Stat 433

STOCHASTIC PROCESSES

University of Waterloo

Course notes by: TC Fraser
Instructor: Yi Shen

tcfraser@tcfraser.com

Version: 1.0

Table Of Contents

Disclaimer

These notes are intended to be a reference for my future self (TC Fraser). If you the reader find these notes useful in any capacity, please feel free to use these notes as you wish, free of charge. However, I do not guarantee their complete accuracy and mistakes are likely present. If you notice any errors please email me at **tcfraser@tcfraser.com**, or contribute directly at **https://github.com/tcfraser/course-notes**. If you are the professor of this course (Yi Shen) and you've managed to stumble upon these notes and would like to make large changes or additions, email me please.

Latest versions of all my course notes are available at **www.tcfraser.com/coursenotes**.

1 DTMC

1.1 Review of Probability

A **random variable (r.v.)** X is a real valued function of the outcomes of a random experiment.

$$X : \Omega \to R$$

Where $\Omega = \{\omega_1, \omega_2, \ldots\}$ is the **sample space** corresponding to all possible outcomes ω_i. The outcomes can in principle be any objects (numbers, strings, etc.). We say that X maps each outcome ω to a real number $\omega \mapsto X(\omega) \in \mathbb{R}$.

A **stochastic process** is a family of random variables $\{X_t\}_{t\in T}$, defined on a common sample space Ω. T is referred to as the index set for the stochastic process which is often understood as time. The index set T can take a discrete spectrum,

$$T = \{0, 1, 2, \ldots\} \qquad \{X_n \mid n = 0, 1, 2, \ldots\}$$

Alternatively, T can take on a continuous spectrum,

$$T = \{t \mid t \geq 0\} = [0, \infty)$$

The **state space** S is the collection of all possible values of X_t's. It is important to understand the distinction of between sample space and state space. Additionally, the state space can either have discrete or continuous spectrum.

A question remains, *Why do we need the family of random variables to be defined on a common sample space?* The answer being that we would like to be able to discuss the joint behaviour of X_t's. If X_1 has domain Ω_1 and X_2 has domain Ω_2 (where $\Omega_1 \neq \Omega_2$), then one can *not* talk about common ideas of correlations and associations between X_1 and X_2. As such we assert that all members of a stochastic process share the same sample space domain Ω.

1.2 Discrete-time Markov Chain

A **discrete-time stochastic process** $\{X_n \mid n \in 0, 1, 2, \ldots\}$ is said to be a **Discrete-time Markov Chain (DTMC)** if the following conditions hold:

1. The state space is at most *countable*[1] (i.e. finite or countable).

$$S = \{0, 1, \ldots, k\} \quad \text{or} \quad S = \{0, 1, 2, \ldots\}$$

2. **Markov Property**: For any $n = 0, 1, 2, \ldots$,

$$P(X_{n+1} = x_{n+1} \mid X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) = P(X_{n+1} = x_{n+1} \mid X_n = x_n)$$

We use capital letters X to denote the random variable and lower case letters x to denote a specific realization or valuation of X. The motivation of the Markov property is that future events $X_{n+1} = x_{n+1}$ are independent of past histories $\{X_i = x_i \mid i = 0, 1, \ldots, n-1\}$ given the immediate past state $X_n = x_n$. The intuition being that the future and the past are probabilistically independent.

Given the present, the future and the past are independent.

[1] Countable meaning there is a one-to-one mapping from the state space to the natural numbers.

1.3 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 MC 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,

$$P_{i,j} \geq 0 \tag{1.2}$$

$$\forall i : \sum_{j \in S} P_{ij} = 1 \tag{1.3}$$

The entries are non-negative because they represent probabilities and the row sums for P are always unitary.

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

$$\forall i, j \in S : P_{ij}^{(n)} \equiv P(X_{n+m} = j \mid X_n = i) = P(X_n = j \mid X_0 = i)$$

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.*

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

Proof. Proof by induction:

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

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

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

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

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

□

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

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

Or equivalently the ***Chapman-Kolmogorov Equation*** *or simply C-K equation,*

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

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 $\alpha_{n,k} \ge 0$ 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:

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

□

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

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

$$P(X_n = x_n \mid X_{n-1} = x_{n-1}, \dots, X_0 = x_0)$$

But by the Markov condition, it must be that,

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

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,

$$P(X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = \alpha_{0,x_0} P_{x_0 x_1} P_{x_1 x_2} \cdots P_{x_{n-1} x_n}$$

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

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

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

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

This means the probabilistic properties of a DTMC are fully characterized by two things:

1. The initial distribution α_0
2. Transition matrix P

1.4 Classification of States

State j is **accessible** from state i (denoted $i \to j$) if there exists $n = 0, 1, \dots$ such that $P^{(n)}_{ij} > 0$. Intuitively, one can transition from state i to state j in finite steps n with positive probability. If i is also accessible from j, then we say i and j **communication**, denoted as $i \leftrightarrow j$.

$$i \leftrightarrow j \Leftrightarrow \exists m, n \ge 0, P^{(m)}_{ij} > 0, P^{(n)}_{ji} > 0$$

Theorem 4. *The binary communication relation "$\leftrightarrow$" is in fact a equivalence relation:*

- *Reflexivity* $i \leftrightarrow i$
- *Symmetry* $i \leftrightarrow j \implies j \leftrightarrow i$
- *Transitivity* $i \leftrightarrow j, j \leftrightarrow k \implies i \leftrightarrow k$

Proof. First, reflexivity is easy to prove by definition. Let $n = 0$ and recognize that $P^{(0)}_{ii}$ has a certain probability by definition,

$$P^{(0)}_{ii} = 1 \implies i \leftrightarrow i$$

Second, symmetry follows by definition,

$$P^{(m)}_{ij} > 0, P^{(n)}_{ji} > 0 \Leftrightarrow P^{(n)}_{ji} > 0, P^{(m)}_{ij} > 0$$

Third, transitivity can be proving by letting m and n be the unknown quantifiers:

$$\exists m \quad P^{(m)}_{ij} > 0, \exists n \quad P^{(n)}_{jk} > 0$$

Then by the CK equation eq. (1.4),

$$P_{ik}^{(m+n)} = \sum_{l \in S} P_{il}^{(m)} P_{lk}^{(n)}$$

Let $l = j$ be a single, fixed entry in the summation,

$$P_{ik}^{(m+n)} \geq P_{ij}^{(m)} P_{jk}^{(n)} > 0$$

Therefore we have that k is accessible from i ($i \to j$). Analogously we have that $i \to j$ therefore $i \leftrightarrow k$.

□

The communication equivalence relations then divides the state space S into different equivalence classes. That is, the states in one class comm with each other; the states in different classes do not comm. The equivalent classes form a *partition* of the state space S.

The family $\{S_1, S_2, \dots S_n\}$ is a **partition** of S if,

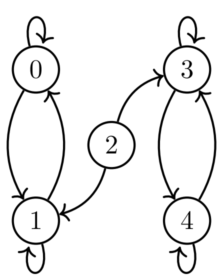
1. $S_i \subset S \mid \forall i \in 1, 2, \dots, n$
2. $S_i \cap S_j \neq \emptyset$ for all $i \neq j$
3. $\bigcup_i S_i = S$

We can find the equivalent classes by drawing a graph where the states in S are the nodes of the graph and a directed edge is placed going from i to j if j is accessible from i in one-step: $P_{ij} > 0$. Then identifying the the equivalent classes corresponds to identifying the loops of this graph within one step.

Example 1. As an example, consider the transition matrix P as follows.

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0.2 & 0.8 & 0 & 0 & 0 \\ 0.6 & 0.4 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.7 & 0.3 \\ 0 & 0 & 0 & 0.1 & 0.9 \end{pmatrix} \end{array}$$

The associated one-step accessibility graph is then,



Where the loops of $S = \{0, 1, 2, 3, 4\}$ form the following partition,

$$S_1 = \{0, 1\} \quad S_2 = \{2\} \quad S_3 = \{3, 4\}$$

These equivalent classes are useful for Markov chains because it allows one to separate the behaviour of the equivalence classes and study them individually. A MC which has only one equivalent class is called **irreducible**.

Furthermore, let us define the **period** of state i as,

$$d(i) = \gcd\{n \in \mathbb{Z}^+ \mid P_{ij}^n > 0\}$$

Additionally, if $P_{ii}^n = 0$ holds for all $n > 0$, we say that $d(i) = \infty$. If the period of i happens to be $d(i) = 1$ then the state i is said to be **aperiodic**. Alternatively, locus of steps that we can go back by are *co-prime*. A MC is called aperiodic if all its states S are aperiodic.

The period of a state is useful do to the following theorem,

Theorem 5. *The period of a state is a class property. If $i \leftrightarrow j$, then $d(i) = d(j)$.*

Proof. If $i = j$ we are already done. If $i \neq j$, since $i \leftrightarrow j$, then $\exists n, m$ such that,

$$P_{ij}^n > 0 \quad P_{ji}^m > 0$$

Then for any l such that $P_{jj}^l > 0$,

$$P_{ii}^{n+m+l} \geq P_{ij}^n P_{jj}^l P_{ji}^m \tag{1.6}$$

Because $P_{ij}^n P_{jj}^l P_{ji}^m$ happens to be a specific way for P_{ii}^{n+m+l} to occur. Since $i \leftrightarrow j$ and l was chosen carefully,

$$P_{ii}^{n+m+l} > 0$$

Moreover, we also have that,

$$P_{ii}^{n+m} \geq P_{ij}^n P_{ji}^m \tag{1.7}$$

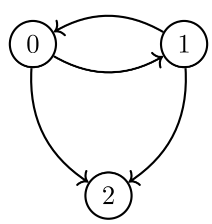
Since $d(i)$ divides both $n + m$ and $n + m + l$ by eqs. (1.7) and (1.6), then $d(i)$ also divides l. This holds for all l such that $P_{ii}^l > 0$. This implies that $d(i)$ is a common divisor of $\{l \mid P_{jj}^l > 0\}$ an thus $d(i)$ divides,

$$d(j) = \gcd\{l \mid P_{jj}^l > 0\}$$

By symmetry $d(j)$ divides $d(i)$. Therefore $d(i) = d(j)$.

□

Remark 1. It is important to note that $d(i) = k \not\Rightarrow P_{ii}^{(k)} > 0$. As a counterexample consider the following one step accessibility graph,



Evidently $P_{00} = 0$ but we have $d(0) = 1$ because $d(0) = \gcd\{2, 3, \ldots\}$.

Remark 2. If the MC is irreducible (having only one class) then all the states have the same period. In this case we ascribe the entire MC the period $d(i)$ where $i \in S$.