Lecture 14

I now define stochastic independence of events and random variables. The intuition is that two events are independent if the occurrence of one has no bearing on the likelihood of the other. For instance, if you flip a coin twice in succession, the probability of heads on the second toss does not depend on whether heads occurred on the first toss. The probability should be one half. If the coin is truly fair and the successive tosses are not linked in any way, then the probability of heads on the 101st toss should be one half, even if all the first 100 tosses came up heads.

Let \((S, \mathcal{F}, P)\) be a probability space. Two events \(A\) and \(B\) in \(\mathcal{F}\) are said to be stochastically independent if \(P(A \cap B) = P(A) P(B)\). For instance, suppose that \(A\) is the occurrence of heads on the first of two tosses of a fair coin and let \(B\) be the occurrence of heads on the second toss. If the tosses are truly independent, then the probability of heads on both tosses, the event \(A \cap B\), is \(1/4\), which is the product of the probabilities of \(1/2\) of having heads on each toss.

If \(A\) and \(B\) are events in \(\mathcal{F}\) such that \(P(B) > 0\), the probability of \(A\) conditional on \(B\) is defined as \(P(A|B) = \frac{P(A \cap B)}{P(B)}\). If \(A\) and \(B\) are stochastically independent, then \(P(A|B) = P(A)\), so that knowledge that \(B\) occurs does not affect the probability of \(A\).

If \(x : S \rightarrow \mathbb{R}\) and \(y : S \rightarrow \mathbb{R}\) are random variables, then \(x\) and \(y\) are stochastically independent if for any Borel subsets \(A\) and \(B\) of \(\mathbb{R}\), the event \(x^{-1}(A) = \{s \in S \mid x(s) \in A\}\) is stochastically independent of the event \(y^{-1}(B) = \{s \in S \mid y(s) \in B\}\). It follows that if \(x\) and \(y\) are independent, then \(E(xy) = (Ex)(Ey)\), where \(xy(s) = x(s)y(s)\), for all \(s \in S\), and provided \(x, y,\) and \(xy\) are integrable. This assertion is easy to verify when \(x\) and \(y\) are indicator functions of measurable sets, for let \(x = \mathbb{1}_A\) and \(y = \mathbb{1}_B\). If \(\mathbb{1}_A\) and \(\mathbb{1}_B\) are independent, then \(A\) and \(B\) must be independent, since \(A = \mathbb{1}_A^{-1}(1)\) and \(B = \mathbb{1}_B^{-1}(1)\) and so

\[E(xy) = E(\mathbb{1}_A \mathbb{1}_B) = E(\mathbb{1}_A \mathbb{1}_B) = P(A \cap B) = P(A) P(B) = E(\mathbb{1}_A) E(\mathbb{1}_B) = E(x) E(y).\]

We can use this equation to verify that \(E(xy) = E(x)E(y)\) for independent simple functions \(x\) and \(y\). In order to see how, assume that \(x\) and \(y\) are bounded and let us form sequences of simple functions \(x_M\) and \(y_M\) that approximate \(x\) and \(y\) asymptotically, as in example 13.2. That is, let \(b\) be a positive number such that \(|x(s)| \leq b\) and \(|y(s)| \leq b\), for all \(s \in S\). For each positive integer \(M\), let \(A_{M,m} = \{s \in S \mid -b + \frac{m}{M}b \leq x(s) < -b + \frac{m+1}{M}b\}\), for \(m = 0, 1, \ldots, 2M - 1\), and let \(B_{M,m} = \{s \in S \mid -b + \frac{m}{M}b \leq y(s) < -b + \frac{m+1}{M}b\}\), for \(m = 0, 1, \ldots, 2M - 1\), and let
\[ y = \sum_{m=0}^{2M-1} (-b + \frac{mb}{M}) \chi_{A_{M,m}}. \]  
Because \( x \) and \( y \) are independent, each of the sets \( A_{M,m} \) is independent of the sets \( B_{M,n} \), for \( n = 0, 1, \ldots, 2M-1 \). The functions \( x, y \) converge to \( xy \) as \( M \) goes to infinity, since, for all \( s \in S \)

\[
|x_M(s) y_M(s) - x(s) y(s)| \leq |x_M(s) y_M(s) - x(s) y(s)| + |x_M(s) y(s) - x(s) y(s)| \\ 
\leq |x_M(s)||y(s)| + |x_M(s) - x(s)||y(s)| \\ 
\leq 2b^2 \left( \frac{2}{M} \right) = \frac{4b^2}{M},
\]

which goes to 0 as \( M \) goes to infinity. Also

\[
\int_S |x_M(s) y_M(s) - x(s) y(s)| P(ds) \\
\leq \int_S |x_M(s) y_M(s) - x(s) y(s)| P(ds) + \int_S |x_M(s) y(s) - x(s) y(s)| P(ds) \\
\leq \int_S |x_M(s)||y(s)| P(ds) + \int_S |x(s) - x(s)||y(s)| P(ds) \\
\leq \int_S \max \left( \frac{2b}{N}, \frac{2b}{M} \right) P(ds) + \int_S \max \left( \frac{2b}{N}, \frac{2b}{M} \right) P(ds) \\
= 2b \max \left( \frac{2b}{N}, \frac{2b}{M} \right)
\]

which goes to 0 as \( N \) and \( M \) go to infinity. Therefore by the definition of the integral,

\[
\int_S x(s) y(s) P(ds) = \lim_{M \to \infty} \int_S x_M(s) y_M(s) P(ds).
\]

Similarly \( \int_S x(s) P(ds) = \lim_{M \to \infty} \int_S x_M(s) P(ds) \) and \( \int_S y(s) P(ds) = \lim_{M \to \infty} \int_S y_M(s) P(ds) \). So

\[
\int_S x(s) y(s) P(ds) = \lim_{M \to \infty} \int_S x_M(s) y_M(s) P(ds) \\
= \lim_{M \to \infty} \int_S \left[ \sum_{m=0}^{2M-1} \left( -b + \frac{mb}{M} \right) \chi_{A_{M,m}}(s) \right] \left[ \sum_{n=0}^{2M-1} \left( -b + \frac{nb}{M} \right) \chi_B(s) \right] P(ds) \\
= \lim_{M \to \infty} \int_S \sum_{m=0}^{2M-1} \sum_{n=0}^{2M-1} \left( -b + \frac{mb}{M} \right) \left( -b + \frac{nb}{M} \right) \chi_{A_{M,m}}(s) \chi_B(s) P(ds)
\]
\[
\lim_{M \to \infty} \sum_{m=0}^{2M-1} \sum_{n=0}^{2M-1} \left( -b + \frac{mb}{M} \right) \left( -b + \frac{nb}{M} \right) \int_{S} \mathcal{X}_{A_{M,m}^{(d)} B_{M,n}^{(d)}} (s) \ P(\ ds) \\
= \lim_{M \to \infty} \sum_{m=0}^{2M-1} \sum_{n=0}^{2M-1} \left( -b + \frac{mb}{M} \right) \left( -b + \frac{nb}{M} \right) \ P(\ A_{M,m} \cap B_{M,n}) \\
= \lim_{M \to \infty} \sum_{m=0}^{2M-1} \sum_{n=0}^{2M-1} \left( -b + \frac{mb}{M} \right) \left( -b + \frac{nb}{M} \right) \ P(\ A_{M,m}) P(\ B_{M,n}) \\
= \lim_{M \to \infty} \left( -b + \frac{mb}{M} \right) \left( -b + \frac{nb}{M} \right) \left( \sum_{n=0}^{2M-1} \ P(\ B_{M,n}) \right) \\
= \lim_{M \to \infty} \sum_{m=0}^{2M-1} \left( -b + \frac{mb}{M} \right) \int_{S} \mathcal{X}_{A_{M,m}^{(d)}} (s) \ P(\ ds) (\sum_{n=0}^{2M-1} \left( -b + \frac{nb}{M} \right) \int_{S} \mathcal{X}_{B_{M,n}^{(d)}} (s) \ P(\ ds) \\
= \lim_{M \to \infty} \int_{S} x(s) \ P(\ ds) \left( \sum_{m=0}^{M} \left( -b + \frac{mb}{M} \right) \int_{S} \mathcal{X}_{A_{M,m}^{(d)}} (s) \ P(\ ds) \right) \\
\int_{S} y(s) \ P(\ ds) \\
= \int_{S} x(s) \ P(\ ds) \int_{S} y(s) \ P(\ ds).
\]

In conclusion,
\[
\int_{S} x(s) \ P(\ ds) \int_{S} y(s) \ P(\ ds) = \int_{S} x(s) \ P(\ ds) \int_{S} y(s) \ P(\ ds),
\]

if \( x \) and \( y \) are bounded independent random variables. This argument must be modified somewhat if \( x \) and \( y \) are unbounded but integrable.

If \( x \) and \( y \) are independent random variables, then \( \text{Var}(x + y) = \text{Var}(x) + \text{Var}(y) \), for

\[
\text{Var}(x + y) = E(x + y)^2 - (Ex + Ey)^2 \\
= E(x^2 + 2xy + y^2) - (Ex)^2 - 2(Ex)(Ey) - (Ey)^2 \\
= E(x^2) + 2(Ex)(Ey) + E(y^2) - (Ex)^2 - 2(Ex)(Ey) - (Ey)^2 \\
= E(x^2) - (Ex)^2 + E(y^2) - (Ey)^2 \\
= \text{Var}(x) + \text{Var}(y).
\]
Suppose that $\mathcal{F}$ is a $\sigma$-field over $S$ that is contained in $\mathcal{F}_0$. That is, any $A \in \mathcal{F}$ belongs to $\mathcal{F}_0$. Such a $\sigma$-field is said to be a sub $\sigma$-field of $\mathcal{F}_0$. A random variable $x : S \to \mathbb{R}^n$ that is measurable with respect to $\mathcal{F}_0$ is said to be independent of $\mathcal{F}$, if every $A \in \mathcal{F}$ is independent of every $C \in \mathcal{F}_0$ of the form $C = x^{-1}(B)$, where $B$ is a Borel subset of $\mathbb{R}^n$. That is, $\mathcal{F}_0$ is independent of $x$, for every $A \in \mathcal{F}_0$. If $\mathcal{F}$ and $x$ are independent, then

$$\int x(s) P(ds) = \int \mathcal{F}_0(s) \cdot x(s) P(ds) = \int \mathcal{F}_0(s) P(ds) \int x(s) P(ds),$$

for all $A \in \mathcal{F}$.

The definition of independence may be extended to a sequence of events or random variables. Events in a sequence $E_1, E_2, \ldots$ are said to be stochastically mutually independent if

$$P(E_{n_1} \cap E_{n_2} \cap \ldots \cap E_{n_K}) = P(E_{n_1}) P(E_{n_2}) \ldots P(E_{n_K}),$$

for any set of distinct positive integers $n_1, n_2, \ldots, n_K$ such that $2 \leq K < \infty$. Random variables $x_{n} : S \to \mathbb{R}$, where $n = 1, 2, \ldots$, are stochastically mutually independent if for any Borel subsets $A_1, A_2, \ldots$ of $\mathbb{R}$, the events $E_n$ are stochastically mutually independent, where

$$E_n = \{ s \in S | x_n(s) \in A_n \},$$

for all $n$.

You might imagine that the random variables $x_1, x_2, \ldots, x_N$ are mutually independent if they are pairwise independent in that $x_k$ and $x_n$ are independent for any $k$ and $n$ that are not equal. Examples show, however, that a set of random variables can be pairwise independent and yet not mutually independent.

It is easy to see that if $x_1, x_2, \ldots, x_N$ are mutually independent, then

$$\text{Var}(x_1 + x_2 + \ldots + x_N) = \text{Var}(x_1) + \text{Var}(x_2) + \ldots + \text{Var}(x_N).$$

The random variables $x_n : S \to \mathbb{R}$, for $n = 1, 2, \ldots$ are said to be identically distributed if they all have the same distribution (or cumulative distribution function) on $\mathbb{R}$. The next theorem is central to probability theory.

**Theorem 14.1**: (Strong Law of Large Numbers) Let $(S, \mathcal{F}, P)$ be a probability space. If $x_n : S \to \mathbb{R}$, $n = 1, 2, \ldots$ is a sequence of random variables that are stochastically mutually
independent and identically distributed and such that \( \mathbb{E}|x_1| < \infty \), then with probability 1

\[
\lim_{N \to \infty} \frac{x_1(s) + x_2(s) + \ldots + x_N(s)}{N} = \mathbb{E}_1.
\]

The statement that \( \lim_{N \to \infty} \frac{x_1(s) + x_2(s) + \ldots + x_N(s)}{N} = \mathbb{E}_1 \) with probability 1 means that \( \left\{ s \in S \mid \lim_{N \to \infty} \frac{x_1(s) + x_2(s) + \ldots + x_N(s)}{N} = \mathbb{E}_1 \right\} \) is measurable, i.e., belongs to \( \mathcal{E} \), and has probability 1. The law of large numbers makes a connection between averages and probabilities. For instance, let \( S \) be set of possible sequences of heads and tails on successive tosses of a fair coin. That is,

\[ S = \{ s = (s_1, s_2, \ldots) \mid s_n = \text{H or T, for all } n \}. \]

Let

\[ x_n(s) = \begin{cases} 1, & \text{if } s_n = \text{H}, \\ 0, & \text{if } s_n = \text{T}. \end{cases} \]

That is, \( x_n(s) = 1 \), if the \( n \)th toss comes up heads, and \( x_n(s) = 0 \), if the \( n \)th toss comes up tails.

By the law of large numbers,

\[
\lim_{N \to \infty} \frac{x_1(s) + x_2(s) + \ldots + x_N(s)}{N} = \frac{1}{2}
\]

with probability 1 as the number of tosses goes to infinity. Hence the average number of times a head appears converges to \( 1/2 \), which is the probability of heads on one toss.

Related to the above is the central limit theorem, which gives some control on how far the averages in the law of large numbers differ from the asymptotic mean, \( \mathbb{E}_1 \). In order to state this theorem, I need to define weak convergence of cumulative distribution functions. Let \( F \) and \( F_1, F_2, \ldots \) be cumulative distribution functions on \( \mathbb{R} \). The sequence \( F_1, F_2, \ldots \) is said to converge weakly to \( F \) if

\[
\lim_{n \to \infty} F_n(r) = F(r),
\]

for every point \( r \) at which \( F \) is continuous.
Theorem 14.2: (Central Limit Theorem) Let \( x \colon S \to R \), where \( n = 1, 2, \ldots \), be a sequence of random variables that are stochastically mutually independent and identically distributed and such that \( \mathbb{E} x = \mu \) and \( \text{Var}(x) = \sigma^2 \) exist and hence are finite. For \( N = 1, 2, \ldots \), let

\[
X_N(s) = \frac{(x_1(s) - \mu) + (x_2(s) - \mu) + \ldots + (x_N(s) - \mu)}{\sigma \sqrt{N}}
\]

and let \( F_N \) be the cumulative distribution function of the random variable \( X_N \). Then \( F_N \) converges weakly to \( F \), where \( F \) is the cumulative distribution function of the normal distribution \( N(0, 1) \) with mean \( 0 \) and variance \( 1 \).

Notice that the assumption that \( \text{Var}(x) < \infty \) implies that \( \mathbb{E} x_i^2 < \infty \), so that the strong law of large numbers applies under conditions in which the central limit theorem applies.

**Stochastic Processes**

**Definition:** A stochastic process is a family of random variables \( \{ x_t \} _{t \in T} \), where \( T \) is an index set and the random variables \( x_t \) are all defined on a common state space \( S \). More specifically, there is a probability space \( (S, \mathcal{S}, P) \) such that \( x_t : S \to R \) is measurable with respect to \( \mathcal{S} \), for all \( t \in T \).

For our purposes, \( T \) is either the set of non-negative integers or the set of non-negative numbers. I will take the set of states \( S \) to be the set of all functions \( s : T \to R \). Then the random variable \( x_t : S \to R \) is the function such that \( x_t(s) = s(t) \), for every function \( s : T \to R \) in \( S \). For each \( t \), let \( \mathcal{O}_t = \{ x^{-1}(B) \mid B \text{ is a Borel subset of } R \} \). The set of measurable subsets, \( \mathcal{S}_t \), of \( S \) is usually taken to be the smallest \( \sigma \)-field that contains \( \mathcal{O}_t \), for all \( t \). Let \( \mathcal{S}_r \) be the smallest \( \sigma \)-field that contains \( \mathcal{O}_r \), for all \( r \geq t \). Then \( \mathcal{S}_r \) is contained in \( \mathcal{S}_t \) and represents the information available from the process \( x \) at time \( t \) from the process \( \{ x_t \}_{t \in T} \).

If \( x : S \to R \) is a random variable and \( \mathcal{O} \) is a \( \sigma \)-field contained in \( \mathcal{S}_r \), then \( x \) is said to be measurable with respect to \( \mathcal{O} \) if \( x^{-1}(B) \in \mathcal{O} \), for every Borel subset \( B \) of \( R \). By the definition of \( \mathcal{S}_r \), the random variable \( x : S \to R \) is measurable with respect to \( \mathcal{S}_r \), for all \( t \geq r \).

For each finite set of distinct indices \( \{ t_1, t_2, \ldots, t_n \} \), where \( t_n \in T \), for all \( n \), the function \( \{ x_{t_1}(s), x_{t_2}(s), \ldots, x_{t_n}(s) \} \) maps the set of states, \( S \), to \( R^n \). The usual way to define a probability distribution for the stochastic process \( \{ x_t \}_{t \in T} \) is to define a probability
distribution for \((x_{i_1}, x_{i_2}, \ldots, x_{i_N})\) on \(\mathbb{R}^N\), for all finite subsets \((t_{i_1}, t_{i_2}, \ldots, t_{i_N})\) of indices. This distribution in turn induces a probability distribution on a sub \(\sigma\)-field of \(S\). If the assignment of distributions is done in a coherent way for all finite subsets of indices, then by a theorem of Kolmogorov there exists a probability distribution on \(S\) that induces the distributions defined on sub \(\sigma\)-fields of \(S\) for all finite subsets of distinct random variables \((x_{i_1}, x_{i_2}, \ldots, x_{i_N})\). The existence of the probability on \(S\) is a complicated topic that I do not pursue.

**Definition:** A Wiener process is a stochastic process \((W_t)_{t \geq 0}\) with state space \(S\) and such that

1) for each finite subset of non-negative numbers \((t_{i_1}, t_{i_2}, \ldots, t_{i_N})\) such that \(t_{i_1} < t_{i_2} < \ldots < t_{i_N}\), the random variable \((W_{t_{i_1}}, W_{t_{i_2}}, \ldots, W_{t_{i_N}}) : S \to \mathbb{R}^N\) has a multivariate normal distribution,

2) for all \(t\), \(W_t\) has the normal distribution \(N(0, t)\), (i.e., \(EW_t = 0\), \(\text{Var}(W_t) = t\), and \(W_t\) is normally distributed),

3) for each \(r\) and \(t\) such that \(0 < r < t\), \(W_t - W_r\) has the normal distribution \(N(0, t - r)\) and \(W_t - W_r\) is independent of \(S\) and hence of the random variable \(W_r\).

These three conditions provide enough information to define a probability distribution on \(S\), so that a Wiener process exists. I do not prove this assertion. As before \(S = \{s : [0, \infty) \to \mathbb{R}\}\), so that \(W_t(s) = s(t)\), for all \(s\) and \(t\).

A standard bit of notation is to use the symbol \(dW_t\) to denote the random variable \(W_{t+dt} - W_t\), where \(dt\) is an infinitesimal and positive interval of time. That is, \(dW_t\) is normally distributed with mean 0 and variance \(dt\). That is, its distribution is \(N(0, dt)\).

The Wiener process may be interpreted as the limit result of tiny random shocks up and down and spread uniformly over time. One way to grasp this intuition is to fix attention on a time \(t\), where \(t > 0\), and to imagine that the interval \([0, t]\) is divided into \(N\) intervals of length \(t/N\). Imagine also that the shocks occur at times \(nt/N\), for \(n = 1, \ldots, N\). Suppose that the \(n\)th shock is expressed by a random variable \(x^n\), which equals either \(\sqrt{t/N}\) or \(-\sqrt{t/N}\), each with probability 1/2. Assume that the random variables \(x^1, x^2, \ldots, x^N\) are mutually independent. Then \(E x^n = 0\) and \(\text{Var}(x^n) = t/N\). Because \(x^1, x^2, \ldots, x^N\) are mutually independent,

\[
\text{Var}(x^1 + x^2 + \ldots + x^N) = \sum_{n=1}^{N} \text{Var}(x^n) = \frac{Nt}{N} = t.
\]

7
Notice that

\[ \sum_{n=1}^{N} x_n^N = \sqrt{t} \sum_{n=1}^{\lfloor t/N \rfloor} x_n^N. \]

By the central limit theorem given above, the distribution of the random variables \( \sqrt{t} \sum_{n=1}^{\lfloor t/N \rfloor} x_n^N \) converges weakly to \( N(0, 1) \). Hence if we imagine that \( \sqrt{t} \sum_{n=1}^{\lfloor t/N \rfloor} x_n^N \) converges to some random variable \( x \) with distribution \( N(0, 1) \), then \( \sqrt{t} \sum_{n=1}^{\lfloor t/N \rfloor} x_n^N \) converges to \( \sqrt{t} x \), which has distribution \( N(0, t) \). That is, the distribution of \( \sum_{n=1}^{N} x_n^N \) converges weakly to the distribution of \( W_t \), the Wiener process at time \( t \).

This rather contrived example illustrates a larger point. Suppose that we build up a stochastic process as a limit

\[ x(t, s) = \lim_{N \to \infty} \sum_{n=1}^{N^{(1)}} \Delta x_n^{N}(n/N, s), \]

where \( N(t) \) is the largest positive integer such that \( N(t)/N \leq t \) and where the random variables \( \Delta x_n^{N}(n/N, s) \), for \( n = 1, ..., N(t) \), are stochastically mutually independent and are identically distributed and have mean 0 and variance \( t/N \). Then by the central limit theorem, \( x(t, s) \) is normally distributed with mean 0 and variance \( t \). It is not necessary for this result that the \( \Delta x_n^{N}(n/N, s) \) be normally distributed. Heuristically we may think of \( \Delta x_n^{N}(n/N, s) \) as \( dx(t, s) \) and write

\[ x(t, s) = \int_0^t dx(r, s) dr. \]

One of the inspirations of the Wiener process comes from the observation of tiny particles in air or water that jiggle about irregularly, a motion attributed to bombardment by molecules of the air or water. The observation is usually attributed to the findings in 1827 of a Scottish botanist, Robert Brown, though the same observation with the same explanation was
made in ancient times. The Wiener process represents the limit result of the time path of such jiggling, though in just one dimension.

A curious fact about Wiener processes is that their paths $W_t(s)$ are continuous functions of time with probability 1.

**Theorem 14.3:** If $W_t$ is a Wiener process, then $W_t(s)$ is a continuous function of $t$ with probability 1.

Another interesting fact about Wiener processes is that they have unbounded variation with probability 1.

**Definition:** The variation of a function $f: [0, T] \to \mathbb{R}$ is

$$
\sup \left\{ \sum_{n=1}^{N} \left| f(t_{n+1}) - f(t_n) \right| \mid 0 = t_1 < t_2 < \ldots < t_N = T \text{ and } N \text{ is a positive integer} \right\}.
$$

**Theorem 14.4:** If $W_t$ is a Wiener process, then over any interval $[0, T]$ with $T > 0$, $W_t(s)$ has infinite variation as a function of $t$ with probability 1.

What is termed *Brownian motion* is a slight generalization of the Wiener process and is defined as a stochastic process $X_t(s)$, for $t$ in $[0, \infty)$, where $X_t(s) = \mu t + \sigma W_t(s)$, $\mu$ is a number, and $\sigma$ is a positive number. It should be clear that $E X_t = \mu t$, $\text{var}(X_t) = \sigma^2 \text{var}(W_t) = \sigma^2 t$, $X_t$ is normally distributed, and if $T > t$, $X_t - X_s$ is independent of $X_t$ and of $S_t$ and is normally distributed with mean $\mu(T-t)$ and variance $\sigma^2(T-t)$.

More general stochastic processes may be defined from the Wiener process by using the stochastic or Itô integral, which I discuss next. Historically one of the motivations for studying integrals was to solve a differential equation, such as

$$
\frac{dx(t)}{dt} = f(x(t), t).
$$

For instance by the fundamental theorem of calculus, the solution of the differential equation

$$
\frac{dx(t)}{dt} = f(t)
$$

is simply $x(t) = a + \int_0^t f(s) \, ds$, where $a$ is a constant number. Imagine now trying to solve a differential equation where the motion is subject to random jiggles. We might write such an equation as
\[
\frac{dx(t, s)}{dt} = f(x(t, s), t) dt + g(x(t, s), t) dW_t(s),
\]

for each s, where \(dW_t(s)\) is understood to be a Wiener process with variance \(dt\), where \(dt\) is an infinitesimal interval of time and \(s\) represents the state. The notation \(dx(t, s)/dt\) is not appropriate, since the function \(x(t, s)\) is not likely to be differentiable with respect to \(t\). So a stochastic differential equation is written as

\[
dx(t, s) = f(x(t, s), t) dt + g(x(t, s), t) dW_t(s).
\]

We can make sense of this expression and hope to find a solution for the equation only if we can somehow integrate the term \(g(x(t, s), t) dW_t(s)\) with respect to time. This integral is called the stochastic or Itô integral. We can simplify the notation by replacing \(g(x(t, s), t)\) by \(g(t, s)\), so that the objective becomes to define

\[
f_T\int_0^t g(t, s) dW_t(s),
\]

where \(W_t(s)\) is a Wiener process.

If we base our intuition about integrals on the definition of the Riemann integral given at the end of lecture 5, the stochastic integral should look something like

\[
\lim_{N \to \infty} \sum_{n=0}^{N-1} g\left(\frac{nT}{N}, s\right) \left[ W_{\frac{n+1}{N}T} - W_{\frac{n}{N}T} \right]. \tag{14.1}
\]

Suppose that \(g(t, s)\) is a continuous function of \(t\) with probability 1. The temptation is then to let the limit in equation 14.1 define \(\int_0^t g(t, s) dW_t(s)\), for every \(s\) such that \(g(t, s)\) and \(W_t(s)\) are continuous with respect to \(t\), since the functions \(W_t(s)\) are continuous with probability 1. The integral would then be what is called the Riemann-Stieltjes integral of \(g(t, s)\) with respect to \(W_t(s)\), for all \(s\) such that \(g(t, s)\) and \(W_t(s)\) were continuous functions of \(t\). The problem with this approach is that the limit in equation 14.1 does not exist with probability 1, because the function \(W_t(s)\) has unbounded variation as a function of \(t\) with probability 1. (The limit would exist if the variation of \(W_t(s)\) were bounded.) Therefore the analogue of theorem 5.18 on the existence of the Riemann integral does not apply, and an entirely different approach is required.

Let
We want to show that as \( N \) goes to infinity the functions \( V_N \) converge in some sense, even if they do not converge for most \( s \). We will do so by considering them to be members of a space of random variables and focusing on the overall behavior of the functions rather than their behavior for particular values of \( s \).

In order to explain the approach, I introduce the space \( L_2(S, \mathcal{S}, P) \), where

\[
L_2(S, \mathcal{S}, P) = \left \{ x : S \rightarrow \mathbb{R} \mid x \text{ is measurable with respect to } \mathcal{S} \text{ and } \int_S x^2(s) \, dP < \infty \right \}.
\]

That is, \( x \in L_2(S, \mathcal{S}, P) \) if its square is integrable or if its variance is finite. If \( x \in L_2(S, \mathcal{S}, P) \), the norm \( ||x||_2 \) is defined by the equation

\[
||x||_2 = \left( \int_S x^2(s) \, dP \right)^{1/2}.
\]

The function \( ||x||_2 \) is a norm on \( L_2(S, \mathcal{S}, P) \) in the sense that \( ||ax||_2 = |a| \, ||x||_2 \), for every number \( a \), and \( ||x + y||_2 \leq ||x||_2 + ||y||_2 \), for all \( x \) and \( y \) in \( L_2(S, \mathcal{S}, P) \). A sequence \( x_n \) in \( L_2(S, \mathcal{S}, P) \) converges to \( x \) or \( \lim_{n \rightarrow \infty} x_n = x \), if \( \lim_{n \rightarrow \infty} ||x - x_n||_2 = 0 \), and the sequence is Cauchy if \( \lim_{n \rightarrow \infty} \sup_{m \geq n} ||x_m - x_n||_2 = 0 \). The property of \( L_2(S, \mathcal{S}, P) \) that we need is that it is complete with respect to the norm \( ||x||_2 \). This means that for any Cauchy sequence \( x_n \) in \( L_2(S, \mathcal{S}, P) \), there exists an \( x \) in \( L_2(S, \mathcal{S}, P) \) such that \( \lim_{n \rightarrow \infty} x_n = x \). (Complete normed vector spaces, such as \( L_2(S, \mathcal{S}, P) \), are called Banach spaces.) Given the importance of the completeness of \( L_2(S, \mathcal{S}, P) \), I present this assertion as a theorem.

**Theorem 14.5:** Every Cauchy sequence in \( L_2(S, \mathcal{S}, P) \) converges to some member of \( L_2(S, \mathcal{S}, P) \).

I now return to the Ito or stochastic integral. Let \( g(t) : S \rightarrow \mathbb{R} \) be a stochastic process, where \( t \) varies over the non-negative numbers. Let \( S, \mathcal{S}, \mathbb{S}_1 \), and \( P \) be as defined earlier.

**Assumptions 14.6:**

1) \( g(t) : S \rightarrow \mathbb{R} \) is measurable with respect to \( \mathbb{S}_1 \), for all \( t \),

11
2) \( g(t) \) belongs to \( L_2(S, \mathcal{S}, P) \), for all \( t \), and

3) the function \( g : [0, \infty) \to L_2(S, \mathcal{S}, P) \) is continuous with respect to the norm \( ||.||_2 \).

The third assumption means that if \( t_n, n = 1, 2, \ldots, \) is a sequence in \([0, \infty)\) that converges to \( t \), then

\[
\lim_{n \to \infty} ||g(t_n) - g(t)||_2 = 0.
\]

This assumption implies that \( ||g(t)||_2 \) is also a continuous function of \( t \), for the triangle inequality for \( ||.||_2 \) implies that

\[
||g(t_n)||_2 - ||g(t)||_2 \leq ||g(t_n) - g(t)||_2,
\]

so that \( \lim_{n \to \infty} ||g(t_n)||_2 - ||g(t)||_2 = 0 \), if \( \lim_{n \to \infty} ||g(t_n) - g(t)||_2 = 0 \).

Notice that the first assumption implies that \( W - W \) is stochastically independent of \( g(\tau) \), if \( t > r \geq \tau \). Therefore

\[
E[g(r)(W_i - W_j)] = Eg(r)E(W_i - W_j) = Eg(r)(0) = 0.
\]

Hence if we take the expected value of the expression in equation 14.2 for \( V_N \), we find that

\[
E[V_N] = E\left[ \sum_{n=0}^{N-1} g\left( \frac{nT}{N} \right) (W_{\frac{n+1}{N}} - W_{\frac{n}{N}}) \right] = \sum_{n=0}^{N-1} E\left[ g\left( \frac{nT}{N} \right) \right] E\left[ (W_{\frac{n+1}{N}} - W_{\frac{n}{N}}) \right] = 0, \quad (14.3)
\]

where I have dropped the variable \( s \) from the functions \( g \) and \( W \). It is important to notice that equation 14.3 would not be true if we wrote the sum in equation 14.2 as

\[
\sum_{n=1}^{N} g\left( \frac{n}{N}, s \right) \left[ W_{\frac{n}{N}}(s) - W_{\frac{n-1}{N}}(s) \right],
\]

for \( W_{\frac{n}{N}}(s) - W_{\frac{n-1}{N}}(s) \) may not be independent of \( g\left( \frac{nT}{N}, s \right) \).

We want to show that the functions \( V_N \) converge in \( L_2(S, \mathcal{S}, P) \). In order to do so, we
must first verify that these functions belong to \( L^2(S,\mathbb{S},\mathbb{P}) \) and then check that the sequence \( V_n \) is Cauchy with respect to \( ||\cdot||_2 \). In order to show that each \( V_n \) belongs to \( L^2(S,\mathbb{S},\mathbb{P}) \), we must show that

\[
\left| \sum_{n=0}^{N-1} g\left(\frac{n}{N}\right) \left( \frac{n+1}{N} - \frac{n}{N} \right) \right|^2 = E\left[ \sum_{n=0}^{N-1} g\left(\frac{n}{N}\right) \left( \frac{n+1}{N} - \frac{n}{N} \right) \right]^2 < \infty.
\]

When we take the square on the right-hand side, we obtain \( N \) terms of the form

\[
E\left[ g\left(\frac{n}{N}\right) \left( \frac{n+1}{N} - \frac{n}{N} \right)^2 \right]
\]

and the other terms have the form

\[
E\left[ g\left(\frac{n}{N}\right) g\left(\frac{m}{N}\right) \left( \frac{n+1}{N} - \frac{n}{N} \right) \left( \frac{m+1}{N} - \frac{m}{N} \right) \right],
\]

where \( m > n \). Since

\[
W_{\frac{n+1}{N}} - W_{\frac{n}{N}}
\]

is independent of

\[
g\left(\frac{n}{N}\right) g\left(\frac{m}{N}\right) \left( \frac{n+1}{N} - \frac{n}{N} \right)
\]

we have that

\[
E\left[ g\left(\frac{n}{N}\right) g\left(\frac{m}{N}\right) \left( \frac{n+1}{N} - \frac{n}{N} \right) \left( \frac{m+1}{N} - \frac{m}{N} \right) \right] = E\left[ g\left(\frac{n}{N}\right) g\left(\frac{m}{N}\right) \left( \frac{n+1}{N} - \frac{n}{N} \right) \right] E\left[ \left( \frac{m+1}{N} - \frac{m}{N} \right) \right] = 0.
\]

Returning to the \( N \) other terms, we see that since \( W_{\frac{n+1}{N}} - W_{\frac{n}{N}} \) is stochastically independent of

\[
g\left(\frac{n}{N}\right),
\]

it follows that

\[
13
\]
\[ E \left[ \left( \frac{N+1}{N} \right)^{nT} \left( W_{\frac{n+1}{N}} - W_{\frac{n}{N}} \right)^2 \right] \]

\[ = E g\left( \frac{nT}{N} \right) E \left( W_{\frac{n+1}{N}} - W_{\frac{n}{N}} \right)^2 = \left[ E g\left( \frac{nT}{N} \right) \right] \frac{T}{N}, \]

where the second equation follows because

\[ E \left( W_{\frac{n+1}{N}} - W_{\frac{n}{N}} \right)^2 = \frac{T}{N}. \]

Therefore

\[ \| V \|_2^2 = \left\{ \sum_{n=0}^{N-1} g\left( \frac{nT}{N} \right) \left( W_{\frac{n+1}{N}} - W_{\frac{n}{N}} \right)^2 \right\}^2 = \frac{T}{N} \sum_{n=0}^{N-1} Eg^2(nT/N) \]

\[ = \frac{T}{N} \sum_{n=0}^{N-1} \| g(nT/N) \|_2^2 < \infty \]

as was to be proved.

Notice that since \( Eg(t)^2 = \| g(t) \|_2^2 \) is a continuous function of \( t \), it follows that the limit as \( N \) goes to infinity of the extreme right-hand side of equation 14.4 is the Riemann integral of \( E(g^2(t)) \) from 0 to \( T \). That is,

\[ \lim_{N \to \infty} \| V \|_2^2 = \int_0^T \| g(t) \|_2^2 dt, \]  

\[ \quad (14.5) \]

where \( V \) is defined by equation 14.2.

I now show that \( V \) is a Cauchy sequence with respect to \( \| . \|_2 \) and do so by imitating the proof of theorem 5.18 in the proof appendix of lecture 5. I must show that for any \( \varepsilon > 0 \), there exists a positive integer \( K \) such that \( \| V_N - V_M \|_2 < \varepsilon \) if \( N \geq K \) and \( M \geq K \).

The proof of theorem 5.18 made use of uniform continuity, so that a first step is to check that uniform continuity applies in the current context. The proof of theorem 5.18 depended on the assertion of lemma 5.17 that a continuous function is uniformly continuous on a compact set. The lemma applied to functions from a subset of \( \mathbb{R}^n \) to a subset of \( \mathbb{R}^k \). We need to apply the same assertion to a continuous function from \([0, \infty)\) to \( L^2(S, \mathcal{S}, \mu) \). It is easy to check that the proof of lemma 5.17 applies in this context. Hence since \( g : [0, \infty) \to L^2(S, \mathcal{S}, \mu) \) is continuous with respect to the norm \( \| . \|_2 \) on \( L^2(S, \mathcal{S}, \mu) \), for every \( \varepsilon > 0 \), there exists a \( \delta > 0 \) such that if \( t \)
and \( r \) belong to \([0, T]\) and \(|t - r| < \delta\), then \(||g(t) - g(r)||_2^2 < \varepsilon^2/4T||.

I continue with the imitation of the proof of theorem 5.18. Let \( \varepsilon \) and \( \delta \) be as in the previous paragraph and let \( K \) be an integer so large that \( K > T/\delta \) and let \( N \) and \( M \) be integers both exceeding \( K \). I show that \(||V - V||_M^2 < \varepsilon||. The numbers \( nT/N \) and \( mT/M \), for \( n = 0, 1, ..., N-1 \) and \( m = 0, 1, ..., M-1 \) form \( Q \) distinct numbers, where \( Q \leq N + M \). Let these numbers be \( t_0, t_1, ..., t_Q \), where \( 0 = t_0 < t_1 < ... < t_Q = T \). Let

\[
I = \sum_{q=0}^{Q-1} g(t_q) \left[ W_{\frac{t_q}{q+1}} - W_{\frac{t_q}{q}} \right] .
\]

It is sufficient to show that \(||V - I||_M^2 < \varepsilon/2 \) and \(||V - I||_Q^2 < \varepsilon/2 \). In order to prove these statements, it is enough to prove that \(||I - V||_N^2 < \varepsilon/2 \), if \( N > T/\delta \). For \( n = 0, 1, ..., N-1 \), let \( nT/N = t_{q(n)} \), so that \( q(n) \leq q < q(n+1) \). Therefore

\[
V - I = \sum_{q=0}^{Q-1} \left[ g(t_q) - g(t_{q+1}) \right] \left[ W_{\frac{t_q}{q+1}} - W_{\frac{t_q}{q}} \right] .
\]

By a close analogue of equation 14.4,

\[
||V - I||_N^2 = \sum_{q=0}^{Q-1} (t_{q+1} - t_q) ||g(t_q) - g(t_{q+1})||_2^2 . \tag{14.6}
\]

Because \( g(t_q) = g\left(\frac{nT}{N}\right) \), for \( n \) such that \( q(n) \leq q < q(n+1) \) and since \( nT/N = t_{q(n)} \leq t_q < t_{q(n+1)} \) = \( (n+1)T/N \) and \( T/N \leq \delta \), it follows that

\[
||g(t_q) - g(t_{q+1})||_2^2 = ||g(nT/N) - g(t_{q+1})||_2^2 < \varepsilon^2/4T .
\]

Therefore

\[
\sum_{q=0}^{Q-1} (t_{q+1} - t_q) ||g(t_q) - g(t_{q+1})||_2^2 < \frac{\varepsilon^2}{4} \sum_{q=0}^{Q-1} (t_{q+1} - t_q) = \frac{\varepsilon^2}{4} . \tag{14.7}
\]

15
Equation 14.6 and inequality 14.7 imply that $\|V - I\|_2 < \varepsilon/2$ and hence complete the proof that the sequence $V_N$ is Cauchy with respect to the norm $\|\cdot\|_2$.

The stochastic integral $\int_0^T g(t) \, dW$ is defined to be $\lim_{N \to \infty} V_N$, where the limit is in $L^2(S, \mathcal{F}, P)$. Notice that $\int_0^T g(t) \, dW$ is itself a stochastic process, since it is a family of random variables indexed by $T$. In fact, it takes values in $L^2(S, \mathcal{F}, P)$. It is important to remember that $\int_0^T g(t) \, dW$ is a random variable. An easy way to recall this is to note that $\int_0^T dW = W(t)$.

Recall that we could not define

$$\lim_{N \to \infty} V(s) = \lim_{N \to \infty} \sum_{n=0}^{N-1} \left( \frac{n}{N} T + \frac{1}{N} \right) \left( W_{n+1}^T(s) - W_n^T(s) \right),$$

for each $s$, because $W(t)$ has unbounded variation as a function of $t$, for $s$ belonging to a set of probability 1. However by considering the limit over all $s$ simultaneously, we end up with a limit in $L^2(S, \mathcal{F}, P)$.