Options and Mathematics: Lecture 15

November 26, 2020

Review of finite probability theory

Let Ω be a set containing a finite number of elements $\omega_1, \omega_2, \ldots, \omega_M$.

We denote Ω as

$$\Omega = \{\omega_1, \ldots, \omega_M\}, \quad \text{or} \quad \Omega = \{\omega_i\}_{i=1,\ldots,M}$$

and call it a **sample space**.

The elements $\omega_i \in \Omega$, $i = 1, \ldots, M$, are called **sample points**. The sample points identify the possible outcomes of an experiment.

Examples

For the experiment "rolling a die" we have

$$\Omega = \{1, 2, 3, 4, 5, 6\} \quad (M = 6),$$

For the experiment "tossing a coin once", we have

$$\Omega = \Omega_1 := \{H, T\} \quad (M = 2),$$

where H stands for "Head" and T for "Tail".

In the experiment "tossing a coin twice" we have

$$\Omega = \Omega_2 := \{(H, H), (H, T), (T, H), (T, T)\} \quad (M = 2^2 = 4)$$

and in the experiment "tossing a coin N times" we have

$$\Omega = \Omega_N := \{\omega = (\gamma_1, \gamma_2, \ldots, \gamma_N);\ \gamma_j = H \text{ or } T,\ j = 1, \ldots, N\} = \{H, T\}^N \quad (M = 2^N).$$

We denote by 2^Ω the **power set** of Ω, i.e., the set of all subsets of Ω.

2^Ω consists of the empty set $\emptyset$, the subsets containing one element, i.e., $\{\omega_1\}, \{\omega_2\}, \ldots, \{\omega_M\}$, which are called **atomic sets**, the subsets containing two elements, i.e.,

$$\{\omega_1, \omega_2\}, \ldots, \{\omega_1, \omega_M\}, \{\omega_2, \omega_3\}, \ldots, \{\omega_2, \omega_M\}, \ldots, \{\omega_{M-1}, \omega_M\},$$

the subsets containing 3 elements and so on, and the set $\Omega = \{\omega_1, \ldots, \omega_M\}$ itself. Thus 2^Ω contains 2^M elements.

For instance

$$2^{\Omega_1} = \{\emptyset, \{H\}, \{T\}, \{H, T\} = \Omega_1\}.$$

The elements of 2^Ω (i.e., the subsets of Ω) are called **events**. They identify possible events that occur in the experiment.

For example

$$\{2, 4, 6\} \equiv [\text{the result of throwing a die is an even number}],$$

$\{(H,H),(T,T)\} \equiv$ [tossing a coin twice gives the same outcome in both tosses].

Let $A, B \in 2^\Omega$ are events.

$A \cup B$ is the event that A *or* B happens

$A \cap B$ is the event that both A *and* B happen.

If the sets $A, B \subset \Omega$ are **disjoint**, i.e., $A \cap B = \emptyset$, the events A and B cannot occur simultaneously.

Probability of events

The atomic set $\{\omega_i\}$ identifies the event that the outcome of the experiment is exactly ω_i.

We want to assign a probability $\mathbb{P}$ to such special events. To this purpose we introduce M real numbers $p_1, p_2, \ldots, p_M$ such that

$$0 < p_i < 1, \text{ for all } i = 1, \ldots, M, \quad \text{and} \quad \sum_{i=1}^{M} p_i = 1.$$

The M-dimensional vector $(p_1, p_2, \ldots, p_M)$ is called a **probability vector**.

We define p_i to be the probability of the event $\{\omega_i\}$, that is

$$\mathbb{P}(\{\omega_i\}) = p_i, \quad i = 1, \ldots, M.$$

Any event $A \in 2^\Omega$ can be written as the disjoint union of atomic events, e.g.,

$$\{\omega_1, \omega_3, \omega_6\} = \{\omega_1\} \cup \{\omega_3\} \cup \{\omega_6\}.$$

This leads to define the probability of the event $A \in 2^\Omega$ as

$$\mathbb{P}(A) = \sum_{i:\omega_i \in A} \mathbb{P}(\{\omega_i\}) = \sum_{i:\omega_i \in A} p_i.$$

We shall also write the definition of $\mathbb{P}(A)$ as

$$\boxed{\mathbb{P}(A) = \textstyle\sum_{\omega \in A} \mathbb{P}(\{\omega\})}$$

In particular

$$\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = \sum_{i=1}^{M} p_i = 1.$$

We also set

$$\mathbb{P}(\emptyset) = 0,$$

which means that it is impossible that the experiment gives no outcome.

The empty set $\emptyset$ *is the only event with zero probability: any other such event is excluded a priori by the sample space.*

At this point every event has been assigned a probability.

Definition 5.1

Given a probability vector $(p_1, \ldots, p_M)$ and a set $\Omega = \{\omega_1, \ldots, \omega_M\}$, the function $\mathbb{P} : 2^\Omega \to [0, 1]$ defined by $\mathbb{P}(\emptyset) = 0$ and

$$\mathbb{P}(A) = \sum_{i:\omega_i \in A} p_i.$$

is called a **probability measure**. The pair $(\Omega, \mathbb{P})$, is called a **finite probability space**.

Example

Definition 5.2

Given $0 < p < 1$, the pair $(\Omega_N, \mathbb{P}_p)$ given by $\Omega_N = \{H, T\}^N$ and

$$\boxed{\mathbb{P}_p(A) = \sum_{\omega \in A} p^{N_H(\omega)}(1-p)^{N_T(\omega)}, \text{ for all } A \in 2^{\Omega_N}},$$

is called the N**-coin toss probability space**. Here $N_H(\omega)$ is the number of H in the sample ω and $N_T(\omega) = N - N_H(\omega)$ is the number of T.

Conditional probability

It is possible that the occurrence of an event A affects the probability that a second event B occurred. For instance, for a fair coin we have $\mathbb{P}_p(\{H, H\}) = 1/4$, but if we know that the first toss is a tail, then $\mathbb{P}_p(\{H, H\}) = 0$. This simple remark leads to the definition of conditional probability.

Definition 5.3

Given two events A, B such that $\mathbb{P}(B) > 0$, the **conditional probability** of A given B is defined as

$$\boxed{\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}}$$

Similarly, if $B_1, B_2, \dots, B_n$ are events such that $\mathbb{P}(B_1 \cap \cdots \cap B_n) > 0$, the conditional probability of A given $B_1, \dots, B_n$ is

$$\mathbb{P}(A|B_1, \dots, B_n) = \frac{\mathbb{P}(A \cap B_1 \cap \cdots \cap B_n)}{\mathbb{P}(B_1 \cap \cdots \cap B_n)}.$$

If the occurrence of B does not affect the probability of occurrence of A, i.e., if $\mathbb{P}(A|B) = \mathbb{P}(A)$, we say that the two events are independent. By the previous definition, the independence property is equivalent to the following.

Definition 5.4

Two events A, B are said to be **independent** if

$$\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)}$$

Similarly, n events $A_1, \ldots, A_n$ are said to be independent if

$$\mathbb{P}(A_{k_1} \cap \cdots \cap A_{k_m}) = \mathbb{P}(A_{k_1}) \cdot \ldots \cdot \mathbb{P}(A_{k_m}),$$

for all $1 \leq k_1 < k_2 < \cdots < k_m \leq n$.

Random Variables

In general the purpose of an experiment is to determine the value of quantities which depend on the outcome of the experiment (e.g., the velocity of a particle, which is determined by successive measurements of its position). We call such quantities random variables.

Definition 5.5

Let $(\Omega, \mathbb{P})$ be a finite probability space. A **random variable** is a function $X : \Omega \to \mathbb{R}$.

If $g : \mathbb{R}^n \to \mathbb{R}$, then the random variable $Y = g(X_1, X_2, \ldots, X_n)$ is said to be measurable with respect to the random variables $X_1, \ldots, X_n$.

Example

Given $A \subset \Omega$, the random variable $\mathbb{I}_A : \Omega \to \{0, 1\}$ given by

$$\mathbb{I}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

is called the **indicator function** of the event A.

Since $\Omega = \{\omega_1, \dots, \omega_M\}$, then a random variable X on a finite probability space is necessarily a **finite random variable**, i.e., it can attain only a finite number of values $x_1, \dots, x_M$, namely

$$X(\omega_i) = x_i, \quad i = 1, \dots, M.$$

The values $x_1, \dots, x_M$ need *not* be distinct.

If $X(\omega_i) = c$, for all $i = 1, \dots, M$, we say that X is a **deterministic constant** (the value of X is independent of the outcome of the experiment).

The **image** of X is the finite set defined as

$$\mathrm{Im}(X) = \{x \in \mathbb{R} \text{ such that } X(\omega) = x, \text{ for some } \omega \in \Omega\},$$

i.e., $\mathrm{Im}(X)$ is the set of possible values attainable by X.

Notation

Given $a \in \mathbb{R}$, we denote

$$\{X = a\} = \{\omega \in \Omega : X(\omega) = a\},$$

which is the event that X attains the value a. Of course, $\{X = a\} = \emptyset$ if $a \notin \mathrm{Im}(X)$. In general, given $I \subseteq \mathbb{R}$, we denote

$$\{X \in I\} = \{\omega \in \Omega : X(\omega) \in I\},$$

which is the event that the value attained by X lies in the set I.

Moreover we denote

$$\{X = a, Y = b\} = \{X = a\} \cap \{Y = b\}, \quad \{X \in I_1, Y \in I_2\} = \{X \in I_1\} \cap \{Y \in I_2\}.$$

The probability that X takes value a is given by

$$\mathbb{P}(X = a) = \mathbb{P}(\{X = a\}) = \sum_{i:X(\omega_i)=a} p_i.$$

If $a \notin \mathrm{Im}(X)$, then $\mathbb{P}(X = a) = \mathbb{P}(\emptyset) = 0$.

More generally, given any open subset I of $\mathbb{R}$, we write

$$\mathbb{P}(X \in I) = \mathbb{P}(\{X \in I\}) = \sum_{i:X(\omega_i)\in I} p_i,$$

which is the probability that the value of X belongs to I.

Example

In the probability space of a fair die consider the random variable

$$X(\omega) = (-1)^\omega, \quad \omega \in \{1, 2, 3, 4, 5, 6\}.$$

Then $X(\omega) = 1$ if ω is even and $X(\omega) = -1$ is ω is odd. Moreover

$$\mathbb{P}(X = 1) = \mathbb{P}(\{2, 4, 6\}) = 1/2, \quad \mathbb{P}(X = -1) = \mathbb{P}(\{1, 3, 5\}) = 1/2,$$

whereas

$$\mathbb{P}(X \neq \pm 1) = \mathbb{P}(\emptyset) = 0.$$

The event $A = \{2, 4, 6\}$ is said to be **resolved** by X, because the occurrence of the event A (i.e., the fact that the outcome of the throw is an even number) is equivalent to X taking value 1.

In general, given a random variable $X : \Omega \to \mathbb{R}$, the events resolved by X are the sets of the form $\{X \in I\}$, for some $I \subseteq \mathbb{R}$. These events comprise the so called **information carried by** X.

Definition 5.6

Let $(\Omega, \mathbb{P})$ be a finite probability space and $X : \Omega \to \mathbb{R}$ a random variable. The function $f_X : \mathbb{R} \to [0, 1]$ defined by

$$\boxed{f_X(x) = \mathbb{P}(X = x)}$$

is called the **probability distribution** of X (or **probability mass function**), while $F_X : \mathbb{R} \to [0, 1]$ given by

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

is called the **cumulative distribution** of X.

Note that $f_X(x)$ is non-zero if only if $x \in \operatorname{Im}(X)$, and that F_X is a non-decreasing function satisfying

$$0 \leq F_X(x) \leq 1, \quad \lim_{x \to -\infty} F_X(x) = 0, \quad \lim_{x \to \infty} F_X(x) = 1.$$

For example, for the random variable $X(\omega) = (-1)^\omega$ defined on the probability space of a fair die we have

$$F_X(x) = \begin{cases} 0, & x < -1, \\ 1/2, & x \in [-1, 1), \\ 1, & x \geq 1. \end{cases}$$

For the applications to the binomial options pricing model, the following probability distribution plays a fundamental role.

Definition 5.6

Given $N \in \mathbb{N}$, $N \geq 1$, and $p \in (0,1)$, a finite random variable X is said to be **binomially distributed** if $\mathrm{Im}(X) = \{0, 1, \ldots, N\}$ and if the probability distribution of X is given by the **binomial distribution**

$$f_X(k) = \phi_{N,p}(k) := \binom{N}{k} p^k (1-p)^{N-k}, \quad k = 0, \ldots, N.$$

For instance, the random variable $X(\omega) = N_H(\omega)$ in the N-coin toss probability space is binomially distributed.

The probability that a random variable X takes value in the interval $[a,b]$ can be written in terms of the distribution of X as

$$\mathbb{P}(a \leq X \leq b) = \sum_{i: X(\omega_i) = x_i \in [a,b]} \mathbb{P}(X = x_i) = \sum_{i: a \leq x_i \leq b} f_X(x_i).$$

In a similar fashion, if $g : \mathbb{R} \to \mathbb{R}$ then

$$\mathbb{P}(a \leq g(X) \leq b) = \sum_{i: g(X(\omega_i)) = g(x_i) \in [a,b]} \mathbb{P}(X = x_i) = \sum_{i: a \leq g(x_i) \leq b} f_X(x_i).$$

Independent random variables

We have seen before that a random variable X carries information.

If $Y = g(X)$ for some (non-constant) function $g : \mathbb{R} \to \mathbb{R}$, then Y carries no more information than X: any event resolved by knowing the value of Y is also resolved by knowing the value of X.

The other extreme case is when two random variables carry independent information.

Definition 5.8

Let $(\Omega, \mathbb{P})$ be a finite probability space.

Two random variables $X_1, X_2 : \Omega \to \mathbb{R}$ are said to be **independent** if the events $\{X_1 \in I_1\}, \{X_2 \in I_2\}$ are independent events, for all sets $I_1 \subseteq \mathrm{Im}(X_1), I_2 \subseteq \mathrm{Im}(X_2)$. This means that

$$\mathbb{P}(X \in I_1, X_2 \in I_2) = \mathbb{P}(X_1 \in I_1)\mathbb{P}(X_2 \in I_2).$$

More generally, n random variables $X_1, \dots, X_n : \Omega \to \mathbb{R}$ are independent if the events $\{X_1 \in I_1\}, \{X_2 \in I_2\}, \dots, \{X_n \in I_n\}$ are independent for all sets $I_1, I_2, \dots, I_n$ such that $I_j \subseteq \mathrm{Im}(X_j)$.

The independence property is linked to the probability defined on the sample space: two random variables may be independent with respect to some probability and not-independent with respect to another. We shall use later the following important result:

Theorem 5.1

Let $X_1, X_2, \dots, X_n$ be independent random variables, $k \in \{1, \dots, n-1\}$ and $g : \mathbb{R}^k \to \mathbb{R}$, $f : \mathbb{R}^{n-k} \to \mathbb{R}$. Then the random variables

$$Y = g(X_1, X_2, \dots, X_k), \quad Z = f(X_{k+1}, \cdots, X_n)$$

are independent.