A Quick Introduction to Stationary and Ergodic Processes

Satyadev Nandakumar

March 26, 2013

1 Random Variables

Material in this chapter is taken from Shiryaev [1]. Let $(\Omega, \mathcal{F}, P)$ be a probability space, where Ω is the sample space, $\mathcal{F}$ is the σ-algebra comprising the events, and P is the probability measure. Recall that a standard example of a measurable space is $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where $\mathcal{B}(\mathbb{R})$ is the smallest σ-algebra containing all the intervals of the form $(infty, a]$, $a \in \mathbb{R}$, and the empty set.

Recall that a *random variable* $X : \Omega \to \mathbb{R}$ is a *measurable function* in the probability space. That is, assuming that the measure space on $\mathbb{R}$ is $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, we require that for every $S \in \mathcal{B}(\mathbb{R})$, $X^{-1}(S) \in \mathbb{F}$. Please consult the notes on measurability to see one intuition behind this definition, and examples of random variables.

The probability distribution μ_X *induced by* X on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is defined as follows. For any set $S \in \mathcal{B}(\mathbb{R})$, we have

$$\mu_X(S) = P(X^{-1}(S)) = P(\{\omega \in \Omega \mid X(\omega) \in S\}).$$

Note that the right hand side is a probability on subsets of Ω. Thus the probability on the domain can be used to define an induced probability on the range.

This is not a very strange concept - for example, when we have a dartboard with three regions, say a red region with a score of 20, a blue region with a score of 10 and a green region with a score of 20, we usually like to think of "the probability that I will get a 20 on a throw". Here we are directly thinking of a probability over the scores. This is naturally defined as probability that the dart lands on a region whose score is 20, that is, either red or green. If

$$\text{score} : \{\text{Red}, \text{Blue}, \text{Green}\} \to \{20, 10\}$$

is the random variable, then we seek the probability of the event $\text{score}^{-1}(\{20\}) = \{\text{Red}, \text{Green}\}$.

When we are discussing this general concept, it is interesting to view a familiar concept - that of a probability distribution - in a new light. First, we recall the definition of a distribution function.

Definition 1.1. The function $F_X : \mathbb{R} \to [0, 1]$ defined by

$$F_X(x) = P\{\omega \in \Omega \mid X(\omega) \leq x\}, \quad x \in \mathbb{R},$$

is called the *distribution function* of X.

When we discuss discrete random variables, we like to talk of $P(X = x)$. This is usually called the probability density function. For discrete random variables, we can use the density function to define the distribution function, and vice versa. (How?)

However, in the continuous case, $P(X = x)$ for a real number x is 0. So the distribution function is more informative than the density function.

Now, note that $F_X(x)$ is $\mu_X((-\infty, x])$, and $(-\infty, x]$ is a set in $\mathcal{B}(\mathbb{R})$. So the distribution function is specifically, the induced probability of some special sets in the σ-algebra over $\mathbb{R}$.

2 Random Vectors

Definition 2.1. A finite sequence of random variables $(X_1(\omega), X_2(\omega), \ldots, X_n(\omega))$ is called an *n-dimensional random vector.*

The induced probability $\mu_{X_1,X_2,\ldots,X_n}(S)$ for any set $S \in \mathcal{B}(\mathbb{R}^n)$ is defined to be $P(X_1^S \cap \cdots \cap X_n^{-1}(S))$.

Just as the distribution function of a random variable is related to the probability induced by it, we can see that the *joint distribution* of the random variables $X_1, \ldots, X_n$ is related to the induced probability of $(X_1, \ldots, X_n)$.

When we have distributions involving multiple random variables, we can now study the relationships between the random variables. An important concept is that of *independence.*

We say that the random variables $X_1, \ldots, X_n$ are *independent* if

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\ldots F_{X_n}(x_n).$$

3 Stochastic Process

An infinite sequence of random variables $X_1, X_2, \ldots,$ where each is a real-valued function on the same space Ω is called a *stochastic process.*

For example, let Ω be the set of infinite binary sequences. Let $X_i(\omega)$ be the i^{th} coordindate of ω. Then $X_0, X_1, \ldots$ is a stochastic process.

We say that a stochastic process $X_0, X_1, \ldots$ is *independent* if for every $k \in \mathbb{N}$, for every set of natural numbers $\{i_1, i_2, \ldots, i_k\}$, the set of random variables $X_{i_1}, \ldots, X_{i_k}$ are independent.

An *independent* stochastic process is said to be an *independent, identically distributed* (i.i.d. for short) stochastic process, if $F_{X_i} = F_{X_j}$ for all $i, j \in \mathbb{N}$.

Exercise

1. Consider the experiment of tossing a fair coin "infinitely many times". Let $X_i : \{\text{Heads}, \text{Tails}\} \to \{0, 1\}$ be the random variable defined by $X(\text{Heads}) = 1$ and $X(\{Tails\}) = 0$. Since it is a fair coin, $P(X_i = 1) = 0.5$ for any $i \in \mathbb{N}$. Define the joint distributions and show that $X_0, X_1, \ldots$ is an independent, identically distributed stochastic process.

2. Consider the unit interval $[0, 1)$. For a real number x, let $\langle x \rangle$ denote the fractional part of x, that is $x - \lfloor x \rfloor$. For example, $\langle 1.5 \rangle = 0.5$.

   Let $X_i : [0, 1) \to \{0, 1\}$ be defined as follows.

   $$X_i(\omega) = \begin{cases} 0 & \text{if } \langle 2^i\omega \rangle \in [0, 0.5) \\ 1 & \text{if } \langle 2^i\omega \rangle \in [0.5, 1). \end{cases}$$

   Intuitively, $X_i(\omega)$ is the i^{th} bit in the binary expansion of ω.

   Show that $X_0, X_1, \ldots$ is an independent process. This exercise shows that we can study "deterministic" processes in the probabilistic setting, together with concepts like independence.

3. Construct an independent binary valued stochastic process that is not i.i.d.

4 Stationary Stochastic Process

Independence is quite a strong assumption in the study of stochastic processes, and when we want to apply theorems about stochastic processes to several phenomena, we often find that the process at hand is not independent.

At the same time, we would like to retain some of the nice properties of stochastic processes like the law of large numbers. It is clear that a purely deterministic process will not obey the law of averages - for example, if an experiment always yields 0, then it will not yield 1s half of the time, and the law of large numbers fails badly. Is there a middle ground where slightly dependent sequence of processes still retain some of the probabilistic features of independent processes - for example, for a binary-valued stochastic process which is not independent, could we still say that with high probability, we will get zeroes half of the time, asymptotically?

There are several such important settings which simultaneously admit dependence, and many probabilistic laws of independent processes. The most important is that of a stationary stochastic process.

Definition 4.1. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X = (X_0, X_1, \dots)$ be a stochastic process. For $k \in \mathbb{N}$, $T^k X = (X_k, X_{k+1}, \dots)$ be the stochastic proces representing the k^{th} "left-shift of X".

X is *stationary* if the probability distributions of X and $T^k X$ are the same for every $k \geq 1$:

$$P(X \in B) = P(T^k X \in B) \quad \forall B \in \mathcal{B}(\mathbb{R}^\infty)\ \forall k \in \mathbb{N}.$$

An equivalent definitions using finite-dimensional distributions is as follows.

Note that for a binary i.i.d. process, $P(01) = P(0)P(1) = P(10)$, but for a stochastic process, $P(01)$ can be different from $P(10)$. This gives an intuition to the kind of generalization possible using stationarity.

There is a version of the law of large numbers applicable to the set of stationary processes, called the *Ergodic Theorem.* To introduce this, we now view stationary processes via a slightly different viewpoint.

4.1 Measure-Preserving Transformations

Exercises

1. Show that every i.i.d. process is stationary.

5 Ergodic Processes

References

[1] A. N. Shiryaev. *Probability.* Graduate Texts in Mathematics v.95. Springer, 2 edition, 1995.