Markov chain in discrete time

Math 478

April 1, 2019

1 Transition probability

Let $X_k, k = 0, 1, \cdots$ be a stochastic process in discrete time. An important ingredient to understand the evolution of X_k is its *conditional* transition probability given its past history:

$$P(X_{n+1} \in E_{n+1} \Big| X_n \in E_n, X_{n-1} \in E_{n-1}, \cdots, X_0 \in E_0),$$

where $E_i, i = 1, \cdots, n+1$ are subsets of the state space (the space where X_k takes value). If this transition probability only depends on the most recent history:

$$P(X_{n+1} \in E_{n+1} \Big| X_n \in E_n, X_{n-1} \in E_{n-1}, \cdots, X_0 \in E_0) = P(X_{n+1} \in E_{n+1} \Big| X_n \in E_n),$$

we say X_k is Markov. In general, a Markov process X_t (with respect to a filtration $\mathcal{F}_t^Y$ generate by a random variable Y possibly different from X) can be stated succinctly as any stochastic process that satisfies for any $s < t$ and any function f

$$E(f(X_t) \Big| \mathcal{F}_s^Y) = E(f(X_t) \Big| Y_s),$$

where $\mathcal{F}_s^Y$ represents all the information given by $Y_u, 0 \le u \le s$.

In the discrete time setting, one may potentially have two definitions for Markov property:

$$E(f(X_{n+1}) \Big| X_n, X_{n-1}, \cdots, X_0) = E(f(X_{n+1}) \Big| X_n) \ \ \forall n,$$

or

$$E(f(X_n) \Big| X_m, X_{m-1}, \cdots, X_0) = E(f(X_n) \Big| X_m) \ \ \forall m < n.$$

The second definition obviously is more general than the first one. It turns out that the two are equivalent, as we shall see.

Example 1.1. *Random walk as a discrete time Markov process*

Let X_i be iid and $S_k = \sum_{i=1}^{k} X_i$. Then for $m < n$

$$\begin{aligned} E(S_n|S_m, S_{m-1}, \cdots S_0) &= E(S_m + \sum_{i=m+1}^{n} X_i|S_m, S_{m-1}, \cdots S_0) \\ &= S_m + E(\sum_{i=m+1}^{n} X_i|S_m, S_{m-1}, \cdots S_0) \\ &= S_m + E(\sum_{i=m+1}^{n} X_i) \\ &= E(S_n|S_m). \end{aligned}$$

A Markov chain is a stochastic process whose state space is discrete. In this case, we can capture its dynamics if we know for any n, i, j

$$P(X_{n+1} = j \Big| X_n = i, X_{n-1} = i_{n-1}, \cdots, X_0 \in i_0) = P(X_{n+1} = j \Big| X_n = i),$$

The RHS of this equation is a quantity that only depends on i, j, n. We make the additional assumption that it is independent of n (which is referred to as a time homogeneous Markov chain). We assume all the Markov chain we discuss is time homogeneous unless otherwise mentioned. Then we can denote

$$P(X_{n+1} = j \Big| X_n = i) = P_{ij}.$$

If we make one last assumption that the state space is finite with size n then P_{ij} can be organized into a $n \times n$ matrix, referred to as a transition matrix. Such matrix obviously has non-negative entries. Since

$$\sum_j P_{ij} = \sum_j P(X_{n+1} = j \Big| X_n = i) = 1,$$

the sum of all row entries in a transition matrix is 1. Additionally, since $P_{ij} \neq P_{ji}$ in general, a transition matrix is NOT necessarily symmetric (in distinction to say, a correlation matrix). Finally, if a state i is such that $P_{ii} = 1$ (and thus $P_{ij} = 0$ for any $i \neq j$) then it is referred to as an absorbing state.

Remark 1.2. *If we're willing to entertain the notion of a (countably) infinite dimensional matrix, then the same organization of information can be applied to a Markov chain with infinite state space. In fact, the matrix notation is only used to facilitate the computation (see next section for the Chapman-Kolmogorov result) and to apply linear algebra results where appropriate. If we take care to be precise, the parallels of those may be applied to the infinite dimensional case as well.*

Example 1.3. *Random walk as a time homogeneous Markov chain*

Let S_k be a random walk. Then

$$P(S_{k+1} = j | S_k = i) = P(X_{k+1} = j - i) = P(X_k = j - i) = P(S_k = j | S_{k-1} = i),$$

since the X_k's are identical. Thus we can denote $P_{ij} = P(S_{k+1} = j | S_k = i)$ without worrying about the dependence on k.

Example 1.4. *Transforming the state space to make a process Markov*

Some process doesn't seem to be Markov at first glance. For example, if the weather tomorrow depends on both the weather today and yesterday, then it is clear that a day to day weather status is not a Markov process. On the other hand, we can transform our state space to represent the weather status of *2 consecutive days*. Then this new process is Markov since clearly the weather of *today and tomorrow* depends on the weather of *today and yesterday*.

Example 1.5. *Transition probability of a birth-death process*

Suppose the numbers of families that check into a hotel on successive days are independent Poisson RVs with mean λ. Also suppose that the number of days that a family stays in the hotel is a Geometric(p) RV (that is a family checks out the next day with probability p independently of how long they have stayed. Suppose all families act independently of one another. Show that X_n, the number of families staying at the hotel on day n is a Markov process. Find its transition probability.

Ans: Let Y_n be the number of newly arrived families on day n and Z_n be the number of checked out families on day n (e.g. on day 0 no family checks out and some family checks in). Then it is clear that Y_n, Z_n are independent of $X_i, i < n$. Moreover,

$$X_{n+1} = X_n + Y_{n+1} - Z_{n+1}.$$

Thus X_n is a random walk with step size $Y_{n+1} - Z_{n+1}$. This is why it is a Markov chain. To find the transition probability $P_{ij} = P(X_{n+1} = j | X_n = i)$ we need to understand the distribution of $(Y_{n+1}, Z_{n+1}) | X_n$. It is clear that Y_{n+1} is a Poisson (λ) RV independent of X_n. On the other hand, $Z_{n+1} | X_n$ is a Binomial(X_n, p). Therefore,

$$P(X_{n+1} = j | X_n = i) = P(Y - Z = j - i),$$

where Y, Z are independent, Y is a Poisson λ and Z is a Binomial(i,p). We have

$$\begin{aligned} P(Y - Z = j - i) &= \sum_k P(Y - Z = j - i | Z = k) P(Z = k) \\ &= \sum_{k=\max(i-j,0)}^{i} P(Y = j - i + k) P(Z = k) \\ &= \sum_{k=\max(i-j,0)}^{i} e^{-\lambda} \frac{\lambda^{j-i+k}}{(j-i+k)!} \binom{i}{k} p^k (1-p)^{i-k}, \end{aligned}$$

where we sum k from $\max(i - j, 0)$ to guarantee that $j - i + k \geq 0$.

2 Chapman-Kolmogorov Equations

Let X_k be a Markov chain with transition matrix P. We want to compute

$$P(X_k = j | X_{k-2} = i)$$

in terms of P. From the law of total probability:

$$\begin{aligned} P(X_k = j, X_{k-2} = i) &= \sum_n P(X_k = j, X_{k-1} = n, X_{k-2} = i) \\ &= \sum_n P(X_k = j | X_{k-1} = n, X_{k-2} = i) P(X_{k-1} = n, X_{k-2} = i) \\ &= \sum_n P(X_k = j | X_{k-1} = n, X_{k-2} = i) P(X_{k-1} = n | X_{k-2} = i) P(X_{k-2} = i). \end{aligned}$$

Thus it is clear that

$$P(X_k = j | X_{k-2} = i) = \sum_k P_{ik} P_{kj},$$

which is a matrix product entry. Note that the above can be an infinite sum (why must it converge in this case?) This justifies us to denote

$$P_{ij}^2 \triangleq P(X_k = j | X_{k-2} = i) = (P \times P)_{ij}.$$

One can easily see by induction how one can extend the above procedure to conclude that

$$P_{ij}^n \triangleq P(X_{k+n} = j | X_k = i) = (P^n)_{ij},$$

where the RHS is understood as the matrix P raised to the nth power and the LHS is a definition. Finally, for any s, t such that $s + m + n = t$ we have

$$\begin{aligned} P(X_t = j, X_s = i) &= \sum_k P(X_t = j, X_{t-n} = k, X_s = i) \\ &= \sum_k P(X_t = j | X_{t-n} = k, X_s = i) P(X_{t-n} = k | X_s = i) P(X_s = i). \end{aligned}$$

Thus

$$\begin{aligned} P_{ij}^{m+n} &= \sum_k P_{ik}^m P_{kj}^n \\ &= \sum_k P_{ik}^n P_{kj}^m (\text{ by a similar reasoning as above }), \end{aligned}$$

which is commutativity of matrix products. This is referred to as the Chapman-Kolmogorov equation.

Example 2.1. *Unconditional distribution of a Markov chain*

The entry P_{ij}^n gives the conditional probability $P(X_n = j | X_0 = i)$. On the other hand, suppose we are given an initial distribution vector π such that $\pi_i = P(X_0 = i)$. Then we have

$$\begin{aligned} P(X_n = j) &= \sum_i P(X_n = j | X_0 = i) P(X_0 = i) \\ &= \sum_i \pi_i P_{ij}^n. \end{aligned}$$

In terms of matrix notation, if we denote π^n as the distribution vector of X_n then $(\pi^n)^T = \pi^T P^n$. That is we need to **left** multiply the initial distribution vector with the transition matrix to get to the *unconditional* distribution of the chain at a future time. The transpose is because we conventionally think of a vector as a column vector.

Example 2.2. *Markov chain distribution at any time point*

An urn always contains 2 balls with possible colors red and blue. At each turn, a ball is selected at random and replaced with a ball of the other color with probability 0.2 and kept the same with probabiltiy 0.8. Initially both balls are red. Find the probability that the fifth ball *selected* is red.

Ans: There are several ways to approach this problem. It may seem natural to model the color of the ball selected. But this is not the best idea, nor is it necessarily a Markov chain, see Ross problem 4.9 ! A better idea is to model the color composition of the balls in the urn at any time. This gives us 3 possibilities : RR, RB, BB, with corresponding states $0, 1, 2$ respectively. Some transition probability is obvious : $P_{02} = P_{20} = 0$. On the other hand

$$\begin{aligned}
P_{01} &= 0.2, P_{00} = 0.8 \\
P_{10} &= .1, P_{11} = 0.8, P_{12} = .1 \\
P_{21} &= 0.2, P_{22} = 0.8
\end{aligned}$$

This gives us the transition matrix

$$P = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0.1 & 0.8 & 0.1 \\ 0 & 0.2 & 0.8 \end{bmatrix}.$$

Initially both balls are red means we start with the initial distribution

$$v_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

The comment here is that we can start with any initial distribution, such as

$$v_0 = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}.$$

Using $v_0^T = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$, the distribution of the urn at time 4 is

$$v_0^T P^4 = \begin{bmatrix} 0.4872 \\ 0.4352 \\ 0.0776 \end{bmatrix},$$

which is the (transpose of) the first row of P^4. Note how $P_{02}^4 \neq 0$ while $P_{02} = 0$. Note also that this is the *unconditional distribution* of the states of the urn at step 4, since we've already **left** multiplied with a given initial distribution.

Picking a ball randomly from the 0th state has 100% chance of being a red while picking from the 1st state has a 50% chance. Thus the answer is

$$0.4872 + 0.4352 \times 0.5 = 0.7048.$$

Example 2.3. *A multidimensional Markov chain*

Consider a Markov chain X with r possible states and assume an individual changes states in accordance with this chain with transition probabilities P_{ij}. The individuals are assumed to move through the system independently. Suppose the numbers of people initially in states $1, 2, \cdots$ are independent Poisson RVs with means $_1, \lambda_2, \cdots, \lambda_r$ respectively. Find the joint distribution of the numbers of individuals in states $1, 2, \cdots, r$ at some time n.

Ans: First, note that this problem does not give an initial distribution of the Markov chain X. We are in fact interested in another multidimensional Markov chain Y (whose dimension is r) that keeps track of the number of individuals at each state of X (Why is Y Markov and what is its transition probabilitiy?). The problem thus gives the initial distribution of Y (what is it?) and asks us to find the general distribution of Y at any time n.

Let's analyze the first step for $n = 1$. From the point of view of a fixed state j, the number of potential individuals coming from state i is Poisson(λ_i). Of those, there are two types : the ones that actually make it and the ones that don't (think of the yoga studio problem with male and female customers again). For any individual from state i, the probability that they make it to state j is P_{ij}. Therefore the distribution of individuals coming to state j from state i is Poisson($\lambda_i P_{ij}$). A similar argument applies to all other states $r \neq i$. Therefore, the distribution of individuals at state j after one step is Poisson($\sum_i \lambda_i P_{ij}$).

A similar argument also applies to n time step. Thus the distribution of Y_{nj} the number of individuals at state j after n steps is Poisson($\sum_i \lambda_i P_{ij}^n$).

Example 2.4. *Birth-death process example revisited*

In the hotel example above, find the expected number of families in the hotel on day n given that there were i families on the first day.

Ans : The question asks for $E(X_n|X_0 = i)$. We have

$$\begin{aligned} E(X_n|X_0) &= E(E(X_n|X_{n-1})|X_0) \\ &= E(X_{n-1} + \lambda - pX_{n-1}|X_0) \\ &= E((1-p)X_{n-1} + \lambda|X_0). \end{aligned}$$

Repeating this relation one more time gives

$$\begin{aligned} E(X_n|X_0) &= E((1-p)((1-p)X_{n-2} + \lambda) + \lambda|X_0) \\ &= E((1-p)^2 X_{n-2} + \lambda(1 + (1-p))|X_0). \end{aligned}$$

Thus we can see that

$$\begin{aligned} E(X_n|X_0) &= \lambda\sum_{i=0}^{n-1}(1-p)^i + (1-p)^n X_0 \\ &= \lambda\frac{1-(1-p)^n}{p} + (1-p)^n X_0. \end{aligned}$$

Example 2.5. *Transforming a transition matrix using absorbing state*

Balls are successively distributed equally likely among 8 urns. What is the probability that there are exactly 3 nonempty urns after 9 balls have been distributed?

Ans: This should remind us of the key collision problem. It is clear that we want to let X_k be the number of *nonempty* urns at time k, with state space $i = 1, \cdots, 8$. Then we see that

$$\begin{aligned} P(X_{k+1} = i|X_k = i) &= \frac{i}{8} \\ P(X_{k+1} = i+1|X_k = i) &= \frac{8-i}{8}. \end{aligned}$$

We can write down the 8×8 transition matrix and take it to the 9th power. On the other hand, this is a non-decreasing Markov chain, and so we can treat state 4 and above as absorbing state (since we're only interested in state 3 probability). The modified transition matrix with 4 states is:

$$P = \begin{bmatrix} 1/8 & 7/8 & 0 & 0 \\ 0 & 2/8 & 6/8 & 0 \\ 0 & 0 & 3/8 & 5/8 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

with initial distribution

$$v_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Raising the transition matrix to the 8th power (instead of 9th since we do not include the state $i = 0$ into the transition matrix) and look at the first row, we have

$$v_0^T P^8 = \begin{bmatrix} 0 \\ 0 \\ 0.00757 \\ 0.9923 \end{bmatrix}.$$

Thus the probability that there are exactly 3 non-empty urns after 9 balls have been distributed is 0.00757. Note that it is almost guaranteed that 4 or more urns are non-empty after 9 balls have been distributed.

3 Absorbing probability before time N

The last example in the previous section gives a general technique to approach the following question : given a sub-collection of states $\mathcal{A}$, and a starting state $i \notin \mathcal{A}$ we want to ask the probability that X_k reaches $\mathcal{A}$ before a time horizon N. That is

$$P(X_k \in \mathcal{A}, \text{ for some } k = 1, 2, \cdots, N | X_0 = i) =?$$

This question is not easy to keep track of in terms of the original transition matrix. On the other hand, we can transform the original Markov chain X_k into a more suitable one Y_k where the state space of Y is the same as X for those not in $\mathcal{A}$ plus one absorbing "state" $\mathcal{A}$. Let P^X, P^Y be the transition matrices of X, Y respectively. Then

$$\begin{aligned} P^Y_{ij} &= P^X_{ij}, \quad i, j \notin \mathcal{A} \\ P^Y_{i\mathcal{A}} &= \sum_{j \in \mathcal{A}} P^X_{ij}, i \notin \mathcal{A} \\ P^Y_{\mathcal{A}\mathcal{A}} &= 1. \end{aligned}$$

The original probability in question can be answered as

$$P(X_k \in \mathcal{A}, \text{ some } 1 \leq k \leq N | X_0 = i) = P(Y_N = \mathcal{A} | Y_0 = i) = (P^Y)^N_{i\mathcal{A}}.$$

Example 3.1. *Understanding the absorbing probability*

Suppose a Markov chain X_k with 5 possible states $\{1, 2, 3, 4, 5\}$ has the following transition probability matrix

$$\begin{bmatrix} 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 0 & 1/3 & 1/6 & 1/3 & 1/6 \\ 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 \end{bmatrix}$$

We are interested in the questions of the type

$$P(X_k \geq 3, \text{ some } k = 1, 2, \cdots, n | X_0 = 1),$$

for certain value of n. From the discussion above, we view states $\{3,4,5\}$ as absorbing and for convenience just relabel it as state 3. We'll also refer to this "new" chain as Y_k. The new transition matrix becomes

$$\begin{bmatrix} 1/5 & 1/5 & 3/5 \\ 0 & 1/3 & 2/3 \\ 0 & 0 & 1 \end{bmatrix}$$

and answer to the above question in general is just $(Q^n)_{13}$. That is we also assert that

$$\begin{aligned} P(X_k \geq 3,\ \text{some } k = 1,2,\cdots,n|X_0 = 1) &= P(Y_n = 3|Y_0 = 1) \\ &= P(Y_k = 3,\ \text{some } k = 1,2,\cdots,n|Y_0 = 1). \end{aligned}$$

Let's take a closer look into why this is true.

For $n = 1$

$$P(X_1 \geq 3|X_0 = 1) = \sum_{j=3}^{5} P_{1j} = Q_{13} = P(Y_1 = 3|Y_0 = 1)$$

For $n = 2$

$$\begin{aligned} P(X_1 \geq 3 \text{ or } X_2 \geq 3|X_0 = 1) &= P(X_1 \geq 3|X_0 = 1) + P(X_2 \geq 3, X_1 < 3|X_0 = 1) \\ &= \sum_{j=3}^{5} P_{1j} + \sum_{i=1}^{2}\sum_{j=3}^{5} P_{1i}P_{ij} \\ &= \sum_{j=3}^{5} P_{1j} + \sum_{i=1}^{2} P_{1i} \sum_{j=3}^{5} P_{ij} \\ &= Q_{13} + \sum_{i=1}^{2} Q_{1i}Q_{i3} \\ &= \sum_{i=1}^{3} Q_{1i}Q_{i3} \text{ because } Q_{33} = 1 \\ &= (Q^2)_{13} = P(Y_2 = 3|Y_9 = 1). \end{aligned}$$

Suppose this is true up to case n. For case $n+1$

$$\begin{aligned} P(X_k \geq 3,\ \text{some } k = 1,2,\cdots,n+1|X_0 = 1) &= P(X_{n+1} \geq 3; X_i < 3, i = 1,2,\cdots,n|X_0 = 1) \\ &+ P(X_k \geq 3,\ \text{some } k = 1,\cdots,n|X_0 = 1). \end{aligned}$$

Now

$$
\begin{aligned}
P(X_{n+1} \geq 3; X_i < 3, i = 1, 2, \cdots, n | X_0 = 1) &= P(Y_{n+1} = 3; Y_i < 3, i = 1, 2, \cdots, n | Y_0 = 1) \\
P(X_k \geq 3, \text{ some } k = 1, \cdots, n | X_0 = 1) &= P(Y_n = 3 | Y_0 = 1) = (Q^n)_{13}
\end{aligned}
$$

Because state 3 is absorbing for Y

$$
\begin{aligned}
P(Y_{n+1} = 3; Y_i < 3, i = 1, 2, \cdots, n | Y_0 = 1) &= P(Y_{n+1} = 3, Y_n < 3 | Y_0 = 1) \\
&= P(Y_{n+1} = 3 | Y_n = 2) P(Y_n = 2 | Y_0 = 1) \\
&+ P(Y_{n+1} = 3 | Y_n = 1) P(Y_n = 1 | Y_0 = 1) \\
&= (Q^n)_{11} Q_{13} + (Q^n)_{12} Q_{23}.
\end{aligned}
$$

Putting it together, we have

$$
\begin{aligned}
P(X_k \geq 3, \text{ some } k = 1, 2, \cdots, n+1 | X_0 = 1) &= (Q^n)_{13} Q_{33} + (Q^n)_{11} Q_{13} + (Q^n)_{12} Q_{23} \\
&\quad (\text{ again because } Q_{33} = 1) \\
&= (Q^{n+1})_{13}.
\end{aligned}
$$

Finally we want to address the question

$$
P(X_k \geq 3, \text{ some } k = 1, 2, \cdots, n | X_0 = j), j = 3, 4, 5.
$$

Clearly the previous technique described in this example does not work, because states $\{3, 4, 5\}$ are not actually absorbing for X_k. The solution is to compute what happens to X_1 and condition on that. For simplicity suppose $j = 3$. We have

$$
\begin{aligned}
P(X_k \geq 3, \text{ some } k = 1 \cdots, n | X_0 = j) &= P(X_k \geq 3, \text{ some } k = 2, \cdots, n, X_1 < 3 | X_0 = j) \\
&+ P(X_1 \geq 3 | X_0 = j).
\end{aligned}
$$

Now

$$
\begin{aligned}
P(X_k \geq 3, \text{ some } k = 2, \cdots, n, X_1 < 3 | X_0 = j) &= \sum_{i=1}^{2} P(X_k \geq 3, \text{ some } k = 2, \cdots, n | X_1 = i) \times \\
&\quad P(X_1 = i | X_0 = j)
\end{aligned}
$$

and we can use our previous technique to deal with this computation.

Example 3.2. A pensioner receives 2 (thousand) dollars at the beginning of each month. He spends an amount of $i = 1, 2, 3, 4$ during the month with probability p_i, independent of how much he currently has. Also if he has more than 3 at the end of a month he gives the excess amount to his son. Suppose he has 5 at the beginning of a month. What is the probability that he has 1 or less within the following 4 months?

Ans: Let X_k be the amount he has at the beginning of each month. X_k is like a random walk with step size $-2, -1, 0, 1$ with the difference that $X_k \leq 3$. In fact we have $P_{33} = p_1 + p_2$. We treat state 1 as an absorbing state. Then X has only $i = 1, 2, 3$ as possible states and we have the following transition matrix

$$P = \begin{bmatrix} 1 & 0 & 0 \\ p_3 + p_4 & p_2 & p_1 \\ p_4 & p_3 & p_1 + p_2 \end{bmatrix}.$$

Finally as he starts at 5 during the first month he will end up with 3. Thus we have

$$v_0 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

The answer is either $v_0^T P^3$ or $v_0^T P^4$, depending on how one counts "the following 4 months."

Example 3.3. *Absorbing probability with constraint on the final step*

A Markov chain has 5 states $i = 1, 2, 3, 4, 5$. What is

$$P(X_4 = 2, X_3 \leq 2, X_2 \leq 2, X_1 \leq 2 | X_0 = 1)?$$

Ans: This question clearly refers to states $\{3, 4, 5\}$ as an absorbing state with an additional specification that the chain ends up at state 2 at step 4. First, it seems easier to compute

$$P(X_4 \leq 2, X_3 \leq 2, X_2 \leq 2, X_1 \leq 2 | X_0 = 1)$$

because the complement of that is

$$P(X_k > 2, \text{ for some } k = 1, \cdots, 4 | X_0 = 1).$$

The verbal interpretation of the above quantity is the probability that the chain ever enters states 3, 4 , 5 in steps 1 to 4, starting at state 1. The new transition matrix is

$$Q = \begin{bmatrix} p_{11} & p_{12} & \sum_{i=3}^{5} p_{1i} \\ p_{21} & p_{22} & \sum_{i=3}^{5} p_{2i} \\ 0 & 0 & 1 \end{bmatrix}.$$

Then

$$P(X_4 \le 2, X_3 \le 2, X_2 \le 2, X_1 \le 2 | X_0 = 1) = 1 - (Q^4)_{13}.$$

On the other hand, since $X_4 = 2$ is the last step in consideration, we actually can use the same set up as above, treating $\{3, 4, 5\}$ as an absorbing state. In fact, it is just

$$P(X_4 = 2, X_3 \le 2, X_2 \le 2, X_1 \le 2 | X_0 = 1) = (Q^4)_{12}.$$

The verbal interpretation of the above quantity is the probability that the chain never enters into states 3, 4 , 5 in steps 1 to 4 AND also end up at step 2 in state 4. Note how the question will be very different if we were to ask

$$P(X_4 \le 2, X_3 = 2, X_2 \le 2, X_1 \le 2 | X_0 = 1) =?$$

Example 3.4. *Absorbing probability with general constraint on the last step*

We see from the previous example that if $\mathcal{A}$ is an absorbing state then for $i, j \notin A$

$$P(X_m = j, X_k \notin A, k = 1, \cdots, m-1 | X_0 = i) = (Q^m)_{ij}.$$

What about the case when $j \in \mathcal{A}$ or $i \in \mathcal{A}$? (remember here that $\mathcal{A}$ is only formally an absorbing state so if $i \in \mathcal{A}$ the original chain can "escape" from $\mathcal{A}$)

First, suppose $j \in \mathcal{A}$ and $i \notin \mathcal{A}$. Then Q^{m-1} represents all possible outcomes of the chain at the $m-1$ step, not getting absorbed in $\mathcal{A}$. All we need to do is from the step $m-1$ compute the probability of arriving at state j from all states outside $\mathcal{A}$ using the original transition matrix P. That is

$$P(X_m = j, X_k \notin A, k = 1, \cdots, m-1 | X_0 = i) = \sum_{r \notin \mathcal{A}} Q_{ir}^{m-1} P_{rj}.$$

Now suppose $i \in A$. In this case we can move to an initial distribution vector v_1 where v_1 is the probability of moving from state i to a state j outside of $\mathcal{A}$ in one step.Then we can repeat the above procedure with $m-1$ steps. Namely the answer is

$$\begin{aligned} & \sum_{r \notin \mathcal{A}} P(X_m = j, X_k \notin A, k = 2, \cdots, m | X_1 = r) P_{ir} \\ = & \sum_{r \notin \mathcal{A}} P(X_{m-1} = j, X_k \notin A, k = 1, \cdots, m-1 | X_0 = r) P_{ir} \end{aligned}$$