Options and Mathematics: Lecture 16

November 27, 2020

Review of finite probability theory

Expectation and Variance

We may think of the expectation of X as an estimate on the average value of X and the variance of X as a measure of how far is this estimate from to the precise value of X.

Definition 5.9

Let $(\Omega, \mathbb{P})$ be a finite probability space and $X : \Omega \to \mathbb{R}$ a random variable. The **expectation** (or **expected value**) of X is defined by

$$\mathbb{E}[X] = \sum_{i=1}^{M} X(\omega_i)\mathbb{P}(\omega_i).$$

We shall write the definition of $\mathbb{E}[X]$ also as

$$\boxed{\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\})}$$

Example In the N-coin toss probability space $(\Omega_N, \mathbb{P}_p)$ we have

$$\mathbb{E}_p[X] = \sum_{\omega \in \Omega_N} X(\omega) p^{N_H(\omega)} (1-p)^{N_T(\omega)},$$

where $N_H(\omega)$ is the number of heads and $N_T(\omega) = N - N_H(\omega)$ is the number of tails in the N-toss $\omega \in \Omega_N$.

We can rewrite the definition of expectation as

$$\mathbb{E}[X] = \sum_{x \in \mathrm{Im}(X)} x \, \mathbb{P}(X = x),$$

or equivalently

$$\boxed{\mathbb{E}[X] = \textstyle\sum_{x \in \mathrm{Im}(X)} x f_X(x)}$$

The importance of the previous formula is that it allows to compute the expectation of X from its distribution, without any reference to the original probability space.

Example

If we are told that a random variable X takes the following values:

$$X = \begin{cases} 1 & \text{with probability } 1/4 \\ 2 & \text{with probability } 1/4 \\ -1 & \text{with probability } 1/2 \end{cases},$$

then we can compute $\mathbb{E}[X]$ as

$$\mathbb{E}[X] = 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} - 1 \cdot \frac{1}{2} = \frac{1}{4}.$$

Some simple properties of the expectation are collected in the following theorem.

Theorem 5.2

Let X, Y be random variables on a finite probability space $(\Omega, \mathbb{P})$, $g : \mathbb{R} \to \mathbb{R}$, $a, b \in \mathbb{R}$. The following holds:

1. $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ (linearity).
2. If $X \geq 0$ and $\mathbb{E}[X] = 0$, then $X = 0$.
3. If X, Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
4. If $Y = g(X)$, i.e., if Y is X-measurable, then

$$\mathbb{E}[Y] = \sum_{x \in \mathrm{Im}(X)} g(x) f_X(x). \tag{1}$$

Definition 5.10

Let $(\Omega, \mathbb{P})$ be a finite probability space. The **variance** of a random variable $X : \Omega \to \mathbb{R}$ is defined by

$$\mathrm{Var}[X] = \mathbb{E}[(\mathbb{E}[X] - X)^2].$$

Using the linearity of the expectation, we obtain easily the formula

$$\boxed{\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2}$$

Remarks

- The variance of a random variable is always non-negative and it is zero if and only if the random variable is a deterministic constant. Hence we may also interpret the variance as a measure of the "randomness" of a random variable.

- $\mathrm{Var}[aX] = a^2\mathrm{Var}[X]$ holds for all constants $a \in \mathbb{R}$, and

  $$\mathrm{Var}[X+Y] = \mathbb{E}[(X+Y)^2]-\mathbb{E}[X+Y]^2 = \mathrm{Var}[X]+\mathrm{Var}[Y]+2(\mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]).$$

  It follows by Theorem 3(3) that the variance of the sum of two independent random variables is the sum of their variance

Using (3) in Theorem 5.2 with $g(x) = x^2$, we can rewrite the definition of variance in terms of the distribution function of X as

$$\mathrm{Var}[X] = \sum_{x\in\mathrm{Im}(X)} x^2 f_X(x) - \left(\sum_{x\in\mathrm{Im}(X)} x f_X(x)\right)^2,$$

which allows to compute $\mathrm{Var}[X]$ without any reference to the original probability space.

For instance for the random variable on page 2 we find

$$\mathrm{Var}[X] = 1\cdot\frac{1}{4} + 4\cdot\frac{1}{4} + 1\cdot\frac{1}{2} - \left(\frac{1}{4}\right)^2 = \frac{27}{16}.$$

Example: mean of log return and volatility of the binomial stock price

Let $0 = t_0 < t_1 < \cdots < t_N = T$ be a partition of the interval $[0, T]$ with $t_i - t_{i-1} = h$, for all $i = 1, \ldots, N$.

Given $u > d$ and $p \in (0, 1)$, consider a random variable X such that $X = u$ with probability p and $X = d$ with probability $1 - p$.

We may think of X as being defined on $\Omega_1 = \{H, T\}$, with $X(H) = u$ and $X(T) = d$.

The binomial stock price at time t_i can be written as $S(t_i) = S(t_{i-1}) \exp(X)$.

Hence the log-return R of the stock in the interval $[t_{i-1}, t_i]$ is

$$R = \log S(t_i) - \log S(t_{i-1}) = \log \frac{S(t_i)}{S(t_{i-1})} = X.$$

It follows that the expectation and the variance of the log-return of the stock in the interval $[t_{i-1}, t_i]$ are

$$\mathbb{E}[R] = \mathbb{E}[X] = (pu + (1-p)d),$$

$$\mathrm{Var}[R] = \mathrm{Var}[X] = [pu^2 + (1-p)d^2 - (pu + (1-p)d)^2)] = p(1-p)(u-d)^2.$$

Thus the parameters α, σ^2 in the binomial model can be rewritten as

$$\boxed{\alpha = \frac{1}{h}\mathbb{E}[R], \quad \sigma^2 = \frac{1}{h}\mathrm{Var}[R]}$$

It is part of our assumptions on the binomial model that the parameters α and σ are the same for every interval $[t_{i-1}, t_i]$ of the partition.

Conditional expectation

If X, Y are independent random variables, knowing the value of Y does not help to estimate the random variable X.

However if X, Y are not independent, then we can use the information carried by Y to find an estimate of X which is better than $\mathbb{E}[X]$. This leads to the important concept of **conditional expectation**.

Definition 5.14

Let $(\Omega, \mathbb{P})$ be a finite probability space, $X, Y : \Omega \to \mathbb{R}$ random variables and $y \in \mathrm{Im}(Y)$. The expectation of X conditional to $Y = y$ (or given the event $\{Y = y\}$) is defined as

$$\boxed{\mathbb{E}[X|Y = y] = \textstyle\sum_{x \in \mathrm{Im(X)}} \mathbb{P}(X = x|Y = y)\, x}$$

where $\mathbb{P}(X = x|Y = y)$ is the conditional probability of the event $\{X = x\}$, given the event $\{Y = y\}$.

The random variable

$$\mathbb{E}[X|Y] : \Omega \to \mathbb{R}, \quad \mathbb{E}[X|Y](\omega) = \mathbb{E}[X|Y = Y(\omega)]$$

is called the expectation of X conditional to Y.

In a similar fashion one defines the conditional expectation with respect to several random variables, i.e., $\mathbb{E}[X|Y_1 = y_1, Y_2 = y_2, \ldots, Y_N = y_N]$ and $\mathbb{E}[X|Y_1, \ldots, Y_N]$.

Example

In the probability space of a fair die, consider

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

Note that $\mathrm{Im}(Y) = \{0, 6, 24, 60\}$. Then we compute

$$\begin{aligned}
\mathbb{E}[X|Y=0] &= \mathbb{P}(X=1|Y=0) - \mathbb{P}(X=-1|Y=0) \\
&= \frac{\mathbb{P}(X=1, Y=0)}{\mathbb{P}(Y=0)} - \frac{\mathbb{P}(X=-1, Y=0)}{\mathbb{P}(Y=0)} \\
&= \frac{\mathbb{P}(\{2\})}{\mathbb{P}(\{1,2,3\})} - \frac{\mathbb{P}(\{1,3\})}{\mathbb{P}(\{1,2,3\})} = -1/3.
\end{aligned}$$

Similarly we find

$$\mathbb{E}[X|Y=6] = 1, \quad \mathbb{E}[X|Y=24] = -1, \quad \mathbb{E}[X|Y=60] = 1,$$

hence $\mathbb{E}[X|Y]$ is the random variable

$$\mathbb{E}[X|Y](\omega) = \begin{cases} -1/3 & \text{if } \omega = 1, 2 \text{ or } 3 \\ 1 & \text{if } \omega = 4 \text{ or } 6 \\ -1 & \text{if } \omega = 5. \end{cases}$$

The following theorem collects a few important properties of the conditional expectation that will be used later on.

Theorem 5.3

Let $X, Y, Z : \Omega \to \mathbb{R}$ be random variables on the finite probability space $(\Omega, \mathbb{P})$. Then

(0) The random variable $\mathbb{E}[X|Y]$ is Y-measurable;

(1) The conditional expectation is a linear operator, i.e.,

$$\mathbb{E}[\alpha X + \beta Y|Z] = \alpha\mathbb{E}[X|Z] + \beta\mathbb{E}[Y|Z],$$

for all $\alpha, \beta \in \mathbb{R}$;

(2) If X is independent of Y, then $\mathbb{E}[X|Y] = \mathbb{E}[X]$;

(3) If X is measurable with respect to Y, i.e., $X = g(Y)$ for some function g, then $\mathbb{E}[X|Y] = X$;

(4) $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$;

(5) If X is measurable with respect to Z, then $\mathbb{E}[XY|Z] = X\mathbb{E}[Y|Z]$;

(6) If Z is measurable with respect to Y then $\mathbb{E}[\mathbb{E}[X|Y]|Z] = \mathbb{E}[X|Z]$.

These properties remain true if the conditional expectation is taken with respect to several random variables.

Remarks

- The interpretation of (2) is the following: If X is independent of Y, then the information carried by Y does not help to improve our estimate on X and thus our best estimate for X remains $\mathbb{E}[X]$.
- The interpretation of (3) is the following: if X is measurable with respect of Y, then by knowing Y we also know X and thus our best estimate on X is X itself.

Stochastic processes

Let $(\Omega, \mathbb{P})$ be a finite probability space and $T > 0$.

A one parameter family of random variables, $X(t) : \Omega \to \mathbb{R}$, $t \in [0, T]$, is called a **stochastic process**.

We denote the stochastic process by $\{X(t)\}_{t \in [0,T]}$ and by $X(t, \omega)$ the value of the random variable $X(t)$ on the sample $\omega \in \Omega$.

For each fixed $\omega \in \Omega$, the curve $t \to X(t, \omega)$, is called a **path** of the stochastic process.

We shall refer to the parameter t as the time variable, as this is what it represents in most applications.

If $X(t, \omega) = C(t)$, for all $\omega \in \Omega$, i.e., if the paths are the same for all sample points, we say that the stochastic process is a **deterministic function** of time.

If t runs over a (possibly finite) discrete set $\{t_0, t_1, \dots\} \subset [0, T]$, then we say that the stochastic process is **discrete**.

Note that a discrete stochastic process is equivalent to a sequence of random variables:

$$\{X_0, X_1, \dots\}, \quad \text{where } X_i = X(t_i),\ i = 0, 1, \dots.$$

If the discrete stochastic process is finite, i.e., if it runs only for a finite number N of time steps, we shall denote it by $\{X_n\}_{n=0,\dots,N}$ and call it a N**-period process**. If it runs for infinitely many steps we denote it by $\{X_n\}_{n \in \mathbb{N}}$.

Definition 5.15

Let $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ be two discrete stochastic processes on a finite probability space.

The process $\{Y_n\}_{n\in\mathbb{N}}$ is said to be **measurable** with respect to $\{X_n\}_{n\in\mathbb{N}}$ if for all $n \in \mathbb{N}$ there exists a function $g_n : \mathbb{R}^{n+1} \to \mathbb{R}$ such that $Y_n = g_n(X_0, X_1, \ldots, X_n)$.

If $Y_n = h_n(X_0, \ldots, X_{n-1})$ for some function $h_n : \mathbb{R}^n \to \mathbb{R}$, then $\{Y_n\}_{n\in\mathbb{N}}$ is said to be **predictable** from the process $\{X_n\}_{n\in\mathbb{N}}$.

Example: The random walk.

Consider the following (discrete and finite) stochastic process $\{X_n\}_{n=1,\ldots,N}$ defined on the N-coin toss probability space $(\Omega_N, \mathbb{P}_p)$:

$$\omega = (\gamma_1, \ldots, \gamma_N) \in \Omega_N, \quad X_n(\omega) = \begin{cases} 1 & \text{if } \gamma_n = H \\ -1 & \text{if } \gamma_n = T \end{cases}$$

Clearly, the random variables $X_1, \ldots, X_N$ are independent and identically distributed (**i.i.d**), namely

$$\mathbb{P}_p(X_n = 1) = p, \quad \mathbb{P}_p(X_n = -1) = 1 - p, \quad \text{for all } n = 1, \ldots, N.$$

Hence

$$\mathbb{E}[X_n] = 2p - 1, \quad \mathrm{Var}[X_n] = 4p(1-p), \quad \text{for all } n = 1, \ldots, N.$$

Now, for $n = 1, \ldots, N$, let

$$M_0 = 0, \quad M_n = \sum_{i=1}^{n} X_i.$$

The stochastic process $\{M_n\}_{n=0,\ldots,N}$ is called the **(N-period) random walk**.

It satisfies

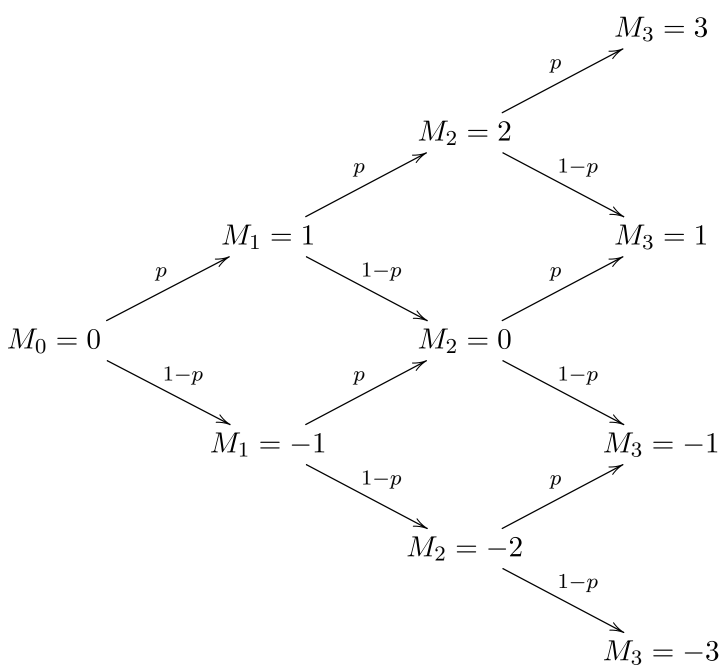
$$\mathbb{E}[M_n] = n(2p-1), \quad \text{for all } n = 0, \ldots, N.$$

Moreover, being the sum of independent random variables, the random walk has variance given by

$$\mathrm{Var}[M_0] = 0, \quad \mathrm{Var}[M_n] = \mathrm{Var}(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{Var}[X_i] = 4np(1-p).$$

When $p = 1/2$, the random walk is said to be **symmetric**. In this case $\{M_n\}_{n=0,\ldots,N}$ satisfies $\mathbb{E}[M_n] = 0$, $n = 0, \ldots, N$ and $\mathrm{Var}[M_n] = n$. When $p \neq 1/2$, $\{M_n\}_{n=0,\ldots,N}$ is called **asymmetric** random walk, or random walk with **drift**.

If $M_n = k$ then M_{n+1} is either $k+1$ (with probability p), or $k-1$ (with probability $1-p$). Hence we can represent the paths of the random walk by using a binomial tree, as in the following example for $N = 3$:



Martingales

A martingale is a stochastic process which has no tendency to rise or fall. The precise definition is the following.

Definition 5.16

A discrete stochastic process $\{X_n\}_{n\in\mathbb{N}}$ on the finite probability space $(\Omega, \mathbb{P})$ is called a **martingale** if

$$\mathbb{E}[X_{n+1}|X_0, X_1, \ldots, X_n] = X_n, \quad \text{for all } n \in \mathbb{N}.$$

Interpretation: The variables $X_0, X_1, \ldots, X_n$ contains the information obtained by "looking" at the stochastic process up to the step n. For a martingale process, this information is not enough to estimate whether, in the next step, the process will raise or fall.

Remarks

1. The martingale property depends on the probability being used: if $\{X_n\}_{n\in\mathbb{N}}$ is a martingale in the probability $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ is another probability measure on the sample space Ω, then $\{X_n\}_{n\in\mathbb{N}}$ need not be a martingale with respect to $\widetilde{\mathbb{P}}$.

2. Using property 4 in Theorem 5.3 we obtain

$$\mathbb{E}[X_{n+1}] = \mathbb{E}[X_n], \quad \text{for all } n \in \mathbb{N}.$$

   Thus, iterating, $\mathbb{E}[X_n] = \mathbb{E}[X_0]$, for all $n \in \mathbb{N}$, i.e., *martingales have constant expectation.*

Example

Next we show that the *symmetric* random walk is a martingale.

Using the linearity of the conditional expectation we have, for all $n = 0, \dots, N-1$,

$$\begin{aligned}\mathbb{E}[M_{n+1}|M_0,\dots,M_n] &= \mathbb{E}[M_n + X_{n+1}|M_0,\dots,M_n] \\ &= \mathbb{E}[M_n|M_0,\dots,M_n] + \mathbb{E}[X_{n+1}|M_0,\dots,M_n].\end{aligned}$$

As M_n is measurable with respect to $M_0, \dots, M_n$, then

$$\mathbb{E}[M_n|M_0,\dots,M_n] = M_n$$

see Theorem 5.3(3).

Moreover, as X_{n+1} is independent of $M_0, \dots, M_n$, Theorem 5.3(2) gives

$$\mathbb{E}[X_{n+1}|M_0,\dots,M_n] = \mathbb{E}[X_{n+1}] = 0$$

It follows that $\mathbb{E}[M_{n+1}|M_0,\dots,M_n] = M_n$, i.e., the symmetric random walk is a martingale.

However the asymmetric random walk ($p \neq 1/2$) is *not* a martingale, as it follows by the fact that its expectation $\mathbb{E}[M_n] = n(2p-1)$ is not constant (it depends on $n \in \mathbb{N}$).