光华管理学院 Guanghua School of Management

Mathematical Methods in Finance

Lecture 2: Conditional Expectation and Introduction to Stochastic Processes

Fall 2013

Overview

- Conditional expectation
- Stochastic processes, e.g. random walk and Poisson processes, etc.

Conditioning: Motivation and Traditional Version

- Motivation: in finance, we usually need to consider conditional behavior (distribution) of asset price given that at an earlier time, e.g. how will the IBM stock price distribute on Oct. 28th assuming that its price at Oct. 1st **were** known.
- Traditional version: $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.
- Consider $A = \{X \leq x\}$ where X is a continuous R.V.
- Conditional CDF: $F(x|B) = \mathbb{P}(X \leq x|B)$
- Conditional PDF: $f(x|B)$ such that $F(x|B) = \int_{-\infty}^{x} f(u|B)du$
- Conditional Expectation: $\mathbb{E}(X|B) = \int_{-\infty}^{+\infty} uf(u|B)du$
- Similarly, we can consider whatever we had for unconditional cases, e.g. conditional variance, etc.
- If X is a discrete R.V., we just change the $\int$ to $\sum$ properly

Conditioning Induced by a Random Variable

- Let $B = \{Y = y\}$ where Y is a R.V.
- If Y is a Discrete R.V. and $\mathbb{P}(B) \neq 0$, just follow previous slides
- If Y is a Continuous R.V., $\mathbb{P}(B) = 0$. Problematic with the division!
- Overcome this by considering $B_\epsilon = \{Y \in (y - \epsilon, y + \epsilon)\}$ and then letting $\epsilon \to 0$
- Conditional CDF:

$$\mathbb{P}(X \leq x|Y = y) = F(x|Y = y) = \int_{-\infty}^{x} f(u|Y = y)du$$

- Conditional PDF:

$$f(u|Y = y) = \frac{f(x, y)}{f_Y(y)}$$

Conditional Expectation: from Classical to Modern Version

Classical definition: a number!

- If X is a continuous R.V., $\mathbb{E}(X|Y=y) = \int_{-\infty}^{+\infty} u f(u|Y=y) du$
- If X is a discrete R.V., $\mathbb{E}(X|Y=y) = \sum_x x\mathbb{P}(x|Y=y)$

Example: Two independent dice are rolled and Y denotes the value of the first roll and X denotes the sum of the two rolls, then we calculate that

$$\mathbb{E}(X|Y=y) = y + \frac{7}{2}.$$

Now, can we consider **random variable** relying on the randomness of the condition in Y (substituting y by Y):

$$\mathbb{E}(X|Y) = Y + \frac{7}{2}.$$

By doing so, we average something random based on something random. (Think about a forward started contract!)

Conditional Expectation: A Further Development

Definition: Given two random variables X and Y, we define the conditional expectation $\mathbb{E}(X|Y)$ as a random variable obtained from

$$\mathbb{E}(X|Y) = g(Y),$$

where $g(y) := \mathbb{E}(X|Y=y)$.

Conditioning on more information (cautious! more abstract):

Let $\sigma(Y)$ be "all the events generated by Y". Formally, $\sigma(Y)$ is the "smallest σ-algebra containing all events like $\{\omega \in \Omega : Y(\omega) \leq y\}$", can we define $\mathbb{E}(X|\sigma(Y))$?

Can we further define a version of conditional expectation in a more informatively way? say, $\mathbb{E}(X|\mathcal{G})$ where $\mathcal{G}$ is the $\sigma-$algebra (just information)?

Conditional Expectation: Formal Mathematical Version

- **Definition**: Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a sub-$\sigma$-algebra $\mathcal{G} \subseteq \mathcal{F}$, and a RV X s.t. either $X \geq 0$ or $|E(X)| < +\infty$. Then the conditional expectation of X given $\mathcal{G}$, denoted by $E(X|\mathcal{G})$, is any RV that satisfies
  - (i) (**Measurability**) $E(X|\mathcal{G})$ is $\mathcal{G}$-measurable, and
  - (ii) (**Partial Averaging**) $\mathbb{E}[1_A E(X|\mathcal{G})] = \mathbb{E}(1_A X)$ for all $A \in \mathcal{G}$. Here, 1_A is a random variable, called indicator, defined as: $1_A(\omega) = 1$, if $\omega \in A$; $1_A(\omega) = 0$, otherwise.
- The mathematical issue of this definition is too abstract to us. We just take it for granted!
- In particular, when $\mathcal{G} = \sigma(Y)$, we can prove that

$$\mathbb{E}(X|\sigma(Y)) = \mathbb{E}(X|Y).$$

  i.e. the formal mathematical version and the classical version agree!
- $\mathbb{E}(X|Y)$ can be calculated from conditional distributions, i.e.

$$E(X|Y = y) = g(y) \Longrightarrow E(X|Y) = E(X|Y = y)|_{y=Y} = g(Y)$$

Some Properties

- (**Linearity**) $\mathbb{E}(c_1 X + c_2 Y|\mathcal{G}) = c_1\mathbb{E}(X|\mathcal{G}) + c_2\mathbb{E}(Y|\mathcal{G})$.
- (**Expectation and conditioning)** $\mathbb{E}[\mathbb{E}(X|\mathcal{G})] = \mathbb{E}X$
- **(Taking out what is known) If $\sigma(X) \subset \mathcal{G}$ (X is $\mathcal{G}$-measurable), then $\mathbb{E}(XY|\mathcal{G}) = X\mathbb{E}(Y|\mathcal{G})$.**
- **(Iterated conditioning) If $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$, then $\mathbb{E}\left[\mathbb{E}(X|\mathcal{G})|\mathcal{H}\right] = \mathbb{E}(X|\mathcal{H})$.**
- **(Independence) If X is independent of $\mathcal{G}$, then $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$.**
- **(Independence) If X is $\mathcal{G}$ measurable (i.e. $\sigma(X) \subset \mathcal{G}$) and Y is independent of $\mathcal{G}$, then $\mathbb{E}[f(X, Y)|\mathcal{G}] = g(X)$, where $g(x) = \mathbb{E}f(x, Y)$.**
- **(Conditional Jensen's Inequality) If $\phi(x)$ is convex, then $\mathbb{E}(\phi(X)|\mathcal{G}) \geq \phi(\mathbb{E}(X|\mathcal{G}))$.**

Understanding the Properties by Simply Taking $\mathcal{G} = \sigma(Y)$

Assume the information is generated by a random variable Y, i.e. $\mathcal{G} = \sigma(Y)$. Classical probability is helpful for us to understand these properties. For example,

$$\begin{aligned} E(E(X|\mathcal{G})) =& E(E(X|\sigma(Y))) \\ =& E(E(X|Y)) = E(g(Y)) \\ =& \int_{-\infty}^{\infty} g(y) f_Y(y) dy \\ =& \int_{-\infty}^{\infty} E(X|Y=y) f_Y(y) dy = EX. \end{aligned} \tag{1}$$

Similarly, try to understand other properties (excellent exercises)!

Conditional Expectation as the Optimal Projection

- Let X be a $\mathcal{F}$-measurable (random variable) R.V..
- For $\mathcal{G} \subset \mathcal{F}$, denote by $L^2(\mathcal{G})$ the collection of $\mathcal{G}$-measurable R.V. satisfying that $\mathbb{E}Z^2 < \infty$.
- Question: for what Z, do we have

$$\min_{Z \in L^2(\mathcal{G})} \mathbb{E}(X - Z)^2?$$

- Answer:

$$\mathbb{E}(X - \mathbb{E}(X|\mathcal{G}))^2 = \min_{Z \in L^2(\mathcal{G})} \mathbb{E}(X - Z)^2?$$

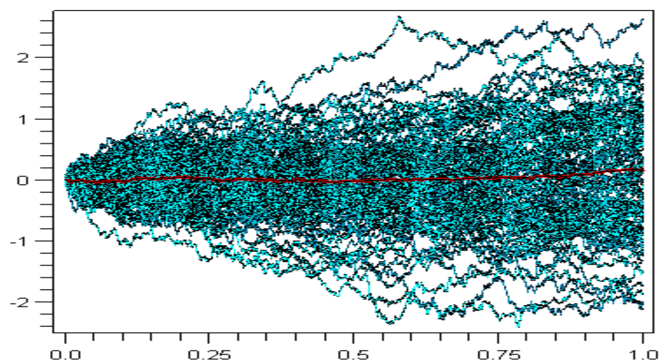
- i.e. the mean square distance is attained by the conditional expectation
- Analogy: the projection of X

Stochastic Process

- A collection of random variables:

$$\{X(t), 0 \leq t \leq T\}$$

- For emphasizing the dependence of both time and random events, we heuristically write $X(t, \omega)$.
- Understanding:
  - A function of two variables $(\omega, t) \in \Omega \times (0, T]$
  - Fix any $t \in (0, T]$, $X(t, \cdot)$ is a random variable
  - Fix any $\omega \in \Omega$, $X(\cdot, \omega)$ is a realization (path) of the process



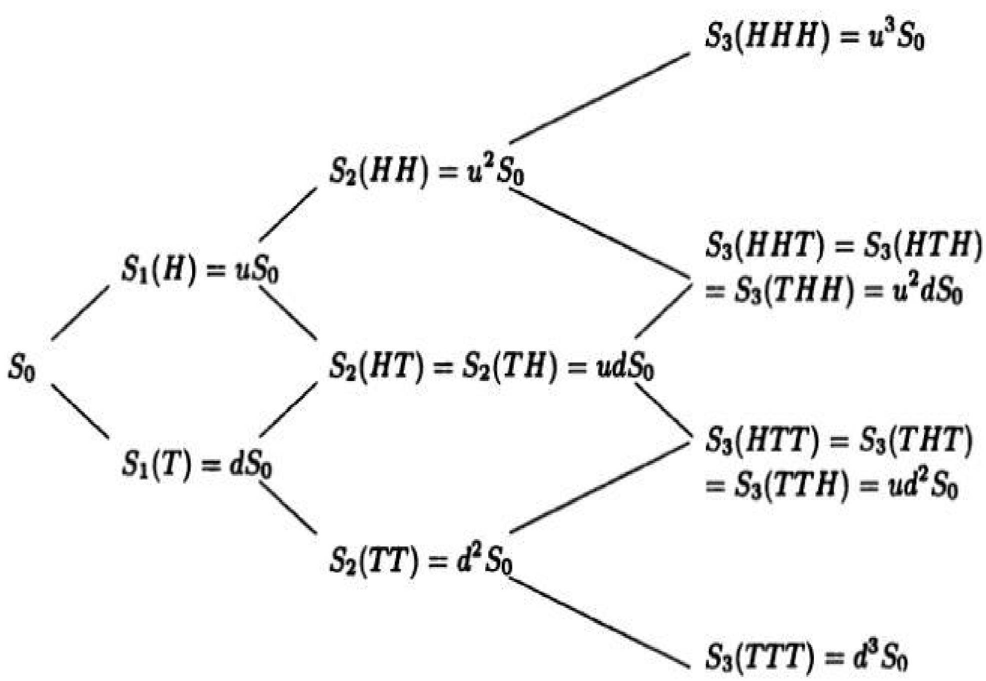
Stochastic Process: Formal Definition

- **Definition**: Let Ω be a nonempty set and T be a fixed positive number. Assume that for each $t \in (0, T]$, there exists a σ-algebra (information) $\mathcal{F}(t)$, and that for any $s \leq t$, $\mathcal{F}(s) \subseteq \mathcal{F}(t)$. Then we call the collection of σ-algebra $\mathcal{F}(t)$, $0 \leq t \leq T$, a filtration.
- **Definition**: Let Ω be a nonempty set equipped with a filtration $\mathcal{F}(t)$, $0 \leq t \leq T$. Let $\{X(t); 0 \leq t \leq T\}$ be a collection of RVs indexed by $t \in [0, T]$. We say this collection of RVs is an adapted (continuous time) stochastic process if for each t, we have $X(t)$ is $\mathcal{F}(t)$-measurable.
- Similarly define filtration (denoted by $\mathcal{F}_n$) and stochastic process (denoted by $\{X_n : n \in \mathbb{N}\}$) on discrete time
- Note: the state space (all possible values) of $X(t)$ (or X_n) is continuous or not is a different concept.

Stochastic Process: Heng Seng Index as an Example

An Example: Binomial Lattice for Option Pricing

An illustration of the filtration: $N = 3$



An Example: Binomial Lattice for Option Pricing

Let us recast a N-period binomial lattice model. A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $\{\mathcal{F}_n\}$ can be specified as follows:

- $\Omega = \{(\omega_1\omega_2\cdots\omega_N) : \omega_i = \mathrm{H} \text{ or } \mathrm{T} \text{ for any } i = 1, 2, \cdots, N\}$: the collection of results of N coin tosses
- $\mathcal{F}$ (collection of all possible events) is a σ-algebra generated by all subsets of Ω.
- Probability measure: $\forall \omega = (\omega_1\omega_2\cdots\omega_N) \in \Omega$,

$$\mathbb{P}(\{\omega\}) := p^{\sum_{i=1}^{N} I_{\{\omega_i=H\}}} q^{\sum_{i=1}^{N} I_{\{\omega_i=T\}}}.$$

  $\forall A \in \mathcal{F}_N$, $\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$.
- Consider a filtration $\{\mathcal{F}_n\}$, $n = 1, 2, \cdots, N$. $\mathcal{F}_n$ is a σ-algebra generated by the "information" up to n. We thus have $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ and $\mathcal{F}_N = \mathcal{F}$.

An Example: Binomial Lattice for Option Pricing

An illustration of the filtration: $N = 3$

- Information up to the first day: $\mathcal{F}_1 = \{\phi, \Omega, A_H, A_T\}$. Here

$$\begin{aligned} A_H =& \{HHH, HHT, HTH, HTT\}, \\ A_T =& \{THH, THT, TTH, TTT\}. \end{aligned} \tag{2}$$

- Information up to the second day: $\mathcal{F}_2 = \{\phi, \Omega, A_H, A_T, A_{HH}, A_{HT}, A_{TH}, A_{TT}, A_{HH}^c, A_{HT}^c, A_{TH}^c, A_{TT}^c, A_{HH} \cup A_{TH}, A_{HH} \cup A_{TT}, A_{HT} \cup A_{TH}, A_{HT} \cup A_{TT}\}$. Here

$$\begin{aligned} A_{HH} =& \{HHH, HHT\}, \quad A_{HT} = \{HTH, HTT\}, \\ A_{TH} =& \{THH, THT\}, \quad A_{TT} = \{TTH, TTT\}. \end{aligned} \tag{3}$$

- Information up to the third day: $\mathcal{F}_3 =$ the set of all subsets of $\Omega = \mathcal{F}$

Stochastic Process: Some Concepts and Properties

- Finite dimensional distributions: the distribution of $(X(t_1), X(t_2), ..., X(t_n))$, where $t_1 < t_2 < ... < t_n$
- Covariance function: $c(t, x) = cov(X(t), X(s))$
- A process is stationary: iff the finite dimensional distribution is translate invariant, i.e.

  $$(X(t_1), X(t_2), ..., X(t_n)) =^d (X(t_1 + h), X(t_2 + h), ..., X(t_n + h)).$$

  $A =^d B$ means the distribution of A and B are the same.
- A process has stationary increments: iff

  $$X(t) - X(s) =^d X(t + h) - X(s + h).$$

- A process has independent increments: iff

  $$X(t_2) - X(t_1), ..., X(t_n) - X(t_{n-1})$$

  are all independent.

Markov Property

- **Definition**: Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a filtration $\mathcal{F}(t)(\subseteq \mathcal{F})$, $0 \leq t \leq T$, and an adapted stochastic process $X(t)$, $0 \leq t \leq T$. Assume that for all $0 \leq s \leq t \leq T$ and for every function f, there exists another function g such that

  $$\mathbb{E}[f(X(t))|\mathcal{F}(s)] = \mathbb{E}[f(X(t))|X(s)](= g(X(s)))$$

  for some function g and all $0 \leq s \leq t \leq T$, we say the process $X(t)$ enjoys the Markov property.
- A process is Markov if it totally forgets about the past: the future depends on the current state only!
- Markov processes in the discrete case – Markov Chains.
- Note that g depends on $f(\cdot)$, t, and s.

Transition Probability and Transition density

- Describe the transition from one state to another of a Markov process
- For discrete time Markov process with discrete state space, we consider the following transition probability:

$$\mathbb{P}(X_{n+1} = a | X_n = b, X_{n-1} = c, ...) = \mathbb{P}(X_{n+1} = a | X_n = b)$$

- For continuous time Markov process with continuous state space, we consider the transition density: for $0 \le s < t$,

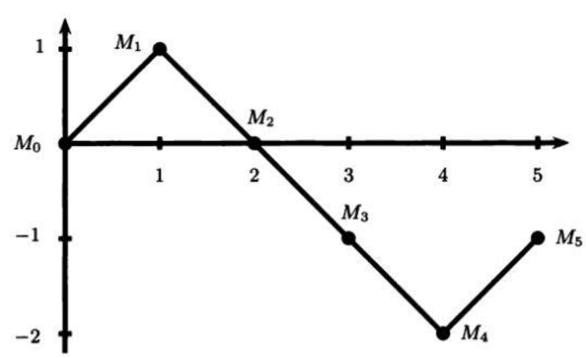
$$p(t, y; s, x) := \frac{d}{dy} P(X(t) \le y | X(s) = x).$$

An example of discrete time process: Simple Random Walk

A simple symmetric random walk

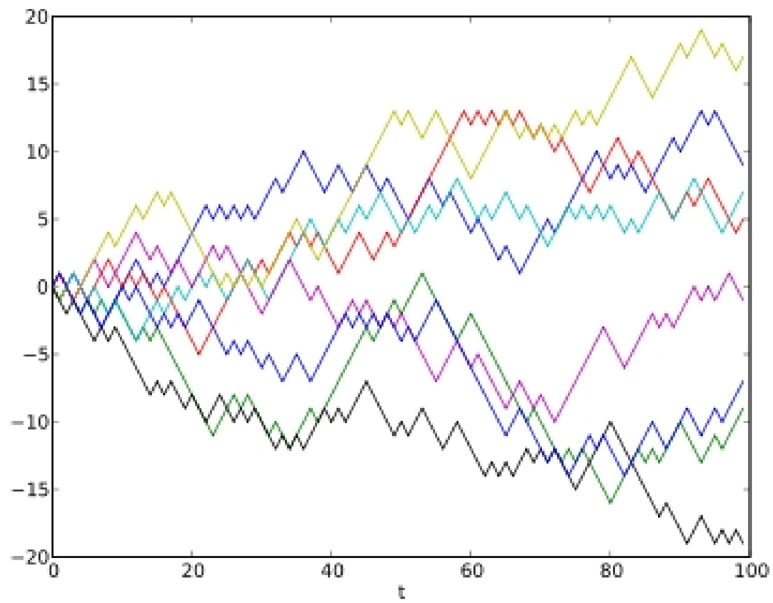
$$W_0 = 0, \quad W_n = \sum_{k=1}^{n} X_k,$$

where $X_k = 1$ or $X_k = -1$ with probability $1/2$.



Five steps of a random walk.

More Sample Paths for a Simple Random Walk



Some Properties of Simple Random Walk

- A simple symmetric random walk

$$W_0 = 0, \quad W_n = \sum_{k=1}^{n} X_k,$$

where $X_k = 1$ or $X_k = -1$ with probability $1/2$.
- A Markov process
- Transition probability:

$$\mathbb{P}(W_{n+1} = s | W_n = r) = \begin{cases} 1/2, & \text{if } s = r + 1; \\ 1/2, & \text{if } s = r - 1. \end{cases}$$

- For $0 < m < n$, conditional expectation: $E(W_n | W_m) = ?$
- Conditional second moments: $E(W_n^2 | \mathcal{F}_m) = ?$ where $\mathcal{F}_m = \sigma(X_1, ..., X_m)$

Simple Random Walk: Gambler's Problem

- An excellent question: Gambler's Ruin

  Suppose you start with n dollars, and make a sequence of bets. For each bet, you win 1 dollar with probability $1/2$, and lose 1 dollar with probability $1/2$. You quit if either you go broke, in which case you lose, or when you reach $n+m$ dollars, what is the probability of your winning?

- Use R.W. to model and solve it by conditioning.

- Later, we solve this more systematically using "martingale"!

An example of continuous time process: Poisson Process

- Model arrivals: e.g. a "jump" in financial market
- A Poisson process $\{N(t)\}$ with intensity λ:
  - N(0)=0
  - stationary and independent increment
  - $N(t+\Delta t) - N(t)$ is a Poisson R.V. with parameter $\lambda \Delta t$

$$\mathbb{P}\left(N(t+\Delta t) - N(t) = k\right) = \frac{\lambda^k \Delta t^k}{k!} e^{-\lambda \Delta t} \tag{4}$$

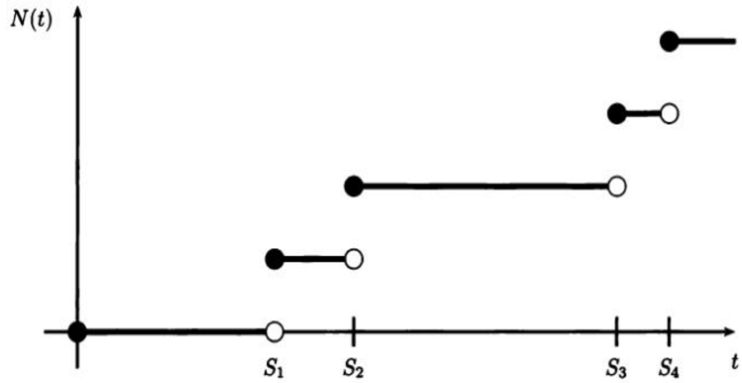
- Constriction from exponential R.V.s,

$$N(t) = \max\left\{n : S_n = \sum_{i=1}^{n} \tau_i \leq t\right\},$$

  where τ_i are I.I.D with exponential distribution: $\mathbb{P}(\tau_i \leq x) = 1 - e^{-\lambda x}$, for $x \geq 0$.
- (4) can be proved using this construction by computing on the density of $\sum_{i=1}^{n} \tau_i$ (an excellent exercise).

Poisson Process

- The path is right continuous



- Expectation of the increment $\mathbb{E}[N(t+\Delta t) - N(t)] = \lambda \Delta t$; this is why λ is called "intensity": arrival per unit time
- What is the variance of the increment?
- Further reading (if you are interested in financial modeling): compounded Poisson process, non-constant intensity, etc.

Supplementary Material

Suggested Reading Material (We only need to focus on the material parallel to our course slides):

- Shreve Vol. II: Section 2.1, 2.2 and 2.3

  Or you can find equivalent material from
- Mikosch: Section 1.2 and 1.4

Suggested Exercises (Do Not Hand In; For Your Deeper Understanding Only)

- Shreve Vol. II: Exercise 2.6, 2.7, 2.8, 2.10