1 Discrete-time Markov chains

1.1 Stochastic processes in discrete time

A *stochastic process* in discrete time $n \in \ = \{0, 1, 2, \ldots\}$ is a sequence of random variables (rvs) $X_0, X_1, X_2, \ldots$ denoted by $\mathbf{X} = \{X_n : n \geq 0\}$ (or just $\mathbf{X} = \{X_n\}$). We refer to the value X_n as the *state* of the process at time n, with X_0 denoting the initial state. If the random variables take values in a discrete space such as the integers $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ (or some subset of them), then the stochastic process is said to be discrete-valued; we then denote the states by i, j and so on. In general, however, the collection of possible values that the X_n can take on is called the *state space*, is denoted by $\mathcal{S}$ and could be, for example, $d-$ dimensional Euclidean space $\mathbb{R}^d$, $d \geq 1$, or a subset of it.

Stochastic processes are meant to model the evolution over time of real phenomena for which randomness is inherent. For example, X_n could denote the price of a stock n days from now, the population size of a given species after n years, the amount of bandwidth in use in a telecommunications network after n hours of operation, or the amount of money that an insurance risk company has right after it pays out its n^{th} claim. The insurance risk example illustrates how "time" n need not really be time, but instead can be a sequential indexing of some kind of events. Other such examples: X_n denotes the amount of water in a reservoir after the n^{th} rain storm of the year, X_n denotes the amount of time that the n^{th} arrival to a hospital must wait before being admitted, or X_n denotes the outcome (heads or tails) of the n^{th} flip of a coin.

The main challenge in the stochastic modeling of something is in choosing a model that has – on the one hand – enough complexity to capture the complexity of the phenomena in question, but has – on the other hand – enough structure and simplicity to allow one to compute things of interest. In the context of our examples given above, we may be interested in computing $P(X_{30} > 50)$ for a stock that we bought for $X_0 = \$35$ per share, or computing the probability that the insurance risk company eventually gets ruined (runs out of money), $P(X_n < 0, \text{for some } n > 0)$, or computing the long-run average waiting time of arrivals to the hospital

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} X_n.$$

As a very simple example, consider the sequential tossing of a "fair" coin. We let X_n denote the outcome of the n^{th} toss. We can take the X_n as $p = 0.5$ Bernoulli rvs, $P(X_n = 0) = P(X_n = 1) = 0.5$, with $X_n = 1$ denoting that the n^{th} flip landed heads, and $X_n = 0$ denoting that it landed tails. We also would assume that the sequence of rvs are independent. This then yields an example of an *independent and identically distributed* (iid) sequence of rvs. Such sequences are easy to deal with for they are defined by a single distribution (in this case Bernoulli), and are independent, hence lend themselves directly to powerful theorems in probability such as the strong law of large numbers and the central limit theorem.

For the other examples given above, however, an iid sequence would not capture enough complexity since we expect some correlations among the rvs X_n. For example, in the hospital example, if the waiting time X_n is very large (and arrivals wait "first-in-first-out") then we would expect X_{n+1} to be very large as well. In the next section we introduce a stochastic process called a Markov chain which does allow for correlations and also has enough structure and simplicity to allow for computations to be carried out. We will also see that Markov chains can be used to model a number of the above examples.

1.2 Definition of a Markov chain

We shall assume that the state space $\mathcal{S} = \mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, the integers, or a proper subset of the integers. Typical examples are $\mathcal{S} = \ = \{0, 1, 2 \ldots\}$, the non-negative integers, or $\mathcal{S} = \{0, 1, 2 \ldots, a\}$, or $\mathcal{S} = \{-b, \ldots, 0, 1, 2 \ldots, a\}$ for some integers $a, b > 0$, in which case the state space is finite.

Definition 1.1 *A stochastic process $\{X_n\}$ is called a Markov chain if for all times $n \geq 0$ and all states $i_0, \ldots, i, j \in \mathcal{S}$,*

$$\begin{aligned} P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) &= P(X_{n+1} = j | X_n = i) \\ &= P_{ij}. \end{aligned} \tag{1}$$

P_{ij} denotes the probability that the chain, whenever in state i, moves next (one unit of time later) into state j, and is referred to as a *one-step transition probability*. The square matrix $\mathbf{P} = (P_{ij}),\ i, j \in \mathcal{S}$, is called the *one-step transition matrix*, and since when leaving state i the chain must move to one of the states $j \in \mathcal{S}$, each row sums to one (e.g., forms a probability distribution): For each $i \in \mathcal{S}$

$$\sum_{j \in \mathcal{S}} P_{ij} = 1.$$

We are assuming that the transition probabilities do not depend on the time n, and so, in particular, using $n = 0$ in (1) yields

$$P_{ij} = P(X_1 = j | X_0 = i).$$

The defining Markov property (1) can be described in words as *the future is independent of the past given the present state.* Letting n be the present time, the future after time n is $\{X_{n+1}, X_{n+2}, \ldots\}$, the present state is X_n, and the past is $\{X_0, \ldots, X_{n-1}\}$. If the value $X_n = i$ is known, then the future evolution of the chain only depends (at most) on i, in that it is stochastically independent of the past values $X_{n-1}, \ldots, X_0$.

> Conditional on the rv X_n, the future sequence of rvs $\{X_{n+1}, X_{n+2}, \ldots\}$ is independent of the past sequence of rvs $\{X_0, \ldots, X_{n-1}\}$.

The Markov property extends to stochastic processes that have non-discrete state spaces as well such as $\mathbb{R}$ or $\mathbb{R}^d$; they are more generally called *Markov Processes*.

Examples of Markov chains

1. *Any independent and identically distributed (iid) sequence:*

   Any iid sequence forms a Markov chain, for if $\{X_n\}$ is iid, then $\{X_{n+1}, X_{n+2}, \ldots\}$ (the future) is independent of $\{X_0, \ldots, X_{n-1}\}$ (the past) given X_n (the present). In fact $\{X_{n+1}, X_{n+2}, \ldots\}$ is independent of $\{X_0, \ldots, X_n\}$ (the past and the present): *For an iid sequence, the future is independent of the past and the present state.* Let $p(j) = P(X = j)$ denote the common probability mass function (pmf) of the X_n. Then $P_{ij} = P(X_1 = j|X_0 = i) = P(X_1 = j) = p(j)$ because of the independence of X_0 and X_1; P_{ij} does not depend on i: Each row of $\mathbf{P}$ is the same, namely the pmf $(p(j))$.

   An iid sequence is a very special kind of Markov chain; whereas a Markov chain's future is allowed (but not required) to depend on the present state, an iid sequence's future does not depend on the present state at all.

2. *Rat in the open maze:* Consider a rat in a maze with four cells, indexed $1-4$, and the outside (freedom), indexed by 0 (that can only be reached via cell 4). The rat starts initially in a given cell and then takes a move to another cell, continuing to do so until finally reaching freedom. We assume that at each move (transition) the rat, independent of the past, is equally likly to choose from among the neighboring cells (so we are assuming that the rat does not learn from past mistakes). This then yields a Markov chain, where X_n denotes the cell visited right after the n^{th} move. $\mathcal{S} = \{0, 1, 2, 3, 4\}$. For example, whenever the rat is in cell 1, it moves next (regardless of its past) into cell 2 or 3 with probability 1/2; $P_{1,2} = P_{1,3} = 1/2$. We assume that when the rat escapes it remains escaped forever after, so we have $P_{0,0} = 1,\ P_{0,i} = 0,\ i \in \{1, 2, 3, 4\}$. The transition matrix is given by

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \end{pmatrix}$$

   State 0 here is an example of an *absorbing* state: Whenever the chain enters state 0, it remains in that state forever after; $P(X_{n+1} = 0 \mid X_n = 0) = P_{00} = 1$. Of interest is determining the expected number of moves required until the rat reaches freedom given that the rat starts initially in cell i. Let $\tau_{i,0} = \min\{n \geq 0 : X_n = 0|X_0 = i\}$, the number of moves required to reach freedom when starting in cell i. We wish to compute $E(\tau_{i,0})$. The trick is to condition on X_1. For example, let us try to compute $E(\tau_{1,0})$; we thus assume that $X_0 = 1$. Then $X_1 = 2$ or 3 w.p. 1/2, and $E(\tau_{1,0}|X_1 = 2) = 1 + E(\tau_{2,0})$; $E(\tau_{1,0}|X_1 = 3) = 1 + E(\tau_{3,0})$. The point is that the rat must take at least 1 step to get out, and if the first step is to cell 2, then by the Markov property, the remaining number of steps is as if the rat started initially in cell 2 and we wish to calculate $E(\tau_{2,0})$, the expected number of steps required to each freedom from cell 2; similarly if $X_1 = 3$. Thus

$$\begin{aligned} E(\tau_{1,0}) &= E(\tau_{1,0}|X_1 = 2)P(X_1 = 2|X_0 = 1) + E(\tau_{1,0}|X_1 = 3)P(X_1 = 3|X_0 = 1) \\ &= (1 + E(\tau_{2,0}))(1/2) + (1 + E(\tau_{3,0}))(1/2) \\ &= 1 + E(\tau_{2,0})(1/2) + E(\tau_{3,0})(1/2). \end{aligned}$$

Using the same trick on each of $E(\tau_{2,0})$,$E(\tau_{3,0})$,$E(\tau_{4,0})$ yields, in the end, four linear equations with the four unknowns. $E(\tau_{40})$ is a little different since it is only from cell 4 that the rat can escape;
$E(\tau_{40}) = \frac{1}{3}(1) + \frac{1}{3}(1 + E(\tau_{20})) + \frac{1}{3}(1 + E(\tau_{30})) = 1 + \frac{1}{3}E(\tau_{20}) + \frac{1}{3}E(\tau_{30})$. The full set of equations is

$$\begin{aligned}
E(\tau_{10}) &= 1 + \frac{1}{2}E(\tau_{20}) + \frac{1}{2}E(\tau_{30}) \\
E(\tau_{20}) &= 1 + \frac{1}{2}E(\tau_{10}) + \frac{1}{2}E(\tau_{40}) \\
E(\tau_{30}) &= 1 + \frac{1}{2}E(\tau_{10}) + \frac{1}{2}E(\tau_{40}) \\
E(\tau_{40}) &= 1 + \frac{1}{3}E(\tau_{20}) + \frac{1}{3}E(\tau_{30})
\end{aligned}$$

Solving (details left to the reader) yields

$$\begin{aligned}
E(\tau_{10}) &= 13 \\
E(\tau_{20}) &= 12 \\
E(\tau_{30}) &= 12 \\
E(\tau_{40}) &= 9
\end{aligned}$$

3. *Random walk:* Let $\{\Delta_n : n \geq 1\}$ denote any iid sequence (called the *increments*), and define

$$X_n \stackrel{\text{def}}{=} \Delta_1 + \cdots + \Delta_n, \; n \geq 1, \; X_0 = 0. \tag{2}$$

The Markov property follows since $X_{n+1} = X_n + \Delta_{n+1}, \; n \geq 0$ which asserts that the future, given the present state, only depends on the present state X_n and an independent (of the past) r.v. Δ_{n+1}.

When $P(\Delta = 1) = p, \; P(\Delta = -1) = 1 - p$, then the random walk is called a *simple random walk*, and can be thought of as representing the step-by-step position of a randomly moving particle on the integer lattice: Each step the particle either moves one unit distance to the right or to the left with probabilities p and $1 - p$ respectively. When $p = 1/2$ the process is called the simple *symmetric* random walk. Since the chain's value can only go up or down by 1 at each step, we see that $P_{i,i+1} = p$, $P_{i,i-1} = 1 - p$ and all other transition probabilities are zero. When $p > 1 - p$ (equivalently, $p > 1/2$), then the random walk is said to have *positive drift* because it can be proved via the strong law of large numbers (SLLN) that $X_n \to \infty$ as $n \to \infty$ wp1.; while when $p < 1 - p$ (equivalently, $p < 1/2$), then the random walk is said to have *negative drift* because it can be proved via the strong law of large numbers (SLLN) that $X_n \to -\infty$ as $n \to \infty$ wp1.

Requiring that $X_0 = 0$ is not necessary, we can start with any deterministic state $X_0 = i$ in which case the process is called a random walk started from state i.

4. *Random walk with restricted state space:* Random walks can also be restricted to stay within a subset of states, $\{0, 1, 2, \ldots\}$, or $\{0, 1, 2, \ldots, N\}$ for example. Let

us consider the simple random walk restricted to stay within $\mathcal{S} = \{0, 1, 2, 3\}$. We must specify what happens when either state 0 or 3 is hit. Let's assume that $P_{0,0} = 1 = P_{3,3}$, meaning that both states 0 and 3 are absorbing. Then the transition matrix is

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

The above can easily be extended to $\mathcal{S} = \{0, 1, 2, \ldots, N\}$, with $P_{0,0} = 1 = P_{N,N}$. Of intrinsic interest for this chain is the probability, when $X_0 = i$, $0 < i < N$, that the chain first hits N (the highest value), before hitting 0 (the lowest value). This is known as the *gambler's ruin problem*, because we can imagine a gambler starting off with i dollars, and then at each gamble either winning a dollar (probability p) or losing a dollar (probability $1-p$). With X_n denoting the total fortune at the end of the n^{th} gamble, the gambler's objective is to reach a fortune of N before ruin (running out of money).

One can also imagine using a random walk to model asset prices over time; $X_n =$ the price per share at the end of the n^{th} day for example. Then one can use the solution to the gambler's ruin problem to compute such things as the probability that the price, currently at \$50, will ever hit \$100 before hitting \$25 (which could be used to analyze barrier options say).

5. *Population model:* Consider a species in which each individual independent of one another gives birth to k new individuals with probability p_k, $k \geq 0$ (this is called the progeny distribution). Let $X_0 > 0$ denote the initial number of individuals, and let $Y_{0,1}, \ldots, Y_{0,X_0}$ denote the corresponding (iid) number of progeny of each such individual; $P(Y_{0,m} = k) = p_k$, $k \geq 0$. Then $X_1 \stackrel{\text{def}}{=} \sum_{m=1}^{X_0} Y_{0,m}$ denotes the population size of the first generation. In general, letting $Y_{n,m}$ denote the (iid) number of progeny of the m^{th} individual from the n^{th} generation yields the recursive stochastic process

$$X_n = \sum_{m=1}^{X_{n-1}} Y_{n-1,m}, \; n \geq 1,$$

denoting the population size of the n^{th} generation. This is known as the Galton-Watson model, and it forms a Markov chain on $\mathcal{S} = \{0, 1, 2, \ldots\}$ with 0 as an absorbing state. For a given initial state and a given progeny distribution, it is of interest to compute the probability that the population will become extinct (e.g., $X_n = 0$ for some n).

6. *Products of iid rvs: The Binomial Lattice Model:* If $\{Y_k : k \geq 1\}$ is an iid sequence of rvs, then, with X_0 the initial value (assumed independent of the sequence $\{Y_k\}$,

$$X_n = X_0 Y_1 \times Y_2 \times \cdots Y_n, \; n \geq 0,$$

defines a Markov chain. This is easily seen since the following recursion holds

$$X_{n+1} = X_n Y_{n+1}, \; n \geq 0,$$

which (as in our proof that random walks are Markovian) asserts that the future, given the present state, only depends on the present state X_n and an independent (of the past) r.v. Y_{n+1}. We would usually allow the state space to be something other than the integers, so in general $\{X_n\}$ forms a Markov process. A famous example is the *Binomial Lattice Model* for modeling the price per share of risky assets (at the end of each day (say)) such as stocks:

$$S_n = S_0 Y_1 \times Y_2 \times \cdots Y_n, \ n \geq 0,$$

where the distribution of the Y_i are given by the two-point distribution $P(Y = u) = p, \ P(Y = d) = 1 - p$, where u, d are given parameters satisfying $0 < d < 1 + r < u$, with r the current interest rate. The initial price per share is S_0 and then its price either goes "up" to uS_0 or "down" to dS_0 yielding S_1. Each day it is as if a coin is flipped to determine what happens next, independent of the past. We will study this model in detail later. The inequality $0 < d < 1 + r < u$ means that if the stock goes down, what we really mean is that it does worse than placing money in the bank at interest rate r, and when it goes up, it does better than placing money in the bank at interest rate r. For example if you put S_0 in the bank today, then tomorrow you would have $S_0(1 + r)$, and after n days you would have $S_0(1 + r)^n$. Alternatively, if you bought a share of the stock at price S_0 today, then after n days it would be worth $S_n = S_0 Y_1 \times Y_2 \times \cdots Y_n$ which is a rv. The bank investment is *risk-free* because there is no randomness, but the stock investment is risky since it has randomness: it might go down many times yielding a bad investment as compared to the bank investment, or it might go up many times yielding a great investment.

1.3 Markov chains as recursions

Proposition 1.1 *Let $f(x, u)$ be a (real-valued) function of two variables and let $\{U_n : n \geq 0\}$ be an iid sequence of random variables. We let U denote a typical such random variable.*

Then the recursion

$$X_{n+1} = f(X_n, U_n), \ n \geq 0, \tag{3}$$

defines a Markov chain. (We of course must specify X_0, making sure it is chosen independent of the sequence $\{U_n : n \geq 0\}$.) The transition probabilities are given by $P_{ij} = P(f(i, U) = j)$.

Proof : It is immediate almost by definition: Given $X_n = i$, $X_{n+1} = f(i, U_n)$ only depends on i and some independent (of the past) random variable U_n; hence the Markov property holds.

The transition probabilities are determined via $P(X_{n+1} = j | X_n = i) = P(f(i, U_n) = j) = P(f(i, U) = j)$. ∎

It turns out that the converse is true as well:

Proposition 1.2 *Every Markov chain can in fact be represented in the form of a recursion*

$$X_{n+1} = f(X_n, U_n), \ n \geq 0,$$

for some function f and some iid sequence $\{U_n\}$. The sequence $\{U_n\}$ can be chosen to be iid with a uniform distribution over the unit interval $(0, 1)$.

Proof : The proof is a consequence of what is called the *inverse transform method* from simulation: For any cumulative distribution function (cdf) $F(x) = P(X \leq x),\ x \in \mathbb{R}$, we can construct a rv X distributed as F by first taking a uniform rv U over $(0, 1)$, and then defining $X \stackrel{\text{def}}{=} F^{-1}(U)$, where $F^{-1}(y)$ is the generalized inverse function of F defined as $F^{-1}(y) = \min\{x : F(x) \geq y\},\ y \in [0, 1]$. (If F is continuous, then this reduces to the inverse function.) To use this: Consider a MC with transition matrix $P = (P_{ij})$. For each state $i \in \mathcal{S}$, define $F_i(x) = P(X_1 \leq x \mid X_0 = i),\ x \in \mathbb{R}$, the cdf of the i^{th} row of P. Let $F_i^{-1}(y),\ y \in [0, 1]$ denote its generalized inverse. Define $f(i, u) = F_i^{-1}(u),\ i \in \mathcal{S},\ u \in [0, 1]$. Now take $\{U_n\}$ as iid with a uniform distribution over the unit interval $(0, 1)$. Applying Proposition 1.1 to this f with our iid uniforms $\{U_n\}$ yields that the process $\{X_n\}$ defined by (3) is a MC. The inverse transform method ensures that it has the same transition matrix P as the original MC: For each i, the random variable $F_i^{-1}(U)$ has the same distribution as the i^{th} row of P. ■

1.3.1 Recursive Examples

Here we illustrate Proposition 1.2 with some examples.

1. *Random walk:* The random walk with iid increments $\{\Delta_n : n \geq 1\}$, defined in (2) was already seen to be in recusive form, $X_{n+1} = X_n + \Delta_{n+1}$. Letting $U_n = \Delta_{n+1},\ n \geq 0,\ f(x, u) = x + u$ is the desired function. Thus $P_{ij} = P(i + \Delta = j) = P(\Delta = j - i)$.

2. *Max and Min of iid sequences:* For $\{Y_n : n \geq 0\}$ any iid sequence, both $M_n = \max\{Y_0, \ldots, Y_n\}$ and $m_n = \min\{Y_0, \ldots, Y_n\}$ are Markov chains: $U_n = Y_{n+1}$ and $f(x, u) = \max(x, u),\ f(x, u) = \min(x, u)$ respectively yields the desired recursive representation.

   We now compute the transition probabilities for M_n above. Suppose that $j > i$. Then $P_{ij} = P(M_{n+1} = j | M_n = i) = P(\max(i, Y_{n+1}) = j) = P(Y = j)$, (where Y denotes a typical Y_n). Note that if $j < i$, then $P(M_{n+1} = j | M_n = i) = 0$ since the maximum can never decrease in value.

   Finally, if $j = i$, then $P(M_{n+1} = i | M_n = i) = P(Y \leq i)$; the maximum remains constant at its current value i if the next Y value is less than or equal to i. A similar analysis yields the transition probabilities for m_n.

3. *The Binomial Lattice Model (BLM):* As we saw when defining the BLM, $S_{n+1} = S_n Y_{n+1}$ and so letting $U_n = Y_{n+1},\ n \geq 0,\ f(x, u) = xu$ is the desired function.

1.4 Chapman-Kolmogorov equations and n-step transition probabilities

Given a Markov chain $\{X_n\}$ with transition matrix $\mathbf{P}$, it is of interest to consider the analogous *n-step transition matrix* $\mathbf{P}^{(n)} = (p_{ij}^n),\ n \geq 1$, where $P_{ij}^n \stackrel{\text{def}}{=} P(X_{m+n} = j | X_m = i)$, a n-step transition probability, denotes the probability that n time units later the chain will be in state j given it is now (at time m) in state i. Since transition probabilities do not depend on the time $m \geq 0$ at which the initial condition is chosen, we can without loss of generality choose $m = 0$ and write $P^n_{ij} = P(X_n = j|X_0 = i)$. Also note that $P^{(1)} = P$.

For example, for the rat in the maze chain (Example 1), $P^2_{11} = P(X_2 = 1|X_0 = 1)$ denotes the probability that the rat, starting initially in cell 1, is back in cell 1 two steps later. Clearly, this can happen only if the rat goes to cell 2 then back to cell 1, or goes to cell 3 then back to cell 1, yielding $P^2_{11} = P(X_1 = 2, X_2 = 1|X_0 = 1) + P(X_1 = 3, X_2 = 1|X_0 = 1) = 1/4 + 1/4 = 1/2$. It turns out that in general, computing $\mathbf{P}^{(n)}$ is accomplished via matrix multiplication:

Proposition 1.3

$$\mathbf{P}^{(n)} = \mathbf{P}^n = \mathbf{P} \times \mathbf{P} \times \cdots \times \mathbf{P},\ n \geq 1;$$

$\mathbf{P}^{(n)}$ *is equal to* $\mathbf{P}$ *multiplied by itself* n *times.*

For example, taking the $\mathbf{P}$ in Example 1 for the rat in the maze chain and squaring yields

$$\mathbf{P}^2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ 1/6 & 0 & 5/12 & 5/12 & 0 \\ 1/6 & 0 & 5/12 & 5/12 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \end{pmatrix};$$

in partucular $P^2_{11} = 1/2$ as was derived by direct calculation before.

A proof of Proposition 1.3 is based on the *Chapman-Kolmogorov equations:*

Proposition 1.4 (Chapman-Kolmogorov) *For any* $n \geq 0,\ m \geq 0,\ i \in \mathcal{S},\ j \in \mathcal{S}$,

$$P^{n+m}_{i,j} = \sum_{k \in \mathcal{S}} P^n_{i,k} P^m_{k,j}.$$

The above is derived by first considering what state the chain is in at time n: Given $X_0 = i$, $P^n_{i,k} = P(X_n = k|X_0 = i)$ is the probability that the state at time n is k. But then, given $X_n = k$, the future after time n is independent of the past, so the probability that the chain m time units later (at time $n + m$) will be in state j is $P^m_{k,j}$, yielding the product, $P^n_{i,k} P^m_{k,j} = P(X_n = k, X_{n+m} = j|X_0 = i)$. Summing up over all k yields the result. A rigorous proof of Proposition 1.4 is given next:

Proof :

$$\begin{aligned}
P_{i,j}^{n+m} &= P(X_{n+m} = j | X_0 = i) \\
&= \sum_{k \in \mathcal{S}} P(X_{n+m} = j, X_n = k | X_0 = i) \\
&= \sum_{k \in \mathcal{S}} \frac{P(X_{n+m} = j, X_n = k, X_0 = i)}{P(X_0 = i)} \\
&= \sum_{k \in \mathcal{S}} \frac{P(X_{n+m} = j | X_n = k, X_0 = i) P(X_n = k, X_0 = i)}{P(X_0 = i)} \\
&= \sum_{k \in \mathcal{S}} \frac{P(X_n = k, X_0 = i) P_{k,j}^m}{P(X_0 = i)} \\
&= \sum_{k \in \mathcal{S}} P_{i,k}^n P_{k,j}^m,
\end{aligned}$$

where, in the second to last equality we used the Markov property to conclude that $P(X_{n+m} = j | X_n = k, X_0 = i) = P(X_{n+m} = j | X_n = k) = P(X_m = j | X_0 = k) = P_{k,j}^m$. ■

Proof :[Proposition 1.3] When $n = m = 1$ Chapman-Kolmogorov yields

$$P_{i,j}^2 = \sum_{k \in \mathcal{S}} P_{i,k} P_{k,j}, \ i \in \mathcal{S}, \ j \in \mathcal{S},$$

which in matrix form asserts that $\mathbf{P}^{(2)} = \mathbf{P}^2$. Similarly when $n = 1$ and $m = 2$ Chapman-Kolmogorov yields

$$P_{i,j}^3 = \sum_{k \in \mathcal{S}} P_{i,k} P_{k,j}^2,$$

which in matrix form asserts that $\mathbf{P}^{(3)} = \mathbf{P} \times \mathbf{P}^{(2)}$. But since $\mathbf{P}^{(2)} = \mathbf{P}^2$, we conclude that $\mathbf{P} \times \mathbf{P}^{(2)} = \mathbf{P} \times \mathbf{P}^2 = \mathbf{P}^3$. The proof is completed by induction: Suppose that $\mathbf{P}^{(l)} = \mathbf{P}^l$ for some $l \geq 2$. Then Chapman-Kolmogorov with $n = 1$ and $m = l$ yields $\mathbf{P}^{(l+1)} = \mathbf{P} \times \mathbf{P}^{(l)}$ which by the induction hypothesis is the same as $\mathbf{P} \times \mathbf{P}^l = \mathbf{P}^{l+1}$. ■