Lecture 2: Preliminaries and Brownian Motion

January 20, 2009
Preliminaries for dealing with continuous random processes.
Brownian motions.
Our main reference for this lecture is Chapter 2 of the textbook.
Finite probability space

- In coin tossing example with finite tosses, $\Omega$ represents a finite sample space.
- Algebras $\mathcal{F}_i$ represents information that one has after $i$th game.
- Algebra $\mathcal{F} = \bigcup_{i}^{n} \mathcal{F}_i$ represents information after all $n$ games.
- Each subset $A \in \mathcal{F}$ represents a particular event whose probability is $P(A)$.
- The triple $(\Omega, \mathcal{F}, P)$ is called a probability space.
- In our example $\mathcal{F} = 2^\Omega$. 
If we toss infinitely many times, $\Omega$ becomes an infinite set.

We need $\mathcal{F} = \bigcup_i^{\infty} \mathcal{F}_i$ to be closed to the union of infinite sequence of subsets of $\mathcal{F}$.

Such $\mathcal{F}$ is called a $\sigma$-algebra.

Note that $\mathcal{F}$ is also closed for intersection of infinite sequence by de Morgan’s law.
Let $\Omega$ be a sample space and let $\mathcal{F}$ be a collection of subsets of $\Omega$. We say $\mathcal{F}$ is a $\sigma$-algebra provide
(i) for any $A \in \mathcal{F}$, $A^c = \Omega \setminus A \in \mathcal{F}$ and
(ii) for any sequence $A_i \in \mathcal{F}$, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

This is a mathematical model for information.
Let $\Omega$ be a sample space and let $\mathcal{F}$ be a $\sigma$-algebra of subsets (events) of $\Omega$.

We call $(\Omega, \mathcal{F})$ a measure space.

A function $P : \mathcal{F} \rightarrow [0, 1]$ is called a probability measure if $P(\emptyset) = 0$, $P(\Omega) = 1$ and, for any sequence of disjoint sets $A_i \in \mathcal{F}$,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

The triple $(\Omega, \mathcal{F}, P)$ is called a probability space.
Almost surely

If event $A \in \mathcal{F}$ satisfies $P(A) = 1$ we say $A$ happens almost surely or in brief a.s.
Let $\Omega$ be a sample space and let $\mathcal{U}$ be a collection of subsets of $\Omega$.

We use $\sigma(\mathcal{U})$ to denote the $\sigma$-algebra generated by $\mathcal{U}$ – the smallest $\sigma$-algebra containing $\mathcal{U}$.

This $\sigma$-algebra can be “constructed” by

$$\sigma(\mathcal{U}) = \bigcap_{\mathcal{A}} \{ \mathcal{U} \subset \mathcal{A} : \mathcal{A} \text{ is a } \sigma - \text{algebra} \}.$$  

(Exercise 2.3)
Examples

- Let $\Omega = \{1, 2, 3\}$.
- Then $\mathcal{F} = \{\emptyset, \Omega, \{1\}, \{2, 3\}\}$ is a $\sigma$-algebra,
- and $\mathcal{G} = \{\emptyset, \Omega, \{2\}\}$ is not.
- In fact, $\sigma(\mathcal{G}) = \{\emptyset, \Omega, \{2\}, \{1, 3\}\}$.
- What is $\sigma(\{\{2, 3\}\})$?
Borel algebra

Let $\mathcal{U}$ be all the open sets in $\mathbb{R}^n$. $\mathcal{B} := \sigma(\mathcal{U})$ is called the Borel $\sigma$-algebra.
Random variable

Let $(\Omega, \mathcal{F}, P)$ be a probability space. A mapping $X : \Omega \rightarrow R^n$ measurable with respect to $\mathcal{F}$, i.e., for any open set $A$ in $R^n$, $X^{-1}(A) \in \mathcal{F}$, is called a random variable.

For random variable that takes only countably number of values $a_i, i = 1, 2, \ldots$ one needs only to check $X^{-1}(a_i) \in \mathcal{F}$ (Exercise 2.1(a)).
Examples

Let $\Omega = \{1, 2, 3\}$ and $\mathcal{F} = \{\emptyset, \Omega, \{1\}, \{2, 3\}\}.$

Define $Y(1) = 1,$ $Y(2) = 0$ and $Y(3) = -1$ then $Y$ is not a random variable.

Define $Z(1) = 1$ and $Z(2) = Z(3) = 0$ then $Z$ is a random variable.
Let $X : \Omega \to \mathbb{R}^n$ be a random variable and let $\mathcal{B}$ be the Borel $\sigma$-algebra. Then

$$\sigma(X) = \{X^{-1}(F) : F \in \mathcal{B}\}$$

is a $\sigma$-algebra, called the $\sigma$-algebra generated by $X$. 
The $\sigma$-algebra generated by $X$ describes information contained in $X$. The Doob-Dynkin Lemma is about information relationship between two random variables.

**Doob-Dynkin Lemma**

Let $X, Y : \Omega \to \mathbb{R}^n$ be two functions. Then $Y$ is $\sigma(X)$ measurable if and only if there exists a Borel measurable function $g : \mathbb{R}^n \to \mathbb{R}^n$ (for any $A \in \mathcal{B}, g^{-1}(A) \in \mathcal{B}$) such that $Y = g(X)$. 

Lecture 2: Preliminaries and Brownian Motion

FE 5204 Stochastic Differential Equations
Simple random variables

Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space. A random variable \(X\) on \(\Omega\) is simple if it has only finite number of outcomes. Alternatively,

\[
X(\omega) = \sum_{i=1}^{k} x_i \chi_{F_i}(\omega),
\]

where \(F_i\) are disjoint sets in \(\mathcal{F}\).
Probability integration of simple random variables

Let $X$ be a simple random variable on $\Omega$ defined by

$$X(\omega) = \sum_{i=1}^{k} x_i \chi_{F_i}(\omega),$$

where $F_i$ are disjoint sets in $\mathcal{F}$. Then

$$\int_{\Omega} X(\omega) dP(\omega) := \sum_{i=1}^{k} x_i P(F_i).$$
Let $X$ be a random variable on $\Omega$. Suppose that there exists a sequence of simple random variables $X_j$ such that $X_j \to X$ with probability 1 on $\Omega$ and $\int_{\Omega} X_j(\omega)dP(\omega)$ converges. Then we defined

$$X(\omega) = \lim_{j \to \infty} \int_{\Omega} X_j(\omega)dP(\omega).$$

(This gives a way of calculation. Use it for Exercise 2.1(b),(c),(d).)
Technicalities

There are usually many sequence of simple random variables $X_j$ such that $X_j \rightarrow X$ with probability 1. In order to justify the above definition one has to show if $\int_{\Omega} X_j(\omega)dP(\omega)$ converges for one of such sequences then all the other sequences leads to integration that converges to the same limit.
Expectation

When \( \int_\Omega |X(\omega)|dP(\omega) < \infty \) the number

\[
E[X] = \int_\Omega X(\omega)dP(\omega)
\]

is called the expectation of \( X \).
The expectation can also be calculated by

$$E[X] = \int_{\mathbb{R}^N} xd\mu_X(x),$$

where $\mu_X(B) = P(X^{-1}(B))$ for any Borel set $B$.

This is a generalization of the concept of expectation when $\Omega$ is finite.

We can also use this formula to deal with Exercise 2.1 (b), (c), (d).
Variance

- To generalize variance we need space

\[ L^2(P) := \{ X : \Omega \rightarrow \mathbb{R}^N : \int_{\Omega} |X(\omega)|^2 dP(\omega) < \infty \}, \]

- with an inner product defined by

\[ \langle X, Y \rangle = \mathbb{E}[XY], \quad X, Y \in L^2(P). \]

- For \( X, Y \in L^2 \), their covariance is

\[ \text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]. \]

- For \( X \in L^2(P) \) its variance is \( \text{Var}(X) = \text{cov}(X, X) \).

- Again note here we need an additional technical assumption to restrict the range in which the concept is applicable.
Independence

- (two events) Events $A, B \in \mathcal{F}$ are independent if
  \[ P(A \cap B) = P(A)P(B). \]

- (multiple events) Events $A_i \in \mathcal{F}, i = 1, 2, \ldots$ are independent if
  \[ P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2})\ldots P(A_{i_k}). \]

- (multiple collection of events) Collection of events $A_i \in \mathcal{F}, i = 1, 2, \ldots$ are independent if for all choices $A_{i_j} \in A_{i_j}$,
  \[ P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2})\ldots P(A_{i_k}). \]
Independence of random variables

- Two random variables $X$ and $Y$ are independent if $\sigma(X)$ and $\sigma(Y)$ are independent.
- For $X, Y \in L^2(P)$, this implies that

$$E[XY] = E[X]E[Y]$$

or equivalently $\text{cov}(X, Y) = 0$.
- The condition can be weakened but we will not pursue the technical details.
- Random variables $X_i, i = 1, 2, \ldots$ are independent if any finite subfamily of $\sigma(X_i), i = 1, 2, \ldots$ are independent.
A continuous stochastic process

**Stochastic process**

Let \((\Omega, \mathcal{F}, P)\) be a probability space and let \([0, T]\) be an interval. \((X_t), t \in [0, T]\) is a stochastic process if for every \(t\), \(X_t\) is a random variable on \((\Omega, \mathcal{F}, P)\).
Path

For each fixed $\omega \in \Omega$,

$$t \to X_t(\omega)$$

is called a path of the stochastic process $(X_t)$. 
A stochastic process \((X_t)\) can also be viewed as a two variable function
\[(t, \omega) \rightarrow X_t(\omega) = X(t, \omega).\]

Usually we will need a certain properties of this function. We often assume every path is continuous which ensures the joint measurability of \(X(t, \omega)\).
Let \((\Omega, \mathcal{F}, P)\) be a probability space and let \([0, T]\) be an interval. \((\mathcal{F}_t), t \in [0, T]\) is a filtration if for every \(t\), \(\mathcal{F}_t \subset \mathcal{F}\) is a \(\sigma\)-algebra and, for any \(s < t\),

\[
\mathcal{F}_s \subset \mathcal{F}_t.
\]

Again, \(\mathcal{F}_t\) represents information available up to time \(t\).
Adapted stochastic process

Let \((F_t), t \in [0, T]\) is a filtration on probability space \((\Omega, F, P)\).
We say a stochastic process \((X_t)\) is \((F_t)\)-adapted provided that, for every \(t\), \(X_t\) is \(F_t\) measurable.

Note: there is no corresponding concept for predictable.
Let \((X_t), t \in [0, T]\) be a stochastic process on probability space \((\Omega, \mathcal{F}, P)\). Define \(\mathcal{F}_t = \sigma(\{X_s^{-1}(F) : s \in [0, t], F \in \mathbb{B}\})\), where \(\mathbb{B}\) is all the Boral sets in \(\mathbb{R}^n\). Then \((\mathcal{F}_t)\) is a filtration and \((X_t)\) is \(\mathcal{F}_t\)-adapted.
History

1. Named after Scottish botanist Robert Brown who in 1828 observed this motion from pollen suspended in liquid.
2. Louis Bachelier use it to model stock market in 1900, wrongly for allowing negative stock price.
3. Paul Samuelson gave the widely used geometric Brownian motion model for stock price movements in 1965, which do not allow any price jump.
4. “All models are wrong. Some are wronger than others.”
Definition of 1-dimensional Brownian motion (Wiener)

A stochastic process \( \{B_t : t \in [0, T)\} \) is called a Brownian motion starting from \( x \) if

1. \( B_0 = x \),
2. for \( 0 \leq t_1 < t_2 < \ldots < t_k \leq T \), the random variables
   \[
   B_{t_2} - B_{t_1}, B_{t_3} - B_{t_2}, \ldots, B_{t_k} - B_{t_{k-1}}
   \]
   are independent,
3. for \( 0 \leq s \leq t \leq T \), \( B_t - B_s \) has Gaussian distribution with mean 0 and variance \( t - s \),
4. for \( \omega \) in a set of probability one, the path \( B_t(\omega) \) is continuous.
Multi-dimensional Brownian motion

A vector stochastic process \( \{ B_t : t \in [0, T] \} \) in \( \mathbb{R}^n \) is called a Brownian motion starting from \( x = (x^1, x^2, \ldots, x^n) \) if 

\[
B_t = (B^1_t, B^2_t, \ldots, B^n_t)
\]

where \( B^i_t, i = 1, 2, \ldots, n \) are independent 1-dimensional Brownian motion starting from \( x^i \).

If \( x = 0 \), we say \( (B_t) \) is a standard Brownian motion. A Brownian motion starting from \( x \) can be viewed as a standard Brownian motion shifted by \( x \).
Existence

1. Existence of a Brownian motion needs to be justified.
2. By and large, there are two ways to do it:
3. by construction (pioneered by Wiener, see e.g. Steele’s book), or
4. by Kolmogorov’s extension theorem (if time permits, otherwise see e.g. [T]).
Brownian motions are not uniquely defined, but their effects are equivalent. We usually pick a ‘convenient’ version.
Let $\mathbf{v} = [v^1, \ldots, v^d]^\top$ be a $d$-dimensional random vector. We define the mean vector and the covariance matrix by $\mu = \mathbf{E}[\mathbf{v}] = [\mu_1, \ldots, \mu_d]^\top$, and

$$
\Sigma = \begin{bmatrix}
\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1d} \\
\sigma_{21} & \sigma_{22} & \cdots & \sigma_{2d} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{d1} & \sigma_{d2} & \cdots & \sigma_{dd}
\end{bmatrix},
$$

where $\sigma_{ij} = \text{cov}(v_i, v_j)$. 
Gaussian distribution

A $d$-dimensional random vector $\mathbf{v}$ is Gaussian with mean $\mu$ and covariance matrix $\Sigma$ if the density of $\mathbf{v}$ is given by

$$f(x) = (2\pi)^{-d/2}(\det \Sigma)^{-1/2}\exp\left(-\frac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu)\right).$$
Two simple and important facts of Gaussian distribution

- Gaussian random vectors are completely determined by its mean and covariance.
- The components of a Gaussian random vector are independent if and only if $\Sigma$ is diagonal.
Characteristic functions

Let \( v \) be a \( d \)-dimensional random vector. Then

\[
\phi(\theta) = \mathbb{E}[\exp(i\theta^\top v)]
\]

is the characteristic function of \( v \) which uniquely determines the distribution of a random vector. By direct computation we have

**Characteristic function of a Gaussian vector**

Let \( v \) be a \( d \)-dimensional Gaussian random vector \( v \) with mean \( \mu \) and covariance matrix \( \Sigma \). Then

\[
\phi(\theta) = \mathbb{E}[\exp(i\theta^\top v)] = \exp \left( i\theta^\top \mu - \frac{1}{2} \theta^\top \Sigma \theta \right).
\]
Gaussian characterization

Let $\nu$ be a $d$-dimensional random vector. Then $\nu$ is Gaussian if and only if for any $\theta \in R^d$, $\theta^T \nu$ is a univariate Gaussian random variable.
Let $X_t$ be a stochastic process. If, for any $0 \leq t_1 < t_2 < \ldots < t_k$, $(X_{t_1}, X_{t_2}, \ldots, X_{t_k})$ is a $k$-dimensional Gaussian random vector then we say $(X_t)$ is a Gaussian process.

A Gaussian process is completely determined by the mean and covariance functions $\mu(t) = \mathbf{E}[X_t]$ and $f(s, t) = \text{cov}(X_s, X_t)$. 
A standard 1-dimensional Brownian motion \((B_t)\) is a well behaved Gaussian process. For \(s < t\), its covariance is
\[
\text{cov}(B_s, B_t) = \mathbb{E}[B_t - B_s + B_s, B_s] = \mathbb{E}[B_s^2] = s.
\]
In general,
\[
\text{cov}(B_s, B_t) = s \wedge t.
\]
The converse is also true:

**Covariance of a Brownian motion**

Let \((X_t)\) be a Gaussian process with continuous paths and \(E[X_t] = 0\) for all \(t \in [0, T]\). Then \((X_t)\) is a standard Brownian motion if and only if \(\text{cov}(X_s, X_t) = s \wedge t\).

**Proof.** We need to show that \(X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, \ldots, X_{t_k} - X_{t_{k-1}}\) are independent. This can be done by expanding, for \(i < j\),

\[
E[(X_{t_i} - X_{t_{i-1}})(X_{t_j} - X_{t_{j-1}})]
\]

to find

\[
E[X_{t_i}X_{t_j}] - E[X_{t_i}X_{t_{j-1}}] - E[X_{t_{i-1}}X_{t_j}] + E[X_{t_{i-1}}X_{t_{j-1}}] = t_i - t_i - t_{i-1} + t_{i-1} = 0.
\]
For a standard 1-dimensional Brownian motion \((B_t)\), we can also use its density function to directly calculate higher moments

\[
\mathbb{E}[(B_t - B_s)^3] = 0,
\]

and

\[
\mathbb{E}[(B_t - B_s)^4] = 3(t - s)^2.
\]

Alternatively, one can also use the method in Exercise 2.8.
Let \((B_t)\) be an \(n\)-dimensional standard Brownian motion. Then we have

**Covariance of a Brownian motion**
\[
\text{cov}(B_s, B_t) = n(s \wedge t). \quad \text{In particular,} \quad \mathbf{E}[|B_t - B_s|^2] = n(t - s).
\]

and

**Higher moments**
\[
\mathbf{E}[|B_t - B_s|^4] = n(n + 2)(t - s)^2.
\]
Finite dimensional distribution

For a stochastic process \((X_t)\) and \(t_i, i = 1, 2, \ldots, k,\)

\[
\mu_{t_1, t_2, \ldots, t_k}(F_1 \times F_2 \times \ldots \times F_k) := P(X_{t_1} \in F_1, \ldots, X_{t_k} \in F_k)
\]

defines a measure on \(\mathbb{R}^{nk}\) called a finite dimensional distribution of \((X_t)\). Here \(F_1, \ldots, F_k\) are Borel sets in \(\mathbb{R}^n\).
Kolmogorov’s extension theorem: conditions

For all \( t_i, i = 1, 2, \ldots, k \), let \( \nu_{t_1,t_2,\ldots,t_k} \) be probability measure on \( \mathbb{R}^{nk} \) such that

\[
\nu_{t_{\sigma(1)},t_{\sigma(2)},\ldots,t_{\sigma(k)}}(F_1 \times F_2 \times \ldots \times F_k) = \nu_{t_1,t_2,\ldots,t_k}(F_{\sigma^{-1}(1)}, F_{\sigma^{-1}(2)}, \ldots, F_{\sigma^{-1}(k)})
\]

for all permutations \( \sigma \) on \( \{1, 2, \ldots, k\} \) and

\[
\nu_{t_1,t_2,\ldots,t_k}(F_1 \times F_2 \times \ldots \times F_k) = \nu_{t_1,t_2,\ldots,t_k,t_{k+1,\ldots,t_{k+m}}}(F_1 \times F_2 \times \ldots \times F_k \times F_{k+1} \times \ldots \times F_{k+m}).
\]
Kolmogorov’s extension theorem: conclusions

Then there exists a probability space \((\Omega, \mathcal{F}, P)\) and a \(\mathbb{R}^n\) stochastic process \((X_t)\) on \((\Omega, \mathcal{F}, P)\) such that

\[
\nu_{t_1, t_2, \ldots, t_k}(F_1 \times F_2 \times \ldots \times F_k) := P(X_{t_1} \in F_1, \ldots, X_{t_k} \in F_k),
\]

for all \(t_i\) and all Borel sets \(F_i\).
For \( x \in \mathbb{R}^n \), define

\[
p(t, x, y) = (2\pi t)^{-n/2} \exp\left(-\frac{|x - y|^2}{2t}\right), \quad y \in \mathbb{R}^n, \quad t > 0,
\]

and \( p(0, x, y) = \delta_x(y) \).

For \( t_i, i = 1, 2, \ldots, k \) increasing define,

\[
\nu_{t_1, t_2, \ldots, t_k} (F_1 \times F_2 \times \ldots \times F_k) = \\
\int_{F_1 \times F_2 \times \ldots \times F_k} p(t_1, x, x_1)p(t_2-t_1, x_1, x_2) \ldots p(t_k-t_{k-1}, x_{k-1}, x_k) \, dx_1 \ldots \, dx_k.
\]

Extend this definition to arbitrary \( t_i, i = 1, 2, \ldots, k \) using the permutation rule in Kolmogorov’s theorem.

Rule (2) automatically holds due to \( \int_{\mathbb{R}^n} p(t, x, y) \, dy = 1 \).
Existence of a Brownian motion

Let \( x \in \mathbb{R}^n \). An \( n \)-dimensional Brownian motion starts from \( x \) exists.

**Proof.** Using the Kolmogorov’s extension theorem to the finite dimensional distribution for a Brownian motion defined above.
Homework is an important part of learning SDE. Homework problems are given as exercises following each lecture and those marked with * are optional. The homework of this lecture is due on Jan 27 at the beginning of the lecture. Discussions with me or classmates are encouraged but the final work should be independently completed. I expect that you submit clear and neatly written work with careful justifications for your conclusions.

**Exercises**

[T] 2.1, 2.3, 2.8(a)(b), 2.12, 2.16, 2.18, 2.20.