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 $X : \Omega \to \mathbb{R}$ that assigns a real number $X(\omega)$ to each outcome $\omega \in \Omega$, where Ω 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 3.1. Let X have a continuous, strictly increasing CDF F. Let $U \sim$ 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(λ) is:

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

- Moment generating function of Exponential(λ) is:

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

- TODO more

THEOREM 4.1. 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 6.1 (Markov's Inequality). Let X be a non-negative random variable and suppose that $E(X)$ exists. For any $t > 0$,

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

THEOREM 6.2 (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 $t > 0$.

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

7. Convergence of Random Variables

Let $X_1, X_2, ...$ 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 $\epsilon > 0$,

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

as $n \to \infty$.

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

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

for all t for which F is continuous.

LEMMA 7.1. 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,

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

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

It is clear that as $n \to \infty$, 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 $S = \{0, 1, ..., N\}$ and a countable index state $T = \{t_0, t_1, t_2, ...\}$ is a Markov chain if

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0)$$

$$= P(X_{n+1} = j | X_n = i)$$

$$= P_{ij}$$

This means that the probability of a transition from state i to a state j is completely dependent on the starting and the ending state. The resulting transition matrix is called the **one step transition matrix**, denoted by $\mathbf{P}$.

LEMMA 1.1.

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

Of course, this matrix means that we could represent a Markov Chain by a digraph, or even Petri Nets (out of syllabus). It is also clear that the probability of being at a particular state after n transitions is given by:

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

So, we get

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

The probability of $P(X_n = i)$ is called the **marginal probability distribution** and is given by $\mathbf{\Pi}_i^{(n)}$. Remember that here $\mathbf{\Pi}$ is a row vector, not a column vector.

DEFINITION 6. The n^{th} step transition probability matrix is defined as:

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

Essentially this is asking the question that if initially the state is i, what is the probability that we are at state j after n steps?

LEMMA 1.2.
$$p_{ij}^{(n)} = P_{ij}^n$$

This follows from the calculation of $\mathbf{\Pi}$ that we did before.

THEOREM 1.3 (Chapman-Kolmogorov Equation).
$$p_{jk}^{(m+n)} = \sum p_{rk}^{(n)} p_{jr}^{(m)} = \sum p_{jr}^{(n)} p_{rk}^{(m)}$$

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:

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

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

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

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

THEOREM 1.4. 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 7. The stationary distribution is the row vector $\mathbf{\Pi}$ such that

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

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