A Brief Review of Probability and Stochastic Processes

M.R. Soleymani

In this note, some basic concepts from the theory of probability and stochastic processes are reviewed. The goal is to give a more formal definition of some the concepts you have already encountered in other courses. This note plus Chapter 2 of the text [1] should be sufficient for following the rest of the course. Those of you who wish to get better understanding of these concepts, possibly, for conducting your graduate research, are encouraged to consult specialized books such as [2] and [3].

Length, area, volume as well as probability are examples of a *measure*. A measure is a *set function*, i.e., an assignment of a number $\mu(A)$ to each set A in a certain class of sets. The class of sets on which the measure μ is defined has to have some structure. Let Ω be a set whose points $\omega \in \Omega$ correspond to the possible outcomes of a certain random experiment. Certain subsets of Ω are called events and are assigned probabilities. Roughly speaking, A is an event if the question "does $\omega \in A$?" has a definite yes or no answer after the experiment is performed. Now, if we can answer the question "does $\omega \in A$?" we should be able to answer the question "does $\omega \in A^c$?". Also, if we can determine whether or not $\omega \in A_i$, $i = 1, \ldots, n$, then we should be able to answer whether ω belongs to $\bigcup_{i=1}^n A_i$ (similarly $\bigcap_{i=1}^n A_i$) or not. So, it is not unreasonable to expect that the class of events be closed under complementation nd finite union and finite intersection. And since the answer to the question "does $\omega \in \Omega$?" is always *yes*, it natural to expect that the entire space Ω be an event.

After this intuitive reasoning, we are in a position to introduce a few abstract concepts.

Definition 1: Let $\mathcal{F}$ be a collection of subsets of a set Ω. Then $\mathcal{F}$ is called a *field* (or an *algebra*) iff $\Omega \in \mathcal{F}$ and $\mathcal{F}$ is closed under complementation and finite union, i.e.,

i) $\Omega \in \mathcal{F}$.
ii) If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$.
iii) If $A_i \in \mathcal{F}, \_i = 1, \ldots, n$ then $\bigcup_{i=1}^n A_i \in \mathcal{F}$.

It is easy to show that $\mathcal{F}$ is also closed under intersection. Let $A_1, \ldots, A_n \in \mathcal{F}$, then,

$$\bigcap_{i=1}^{n} A_i = \left(\bigcup_{i=1}^{n} A_i\right)^c \in \mathcal{F}.$$

If $\mathcal{F}$ is closed under *countable* union, i.e., if (iii) is replaced with

iv) If $A_1, A_2, \ldots \in \mathcal{F}$, then, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$,

then $\mathcal{F}$ is called a $\sigma - field$ (or a $\sigma - algebra$).

Definition 2: The class of *Borel sets* of Ω denoted by $\mathcal{B}(\Omega)$ or simply $\mathcal{B}$, is defined as the smallest $\sigma-$field of subsets of Ω .

Example 1: If $\Omega = R^1$, i.e., the set of real numbers, then any set containing all open (or equivalently closed) intervals on R^1 and all their finite and countable unions is a $\sigma-$field. The smallest $\sigma-$field, i.e., the class of Borel sets (called the *Borel field*) , $\mathcal{B}(R^1)$ or $\mathcal{B}_1$ is the intersection of all such $\sigma-$fields. In other words, any interval or any union (finite or countable) of intervals is a Borel set in the real line.

Example 2: Denote by R^∞ the space consisting of all infinite sequences $(x_1, x_2, \ldots)$ of real numbers. In R^∞, an *n-dimensional rectangle* is defined as

$$\{x \in R^\infty; x_1 \in I_1, \ldots, x_n \in I_n\},$$

where $I_1, \ldots, I_n$ are intervals on R^1. The Borel field, $\mathcal{B}_\infty$ is the smallest $\sigma-$field of subsets of R^∞ containing all finite dimensional rectangles. $\mathcal{B}_\infty$ is formed by intersecting of all sets containing all the finite dimensional rectangles and their corresponding (finite and infinite) unions.

Now, we are in possession of the tools needed for defining the probability space, random variables and random processes.

Definition 3: A *probability space* consists of a triplet $(\Omega, \mathcal{F}, P)$ where:

i) Ω is called a *sample space* and is a set of points ω called *sample points.*

ii) $\mathcal{F}$ is a $\sigma-$field of subsets of Ω. These subsets are called *events.*

iii) $P(.)$ is a probability measure on $\mathcal{F}$.

Random variables are formed by assigning real numbers to the outcomes of a random experiment. As such, a *random variable* X is a function $X(\omega)$: $\Omega \to R^1$. We usually, want to know the probability of events involving X. For example, we may want to know the probability that X belongs to a set of real numbers B. So, we are looking for $P\{\omega : X(\omega) \in B\}$. For this to make sense, $\{\omega : X(\omega) \in B\}$ has to be an event, i.e., it has to belong to the $\sigma-$field $\mathcal{F}$. More formally, random variables are defined as follows.

Definition 4: A function $X(\omega)$ defined on Ω is called a *random variable* if for every Borel set $B \in \mathcal{B}_1$ in the real line R^1, the set $\{\omega : X(\omega) \in B\}$ is in $\mathcal{F}$. It is said that $X(\omega)$ is a *measurable* function on $(\Omega, \mathcal{F})$.

From the fact that $\mathcal{F}$ is a $\sigma-$field, it is easy to conclude that if $\{\omega : X(\omega) \in I\}$ is in $\mathcal{F}$ for all intervals I, then X must be a random variable. That is, it is sufficient to be able to assign probabilities to all intervals.

Definition 5: The *distribution function* of a random variable X is the function $F_X : R^1 \to [0, 1]$ given by

$$F_X(x) = P\{\omega : X(\omega) \le x\}, \; \forall x \in R^1.$$

The subscript is omitted when there is no risk of confusion and the distribution function is denoted as $F(x)$.

Definition 6: A countable *stochastic (random) process* is a sequence of random variables X_1, X_2, ... defined over a common probability space $(\Omega, \mathcal{F}, P)$.

Definition 7: Given a random process $\{X_n\}$ on $(\Omega, \mathcal{F}, P)$, the n-dimensional *distribution function,* $F_n : R^n \to [0, 1]$ is defined as,

$$F_n(x_1, \ldots, x_n) = P(X_1 < x_1, \ldots, X_n < x_n).$$

Definition 8: A random process $\{X_n\}$ is *stationary* if for every k, the process X_{k+1}, X_{k+1}, ... has the same distribution as X_1, X_2, ..., i.e., for any $B \in \mathcal{B}_\infty$,

$$P\left((X_{k+1},\, X_{k+2},\, ...) \in B\right) = P\left((X_1,\, X_2,\, ...) \in B\right),$$

or, equivalently,

$$P(X_{k+1} < x_1,\, \ldots,\, X_{k+n} < x_n) = P(X_1 < x_1,\, \ldots,\, X_n < x_n),$$

that is,

$$F_{X_{k+1},\ldots,X_{k+n}}(x_1,\, \ldots,\, x_n) = F_{X_1,\ldots,X_n}(x_1,\, \ldots,\, x_n).$$

Exercise 1: Show that a random process $\{X_n\}$ is *stationary* if the process $X_2,\, X_3,\, \ldots$ has the same distribution as $X_1,\, X_2,\, ...$

Consider a probability space $(\Omega,\, \mathcal{F},\, P)$. Let T be a transformation of Ω into itself. The transformation T is called measurable if the inverse image of any set in $\mathcal{F}$ is again in $\mathcal{F}$. That is, if for any set $A \in \mathcal{F}$, we have $T^{-1}A = \{\omega : T\omega \in A\} \in \mathcal{F}$.

Definition 9: A measurable transformation T defined on $\Omega \to \Omega$ is *measure preserving* if $P(T^{-1}A) = P(A),\, \forall A \in \mathcal{F}$.

As an example, consider the shift transformation, $S : \mathrm{R}^\infty \to R^\infty$ defined on $(R^\infty,\, \mathcal{B}_\infty)$ by $S(x_1,\, x_2,\, \ldots) = (x_2,\, x_3,\, \ldots)$. It is easy to show that the shift transformation S is measurable. Using the shift transformation, we can give another definition of stationarity:

Definition 10: A random process $\{X_n\}$ defined on the probability space $(R^\infty,\, \mathcal{B}_\infty,\, P)$ is *stationary* if the shift transformation S is measure preserving, i.e., if $P(S^{-1}B) = P(B),\, \forall B \in \mathcal{B}_\infty$.

From this point on, we focus our attention on the shift transformation $T = S$. By doing so, we limit our coverage of *ergodicity* to stationary processes.

Definition 11: A set B is called *invariant* if $T^{-1}B = B$. That is, transformation of any point $\omega \in B$, denoted $T^{-1}\omega$ is in B.

Definition 12: A stationary random process $\{X_n\}$ is *ergodic* if every invariant set has probability zero or one.

in ENCS6161, you have learned the concept of mean ergodicity. That is , you have seen that for certain processes, called mean ergodic processes, the time average equals the expected value. Using the above definition of ergodicity, we can introduce a stronger version of this phenomenon.

Ergodic Theorem: For any ergodic stochastic process on $(\Omega,\, \mathcal{F},\, P)$ and any function $g(\omega)$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} g(T^k\omega) = E\left[g(\omega)\right],$$

for almost all ω.

As an example, for any set $A \in \mathcal{F}$, take $g(\omega) = \mathcal{X}_A(\omega)$, where $\mathcal{X}_A(\omega)$ is the *characteristic function* set A defined as,

$$\mathcal{X}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}.$$

Then, according to ergodic theorem,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathcal{X}_A(T^k\omega) = P(A).$$

This means that, for almost any ω, the asymptotic portion of points in A is exactly equal to $P(A)$.

Exercise 2: From the ergodic theorem given above, conclude ergodicity in the mean.

Definition 12 and the ergodic theorem relate the empirical and axiomatic aspects of the theory of probability. Definition 12 constrains the structure of an ergodic process. In a very intuitive sense, it says that in an ergodic process, each sample function $X(\omega)$ contains *almost* all the possible sequences (modes) of the process. That is, no sample function gets stuck in a finite number of modes and become repetitive. If we look at the structure of a process from a state space point of view, all states of an ergodic process are reachable from any other state while the state space of a non-ergodic process can be split into parts of positive probability that are not accessible to one another. According to ergodic theorem, the structure of an ergodic process can be used in order to find the probability of any sequence by observing almost any sample function. In other words, if we make N observations on n samples of an ergodic process X_1, X_2, ... then the probability that these samples denoted by vector $\mathbf{X}^n$ take values from a set B is equal to the asymptotic value of the relative frequency of this event, i.e.,

$$\lim_{N\to\infty} \frac{N_n(B)}{N} = P(B),$$

where $N_n(B)$ is the number of times when the event $\{\mathbf{X}^n \in B\}$ has occurred.

In order to clarify the concepts discussed above, we end this note with the following simple exercise.

Exercise 3: Take a random source generation binary sequences, i.e., $\Omega = \{0,1\}^\infty$. Assume that the source has three modes. In the first mode, the source repeats the 3-bit sequence 110, i.e., its output is $110110110\ldots$. In the second mode, the source repeats the sequence 001. Finally, in the third mode the source generates zeros and ones with probability 0.5. The probability of occurrence of the three modes is $1/6$, $1/6$ and $2/3$, respectively. Show that this process (source) is not ergodic. Is the process ergodic in the mean?

Exercise 4: Consider a random source generation binary sequences. Assume that the source has two modes with probabilities p and $1-p$. In the first mode, the source repeats the source output alternates between 0 and 1 i.e., its output is 010101010101 . . .. In the other mode, it generates zeros and ones with probabilities q and $1 - q$, respectively. What is the condition for the source to be ergodic? What is the condition for the source to be ergodic in the mean?

References:

[1] J. G. Proakis, *Digital Communications,* fourth edition, McGraw-Hill, 2001.
[2] R.B. Ash, *Real Analysis and Probability,* Academic Press, 1972.
[3] L. Breiman, *Probability,* Addison Wesley, 1968.