Stochastic Process

Daniel Mao

Contents

Chapter 1

Stochastic Process

1.1 Definitions

Definition (Stochastic Process)**.** *Let $\mathcal{T}$ be an index set. Let $X(t)$ be a random variable. We define a **stochastic process** to be the net $(X(t))_{t\in\mathcal{T}}$.*

Definition (Discrete-Time Stochastic Process)**.** *Let $(X(t))_{t\in\mathcal{T}}$ be a stochastic process. We say that it is a **discrete-time stochastic process** if the index set $\mathcal{T}$ is countable.*

Definition (Markov Property)**.** *Let $\mathcal{S}$ be a state space. Let $(X_n)_{n\in\mathbb{N}}$ be a discrete-time stochastic process. We say that is has the **Markov property** if*

$$\forall n \in \mathbb{N}, \forall x_0..x_{n+1} \in \mathcal{S}, \quad \Pr(X_{n+1} = x_{n+1} \mid (X_n)_{n=0}^{n} = (x_n)_{n=0}^{n}) = \Pr(X_{n+1} = x_{n+1} \mid X_n = x_n).$$

This property states that the conditional distribution of any future state X_{n+1} given the past states $X_0, ..., X_{n-1}$ and the present state X_n is independent of the past states.

i.e., if we know the value taken by te process at a given time, we will not get any additional information about the future behavior of the process by gathering more knowledge about the past.

Definition (Markov Chain)**.** *We define a **Markov chain** to be a discrete-time stochastic process with the Markov property.*

Proposition 1.1.1.

$$\forall n \in \mathbb{N}, \forall j \in \{0..n-1\}, \forall x_0..x_{n+1} \in \mathcal{S}, \quad \Pr(X_{n+1} = x_{n+1} \mid X_n = x_n, (X_i)_{i=1}^{j-1} = (x_i)_{i=1}^{j-1}, (X_i)_{i=j+1}^{n-1} = (x_i)_{i=j+1}^{n-1}) = \Pr$$

Proof.

$$\Pr(X_{n+1} = x_{n+1} \mid X_n = x_n, (X_i)_{i\neq j} = (x_i)_{i\neq j}) \tag{1.1}$$

$$= \frac{\Pr(X_{n+1} = x_{n+1}, X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j}}{\Pr(X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j})} \tag{1.2}$$

$$= \frac{\sum_{x_j=0}^{\infty} \Pr(X_{n+1} = x_{n+1}, X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j}, X_j = x_j)}{\Pr(X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j})} \tag{1.3}$$

$$= \frac{\sum_{x_j=0}^{\infty} \Pr(X_{n+1} = x_{n+1} \mid X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j}, X_j = x_j) \Pr(X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j}, X_j = x_j)}{\Pr(X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j})} \tag{1.4}$$

$$= \frac{\sum_{x_j=0}^{\infty} \Pr(X_{n+1} = x_{n+1} \mid X_n = x_n) \Pr(X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j}, X_j = x_j)}{\Pr(X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j})} \tag{1.5}$$

$$= \Pr(X_{n+1} = x_{n+1} \mid X_n = x_n) \frac{\sum_{x_j=0}^{\infty} \Pr(X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j}, X_j = x_j)}{\Pr(X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j})} \tag{1.6}$$

$$= \Pr(X_{n+1} = x_{n+1} \mid X_n = x_n) \frac{\Pr(X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j})}{\Pr(X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j})} \tag{1.7}$$

$$= \Pr(X_{n+1} = x_{n+1} \mid X_n = x_n). \tag{1.8}$$

That is,

$$\Pr(X_{n+1} = x_{n+1} \mid X_n = x_n, (X_i)_{i \neq j} = (x_i)_{i \neq j}) = \Pr(X_{n+1} = x_{n+1} \mid X_n = x_n).$$

■

Definition (Transition Probability). *Let i and j be a pair of states. Let n be some time step. We define the* ***transition probability*** *from state i at time n to state j at time $n+1$, denoted by $P_{n,i,j}$. to be the conditional probability given by*

$$P_{n,i,j} = \Pr(X_{n+1} = j \mid X_n = i).$$

Definition (Stationary / Homogeneous). *We say that a discrete-time Markov chain is* ***stationary*** *or* ***homogeneous*** *if $\forall i, j \in \mathcal{S}$, $\forall n \in \mathbb{N}$, $P_{n,i,j} = P_{i,j}$ for some $P_{i,j}$.*

Theorem 1 (Chapman-Kolmogorov Equations).

$$P^{(n)} = P^{(m)} P^{(n-m)}.$$

1.2 Accessibility and Communication

Definition (Accessible). *Let i and j be two states. We say that state j is* ***accessible*** *from state i if $\exists n \in \mathbb{N}$ such that $P_{i,j}^{(n)} > 0$.*

Definition (Communicate). *Let i and j be two states. We say that state i and state j* ***communicate*** *if i and j are accessible from each other.*

Proposition 1.2.1. *The communication relation is an equivalence relation. i.e., it is reflexive, symmetric, and transitive.*

Proof. **Transitivity**:

Let i, j, k be states. Assume that $i \leftrightarrow j$ and $j \leftrightarrow k$. We are to prove that $i \leftrightarrow k$. Since $i \leftrightarrow j$, $\exists n \in \mathbb{N}$ such that $P_{i,j}^{(n)} > 0$. Since $j \leftrightarrow k$, $\exists m \in \mathbb{N}$ such that $P_{i,j}^{(m)} > 0$. By the Chapman-Kolmogorov equation, we get

$$P_{i,k}^{(n+m)} = \sum_{l=0}^{\infty} P_{i,l}^{(n)} P_{l,k}^{(m)} \geq P_{i,j}^{(n)} P_{j,k}^{(m)} > 0.$$

That is, $P_{i,k}^{(n+m)} > 0$. So $i \to k$. Similarly, we can show that $k \to i$. So $i \leftrightarrow k$. ■

Proposition 1.2.2. *Let i and j be two states. If state j is not accessible from state i, then*

$$\Pr(\textit{DTMC ever exists state } j \mid X_0 = i) = 0.$$

Proof. Since state j is not accessible from state i, we have $\forall n \in \mathbb{N}$, $P_{i,j}^{(n)} = 0$.

$$\begin{aligned} & \Pr(\text{DTMC ever exists state } j \mid X_0 = i) \\ = \ & \Pr(\bigcup_{n=0}^{\infty} \{X_n = j\} \mid X_0 = i) \leq \sum_{n=0}^{\infty} \Pr(X_n = j \mid X_0 = i) \\ = \ & \sum_{n=0}^{\infty} P_{i,j}^{(n)} = 0. \end{aligned}$$

That is,

$$\Pr(\text{DTMC ever exists state } j \mid X_0 = i) = 0.$$

■

Definition (Communication Class)**.** *We define a **communication class** to the set of states that communicate with each other.*

Definition (Irreducible, Reducible)**.** *We say that a discrete-time Markov chain is **irreducible** if it has only one communication class; and we say that it is **reducible** otherwise.*

1.3 Periodicity

Definition (Period)**.** *Let i be a state. We define the **period** of i, denoted by $d(i)$, to be the number given by*

$$d(i) := \gcd\{n \in \mathbb{Z}_+ : P_{i,i}^{(n)} > 0\}.$$

Definition (Aperiodic)**.** *We say that a state i is **aperiodic** if $d(i) = 1$. We say that a discrete-time Markov chain is **aperiodic** if $d(i) = 1$ for all state i.*

Proposition 1.3.1. *Let i and j be two states. If $i \leftrightarrow j$, then $d(i) = d(j)$.*

Proof. Since $i \leftrightarrow j$, $\exists n \in \mathbb{Z}_+$ such that $P_{i,j}^{(n)} > 0$; $\exists m \in \mathbb{Z}_+$ such that $P_{j,i}^{(m)} > 0$; and $\exists s \in \mathbb{Z}_+$ such that $P_{j,j}^{(s)} > 0$. Note that

$$P_{i,i}^{(n+m)} \geq P_{i,j}^{(n)} P_{j,i}^{(m)} > 0.$$

and

$$P_{i,i}^{(n+s+m)} \geq P_{i,j}^{(n)} P_{j,j}^{(s)} P_{j,i}^{(m)} > 0.$$

So $d(i) \mid (n+m)$ and $d(i) \mid (n+s+m)$. So $d(i) \mid ((n+s+m)-(n+m)) = s$. Since $\forall s \in \mathbb{Z}_+ : P_{j,j}^{(s)} > 0$, $d(i) \mid s$, we get $d(i) \mid d(j)$. Similarly, we have $d(j) \mid d(i)$. So $d(i) = d(j)$. ■

1.4 Transience and Recurrence

Chapter 2

Convergence of Random Variables

2.1 Definitions

Definition (Convergence in Distribution). *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables. Let F_n be the cumulative distribution function of X_n. Let X be a random variable. Let F_X be the cumulative distribution function of X. We say that the sequence $\{X_n\}_{n\in\mathbb{N}}$* ***converges in distribution*** *to X, denoted by $X_n \xrightarrow{d} X$, if $\forall x$ at which F is continuous,*

$$\lim_{n\to\infty} F_n(x) = F_X(x).$$

In this case, we say F_X is the asymptotic distribution of $\{X_n\}_{n\in\mathbb{N}}$.

Definition (Convergence in Probability). *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables. Let X be a random variable. We say that the sequence $\{X_n\}_{n\in\mathbb{N}}$* ***converges in probability*** *to X, denoted by $X_n \xrightarrow{p}$, if*

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} P\big(|X_n - X| \geq \varepsilon\big) = 0.$$

Or equivalently,

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} P\big(|X_n - X| < \varepsilon\big) = 1.$$

Definition (Almost Sure Convergence). *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables. Let X be a random variable. We say that the sequence $\{X_n\}_{n\in\mathbb{N}}$* ***converges almost surely*** *to X if*

$$P(\lim_{n\to\infty} X_n = X) = 1.$$

Definition (Sure Convergence). *Let Ω be a sample space of the underlying probability space. Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables. Let X be a random variable. We say that the sequence $\{X_n\}_{n\in\mathbb{N}}$* ***converges surely*** *to X if*

$$\forall \omega \in \Omega, \quad \lim_{n\to\infty} X_n(\omega) = X(\omega).$$

Definition (Convergence in Mean). *Let $r \geq 1$. Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables. Let X be a random variable. We say that the sequence $\{X_n\}_{n\in\mathbb{N}}$* ***converges in the r^{th} mean*** *to X, denoted by $X_n \xrightarrow{L^r} X$, if the r^{th} absolute moments $\mathbb{E}[|X_n^r|]$ and $\mathbb{E}[|X|^r]$ of X_n and X exists and*

$$\lim_{n\to\infty} \mathbb{E}[|X_n - X|^r] = 0.$$

2.2 Markov's Inequality

Theorem 2 (Markov's Inequality). *Let X be a random variable. Let k and c be arbitrary positive numbers. Then*

$$P\big(|X| \geq c\big) \leq \frac{\mathbb{E}\big[|X|^k\big]}{c^k}.$$

Corollary.

$$P\Big(|X - \mathbb{E}[X]| > k\sqrt{\operatorname{var}[X]}\Big) \leq \frac{1}{k^2}.$$

2.3 Properties

Proposition 2.3.1. *Convergence in probability implies convergence in distribution.*

Proposition 2.3.2. *Almost sure convergence implies convergence in probability.*

Proposition 2.3.3. *Convergence in the r^{th} mean for $r \geq 1$ implies convergence in probability.*

Proposition 2.3.4. *Let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of random variables. Let c be a constant. Then $\{X_i\}_{i\in\mathbb{N}}$ converges to c in distribution if and only if $\{X_i\}_{i\in\mathbb{N}}$ converges to c in probability.*

Sketch Proof.

$$\begin{aligned}
P\big(|X_i - c| \geq \varepsilon\big) &= P\big(X_i \geq c + \varepsilon\big) + P\big(X_i \leq c - \varepsilon\big) \\
&= 1 - P\big(X_i < c + \varepsilon\big) + F_i(c - \varepsilon) \\
&\leq 1 - P\big(X_i \leq c + \varepsilon/2\big) + F_i(c - \varepsilon) \\
&= 1 - F_i(c + \varepsilon/2) + F_i(c - \varepsilon)
\end{aligned}$$

$$
\begin{aligned}
&\lim_{i\to\infty}\left[1 - F_i(c+\varepsilon/2) + F_i(c-\varepsilon)\right] \\
&= 1 - F(c+\varepsilon/2) + F(c-\varepsilon) \\
&= 1 - 1 + 0 \\
&= 0.
\end{aligned}
$$

■

Proposition 2.3.5 (Continuous Map). *Let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of random variables. Let g be a continuous function on the X_i's. Then*

(1) if $X_i \xrightarrow{d} X$, we have $g(X_i) \xrightarrow{d} g(X)$.

(2) if $X_i \xrightarrow{p} c$, we have $g(X_i) \xrightarrow{p} g(c)$.

Proposition 2.3.6 (Slutsky's Theorem). *Let $\{X_i\}_{i\in\mathbb{N}}$ and $\{Y_i\}_{i\in\mathbb{N}}$ be sequences of random variables. Suppose $X_i \xrightarrow{d} X$ for some random variable X and $Y_i \xrightarrow{p} c$ for some constant c. Then*

(1) $X_i + Y_i \xrightarrow{d} X + c$.

(2) $X_iY_i \xrightarrow{d} cX$.

(3) $X_i/Y_i \xrightarrow{d} X/c$.

2.4 Law of Large Numbers

Theorem 3 (Strong Law of Large Nubmers). *Let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables. Suppose that $\mathbb{E}[X_i] = \mu$ for some $\mu \in \mathbb{R}$ for all $i \in \mathbb{N}$. Then their cumulative average $\bar{X}_n$ converges almost surely to μ. That is,*

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \overset{almostsurely}{\longrightarrow} \mu.$$

Chapter 3

Markov Decision Process

Chapter 4

Poisson Process

4.1 Homogeneous Poisson Process

4.1.1 Definitions

Definition (Homogeneous Poisson Process). *We say a counting process is a* ***homogeneous Poisson counting process*** *with rate* $\lambda > 0$ *if it has the following three properties:*

- $N(0) = 0$*;*
- *it has independent increments; and*
- *the number of events in any interval of length* t *is a Poisson random variable with parameter* λt*.*

Definition (Homogeneous Poisson Process). *We say a point process is a* ***homogeneous Poisson point process*** *with rate* $\lambda > 0$ *if the following two conditions hold:*

- *The probability* $\mathbb{P}\{N(a,b] = n\}$ *of the number* $N(a,b]$ *of points of the process in the interval* $(a,b]$ *being equal to some counting number* n *is given by*

  $$\mathbb{P}\{N(a,b] = n\} = \frac{[\lambda(b-a)]^n}{n!} e^{-\lambda(b-a)}.$$

  i.e. the number of arrivals in each finite interval has a Poisson distribution.

- *For any positive integer* k *and non-overlapping intervals* $(a_1, b_1], ..., (a_k, b_k]$*,*

  $$\mathbb{P}\left\{\bigwedge_{i=1}^{k} N(a_i, b_i] = n_i\right\} = \prod_{i=1}^{k} \mathbb{P}\{N(a_i, b_i] = n_i\}.$$

  i.e. the number of arrivals in disjoint intervals are independent random variables.