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.

$$X : \Omega \to \mathbb{R}$$

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 possible outcomes. 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 th 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.

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,

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

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

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,

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

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

The row sum for P is 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,

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

There is a simple relation between $P^{(n)}$ and 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 $\forall i,j \in S$,

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

$$= \left(P^{n+1}\right)_{ij} \quad \text{Inductive Hypothesis}$$

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

As a corollary, we have obtained that,

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

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

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

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 distribution of X_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,

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

The transition probability matrix gives us a relationship between α_n and α_0,

$$\alpha_n = \alpha_0 \cdot P^n \tag{1.1}$$

The proof eq. (1.1) is quite trivial:

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