{pmatrix} Q & R \\ \mathbf{0} & I \end{pmatrix},$$

where

- I is an $r \times r$ identity matrix,
- $\mathbf{0}$ is an $r \times l$ zero matrix,
- Q is an $l \times l$ matrix with transition probabilities *between transient states*,
- R is an $l \times r$ matrix with transition probabilities *from transient to absorbing states*.

Note that

$$P^n = \begin{pmatrix} Q^n & R(n) \\ \mathbf{0} & I \end{pmatrix}$$

for some matrix $R(n)$.

Time to absorption

Let $\{X_n\}$ be an absorbing Markov chain and let Q be as before. The *fundamental matrix* is defined to be

$$N = (I - Q)^{-1}$$

The entries $n_{i,j}$ of N satisfy

$$n_{i,j} = \text{ the exp. \# of visits of } \{X_n\} \text{ to the state } s_j \text{ when started from state } s_i.$$

Note that $N = I + Q + Q^2 + \ldots.$

Proof. Fix i and let $\{X_n\}$ be a Markov chain started at s_i. Let $Y_n^j = 1$ if $X_n = s_j$, and $Y_n^j = 0$ otherwise. Then,

$$Y^j = Y_0^j + Y_1^j + \ldots$$

is the total number of visits of $\{X_n\}$ to s_j. Hence,

$$\begin{aligned}\mathbf{E}[Y^j] &= \mathbf{E}[Y_0^j] + \mathbf{E}[Y_1^j] + \ldots = I_{i,j} + \mathbf{P}(X_1 = s_j \mid X_0 = s_i) + \mathbf{P}(X_2 = s_j \mid X_0 = s_i) + \ldots \\ &= I_{i,j} + p_{i,j} + p_{i,j}^{(2)} + \ldots = I_{i,j} + q_{i,j} + q_{i,j}^{(2)} + \ldots = n_{i,j},\end{aligned}$$

where $q_{i,j}^{(n)}$ is the i,j entry of the matrix Q^n.

Time to absorption

Let

$$t_i = \text{the expected time before absorption when started from } s_i,$$

and let $\vec{t} = (t_1, t_2, \ldots, t_l)$ be a column vector. Then,

$$\vec{t} = N \cdot \vec{c},$$

where $\vec{c}$ is a column vector of 1's.

Proof. Fix i and let $\{X_n\}$ be a Markov chain started at s_i. Let Z^j be the total number of visits of $\{X_n\}$ to s_j. Let T_i be the time until absorption for $\{X_n\}$. We have $T_i = \sum_j Z^j$, and hence

$$t_i = \mathbf{E}[T_i] = \sum_j \mathbf{E}[Z^j] = \sum_j n_{i,j} = (N \cdot \vec{c})_i.$$

Probability of absorption

Let s_i be a transient state. We define

$$b_{i,j} = \mathbf{P}(\text{ absorption in } s_j \mid X_0 = s_i),$$

and put $B = (b_{i,j})_{1 \le i \le l, 1 \le j \le r}$. Then,

$$B = N \cdot R.$$

Proof. We represent $Q^n_{i,k}$ as a sum over all possible trajectories of $\{X_n\}$ of n steps from a transient state s_i to a transient state s_k. All trajectories of $n+1$ steps from s_i ending at an absorbing state s_j have to have their first n steps between transient states, and their last step between a transient and the absorbing state s_j. Hence, the sum over all such trajectories is

$$\sum_k Q^n_{i,k} R_{k,j} = (Q^n R)_{i,j}.$$

If we sum over all possible numbers of steps, we get

$$B_{i,j} = [(I + Q + Q^2 + \dots)R]_{i,j} = (NR)_{i,j}.$$

Exercise: Compute the vector $\vec{t}$ and the matrix B for the mouse on a table.