Probability I
Lecture 9

Daniel Yu

October 17,2024

Contents

We are now moving from the viewpoint of an analytical perspective of probabiltiy to a stochastic persepctive of probability! This marks the halfway point of the course.

1 Stochastic processes

Definition 1. A **stochastic process** is any sequence of random variables where we think of n as time.

Note. Example
Let there be an elephant that begins in room 1. Each minute, the element chooses a room uniformly among the neighbors that has not visited in the past. Let $X_0 = 1$ with probability 1.

1. Then what is $P[X_0 = 1 \cap X_1 = 2 \cap X_2 = 3]$?

   Proof.

$$\begin{aligned} P[X_0 = 1 \cap X_1 = 2 \cap X_2 = 3] &= P[X_0 = 1] \cdot P[X_1 = 2 \cap X_2 = 3 | X_0 = 1] \\ &= P[X_0 = 1] \cdot P[X_1 = 2 | X_0 = 1] \cdot P[X_2 = 3 | X_0 = 1, X_1 = 2] \\ &= 1 \cdot \frac{1}{3} \cdot 1 \\ &= \frac{1}{3}. \end{aligned}$$

□

This problem is called a **self avoiding walk**. In generally, incredibly hard!
What if we replace the elephant with a forgetful mouse, who acts uniformly at random with no regard to the past?

1. then what is $P[X_0 = 1 \cap X_1 = 2 \cap X_2 = 3]$?

   Proof.

$$\begin{aligned} P[X_0 = 1 \cap X_1 = 2 \cap X_2 = 3] &= P[X_0 = 1] \cdot P[X_1 = 2 \cap X_2 = 3 | X_0 = 1] \\ &= P[X_0 = 1] \cdot P[X_1 = 2 | X_0 = 1] \cdot P[X_2 = 3 | X_0 = 1, X_1 = 2] \\ &\text{in this case the past history is irrelevant!} \\ &= P[X_0 = 1] \cdot P[X_1 = 2 | X_0 = 1] \cdot P[X_2 = 3 | X_1 = 2] \\ &= 1 \cdot \frac{1}{3} \cdot \frac{1}{2} \\ &= \frac{1}{6}. \end{aligned}$$

□

We will define this concept examplified by the forgetful mouse as a **markov chain**.

2 Markov Chain

We will focus on finite state markov chains in this course

Definition 2. A stochastic process is a **markov chain** if

$$P[X_n = j \mid X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_{n-1} = i_{n-1}] = P[X_n = j | X_{n-1} = i_{n-1}].$$

for any set of states $(i_0, \ldots, i_{n-1})$, $i \in \Omega$, $\forall n \geq 0$.

Definition 3. If in addition to the above,

$$P[X_n = j | X_{n-1} = i_{n-1}] = P[X_i = j | X_0 = i_{n-1}].$$

, then the **markov chain** is **time-homogenous**

Remark. Trivially, iid sequences are time-homogenous markov chains.

1. Let $\{W_n\}$ be random walks where $W_n = \sum_{i=1}^{n} X_i$ with $X_i = \pm 1$ with probability $\frac{1}{2}$ and $\{X_i\}$ are iid. Are W_n independent of one another? No, but is the stochastic process $\{W_n\}$ a markov chain?

   Proof. Consider,

$$P[W_{n+1} = K | W_0 = 0, W_1 = i, W_2 = i_2, \dots W_n = i_n] = \begin{cases} \frac{1}{2} \text{ if } k = i_n + 1 \\ \frac{1}{2} \text{ if } k = i_n - 1 \\ 0 \text{ otherwise} \end{cases}.$$

   Then realize that

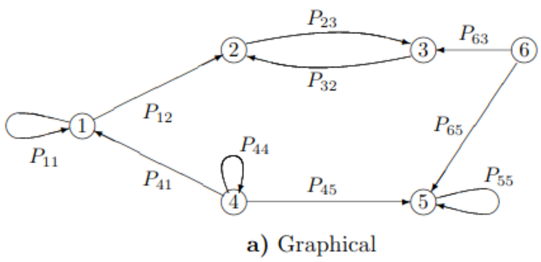
$$P[W_{n+1} = K | W_n = i_n] = \begin{cases} \frac{1}{2} \text{ if } k = i_n + 1 \\ \frac{1}{2} \text{ if } k = i_n - 1 \\ 0 \text{ otherwise} \end{cases}.$$

   And $\{W_n\}$ is a time-homogenous **infinite state** markov chain □

2.1 Finite State Time Homogenous Markov Chain

Proposition 1. We can describe any time-homogenous finite state markov chain with a **transition matric** $P_{|\Omega| \times |\Omega|} = (P_{i,j})$ where

$$P_{i,j} = P[X_1 = j | X_0 = i].$$



$$[P] = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{16} \\ P_{21} & P_{22} & \cdots & P_{26} \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ P_{61} & P_{62} & \cdots & P_{66} \end{bmatrix}$$

b) Matrix

Figure 1: Finite 6 State Transition Matrix

Lemma 1. Properties of a transition matrix P

1. $P_{i,j} \geq 0$
2. $\sum_{j=1}^{|\Omega|} P_{i,j} = \sum_{j \in \Omega} P[X_1 = j \mid X_0 = i] = P[\cup_{j \in \Omega} \{X_1 = j\} \mid X_0 = i] = 1$

This is known as a **row-stochastic matrix**.

Proposition 2. A matrix is **row stochastic** $\Leftrightarrow$ there is a finite state markov chain associated with it $\Leftrightarrow$ there is a weighted directed graph where the weights of outgoing edges from each node sum to 1 (the matrix is the adajency matrix)

Note. Example Consider $P = \begin{pmatrix} 0.2 & 0.5 & 0.3 \\ 0.6 & 0.3 & 0.1 \\ 0.4 & 0.4 & 0.2 \end{pmatrix}$.

1. What is $P(F_0 = 1, F_1 = 2, F_2 = 3, F_3 = 1, F_4 = 1)$?

$$\begin{aligned} P(F_0 = 1, F_1 = 2, F_2 = 3, F_3 = 1, F_4 = 1) &= P(F_0 = 1) \cdot P(F_1 = 2 \mid F_0 = 1) \cdot P(F_2 = 3 \mid F_1 = 2) \\ &\quad \cdot P(F_3 = 1 \mid F_2 = 3) \cdot P(F_4 = 1 \mid F_3 = 1) \\ &= 1 \cdot P_{1,2} \cdot P_{2,3} \cdot P_{3,1} \cdot P_{1,1} \\ &= 1 \cdot 0.5 \cdot 0.1 \cdot 0.4 \cdot 0.2 \\ &= 0.004. \end{aligned}$$

2. What is $P[F_2 = 3|F_0 = 1]$?

$$\begin{aligned} P(F_2 = 3 \mid F_0 = 1) &= \sum_{k=1}^{3} P(F_2 = 3 \mid F_1 = k, F_0 = 1) \cdot P(F_1 = k \mid F_0 = 1) \\ &= \sum_{k=1}^{3} P(F_2 = 3 \mid F_1 = k) \cdot P(F_1 = k \mid F_0 = 1) \\ &= P(F_2 = 3 \mid F_1 = 1) \cdot P(F_1 = 1 \mid F_0 = 1) \\ &\quad + P(F_2 = 3 \mid F_1 = 2) \cdot P(F_1 = 2 \mid F_0 = 1) \\ &\quad + P(F_2 = 3 \mid F_1 = 3) \cdot P(F_1 = 3 \mid F_0 = 1) \\ &= (0.3 \cdot 0.2) + (0.1 \cdot 0.5) + (0.2 \cdot 0.3) \\ &= 0.06 + 0.05 + 0.06 \\ &= 0.17. \end{aligned}$$

Notice that, $\sum_{k=1}^{3} P(F_2 = 3 \mid F_1 = k) \cdot P(F_1 = k \mid F_0 = 1) = \sum_{k=1}^{3} P_{k,3} \cdot P_{1,k} = (P^2)_{1,3}$!

Remark. Notice that for these finite state markov chains, it is simplest to compute joint probabilities!

Definition 4. Let P be the transition matrix of a time-homogeneous Markov chain with m states. The (i, j)-th entry of P^n, denoted $(P^n)_{i,j}$, represents the probability of transitioning from state i to state j in n steps. Formally,

$$(P^n)_{i,j} = P(F_n = j \mid F_0 = i)$$

This can be expressed as:

$$(P^n)_{i,j} = \sum_{k_1=1}^{m} \sum_{k_2=1}^{m} \cdots \sum_{k_{n-1}=1}^{m} P_{i,k_1} P_{k_1,k_2} \ldots P_{k_{n-1},j}$$

where the summation is taken over all possible intermediate states $k_1, k_2, \ldots, k_{n-1}$.

Lemma 2. Let $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_m)$ be the initial distribution, where $\alpha_i = P(F_0 = i)$ for $i = 1, \ldots, m$. Then the probability that the chain is in state j after n steps is given by:

$$P(F_n = j) = (\alpha P^n)_j = \sum_{i=1}^{m} \alpha_i (P^n)_{i,j}$$

Note. We are conditioning over F_0

In this case, αP^n is the row vector obtained by multiplying the initial distribution α by the matrix P^n. The j-th entry of this vector gives the probability that the chain is in state j at time n.

2.2 Recurrent States and Absorbing Markov Chains

Definition 5. A state is called **recurrent** if

$$P[X_n = i \text{ for infinitely many n's} | X_0 = i] = 1$$

. Otherwise, a state is **transient**.

Definition 6. A state is called **absorbing** if $P_{i,i} = 1$. Every absorbing state is recurrent.

Definition 7. A markov chain is called absorbing if $\exists$ at least one absorbing state and it is possible to reach an absorbing state from every non-absorbing state (not necessarily in one step).

Mathematically, let $A \subseteq \Omega : A = \{i : P_{ii} = 1\}$, the set of absorbing states. We say a chain is abosrbing if:

$$A \neq \emptyset.$$

and

$$\forall j \in \Omega \setminus A, \exists a \in A \text{ such that } P[X_n = a | X_0 = j] > 0.$$

Basically, regardless of intial conditions, there is a positive probability that you will eventually reach an absorping state.

Remark. Question: Suppose I start at a transient state j, what is the probability $P[$ I get absorbed evenutally $| X_0 \neq j]$?

Proof. Let i, j be transient states in Ω.

$$P[X_n = j \mid X_0 = i] = P[X_n = j \text{ and } X_k \text{ is transient } \forall k \in \{1, 2, 3, \ldots, n\} \mid X_0 = i]$$

because we know that any intermediate states between X_n and X_0 can't be absorbing.

We can manipulate the transition matrix to have a block structure by permuting the matrix (relabeling nodes):

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

where Q are the transient states, R are the connections from the transient to absorbing states, and I the absorbing states. For example, the matrix:

$$P = \left[\begin{array}{cc|cc} 0.5 & 0.5 & 0 & 0 \\ 0.2 & 0.3 & 0.5 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

Then, we can consider

$$\begin{aligned} (P^n)_{i,j} &= P[X_j = j \mid X_0 = i] \\ &= P[X_n = j \text{ and } X_k \text{ is transient } \forall k \in \{1, 2, 3, \ldots, n\} \mid X_0 = i] \\ &= (Q^n)_{i,j}. \end{aligned}$$

□

Theorem 1. if P represents an absorbing markov chain, then $Q^n \to 0$ the zero matrix. Thus, for any state, $P[$ eventual absorption $\mid X_0 = i] = 1$

Proof. for every transient state j, $\exists m_j$ such that $P[X_{m_j} \in A \mid X_0 = j] = P_j > 0$. We know

1. $m = max_{j \in \Omega \setminus A} m_j \leq \infty$ since finite number of states, so finite number of possible steps (m_j represents the m_jth step)
2. $p = min_{j \in \Omega \setminus A} P_j > 0$ the minimum probability to reach an absorbing state from any starting state

Then for $j \in \Omega \setminus A$,

$$P[\text{ being absorbed by the m} \mid X_0 = j] \geq p$$

. Let $NA(k) = \{$the markov chain has not been absorbed within the first k $\cdot$ m steps$\}$. To prove the thorem, it sufficies to show that for any transient state i,

$$P[NA(k) \mid X_0 = i] \to 0, k \to \infty.$$

We will prove the stronger statement by induction

$$P[NA(k) \mid X_0 = i] \leq (1-p)^k.$$

1. Base Case: $P[NA(1) \mid X_0 = i] \leq (1-p)$ which must be true since $P[NA(1) \mid X_0 = i] = 1 - P[\text{ being absorbed by the m} \mid X_0 = j] \leq 1 - p$
2. Induction Step: Assume $P[NA(k) \mid X_0 = i] \leq (1-p)^k$ for any state i.

$$\begin{aligned}
P[NA(k+1) \mid X_0 + i] &= P[NA(k+1) \cap NA(k) \mid X_0 = i] \\
&= P[NA(k) \mid X_0 = i] \cdot P[NA(k+1) \mid NA(k), X_0 = i] \\
&\leq (1-p)^k \cdot P[NA(k+1) \mid NA(k), X_0 = i] \\
&= (1-p)^k \cdot \sum_{j \in \Omega \setminus A} P[NA(k+1) \mid X_{km} = j, NA(k), X_0 = i] \cdot \\
&P[X_{km} = j \mid NA(k), X_0 = j] \\
&\text{use markov property} \\
&= (1-p)^k \cdot \sum_{j \in \Omega \setminus A} P[NA(k+1) \mid X_{km} = j] \cdot P[X_{km} = j \mid NA(k), X_0 = j] \\
&\text{then let's reset the time by setting } km \text{ to new time } 0 \\
&= (1-p)^k \cdot \sum_{j \in \Omega \setminus A} P[NA(1) \mid X_0 = j] \cdot P[X_{km} = j \mid NA(k), X_0 = j] \\
&\leq (1-p)^k \cdot (1-p) \cdot \sum_{j \in \Omega \setminus A} P[X_{km} = j \mid NA(k), X_0 = j] \\
&\text{if we haven't been absorbed and we go from every possible transient to every other trans} \\
&\leq (1-p)^k \cdot (1-p) \cdot 1 \\
&\leq (-1p)^{k+1}.
\end{aligned}$$

□

Remark. How long does it take to get absorbed? What is:

$$E[T \mid X_0 = i] = ?.$$

Proposition 3. If P is an absorbing markov chain with transient block matrix Q, then $N = (I-Q)^{-1}$ is equal to $I + Q + Q^2 + Q^3 + \ldots$

Proof. Recall $Ax = 0$ has only the trivial solution $x = 0$ then A is invertible. This is because A has full rank and the null space is trivial. So $(I-Q)x = 0$ only when $x = 0$. Then,

$$\begin{aligned}(I-Q)x = 0 &\Leftrightarrow x - Qx = 0\\ &\Leftrightarrow x = Qx\\ &\Leftrightarrow x = Qx = Q^2x = Q^3x = \ldots = Q^nx.\end{aligned}$$

Then since $Q^n \to 0$:

$$X = \lim_{n\to\infty} Q^n x = 0.$$

Let $S_m = I + Q + \ldots + Q^m$.

$$\begin{aligned}(I-Q)\cdot S_m &= (I + Q + Q^2 + \ldots + Q^m) - (Q + Q^2 + \ldots + Q^{m+1})\\ &= I - Q^{m+1}.\end{aligned}$$

Take the limit $m \to \infty$, so:

$$\begin{aligned}(I-Q)S_\infty &= \lim_{m\to\infty} I - Q^{m+1}\\ &= I.\end{aligned}$$

Since inverses are unique: $N = (I-Q)^{-1} = S_\infty$ □

Proposition 4. Let $V(j)$ = number times markov chain visits state j. Then,

$$(N)_{i,j} = E[V(j) \mid X_0 = i].$$

The entries in the inverse of $I - Q$ gives the expected number of times a markov chain starting from i visits j over an infinite number of steps

Proof. Express $V(j)$ as sum of indicators.

$$M_j^n = \begin{cases} 1, x_n = j \\ 0, x_n \neq= j \end{cases}.$$

Using linearity,

$$\begin{aligned}E[V(j) \mid X_0 = i] &= E[\sum_{n=0}^{\infty} M_j^n \mid X_0 = i]\\ &= \sum_{n=0}^{\infty} E[M_j^n \mid X_0 = i]\\ &= \sum_{n=0}^{\infty} P[X_n = j \mid X_0 = i]\\ &= \sum_{n=0}^{\infty} (Q^n)_{i,j}\\ &= (I + Q + Q^2 + \ldots)\\ &= N_{i,j}.\end{aligned}$$

□

Theorem 2. The expected value of the number of steps until it is absorbed from state i is:

$$E[T \mid X_0 = i] = E[\sum_{j \in \Omega \setminus A} V(j) \mid X_0 = i] = \text{ sum of row } i \text{ in } N.$$