Introduction to Stochastic Processes

Assignment 2
Ari Feiglin

Exercise

We are given the following two stochastic matrices

$$P_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{2}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \qquad P_2 = \begin{pmatrix} \frac{1}{5} & \frac{3}{5} & \frac{1}{5} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} \end{pmatrix}$$

(1) Find the recurrent states, and the irreducible classes of each matrix.

(2) For P_1, how does $N_4(4)$ distribute?

(3) For P_2, compute $f_{1\to 4}$ and $f_{6\to 5}$.

(1) In P_1 the recurrent states are $1, 5, 6$. 1 is obviously recurrent since if you get to it, you will always return immediately to it (it is absorbing). If you get to 5, you either go to 5 again or go to 6 and then 5, so you will always return to 5 eventually. And since $5 \to 6$, 6 is also recurrent. $2 \to 1$ but 2 is not reachable from 1, and so 2 is transient. And 5 is reachable from 3 but 3 is not reachable from 5, so 3 is transient and since $4 \to 3$, 4 is also transient. The irreducible classes are $\{1\}$ and $\{5, 6\}$ as they are obviously both closed and connected.

In P_2 the recurrent states are $2, 5, 6$. 2 is recurrent since it is absorbing, and 5 and 6 are for the same reason as before. Since 2 is reachable from 4 but not vice versa, 4 is transient. And since $1 \to 3 \to 4$, 1 and 3 are also transient. The irreducible classes are $\{2\}$ and $\{5, 6\}$.

(2) To compute this we must compute f_4, which is the probability of returning to 4. The probability of not returning to 4 is if we go $4 \to 3 \to 5$ which has a probability of $\mathbb{P}(X_1 = 3 \mid X_0 = 4) \cdot \mathbb{P}(X_2 = 5 \mid X_1 = 3) = \frac{2}{3} \cdot \frac{1}{3} = \frac{2}{9}$. Thus $f_4 = \frac{7}{9}$ and so

$$N_4(4) \sim \mathrm{Geo}(1 - f_4) - 1 = \mathrm{Geo}\left(\frac{2}{9}\right) - 1$$

(3) To go from 1 to 4, we can visit 1 an arbitrary amount of times and then go to 3 and then 4, so the probability is

$$\sum_{n=0}^{\infty} \mathbb{P}(X_{n+2} = 4, X_{n+1} = 3, X_n = \cdots = X_1 = 1 \mid X_0 = 1) = P(3 \to 4) \cdot P(1 \to 3) \cdot \sum_{n=0}^{\infty} P(1 \to 1)^n$$
$$= \frac{1}{10} \cdot \sum_{n=0}^{\infty} \frac{1}{5} = \frac{1}{10} \cdot \frac{5}{4} = \frac{1}{8}$$

Thus $f_{1\to 4} = \frac{1}{8}$. And $f_{6\to 5} = 1$ since 6 and 5 are recurrent and we showed in lecture that this measn $f_{6\to 5} = 1$.

Exercise

Yaron collects pokemon. He wants to collect all N pokemon in a series. In every package there is a pokemon which is uniformly distributed and independent of other packages. Let us denote by X_n the number of different pokemon Yaron has after purchasing n packages.

(1) Show that X_n is a Markov chain over the state space $\{0, 1, \ldots, N\}$ and compute the probability transitions.

(2) Assuming $X_0 = 0$, let $R_k = T_k - T_{k-1}$. Find the distribution of R_k and explain why R_k are all independent.

(3) Find a formula for the expected time until Yaron will collect all N distinct pokemon and show that it is approximately $N \log N$.

(1) If $X_{n-1} = a_{n-1}, \dots, X_0 = a_0$ then the probability that $X_n = a_{n-1}$ is $\frac{a_{n-1}}{N}$ since this is the probability of choosing one of the pokemon already collected. And the probability that $X_n = a_{n-1}+1$ is $\frac{N-a_{n-1}}{N}$, as this is the probability of choosing a new pokemon. This is independent of $X_0, \dots, X_{n-2}$ and so X_n is indeed a Markov chain. And the transition probabilities are

$$\mathbb{P}(X_n = i \mid X_{n-1} = j) = \begin{cases} \frac{j}{N} & i = j \\ \frac{N-j}{N} & i = j+1 \\ 0 & \text{else} \end{cases}$$

(2) R_k is the number of packages opened between finding the $k-1$th and kth distinct pokemon. Let us denote $T_{k-1} = t$ then $X_t = k-1$ and so

$$\begin{aligned} \mathbb{P}(R_k = a) &= \mathbb{P}(T_k = a+t) = \mathbb{P}(X_{a+t} = k, X_{a+t-1} = k-1, \dots, X_t = k-1) \\ &= \mathbb{P}(X_{a+t} = k \mid X_{a+t-1} = k-1) \cdot \mathbb{P}(X_{a+t-1} = k-1 \mid X_{a+t-2} = k-1) \cdots \mathbb{P}(X_{t+1} = k-1 \mid X_t = k-1) \end{aligned}$$

By the previous subquestion this is equal to

$$\frac{N-k+1}{N} \cdot \left(\frac{k-1}{N}\right)^{a-1}$$

and thus $R_k \sim \mathrm{Geo}\big(1 - \frac{k-1}{N}\big)$. R_k are all independent since the time it takes to find the kth pokemon after finding the $k-1$th has no bearing on how long it takes to find the jth after finding the $j-1$th.

(3) This is asking for the expected value of T_N. By definition $T_N = R_N + \cdots + R_1$ and so by linearity of the expected value, and the expected value of a geometric distribution we get

$$\mathbb{E}[T_N] = \sum_{k=1}^{N} \mathbb{E}[R_N] = \sum_{k=1}^{N} \frac{N}{N-k+1} = N \cdot \sum_{k=1}^{N} \frac{1}{k}$$

Now, it is known that $\sum_{k=1}^{N} \frac{1}{k} \sim \log N$ and thus $\mathbb{E}[T_N] \sim N \log N$ as required.