

# Applied Stochastic Processes

2018A7PS0193P



## CHAPTER 1

### Fundamentals

#### 1. Stochastic Processes

DEFINITION 1. A stochastic process is a probability model that describes the evolution of a system evolving randomly in time.

DEFINITION 2. A random variable is a mapping  $X : \Omega \rightarrow \mathbb{R}$  that assigns a real number  $X(\omega)$  to each outcome  $\omega \in \Omega$ , where  $\Omega$  is the sample space.

A stochastic process can be given by a collection of random variables  $\{X(t), t \in T\}$ , where  $T$  is called the **index set**. If  $T$  is countable (observed at discrete times), we get a **discrete time stochastic process**. On the other hand, if  $T$  is uncountable (observed continuously), then we get a **continuous time stochastic process**.

DEFINITION 3. The state space of a stochastic process is defined as the set of all possible values that the random variables  $X(t)$  can assume.

#### 2. Elementary Probability

For a recap of elementary probability, refer to the notes from Applied Statistical Methods.

#### 3. Transformation of Random Variables

LEMMA 1.1. Let  $X$  have a continuous, strictly increasing CDF  $F$ . Let  $U \sim \text{Uniform}(0,1)$ . If  $Y = F^{-1}(U)$ , then  $Y$  also has the CDF  $F$ .

The lemma above allows us to transform  $U$  to any other random variable, as long as it has a continuous and strictly increasing CDF. Say we had an algorithm to define a uniform random variable in the range  $(0,1)$ , now we have a way to generate random variables from a different distribution.

#### 4. Moment Generating Functions

DEFINITION 4. The moment generating function  $\phi(t)$  of the random variable  $X$  is defined for all values  $t$  as  $\phi(t) = E[e^{tx}]$ .

The moment generating functions of some oft-used distributions are as follows:

- Moment generating function of Binomial( $n, p$ ) is:

$$\phi(t) = (pe^t + 1 - p)^n$$

- Moment generating function of Poisson( $\lambda$ ) is:

$$\phi(t) = e^{\lambda(e^t - 1)}$$

- Moment generating function of Exponential( $\lambda$ ) is:

$$\phi(t) = \frac{\lambda}{\lambda - t}$$

- Moment generating function for Normal( $\mu, \sigma^2$ ) is:

$$e^{t\mu + \frac{1}{2}\sigma^2 t^2}$$

- Moment generating function for Uniform( $a, b$ ) is :

$$\frac{e^{tb} - e^{ta}}{t(b - a)}$$

THEOREM 1.2. The moment generating function of the sum of independent random variables is the product of the individual moment generating functions.

So, this means that  $\phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t)$ , as long as  $X \perp Y$  (this notation means that they are independent).

#### 5. Conditional distributions

The conditional probability distribution of  $Y$  given the occurrence of the value  $x$  of  $X$  is given by:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

where  $f_{X,Y}(x, y)$  is the joint distribution and  $f_X(x)$  is the marginal density of  $X$ .

The conditional expectation of  $X$  given  $Y$  is:

$$E(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x, y) dx$$

### 6. Markov's and Chebyshev's Inequality

**THEOREM 1.3** (Markov's Inequality). Let  $X$  be a non-negative random variable and suppose that  $E(X)$  exists. For any  $t > 0$ ,

$$P(X > t) \leq \frac{E(X)}{t}$$

**PROOF.**

$$E(X) = P(X < a) \cdot E(X|X < a) + P(X \geq a) \cdot E(X|X \geq a)$$

Here,  $E(X|X < a)$  is larger than 0 since  $X$  is a non negative random variable.  $E(X|X \geq a)$  is larger than  $a$ , since it only considers values larger than  $a$ . Hence,

$$E(X) \geq P(X < a) \cdot 0 + P(X \geq a) \cdot E(X|X \geq a) \geq a \cdot P(X \geq a)$$

This directly leads to Markov's Inequality. □

**THEOREM 1.4** (Chebyshev's Inequality). Let  $\mu = E(X)$  and  $\sigma^2 = V(X)$ . Then

$$P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}$$

$$P(|Z| \geq k) \leq \frac{1}{k^2}$$

where  $Z = (X - \mu)/\sigma$  and  $t > 0$ .

**PROOF.**

$$\begin{aligned} \sigma^2 &= E((X - \mu)^2) \\ &= E((X - \mu)^2 | k\sigma \leq |X - \mu|) P(k\sigma \leq |X - \mu|) + E((X - \mu)^2 | k\sigma > |X - \mu|) P(k\sigma > |X - \mu|) \\ &\geq (k\sigma)^2 P(k\sigma \leq |X - \mu|) + 0 \cdot P(k\sigma > |X - \mu|) \\ &= k^2 \sigma^2 P(k\sigma \leq |X - \mu|) \end{aligned}$$

Now the inequality follows from dividing by  $k^2 \sigma^2$ . □

Chebyshev's Inequality is a more general version of Markov's Inequality, applicable for any random variable  $X$ .

## 7. Convergence of Random Variables

Let  $X_1, X_2, \dots$  be a sequence of random variables and let  $X$  be another random variable. Let  $F_n$  be the CDF of  $X_n$  and  $F$  be the CDF of  $X$ .

We say that  $X_n$  converges to  $X$  in probability, denoted by  $X_n \xrightarrow{P} X$ , if for every  $\epsilon > 0$ ,

$$P(|X_n - X| > \epsilon) \rightarrow 0$$

as  $n \rightarrow \infty$ .

We say that  $X_n$  converges to  $X$  in distribution, denoted by  $X_n \rightsquigarrow X$ , if

$$\lim_{n \rightarrow \infty} F_n(t) = F(t)$$

for all  $t$  for which  $F$  is continuous.

LEMMA 1.5. If  $X_n \xrightarrow{P} X$ , then  $X_n \rightsquigarrow X$ .

The above lemma is provided without proof, as it is beyond the scope of the course.

EXERCISE. Let  $X_n \sim N(0, \frac{1}{n})$ . Prove that this series converges to

$$F_X(n) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

in distribution.

SOLUTION. Let  $t > 0$ , and define the standard normal variable  $Z_n = \sqrt{n}X_n$ , so  $Z_n \sim N(0, 1)$ . So,

$$\begin{aligned} F_{X_n}(t) &= P(X_n \leq t) \\ &= P(Z_n \leq \sqrt{nt}) \\ &= \int_{-\infty}^{\sqrt{nt}} f(x) dx \end{aligned}$$

where  $f(x)$  is the PDF of  $Z_n$ .

It is clear that as  $n \rightarrow \infty$ , we get the following distribution:

$$F_{X_n}(t) = \begin{cases} 0 & t < 0 \\ 0.5 & t = 0 \\ 1 & t > 0 \end{cases}$$

So,  $F_{X_n} \rightsquigarrow F_X \forall t \in \mathbb{R} - \{0\}$ .

■





## CHAPTER 2

### Markov Chains

#### 1. Introduction

DEFINITION 5. A stochastic process with a finite number of state spaces  $S = \{0, 1, \dots, N\}$  and a countable index state  $T = \{t_0, t_1, t_2, \dots\}$  is a Markov chain if

$$\begin{aligned} P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ = P(X_{n+1} = j | X_n = i) \\ = \mathbf{P}_{ij} \end{aligned}$$

This means that the probability of a transition from state  $i$  to a state  $j$  is completely dependent on  $i$  and  $j$ , and not on any history. The resulting transition matrix is called the **one step transition matrix**, denoted by  $\mathbf{P}$ .

LEMMA 2.1.

$$\sum_{j=0}^{\infty} \mathbf{P}_{ij} = 1 \forall i \in S$$

Of course, this matrix means that we could represent a Markov Chain by a digraph, or even Petri Nets (out of syllabus).

#### 2. State Probabilities

Let  $\mathbf{\Pi}^{(n)}$  be a row vector such that  $\mathbf{\Pi}_i^{(n)}$  is the probability that after  $n$  transitions, we are at state  $i$ . This is known as the **marginal probability distribution**. This vector can be recursively computed by the formula:

$$\mathbf{\Pi}^{(n)} = \mathbf{\Pi}^{(n-1)} \mathbf{P}$$

So, we get

$$\mathbf{\Pi}^{(n)} = \mathbf{\Pi}^{(0)} \mathbf{P}^n$$

$\mathbf{P}^n$  is known as the  $n^{th}$  step transition probability matrix (denoted by  $p^{(n)}$ ), where:

$$\mathbf{P}_{ij}^n = p_{ij}^{(n)} = P(X_n = j | X_0 = i)$$

This means that  $p^{(n)}$  solves the question - after  $n$  transitions starting from  $i$ , what is the probability that I will be in state  $j$ ?

A problem we are facing with this approach is that we have to calculate  $\mathbf{P}^n$  fast enough. It is not enough to use binary exponentiation and find it in  $O(d^3 \log n)$ , since we are using matrices with large size, so while it would be fast in exponent, it would be slow doing matrix multiplication. Instead, we use **Eigendecomposition** (read FDS notes).

Using eigendecomposition, we can decompose  $\mathbf{P}$  into:

$$\mathbf{P} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}$$

Raising it to power of  $n$ , we get the decomposition to be:

$$\mathbf{P}^n = \mathbf{Q}\mathbf{\Lambda}^n\mathbf{Q}^{-1}$$

Since  $\mathbf{\Lambda}$  is a diagonal matrix, we can find the exponent even faster in  $O(d \log n)$ , hence speeding up the process.

THEOREM 2.2 (Chapman-Kolmogorov Equation).

$$p_{ij}^{(n+m)} = \sum_{k=0}^{\infty} p_{ik}^{(n)} p_{kj}^{(m)}$$

PROOF.

$$\begin{aligned} p_{ij}^{(n+m)} &= P(X_{n+m} = j | X_0 = i) \\ &= \sum_{k=0}^{\infty} P(X_{n+m} = j, X_n = k | X_0 = i) \\ &= \sum_{k=0}^{\infty} P(X_{n+m} = j | X_n = k, X_0 = i) P(X_n = k | X_0 = i) \\ &= \sum_{k=0}^{\infty} P_{kj}^m P_{ik}^n \end{aligned}$$

□

THEOREM 2.3. If  $\mathbf{P}$  is a transition matrix for a finite state Markov chain, it has at least one eigenvalue as 1. All the other eigenvalues have an absolute value  $|\lambda_i| \leq 1$ .

DEFINITION 6. The stationary distribution of a Markov Chain is a row vector  $\boldsymbol{\Pi}$  such that

$$\boldsymbol{\Pi} \cdot \mathbf{P} = \boldsymbol{\Pi}$$

So, the stationary distribution  $\mathbf{\Pi}$  is the left eigenvector of  $\mathbf{P}$  such that it's eigenvalue is 1.

DEFINITION 7. If  $\lim_{n \rightarrow \infty} \mathbf{P}^n$  is such that all the rows in it are equal, that row is said to be the limiting distribution.

The limiting distribution may not always exist, but if it does, it is equivalent to the stationary distribution.

### 3. Occupancy Time and First Entrance Time

DEFINITION 8. Let  $N_{ij}^{(n)}$  be the number of times a discrete time Markov Chain visits a state  $j$  starting from state  $i$  over a given time span of  $n$ . The occupancy time for state  $j$  starting from  $i$  is:

$$T_{ij}^{(n)} = E(N_{ij}^{(n)})$$

THEOREM 2.4. The occupancy times matrix  $\mathbf{T}^{(n)} = \sum_{k=0}^n \mathbf{P}^k$ .

DEFINITION 9. Let  $f_{ij}^{(n)}$  be the probability the Markov Chain visits a state  $j$  for the first time starting from a state  $i$  after  $n$  transitions. Then  $f$  is the First Entrance Time Matrix.

Of course, the probability that a system starting at  $i$  will ever reach  $j$  is:

$$F_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}$$

$F_{ij}$  is called the probability of eventual return.

THEOREM 2.5 (First Entrance Theorem).

$$p_{jk}^{(n)} = \sum_{r=0}^n f_{jk}^{(r)} p_{kk}^{(n-r)}, n \geq 1$$

where

$$p_{kk}^{(0)} = 1, f_{jk}^{(0)} = 0, f_{jk}^{(1)} = p_{jk}$$

The proof of this is rigorous and uses the strong Markov Property, so it is not included here.

DEFINITION 10. The mean time of eventual return is the expected number of transitions to go from  $i$  to  $j$ . This is denoted by  $\mu_{ij}$  and is given by:

$$\mu_{ij} = \sum_{n=1}^{\infty} n f_{ij}^{(n)}$$

$\mu_{ii}$  is called the mean recurrence time. This value is not guaranteed to converge.

#### 4. Classification of States

DEFINITION 11 (accessibility). We say that  $i$  reaches  $j$  (or  $j$  is accessible from  $i$ ) if  $P_{ij}^n > 0$  for some  $n$ , and we denote it by  $i \rightarrow j$ .

DEFINITION 12 (communicability). If  $i \rightarrow j$  and  $j \rightarrow i$ , then we write  $i \leftrightarrow j$  and we say that  $i$  and  $j$  communicate.

THEOREM 2.6. The communication relation satisfies the following properties:

- (1)  $i \leftrightarrow i$  (reflexive)
- (2)  $i \leftrightarrow j \Rightarrow j \leftrightarrow i$  (symmetric)
- (3) If  $i \leftrightarrow j$  and  $j \leftrightarrow k$  then  $i \leftrightarrow k$  (transitive)
- (4) The set of states  $\chi$  can be written as a disjoint union of classes  $\chi = \chi_1 \cup \chi_2 \cup \dots$  where two states  $i$  and  $j$  communicate with each other if and only if they are in the same class. (equivalence class)

It is hence an equivalence relation.

REMARK. This has its own mathematical proof from the definition, but it is far more intuitive to think of the strongly connected components in a directed graph.

If all states communicate with each other, then the chain is said to be **irreducible**. A set of states is **closed** if, once you enter that set of states you never leave. In a more mathematical sense, a set  $B$  is closed if for all  $j \in B^C$ , there is no  $i \in B$  such that  $i \rightarrow j$ . A closed set consisting of a single state is called an **absorbing state**.

DEFINITION 13. State  $i$  is recurrent or persistent if

$$P(X_n = i \text{ for some } n \geq 1 | X_0 = i) = 1$$

Otherwise, the state is transient.

This definition means that if a state is recurrent, then if we start at a state  $i$ , we will definitely return to that state after some  $n > 0$  steps, i.e,  $F_{ii} = 1$ . Since this is essentially a recursive process (start from state  $i$ , return to state  $i$ , start again), the process will re-enter  $i$  again and again and again.

However, if a state  $i$  is transient, there is some probability  $p < 1$  ( $p$  is in fact  $F_{ii}$ ) that it will re-enter the state. Hence, the probability that the process will be in state  $i$  for exactly  $n$  time periods is:

$$p^{n-1}(1 - p)$$

A recurrent state is **null** if  $\mu_{ii} = \infty$ . Otherwise it is called **positive** or non null.

**THEOREM 2.7.** A state  $i$  is recurrent if and only if  $\sum_n P_{ii}^n = \infty$ . A state  $i$  is transient if and only if  $\sum_n P_{ii}^n < \infty$

**PROOF.** From our discussion, it is clear that a state  $i$  is recurrent if and only if the expected number of time periods that the process is in state  $i$  is infinite. Let  $I_n$  be such that:

$$I_n = \begin{cases} 1, & \text{if } X_n = i \\ 0, & \text{if } X_n \neq i \end{cases}$$

Hence,  $\sum_{n=0}^{\infty} I_n$  is the number of periods that a process is in a state  $i$ .

$$\begin{aligned} E \left( \sum_{n=0}^{\infty} I_n | X_0 = i \right) &= \sum_{n=0}^{\infty} E(I_n | X_0 = i) \\ &= \sum_{n=0}^{\infty} P(X_n = i | X_0 = i) \\ &= \sum_{n=0}^{\infty} P_{ii}^n \end{aligned}$$

Hence, if the state is recurrent, this value must be infinite. Otherwise, it would be recurrent. □

Note that this theorem also implies that if  $i$  is transient,  $P_{ii}^n \rightarrow 0$  as  $n \rightarrow \infty$ .

**EXERCISE.** Prove that in a finite state Markov Chain, all states cannot be transient

**SOLUTION.** Let us assume we have a Markov Chain where all states are transient. If this were the case, there would be some finite amount of time after which state  $i$ , will never be visited, for every state  $i$ . Hence, after some finite time, no states will be visited. But, the

process must be in some state, so we have arrived at a contradiction. Hence all states cannot be transient. ■

THEOREM 2.8. If state  $i$  is recurrent and  $i \leftrightarrow j$ , then  $j$  is recurrent.

PROOF. To prove this, notice that since  $i \leftrightarrow j$ , there exists integers  $k$  and  $m$  such that  $P_{ij}^k > 0, P_{ji}^m$ . For any integer  $n$ ,

$$P_{jj}^{m+n+k} \geq P_{ji}^m P_{ii}^n P_{ij}^k$$

This is because the LHS is the probability of going from state  $j$  to state  $j$  in  $m+n+k$  steps, but the RHS requires this to occur along a particular path. Summing over all  $n$ ,

$$\sum_{n=0}^{\infty} P_{jj}^{m+n+k} \geq P_{ji}^m P_{ij}^k \sum_{n=0}^{\infty} P_{ii}^n = \infty$$

This is straightforward from theorem 2.7. From that same theorem,  $j$  is recurrent. □

This means that if  $i$  is a recurrent state, then if every state in it's equivalence class is also recurrent, i.e. it is a class property.

THEOREM 2.9. A set  $B$  is closed if and only if  $P_{ij} = 0 \forall j \in B^C, i \in B$ .

REMARK. This comes directly from thinking of the Markov Chain as a graph.

THEOREM 2.10. If  $B$  is a closed set, then  $B$  is the union of equivalence classes in the Markov Chain.

PROOF. Assume this is not true. Then, there is some states  $i \in B$  and  $j \in B^C$  such that  $i \leftrightarrow j$ . By definition of a closed state, there must be no  $j \in B^C$  such that  $i \rightarrow j$ . This generates a contradiction. Hence,  $j \in B$ . □

DEFINITION 14. The period of a state  $i$  is  $d_i = \gcd(\{n : P_{ii}^n > 0\})$ . If  $d_i > 1$ , the state is periodic, and is otherwise aperiodic if  $d_i = 1$ .

Periodicity is a property of a class. Every state in the same class has the same period.

DEFINITION 15. A state is ergodic if it is persistent, non-null and aperiodic. A chain is ergodic if all it's states are ergodic.

Ergodism is a class property, since periodicity and persistency are both class properties.

EXERCISE. Consider a gambler who at each play of the game has probability  $p$  of winning one unit and probability  $q = 1 - p$  of losing one unit. Assuming that successive plays of the game are independent, what is the probability that starting with  $i$  units, the gambler's fortune will reach  $N$  before reaching 0?

SOLUTION. Let  $\{X_n, n = \{0, 1, 2, \dots\}$  be a Markov chain, where  $X_n$  is the player's fortune at time  $n$ . Then, the transition probability  $P$  is given by:

$$P_{00} = P_{NN} = 1$$

$$P_{i,i+1} = p = 1 - P_{i,i-1}, i = 1, 2, \dots, N - 1$$

As such, our Markov chain has 3 classes -  $\{0\}, \{1, 2, 3, \dots, N - 1\}, \{N\}$ . The second class is transient, the others are recurrent. Hence, at some point the gambler will go broke or hit the jackpot of  $N$ . If  $P_i$  is the probability of reaching  $N$  from  $i$  eventually, then:

$$P_i = pP_{i+1} + qP_{i-1}, i = 1, 2, \dots, N - 1$$

$$pP_i + qP_i = pP_{i+1} + qP_{i-1}$$

$$P_{i+1} - P_i = \frac{q}{p}(P_i - P_{i-1})$$

Of course,  $P_0 = 0, P_N = 1$ . From our recurrence,

$$P_i - P_{i-1} = \left(\frac{q}{p}\right)^{i-1} P_1$$

Adding these equations over all  $i$ ,

$$P_i - P_1 = P_1 \left[ \left(\frac{q}{p}\right) + \dots + \left(\frac{q}{p}\right)^{i-1} \right]$$

Hence,

$$P_i = \begin{cases} \frac{1-(q/p)^i}{1-(q/p)} P_1 & \frac{q}{p} \neq 1 \\ iP_1 & \text{if } \frac{q}{p} = 1 \end{cases}$$

Using the fact that  $P_N = 1$ ,

$$P_1 = \begin{cases} \frac{1-(q/p)}{1-(q/p)^N} & \frac{q}{p} \neq 1 \\ \frac{1}{N} & \text{if } \frac{q}{p} = 1 \end{cases}$$

Hence,

$$P_i = \begin{cases} \frac{1-(q/p)^i}{1-(q/p)^N} & p \neq \frac{1}{2} \\ \frac{1}{N} & p = \frac{1}{2} \end{cases}$$

We can see that if  $p \leq \frac{1}{2}$ , the gambler will definitely go broke against an infinitely rich adversary. Otherwise, there is a positive probability of increasing indefinitely. ■





## CHAPTER 3

### Poisson Processes

DEFINITION 16. Let  $X$  be a continuous random variable whose values lie in the non negative real numbers. The probability distribution of  $X$  is **memoryless** if for any non negative real numbers  $t$  and  $s$ , we have:

$$P(X > t + s | X > t) = P(X > s)$$

Hence, the distribution of a “waiting time” until a certain event does not depend on how much time has elapsed already.

THEOREM 3.1. A positive continuous random variable  $X$  is memoryless if and only if  $X \sim \text{Exp}(x; \lambda)$  for some  $\lambda$ .

DEFINITION 17. A stochastic process  $\{N(t), t \geq 0\}$  is said to be a counting process if  $N(t)$  represents the total number of events that occur by time  $t$ .

From it's definition,

- $N(t) \geq 0$
- $N(t)$  is integer valued
- If  $s < t$  then  $N(s) \leq N(t)$
- For  $s < t$ ,  $N(t) - N(s)$  equals the number of events that occur in the interval  $(s, t]$

A counting process has **independent increments** if the numbers of events that occur in disjoint time intervals are independent. A counting process has **stationary increments** if the distribution of the number of events that occur in any time interval only depends on the length of that time interval.

DEFINITION 18. The counting process  $\{N(t), t\}$  is said to be a Poisson Process having rate  $\lambda, \lambda > 0$  if

$$(1) N(0) = 0$$

- (2) The process has independent increments
- (3) The number of events in any interval of length  $t$  is Poisson distributed with mean  $\lambda t$ . That is, for all  $s, t \geq 0$

$$P(N(t+s) - N(s) = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$$

This means it has stationary increments.

The definition of a Poisson process can also be encoded in the form of three postulates, which must be present for a counting process to be a Poisson process:

- (1) **Independence** :  $N(t+h) - N(t)$  is independent of the number of occurrences prior to that interval
- (2) **Homogeneity in time** :  $p_n(t)$  depends only on the length  $t$  of the interval and is independent of where the interval is situated, where:

$$p_n(t) = P(N(t+s) - N(s) = n)$$

- (3) **Regularity** : In an interval of infinitesimal length  $h$ , the probability of exactly one occurrence  $\lambda h + o(h)$  and that of more than one occurrence is  $o(h)$  (wut?)

It is possible to prove that these two definitions are equivalent.

LEMMA 3.2. For a Poisson Process with rate  $\lambda$ ,

$$E[N(t)] = \lambda t$$