Lecture Notes: Jan 13, 2024

In the last lecture we have discussed different one dimensional random variables both discrete and continuous. We have seen that the discrete random variables are characterized through their probability mass function, and the continuous random variables are characterized through their cumulative distribution function. Now we will be discussing about more than one random variables. First we consider a finite number of random variables and then we discuss about an infinite collection of random variables.

Suppose $X_1, \ldots, X_n$ is a set of discrete random variables. The joint probability mass function of $X_1, \ldots, X_n$ is defined as

$$P(X_1 = i_1, \ldots, X_n = i_n) = p_{i_1 i_2 \ldots, i_n},$$

here $i_1 \in A_1, \ldots, i_n \in A_n$, where $A_1, \ldots, A_n$ are countable sets, $p_{i_1 i_2 \ldots, i_n} \geq 0$ and $\displaystyle\sum_{i_1 \in A_1, \ldots, i_n \in A_n} p_{i_1 i_2 \ldots, i_n} = 1$. It is very clear that once we know the joint probability mass function of $X_1, \ldots, X_n$, the joint probability mass function of any subset of $\{X_1, \ldots, X_n\}$ also can be easily obtained. It is important to observe that

$$\begin{aligned} P(X_1 = i_1, \ldots, X_n = i_n) &= P(X_n = i_n | X_{n-1} = i_{n-1}, \ldots, X_1 = i_1) \times \ldots \\ &\quad \times P(X_2 = i_2 | X_1 = i_1) \times P(X_1 = i_1). \end{aligned}$$

The above relation will be used quite extensively later.

Similarly, if $X_1, \ldots, X_n$ is a set of continuous random variables, then the joint cumulative distribution function of $X_1, \ldots, X_n$ is defined as

$$P(X_1 \leq x_1, \ldots, X_n \leq x_n) = F_{X_1, \ldots, X_n}(x_1, \ldots, x_n); \quad -\infty < x_1, \ldots, x_n < \infty.$$

Once the joint cumulative distribution function of $X_1, \ldots, X_n$ is known, then the joint cumulative distribution function of any subset of $\{X_1, \ldots, X_n\}$ is also known, for example

$$P(X_1 \leq x_1, X_2 \leq x_2) = \lim_{x_3 \to \infty, \ldots, x_n \to \infty} F_{X_1, \ldots, X_n}(x_1, x_2, \ldots, x_n); \tag{1}$$

for $-\infty < x_1, x_2 < \infty$. Therefore, it is clear that a finite set of random variables, either discrete or continuous, are completely characterized by either it joint probability mass function or by the joint cumulative distribution function.

Now the natural question comes how to generalize it to an infinite set of random variables. For example we have a collection of random variables $\{X_1, X_2, \ldots\}$, then how do we define the joint probability mass function or the joint cumulative distribution function. It is not straight forward to generalize it for the infinite set of random variables. We take a slightly different approach. For an infinite collection of random variables say $\{X_1, X_2, \ldots\}$, we define the joint probability mass function or the joint cumulative distribution function, depending on whether there are discrete or continuous, of any subset $\{X_{i_1}, \ldots, X_{i_k}\}$, for all k, and for all $i_1, \ldots, i_k$. Note that naturally it has to satisfy certain consistency properties as defined in (1).

Similarly, suppose we have an infinite (uncountable) collection of random variables, say for example $\{X_\alpha; 0 \le \alpha \le 1\}$, then also we can define the joint probability mass function or the joint cumulative distribution function along the same way.

A stochastic process is a collection of random variables, i.e. of the form $\{X_n; n = 0, 1, 2, \ldots\}$ or $\{X_\alpha; \alpha \in T\}$, here T is an index set. The first one is a countable collection, where as the second one is an uncountable collections of random variables. Therefore, it is clear we have four different types of stochastic process, namely (a) discrete time and discrete state space, (b) discrete time and continuous state space, (c) continuous time and discrete state space and (d) continuous time and continuous state space. First we will be discussing about discrete time and discrete state space stochastic process, i.e. X_n's are discrete random variables, and without loss of generality it takes values $0, 1, 2, \ldots$, and the index set $T = \{0, 1, 2, 3, \ldots\}$.

A discrete time and discrete state space stochastic process $\{X_n; n = 0, 1, 2, \ldots\}$ is called a Markov Chain if

$$P((X_{n+1} = i_{n+1} | X_n = i_n, \ldots, X_0 = i_0) = P(X_{n+1} = i_{n+1} | X_n = i_n);$$

for all $i_0, \ldots, i_{n+1}$. The quantity

$$p_{ij}^{n,n+1} = P(X_{n+1} = j | X_n = i)$$

is known as the one step transition probability. If the quantity $p_{ij}^{n,n+1}$ is independent of n, it is called as stationary transition probability. We will be mainly discussing about the stationary transition probability only from now on, i.e. it is assumed from now on that $p_{ij}^{n,n+1}$ is independent of n and it will be denoted by p_{ij} only. Often, we write it as a matrix, i.e.

$$\boldsymbol{P} = \begin{bmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ p_{i0} & p_{i1} & p_{i2} & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

Note that $p_{ij} \geq 0$, and $\sum_{j=0}^{\infty} p_{ij} = 1$, for all $i = 0, 1, \ldots$. A $n \times n$ matrix $\boldsymbol{A} = ((a_{ij}))$, where n can be finite or infinite, is such that $a_{ij} \geq 0$, and $\sum_{j=1}^{n} a_{ij}$ $= 1$, for all $i = 1, \ldots, n$ is known as a transition probability matrix. If $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are transition probability matrices with same order, then $\boldsymbol{A}_1\boldsymbol{A}_2$ is also a transition probability matrix.

For a Stationary Markov Chain if we know the initial distribution, i.e. the distribution of X_0, and the stationary transition probability matrix $\boldsymbol{P}$, then we can compute the joint distribution of any $\{X_{i_1}, \ldots, X_{i_k}\}$. Suppose $P(X_0 = i) = a_{0i}$, where $a_{0i} \geq 0$, and $\sum_{i=0}^{\infty} a_{0i} = 1$. Let us denote the vector $\boldsymbol{a}^0 = (a_{00}, a_{01}, \ldots)$. If $P(X_1 = i) = a_{1i}$, and $\boldsymbol{a}^1 = (a_{10}, a_{11}, a_{12}, \ldots)$, then

$$P(X_1 = i) = \sum_{k=0}^{\infty} P(X_1 = i | X_0 = k) P(X_0 = k) = \sum_{k=0}^{\infty} p_{ki} a_{0k}.$$

$$\boldsymbol{a}^1 = \boldsymbol{a}^0 \boldsymbol{P}.$$

Similarly, if $P(X_n = i) = a_{ni}$, and $\boldsymbol{a}^n = (a_{n0}, a_{n1}, a_{n2}, \ldots)$, then

$$\boldsymbol{a}^n = \boldsymbol{a}^0 \boldsymbol{P}^n.$$

Similarly,

$$\begin{aligned}
P(X_1 = i, X_2 = j) &= \sum_{k=0}^{\infty} P(X_1 = i, X_2 = j, X_0 = k) \\
&= \sum_{k=0}^{\infty} P(X_1 = i, X_2 = j | X_0 = k) \times P(X_0 = k) \\
&= \sum_{k=0}^{\infty} P(X_2 = j | X_1 = i, X_0 = k) \times P(X_1 = i | X_0 = k) P(X_0 = k) \\
&= \sum_{k=0}^{\infty} P(X_2 = j | X_1 = i) \times P(X_1 = i | X_0 = k) P(X_0 = k) \\
&= \sum_{k=0}^{\infty} p_{ij} p_{ki} a_{0k} = p_{ij} \sum_{k=0}^{\infty} p_{ki} a_{0k} = p_{ij} (\boldsymbol{a}^0 \boldsymbol{P})_i.
\end{aligned}$$

Here $(\boldsymbol{a}^0 \boldsymbol{P})_i$ denotes the i element of the vector $(\boldsymbol{a}^0 \boldsymbol{P})$, i.e. $P(X_1 = i)$. Alternatively

$$P(X_2 = j, X_1 = i) = P(X_2 = j | X_1 = i) P(X_1 = i) = p_{ij} (\boldsymbol{a}^0 \boldsymbol{P})_i.$$

Now let us compute

$$\begin{aligned}
P(X_3 = j | X_1 = i) &= \sum_{k=0}^{\infty} P(X_3 = j, X_2 = k | X_1 = i) \\
&= \sum_{k=0}^{\infty} P(X_3 = j | X_2 = k, X_1 = i) \times P(X_2 = k | X_1 = i) \\
&= \sum_{k=0}^{\infty} P(X_3 = j | X_2 = k) \times P(X_2 = k | X_1 = i) \\
&= \sum_{k=0}^{\infty} p_{ik} p_{kj} = (\boldsymbol{P}^2)_{ij}.
\end{aligned}$$

Here $(\boldsymbol{P}^2)_{ij}$ denotes the (i, j)-th element of the matrix $\boldsymbol{P}^2$.