1 Notes for Gongchen

1.1 Transition Probability

The **transition probability** from a state $i \in S$ at time n to state $j \in S$ (at time $n+1$) is given by,

$$P_{n,i,j} \equiv P(X_{n+1} = j \mid X_n = i) \qquad n = 0, 1, 2, \dots \tag{1.1}$$

In full generality, the transition probability could depend on time n but in this course we will restrict ourselves to transition probabilities that *do not* depend on time n ($P_{n,i,j} = P_{i,j}$). We say that the markov chain is **(time-)homogeneous** if this property holds. From now on, this will be our default setting.

The matrix of all transition probabilities $P = \{P_{i,j} \mid i, j \in S\}$ is called the **one-step transition (probability) matrix** for $\{X_n \mid n \in T\}$.

$$P = \begin{pmatrix} P_{00} & P_{01} & \cdots & P_{0j} & \cdots \\ P_{10} & P_{11} & \cdots & P_{1j} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \cdots \\ P_{i0} & P_{i1} & \cdots & P_{ij} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

The one-step transition matrix P has the following properties:

1. The entries of P are non-negative:
$$P_{i,j} \geq 0 \tag{1.2}$$
2. The rows of P sum to unity:
$$\forall i : \sum_{j \in S} P_{ij} = 1 \tag{1.3}$$

The **n-step transition probability** is defined via the homogeneous property,

$$\forall i, j \in S : P_{ij}^{(n)} \equiv P(X_{n+m} = j \mid X_n = i) = P(X_n = j \mid X_m = i)$$

Analogously, the **n-step transition matrix** is the matrix,

$$P^{(n)} = \left\{ P_{ij}^{(n)} \mid i, j \in S \right\}$$

Theorem 1. *There is a simple relation between the n-step transition matrix $P^{(n)}$ and the one step transition matrix P.*

$$P^{(n)} = P^{(n-1)} \cdot P = \underbrace{P \cdot P \cdot \dots \cdot P}_{n} = P^n$$

Proof. Proof by induction:

$$P^{(1)} = P \quad \text{By definition.}$$

We also have $P^{(0)} = P^0 = \mathbf{1}$ is the identity matrix. We now assume $P^{(n)} = P^n$. Then $\forall i, j \in S$,

$$\begin{aligned} P_{ij}^{(n+1)} &= P(X_{n+1} = j \mid X_0 = i) \\ &= \sum_{k \in S} P(X_{n+1} = j, X_n = k \mid X_0 = i) \quad \text{Total probability} \\ &= \sum_{k \in S} \frac{P(X_{n+1} = j, X_n = k, X_0 = i)}{P(X_0 = i)} \\ &= \sum_{k \in S} \frac{P(X_{n+1} = j, X_n = k, X_0 = i)}{P(X_n = k, X_0 = i)} \frac{P(X_n = k, X_0 = i)}{P(X_0 = i)} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k \in S} P(X_{n+1} = j \mid X_n = k, X_0 = i) \cdot P(X_n = k \mid X_0 = i) \quad \text{Conditional total probability} \\
&= \sum_{k \in S} P(X_{n+1} = j \mid X_n = k) \cdot P(X_n = k \mid X_0 = i) \quad \text{Use Markov Property} \\
&= \sum_{k \in S} P_{kj} \cdot P_{ik}^{(n)} \quad \text{Matrix terms} \\
&= \left(P \cdot P^{(n)}\right)_{ij} \quad \text{Matrix product} \\
&= \left(P^{n+1}\right)_{ij} \quad \text{Inductive Hypothesis}
\end{aligned}
$$

There we have proved that $P^{(n+1)} = P^{n+1}$ and so we have a completed the proof that $P^{(n)} = P^n$.

□

This result is very fundamental. We now have a relationship between the n-step transition matrix and the 1-step transition matrix (namely $P^{(n)} = P^n$). It is important to not to be confused by notation ($P^{(n)} = P^n$ is not a tautology). $P^{(n)}$ is a single matrix with entries populated by n-step transition probabilities while P^n is a single matrix multiplied by itself $n - 1$ times.

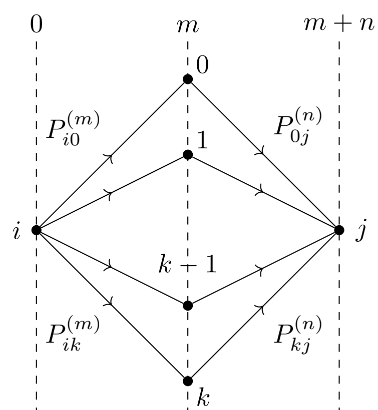
Corollary 2. *As a corollary, we have obtained that,*

$$P^{(n)} = P^{(m)} \cdot P^{(n-m)} \quad \forall 0 \leq m \leq n$$

Or equivalently the ***Chapman-Kolmogorov (C-K) Equation****,*

$$P_{ij}^{(n)} = \sum_{k \in S} P_{ik}^{(m)} P_{kj}^{(n-m)} \quad \forall i, j \in S, \forall 0 \leq m \leq n \tag{1.4}$$

Pictorially the C-K gives reveals the following picture that holds for all Markov chains,



So far, we have only been discussing transition probabilities. We will now divert our attention to actual distributions for a stochastic process.

Let $\alpha_n = (\alpha_{n,0}, \alpha_{n,1}, \ldots)$ be the **probability distribution vector** for X_n at time n.

$$\alpha_{n,k} = P(X_n = k) \qquad \forall k \in S$$

Note that $\alpha_{n,k} \geq 0$ and $\sum_{k \in S} \alpha_{n,k} = 1$ and $n = 0, 1, 2, \ldots$. We also define the initial distribution α_0,

$$\alpha_0 = (P(X_0 = 0), P(X_0 = 1), \ldots)$$

Theorem 3. *The transition probability matrix reveals the following relationship between the distribution α_n at time n and the distribution α_0 at time 0,*

$$\alpha_n = \alpha_0 \cdot P^n \tag{1.5}$$

Proof. The proof eq. (1.5) is quite trivial:

$$\begin{aligned}
\forall j \in S \quad \alpha_{n,j} &= P(X_n = j) \\
&= \sum_{i\in S} P(X_n = j \mid X_0 = i) \cdot P(X_0 = i) \\
&= \sum_{i\in S} \alpha_{0,i} \cdot P^n_{ij} \\
&= \alpha_{0,0} \cdot P^n_{0j} + \alpha_{0,1} \cdot P^n_{1j} + \dots \\
&= (\alpha_0 \cdot P^n)_j
\end{aligned}$$

□

More generally, for any $n = 1, 2, \dots$ the finite dimensional distribution can be obtained from the following process iterative process,

$$\begin{aligned}
&P(X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = \\
&\qquad P(X_0 = x_0)\cdot \\
&\qquad P(X_1 = x_1 \mid X_0 = x_0)\cdot \\
&\qquad P(X_2 = x_2 \mid X_1 = x_1, X_0 = x_0)\cdots \\
&\qquad P(X_n = x_n \mid X_{n-1} = x_{n-1}, \dots, X_0 = x_0)
\end{aligned}$$

But by the Markov condition, it must be that,

$$\begin{aligned}
&P(X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = \\
&\qquad P(X_0 = x_0)\cdot \\
&\qquad P(X_1 = x_1 \mid X_0 = x_0)\cdot \\
&\qquad P(X_2 = x_2 \mid X_1 = x_1)\cdots \\
&\qquad P(X_n = x_n \mid X_{n-1} = x_{n-1})
\end{aligned}$$

First recognize the first term on the RHS ($P(X_0 = x_0) = \alpha_{0,x_0}$), and also the remaining terms are transition probabilities as per eq. (1.1). Therefore it must be that,

$$P(X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = \alpha_{0,x_0} P_{x_0x_1} P_{x_1x_2} \cdots P_{x_{n-1}x_n}$$

Even more generally, for $0 \le t_1 < t_2 < \dots < t_n$,

$$P\big(X_{t_n} = x_{t_n}, X_{t_{n-1}} = x_{t_{n-1}}, \dots, X_{t_1} = x_{t_1}\big) = P(X_{t_1} = x_{t_1})\big(P^{t_2-t_1}\big)_{x_{t_1}x_{t_2}}\big(P^{t_3-t_2}\big)_{x_{t_2}x_{t_3}} \cdots \big(P^{t_n-t_{n-1}}\big)_{x_{t_{n-1}}x_{t_n}}$$

Since $P(X_{t_1} = x_{t_1}) = \alpha_{t_1x_{t_1}} = \sum_{k\in S} \alpha_{0,k} P^{t_1}_{k,x_{t_1}}$,

$$\alpha_{t_1} = \alpha_0 \cdot P^{t_1} \tag{1.6}$$

Remark 1. Equation (1.5) carries a very important interpretation. The probabilistic properties of a Discrete-Time Markov Chain (DTMC) are fully characterized by two things:

1. The initial distribution α_0
2. Transition matrix P

Knowing these two things fully characterizes the distribution α_n for all times n.

1.2 Stationary Distribution (Invariant Distribution)

In this section, we are interested in determining which distributions α_0 remain unchanged for all time $n \in T$.

Definition 1. A probability distribution $\pi = (\pi_0, \pi_1, \cdots)$ is called a **stationary (invariant) distribution** of the DTMC $\{X_n\}_{n=0,1,\cdots}$ with transition matrix P if the following conditions hold,

1. The transition matrix does not change π:
$$\pi = \pi \cdot P \tag{1.7}$$

2. The vector π is a valid probability distribution,
$$\sum_{i \in S} \pi_i = 1 \qquad \pi_i \geq 0 \tag{1.8}$$

Notice that if we posit that π is a probability distribution, then the second condition is already satisfied. Nonetheless, in practice we are able to find candidate π's using the the first condition and then we need to check these candidates against the second condition.

Why are such π's called stationary/invariant distributions? Notice that eq. (1.7) completely answers this question. Assume that the MC starts with initial distribution $\alpha_0 = \pi$ for X_0. In this case, the distribution of X_1 is determined by P,

$$\alpha_1 = \alpha_0 \cdot P$$

But since α_0 is π and π satisfies eq. (1.7),

$$\alpha_1 = \pi \cdot P = \pi$$

The distribution for X_1 is the *same* as the distribution for X_0. This process continues,

$$\alpha_2 = \alpha_1 \cdot P = \pi \cdot P = \pi$$

$$\alpha_n = \alpha_0 \cdot P^n = \pi \cdot P^n = \pi \cdot P^{n-1} = \cdots = \pi$$

Thus if the Markov chain starts with a stationary/invariant distribution then its marginal distribution will *never change*; hence why we refer π as stationary. Also not that this *does not* indicate that the value of X_i does not change over time (it almost certainly will), but its distribution does.

Example 1. Consider an electron with two states: ground (0) and excited (1). Let X_n be the state at time n. At each step, with probability α the MC chains state if it is in the ground state. With probability β the MC will transition to the ground state if it is in the excited state. Then $\{X_n\}_{n=0,1,\ldots}$ is a DTMC and its transition matrix is,

$$P = \begin{array}{c c} & \begin{array}{cc} {\scriptstyle (0)} & {\scriptstyle (1)} \end{array} \\ \begin{array}{c} {\scriptstyle (0)} \\ {\scriptstyle (1)} \end{array} & \left(\begin{array}{cc} 1-\alpha & \alpha \\ \beta & 1-\beta \end{array} \right) \end{array}$$

Now let us solve for the stationary distribution π.

$$\pi = \pi \cdot P \qquad \pi = (\pi_0, \pi_1) \qquad \pi_0 + \pi_1 = 1$$

Therefore,

$$\pi_0 = (1-\alpha)\pi_0 + \beta\pi_1 \tag{1.9}$$
$$\pi_1 = \alpha\pi_0 + (1-\beta)\pi_1 \tag{1.10}$$

However note that these two equations are not linearly independent. This is evident because summing eq. (1.9) with eq. (1.10) results in the trivial statement of $\pi_0 + \pi_1 = \pi_0 + \pi_1$. Nonetheless rearranging eq. (1.9) gives,

$$\alpha\pi_0 = \beta\pi_1 \implies \frac{\pi_0}{\pi_1} = \frac{\beta}{\alpha}$$

This is where we need $\pi_0 + \pi_1 = 1$.

$$\pi_0 = \frac{\beta}{\alpha+\beta} \qquad \pi_1 = \frac{\alpha}{\alpha+\beta}$$

Where $\alpha + \beta$ is considered the normalizing constant.

An important remark: sometimes the candidate distribution is not normalizable. In particular, there are configurations where eq. (1.7) is satisfiable but eq. (1.8) is not. In the above example, there exists a unique stationary distribution,

$$\pi = \left(\frac{\alpha}{\alpha+\beta}, \frac{\beta}{\alpha+\beta}\right)$$

If $\alpha_0 = \pi$ then we know immediately that,

$$P(X_n = 0) = \frac{\beta}{\alpha+\beta} \qquad P(X_n = 1) = \frac{\alpha}{\alpha+\beta} \qquad \forall n = 1, 2, \ldots$$

Remark 2. By the above procedure of solving for stationary distribution is typical.

1. Use eq. (1.7) to get proportions between different components of π.
2. Use eq. (1.8) to normalize π and get exact values.

Remark 3. Note that if $\beta = 2\alpha$ then π is always $(2/3, 1/3)$ regardless the actual value of α.