Mathematical & Computational Finance II
Lecture Notes

Welcome to Measure Theory

September 22 2015
Last update: December 4, 2017

1 Continue the Crash Course on Probability Measures

Definition 1. For $X \in L^p(\Omega\mathcal{F}, \mathbb{P})$, for $1 \leq p < \infty$, define a norm (generalized Euclidean norm on $\mathbb{R}^n$) as

$$\|X\|_p = (\mathbb{E}|X|^p])^{1/p}$$

Let $1 \leq p < \infty$, define $q = \frac{p}{1-p}$ [1]. Then $q \in (1, \infty)$ and $\frac{1}{p} + \frac{1}{q} = 1$. If p and q are conjugates then, for $a, b > 0$,

$$a^{1/p} + b^{1/q} \leq \frac{1}{p}a + \frac{1}{q}b$$

Proposition 1. If p and q are conjugates and $X \in L^p, Y \in L^q$ then,

$$XY \in L^1 \quad \text{and}$$
$$\mathbb{E}[|XY|] \leq \|X\|_p + \|Y\|_p$$

Proposition 2. (Minkowski's Inequality)[2] For $1 \leq p < \infty$, if $X, Y \in L^p$ then

$$X + Y \in L^p \quad \text{and}$$
$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$$

Remarks:

1. $\|\lambda X\|_p = \lambda \|X\|_p$, for $\lambda \in \mathbb{R}$
2. $\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$
3. If $\|X\|_p = 0$ then $|X|_p = 0$ a.s. $\implies X = 0$ a.s.[3]

These remarks give us that $\|\cdot\|_p$ is a norm on L^p. So, we say that L^p is a normed linear space.

[1] That is, q is the conjugate to p.
[2] This is a generalization to the Triangle Inequality.
[3] i.e. $X = 0$ up to an equivalence class a.s.

1.1 L^2 and Conditional Expectation

A linear functional on L^2 is a map $\phi : L^2(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$, or equivalently $\phi : X \to \mathbb{R}$, such ϕ is linear:

$$\phi(\alpha X + \beta Y) = \alpha\phi(X) + \beta\phi(Y) \quad \forall\, X, Y \in L^2, \forall\, \alpha, \beta \in \mathbb{R}$$

A map $\phi : L^2 \to \mathbb{R}$ is bounded if $\exists\, k > 0$ such that

$$|\phi(X)| \leq k \|X\|_2 \quad \forall\, X \in L^2$$

A sequence of random variables $\{X_n\} \in L^p$ converges to $X \in L^p$ if

$$\begin{gathered} X \in L^p \quad \text{and} \\ \|X_n - X\|_p \longrightarrow 0 \quad \text{as } n \longrightarrow \infty \end{gathered}$$

Suppose $\{X_n\} \in L^2$ converges to $X \in L^2$ and ϕ is a bounded linear functional on L^2 then,

$$\begin{aligned} |\phi(X_n) - \phi(X)| &= |\phi(X - X_n)| \quad \text{(by linearity)} \\ |\phi(X_n - X)| &\leq k \|X_n - X\|_2 \longrightarrow 0 \quad (\text{as } n \longrightarrow \infty) \end{aligned}$$

Now we may define

Definition 2. Suppose ϕ is a bounded linear functional on L^2, define

$$\|\phi\| = \inf_{\{X \in L^2 \,:\, \|X\|_2 \neq 0\}} \frac{|\phi|}{\|X\|_2}$$

Aside: $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is what we call a *Hilbert Space*[4] with inner product

$$\langle X, Y \rangle = \int XY \, d\mathbb{P} = \mathbb{E}[XY]$$

Definition 3. A sequence $\{Y_n\}$ in a normed vector space is a Cauchy sequence if

$$\sup_{m \in N} \|y_{n+m} - y_n\| \longrightarrow 0 \quad \text{as } n \longrightarrow \infty$$

That is, we take elements of the sequence arbitrarily far apart and see their norm $\longrightarrow 0$ as $n \longrightarrow 0$.

We say a space is complete if every Cauchy sequence is a convergent sequence.

Theorem: $L^p(\Omega, \mathcal{F}, \mathbb{P})$ is a complete normed vector space.

Why is this important? Whenever you have a Hilbert space this gives you the following theorem...

[4] A Hilbert Space is a *complete* inner product space (Banach Space)[5].

[5] Left undefined for this course.

Theorem: Riesz Representation Theorem. Let $\mathcal{H}$ be a Hilbert space and L be a linear continuous functional on $\mathcal{H}$. Then there exists a unique $y \in \mathcal{H}$ such that

$$L(x) = \langle x, y \rangle \quad \forall x, y \in \mathcal{H} \quad \text{with } \|L\| = \|y\|$$

Definition 4. Let $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$ (i.e. $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$). Then the conditional expectation of X with respect to $\mathcal{G}$ denoted $\mathbb{E}[X|\mathcal{G}]$ is a random variable $Z \in L^2$ satisfying

1. Z is $\mathcal{G}$-measurable.
2. $\mathbb{E}[ZY] = \mathbb{E}[XY] \quad \forall$ bounded $\mathcal{G}$-measurable random variables Y.

Note that Z is a random variable depending on $\omega \in \Omega$ meaning $Z = Z(\omega) = \mathbb{E}[X|\mathcal{G}](\omega)$.

1.1.1 Existence

For fixed $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ the map

$$\begin{aligned} &\phi_X : L^2(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R} \quad \text{or equivalently} \\ &\phi_X : Y \to \mathbb{E}[XY] \end{aligned}$$

is a bounded continuous linear functional on $L^2(\Omega, \mathcal{G}, \mathbb{P})$. So,

$$\exists\, Z \in L^2(\Omega, \mathcal{G}, \mathbb{P}) \quad \text{(by the Riesz Representation Theorem)}$$

Such that

$$\phi_X(Y) = \int XY \, d\mathbb{P} = \langle Z, Y \rangle = \int ZY \, d\mathbb{P}$$

For all $Y \in L^2(\Omega, \mathcal{G}, \mathbb{P})$.
$\therefore Z$ satisfies Definition **??** Conditions **??** and **??**.

1.1.2 Uniqueness

This is trickier to do and is omitted in this course.

1.1.3 Interpretation

$\mathbb{E}[X_2|X_1]$ *really* means $\mathbb{E}[X_2|\sigma(X_1)]$ where the conditional $\sigma(X_1)$ means the information generated by the smallest σ-algebra generated by X_1.

1.1.4 Properties

1. Linear: $\mathbb{E}[\alpha X_1 + \beta X_2|\mathcal{G}] = \alpha\mathbb{E}[X_1|\mathcal{G}] + \beta\mathbb{E}[X_2|\mathcal{G}]$
2. Integrable: $\mathbb{E}[|\mathbb{E}[X|\mathcal{G}]|] < \infty$
3. If $X \geq 0$ then $\mathbb{E}[X|\mathcal{G}] \geq 0$ (in probability a.s.)

4. $\mathbb{E}[a|\mathcal{G}] = a, \quad \forall\, a \in \mathbb{R}$

5. "Taking out what is known": If W is $\mathcal{G}$-measurable (i.e. $W \in \mathcal{G}$) and $\mathbb{E}[|XW|] < \infty$ (i.e. XW is integrable) $\implies \mathbb{E}[XW|\mathcal{G}] = W\mathbb{E}[X|\mathcal{G}]$

   Corollary 1. If X is $\mathcal{G}$-measurable then $\mathbb{E}[X|\mathcal{G}] = X$

   Corollary 2. If X is independent[6] of $\mathcal{G}$ then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$ (i.e. $\mathcal{G}$ gives us no information about X).

   Remember: If $\mathcal{C}$ generates $\mathcal{F}$ then $\{X^-1(c) : c \in \mathcal{C}\}$ generates $\sigma(X)$.

6. The "Tower' Property": If $\mathcal{H} \subseteq \mathcal{G}$ are sub-σ-algebras of $\mathcal{F}$ then

$$\begin{aligned} \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] &= \mathbb{E}[X|\mathcal{H}] \\ &= \mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}] \end{aligned}$$

7. "Jensen's Inequality": If $X \in L^2$ we can show

$$(\mathbb{E}[X|\mathcal{G}])^2 \leq \mathbb{E}[X^2|\mathcal{G}]$$

Here's a nice result as to why conditional expectation is useful

Proposition 3. Let $X \in L^2, g(Y) \in L^2$ (i.e. g is a square integrable function) then

$$\begin{aligned} \mathbb{E}[(X - g(Y))^2] &= \int (X - g(Y))^2 \, d\mathbb{P} \\ &= \int (X - \mathbb{E}[X|\sigma(Y)] + \mathbb{E}[X|\sigma(Y)] - g(Y))^2 \, d\mathbb{P} \quad \text{(add \& subtract the same value)} \\ &= \int (X - \mathbb{E}[X|\sigma(Y)])^2 \, d\mathbb{P} + 2\int (X - \mathbb{E}[X|\sigma(Y)])(\mathbb{E}[X|\sigma(Y)] - g(Y)) \, d\mathbb{P} \\ &\qquad + \int (\mathbb{E}[X|\sigma(Y)] - g(Y))^2 \, d\mathbb{P} \\ &= \mathbb{E}[(X - \mathbb{E}[X|\sigma(Y)])^2] + 2\mathbb{E}[(X - \mathbb{E}[X|\sigma(Y)])(\mathbb{E}[X|\sigma(Y)] - g(Y))] \\ &\qquad + \mathbb{E}[(\mathbb{E}[X|\sigma(Y)] - g(Y))^2] \end{aligned}$$

Note that in the middle term $\mathbb{E}[(X - \mathbb{E}[X|\sigma(Y)])(\mathbb{E}[X|\sigma(Y)] - g(Y))]$ we have $(\mathbb{E}[X|\sigma(Y)] - g(Y)) \in \sigma(Y)$, so

$$\begin{aligned} \mathbb{E}[(X - \mathbb{E}[X|\sigma(Y)])(\mathbb{E}[X|\sigma(Y)] - g(Y))] &= \mathbb{E}[\{(X - \mathbb{E}[X|\sigma(Y)])(\mathbb{E}[X|\sigma(Y)] - g(Y))\}|\sigma(Y)] \\ &= \mathbb{E}[(\mathbb{E}[X|\sigma(Y)] - g(Y))\mathbb{E}[X - \mathbb{E}[X|\sigma(Y)|\sigma(Y)]]] \end{aligned}$$

[6] X is independent of $\mathcal{G}$ if $\forall\, A \in \sigma(X)$ and $\forall\, B \in \mathcal{G} \implies \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. This is something to absorb & dwell on for a moment... but the basic intuition is the same.

But in this expectation we have

$$\begin{aligned}\mathbb{E}[X - \mathbb{E}[X|\sigma(Y)]|\sigma(Y)] &= \mathbb{E}[X|\sigma(Y)] - \mathbb{E}[\mathbb{E}[X|\sigma(Y)]|\sigma(Y)] \quad \text{(by linearity)} \\ &= \mathbb{E}[X|\sigma(Y)] - \mathbb{E}[X|\sigma(Y)] \quad \text{(this is obvious, do we have to elaborate?)} \\ &= 0\end{aligned}$$

So our middle term vanishes leaving our original expectation as

$$\mathbb{E}[(X - \mathbb{E}[X|\sigma(Y)]^2] + \mathbb{E}[(\mathbb{E}[X|\sigma(Y)] - g(Y))^2] \geq \mathbb{E}[X - \mathbb{E}[X|\sigma(Y)]^2]$$

$\therefore \mathbb{E}[X|\sigma(Y)]$ is the best estimator (in L^2) of X that is a function of Y.

2 Stochastic Processes

"A stochastic process is a family of random variables indexed by some set, usually time."

Definition 5. A stochastic process is a map $X : \Omega \times [0, \infty) \to \mathbb{R}^d$

If we fix a $\omega \in \Omega$ and consider the map $t \to X_t(\omega)$ then we are looking at some sample path

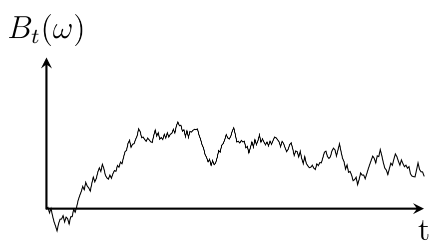

Figure 1: Some realisation of a stochastic process $X_t(\omega)$.

or "realisation" of the process. So, fix t and consider a map $\omega \to X(\omega, t)$. For each fixed t, $X(\omega, t)$ would have some distribution (i.e. the distribution of the "cross section" of $X(\omega, t)$ for fixed t).

Let $B \in \mathcal{B}(\mathbb{R}^d)$ and consider

$$X_s^{-1}(B) = \{\omega \in \Omega : X_s(\omega) \in B\}$$

Define

$$\mathcal{X}_s = \{X_s^{-1}(B) : B \in \mathcal{B}(\mathbb{R}^d)\} \quad \text{(not necessarily a } \sigma\text{-algebra)}$$

Then, let

$$\mathcal{F}_t^X = \sigma(\mathcal{X}_s : 0 \leq s \leq t) \quad \text{(i.e. our } \sigma\text{-algebra generated by } \mathcal{X}_s)$$

Note that $\mathcal{F}_t^X$ contains $\mathcal{F}_s^X$ (i.e. $\mathcal{F}_s^X \subseteq \mathcal{F}_t^X$). We say that $(\mathcal{F}_t^X)_{t \geq 0}$ is a filtration, that is more information is releaved in our σ-algebra as we progress in t.

2.1 Brownian Motion

Definition 6. A d-dimensional Brownian Motion (BM) is a random walk/stochastic process B_t with properties

1. $B_0 = \hat{0} \in \mathbb{R}^d$
2. B_t has independent increments, that is, for $0 = t_0 \leq t_1 \leq \cdots \leq t_n$,

$$B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, \cdots, B_{t_n} - B_{t_{n-1}}$$

   are independent random variables.
3. For $0 \leq s \leq t, B_t - B_s \sim N(0, (t-s)\mathbb{I}^d)$, where $\mathbb{I}^d$ is the d-dimensional identity matrix.
4. **The most important property is that it is** (almost surely) **continuous**[7].

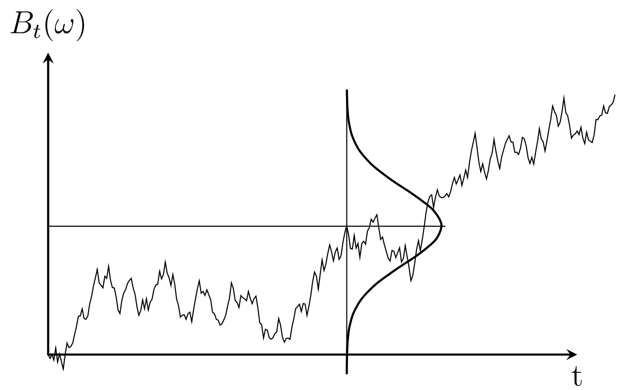

Figure 2: Intervals in one dimensional Brownian Motion are normally distributed with $\mu = 0$ and $\sigma^2 = (t-s)$.

[7] $\mathbb{P}(\{\omega : B_0(\omega) = 0 \text{ and } t \mapsto B_t(\omega) \text{ is continuous}\}) = 1$