Lecture #14

I continue with the discussion of probability theory. If $x : S \to R$ and $y : S \to R$ are random variables, then x and y are stochastically independent if for any Borel subsets A and B of 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, if x and y are indicator functions of measurable sets, for let $x = \mathcal{X}_A$ and $y = \mathcal{X}_B$. Then

$$E(xy) = E(\mathcal{X}_A \mathcal{X}_B) = E(\mathcal{X}_{A\cap B}) = P(A\cap B) = P(A)P(B) = E(\mathcal{X}_A)E(\mathcal{X}_B) = E(x)E(y).$$

You can use this equation to verify that $E(xy) = E(x)E(y)$ for independent simple functions x and y. Since the integral of every integrable function is the limit of the integrals of simple functions that converge to it, it is easy to show that $E(xy) = E(x)E(y)$ whenever x, y, and xy are integrable and x and y are independent.

This assertion is easy to verify directly if S is finite, for suppose that x takes on values $r_1, r_2, \ldots, r_N$ and y takes on values $t_1, t_2, \ldots, t_M$. Then

$$E(xy) = \sum_{n=1}^{N}\sum_{m=1}^{M} r_n t_m P\{s \in S \mid x(s) = r_n,\ y(s) = t_m\}$$

$$= \sum_{n=1}^{N}\sum_{m=1}^{M} r_n t_m P\{s \in S \mid x(s) = r_n\}P\{s \in S \mid y(s) = t_m\}$$

$$= \sum_{n=1}^{N}\sum_{m=1}^{M} r_n P\{s \in S \mid x(s) = r_n\} t_m P\{s \in S \mid y(s) = t_m\}$$

$$= \sum_{n=1}^{N} r_n P\{s \in S \mid x(s) = r_n\} \sum_{m=1}^{M} t_m P\{s \in S \mid y(s) = t_m\} = (Ex)(Ey).$$

If S is infinite and x and y have densities $f : R \to [0, \infty)$ and $g : R \to [0, \infty)$, respectively, then x and y are independent if and only if the density $h : R^2 \to [0, \infty)$ of the random variable $(x, y) : S \to R^2$ has the form $h(r_1, r_2) = f(r_1)g(r_2)$, for all r_1 and r_2. From this fact, it follows that

$$E(xy) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} r_1 r_2 h(r_1, r_2)\, dr_1 dr_2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} r_1 r_2 f(r_1) g(r_2)\, dr_1 dr_2$$

$$= \int_{-\infty}^{\infty} r_1 f(r_1)\, dr_1 \int_{-\infty}^{\infty} r_2 g(r_2)\, dr_2 = (Ex)(Ey).$$

It is true in general that if x and y are stochastically independent, then $E(xy) = (Ex)(Ey)$, though I do not provide a proof.

It follows that if x and y are independent random variables, then $\text{var}(x + y) = \text{var}(x) + \text{var}(y)$, for

$$\text{var}(x + y) = E(x + y)^2 - (Ex + Ey)^2$$

$$= E(x^2 + 2xy + y^2) - (Ex)^2 - 2(Ex)(Ey) - (Ey)^2$$

$$= E(x^2) + 2E(x)E(y) + E(y^2) - (Ex)^2 - 2(Ex)(Ey) - (Ey)^2$$

$$= E(x^2) - (Ex)^2 + E(y^2) - (Ey)^2$$

$$= \text{var}(x) + \text{var}(y).$$

Suppose that $\mathcal{A}$ is a σ-field over S that is contained in $\mathcal{S}$. That is, any $A \in \mathcal{A}$ belongs to $\mathcal{S}$. Such a σ-field is said to be a subfield of $\mathcal{S}$. A random variable $x : S \to R^N$ that is measurable with respect to $\mathcal{S}$ is said to be independent of $\mathcal{A}$, if every $A \in \mathcal{A}$ is independent of every $C \in \mathcal{S}$ of the form $C = x^{-1}(B)$, where B is a Borel subset of R^N. If $\mathcal{A}$ and x and independent, then

$$\int_A x(s)P(ds) = \int_S \mathcal{X}_A(s)x(s)P(ds) = \int_S \mathcal{X}_A(s)P(ds)\int_S x(s)P(ds) = P(A)\int_S x(s)P(ds),$$

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

The definition of independence may be extended to a sequence of events or random variables. Events $E_1, E_2, \ldots$. are said to be <u>stochastically mutually independent</u> 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 $2 \le K < \infty$. Random variables $x_n : S \to R$, where n = 1, 2, ..., are stochastically mutually independent if for any Borel subsets $A_1, A_2, \ldots$ of R, the events $E_1, E_2, \ldots$ are stochastically mutually independent, where $E_n = \{s \in S \mid x_n(s) \in A_n\}$, 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 $x_n : S \to R$, for n = 1, 2, ... are said to be identically distributed if they all have the same distribution (or cumulative distribution function) on R. The next theorem is central to probability theory.

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

$$\lim_{N \to \infty} \frac{x_1(s) + x_2(s) + \ldots + x_n(s)}{N} = Ex_1.$$

The statement that $\lim_{N \to \infty} \frac{x_1(s) + x_2(s) + \ldots + x_n(s)}{N} = Ex_1$ with probability 1 means that $\left\{ s \in S \,\middle|\, \lim_{N \to \infty} \frac{x_1(s) + x_2(s) + \ldots + x_n(s)}{N} = Ex_1 \right\}$ is measurable, i.e., belongs to $\mathbb{S}$, and has probability 1. The law of large numbers makes a connection between averages and probabilities. For instance, let S be set of sequence of heads and tails on successive tosses of a fair coin. That is,

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

Let

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

That is, $x_n(s) = 1$, if and only if the nth toss comes up heads. 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, Ex_1. In order to state this theorem, I need to define weak convergence of cumulative distribution functions. Let F and F_1, F_2, ... be cumulative distribution functions on R. The sequence F_1, F_2, ... 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 $x_n : S \to R$, where n = 1, 2, ... , be a sequence of random variables that are stochastically mutually independent and identically distributed and such that $Ex_1 = \mu$ and $Var(x_1) = \sigma^2$ exist and are finite. For n = 1, 2, ... , let

$$X_N(s) = \frac{(x_1(s) - \mu) + (x_2(s) - \mu) + ... + (x_N(s) - \mu)}{\sigma\sqrt{N}}$$

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 $Var(x_1) < \infty$ implies that $E|x_1| < \infty$, 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 $(x_t)_{t \in T}$ or $((x(s))_{t \in T}$, where T is some index set and the random variables x_t are all defined on some common state space S. More specifically, there is a probability space $(S, \mathfrak{S}, P)$ such that $x_t : S \to R$ is measurable with respect to $\mathfrak{S}$, 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 $\mathcal{C}_t = \{x_t^{-1}(B) \mid B \text{ is a Borel subset of } R\}$. The set of measurable subsets, $\mathfrak{S}$, of S is usually taken to be the smallest σ-field that contains $\mathcal{C}_t$, for all t. Let $\mathfrak{S}_t$ be the smallest σ-field that contains $\mathcal{C}_r$, for all $r \leq t$. Then $\mathfrak{S}_t$ is contained in $\mathfrak{S}$ and represents the information available at time t.

If $x : S \to R$ is a random variable and $\mathcal{C}$ is a σ-field contained in $\mathfrak{S}$, then x is said to be measurable with respect to $\mathcal{Q}$ if $x^{-1}(B) \in \mathcal{Q}$, for every Borel subset B or R. By definition, the random variable $x_r : S \to R$ is measurable with respect to $\mathfrak{S}_t$, 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 $(x_{t_1}, x_{t_2}, \ldots, x_{t_N})$ on R^N, for all finite subsets $(t_1, t_2, \ldots, t_N)$ of indices. If this is done in a coherent way for all finite subsets of indices, then by a theorem of Kolmogorov there exists a probability distribution on $\mathfrak{S}$ that induces the distributions defined for all finite subsets of random variables $(x_{t_1}, x_{t_2}, \ldots, x_{t_N})$. The existence of the probability on $\mathfrak{S}$ is a complicated topic that I do not pursue.

Definition: A Weiner process is a stochastic process $(W_t)_{t \in [0, \infty)}$ with state space S and such that

1) for each finite subset of non-negative numbers $(t_1, t_2, \ldots, t_N)$ such that $t_1 < t_2 < \ldots < t_N$, the random variable $(x_{t_1}, x_{t_2}, \ldots, x_{t_N}) : S \to R^N$ 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 $0 \leq r < t$, $W_t - W_r$, has the normal distribution N(0, t − r) and $W_t - W_r$ is independent of $\mathfrak{S}_r$ and hence of the random variable W_r.

These three conditions provide enough information to define a probability distribution on S, so that a Weiner process exists. I do not prove this assertion. Notice that as before $S = \{s : [0, \infty) \to R\}$, 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. It is understood that dW_t is normally distributed with mean 0 and variance dt That is, its distribution is N(0, dt).

The Weiner 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, ..., N. Suppose that the nth shock is expressed by a random variable x_n^N, which equals either $\sqrt{t/N}$ or $-\sqrt{t/N}$, each with probability 1/2. Assume that the random variables $x_1^N, x_2^N, \ldots, x_N^N$ are mutually independent. Then $Ex_n^N = 0$ and $\mathrm{var}(x_n^N) = t/N$. Because $x_1^N, x_2^N, \ldots, x_N^N$ are mutually independent,

$$\mathrm{var}(x_1^N + x_2^N + \ldots + x_N^N) = \sum_{n=1}^{N} \mathrm{var}(x_n^N) = \frac{Nt}{N} = t.$$

Notice that

$$\sum_{n=1}^{N} x_n^N = \sqrt{t}\,\frac{\sum_{n=1}^{N} x_n^N}{\sqrt{\frac{t}{N}}\sqrt{N}}.$$

By the central limit theorem, the distribution of the random variables $\frac{\sum_{n=1}^{N} x_n^N}{\sqrt{\frac{t}{N}}\sqrt{N}}$ converges weakly to N(0, 1). Hence if we imagine that $\frac{\sum_{n=1}^{N} x_n^N}{\sqrt{\frac{t}{N}}\sqrt{N}}$ converges to some random variable x with distribution N(0, 1), then $\sqrt{t}\,\frac{\sum_{n=1}^{N} x_n^N}{\sqrt{\frac{t}{N}}\sqrt{N}} = \sum_{n=1}^{N} x_n^N$ 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 the W_t, the Weiner process at time t.

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

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

where N(t) is the largest positive integer such that $N(t)/N \le T$ and where the random variables $\Delta x_N(n/N, s)$ 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 $\Delta x_N(n/N, s)$ be normally distributed. Heuristically we my think of $\Delta x_N(n/N, s)$ 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 in the air or water. The observation is normally 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 Weiner process represents the limit result of the time path of such jiggling, though in just one dimension.

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

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

Another curious fact about Weiner processes is that they have bounded variation with probability 1.

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

$$\sup \{ \sum_{n=1}^{N} |f(t_{n+1}) - f(t_n)| \; \Big| \; 0 = t_1 < t_2 < .... < t_N = 1\}.$$

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

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

More general stochastic processes may be defined from the Weiner process by using the stochastic or Ito 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, 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 number. Imagine now trying to solve a differential equation where the motion is subject to random jiggles. We might write such an equation as

$$\frac{dx(t, s)}{dt}dt = f(x(t, s), t)dt + g(x(t, s), t)dW_t(s),$$

for each s, where $dW_t(s)$ is understood to be a Weiner process with mean 0 and variance dt, where dt is an infinitesimal interval of time and s represents the state. The notation dx(t, s)/dt does not apply, because 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 <u>stochastic or Ito integral</u>. We can simplify the notation by replacing g(x(t, s), t) by g(t, s), so that the objective becomes to define

$$\int_0^T g(t, s)dW_t(s),$$

where $W_t(s)$ is a Weiner 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

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

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 Riemann-Stieltjes integral of g(t, s) with respect to $W_t(s)$. 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. (The limit would exist if the variation of $W_t(s)$ were bounded.) Therefore the analogue of theorem 5.15 on the existence of the Riemann integral does not apply, and an entirely different approach is required. Let

$$V_N(s) = \sum_{n=0}^{N-1} g\left(\frac{n}{N}T, s\right)\left[W_{\frac{n+1}{N}T}(s) - W_{\frac{n}{N}T}(s)\right]. \tag{14.2}$$

We want to show that the functions V_N converge in some sense. 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 $L_2(S, \mathfrak{S}, P)$, where

$$L_2(S, \mathfrak{S}, P) = \left\{x : S \to R \mid x \text{ is measurable with respect to } \mathfrak{S} \text{ and } \int_S x^2(s)\, P(ds) < \infty\right\}.$$

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

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

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

<u>Theorem</u> 14.5: Every Cauchy sequence in $L_2(S, \mathfrak{S}, P)$ converges to some member of $L_2(S, \mathfrak{S}, P)$.

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

<u>Assumptions</u>:

1) $g(t) : S \to R$ is measurable with respect to $\mathfrak{S}_t$, for all t,

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

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

The third assumption means that if t_n, n = 1, 2, ... , 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

$$\Big| ||g(t_n)||_2 - ||g(t)||_2 \Big| \le ||g(t_n) - g(t)||_2,$$

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

Notice that the first assumption implies that $W_t - W_r$ is stochastically independent of g(r), if t > r. Hence

$$E[g(r)(W_t - W_r)] = Eg(r)E(W_t - W_r) = Eg(r)(0) = 0.$$

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

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

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

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

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

We want to show that the functions V_N converge in $L_2(S, \mathbb{S}, P)$. In order to do so, we must first verify that these functions belong to $L_2(S, \mathbb{S}, P)$ and then verify that the sequence V_N

is Cauchy. In order to show that each V_N belongs to $L_2(S, \mathbb{S}, P)$, we must show that

$$\left\| \sum_{n=0}^{N-1} g\left(\frac{n}{N}T\right)\left(W_{\frac{n+1}{N}T} - W_{\frac{n}{N}T}\right)\right\|_2^2 = E\left[\sum_{n=0}^{N-1} g\left(\frac{n}{N}T\right)\left(W_{\frac{n+1}{N}T} - W_{\frac{n}{N}T}\right)\right]^2 < \infty.$$

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

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

and the other terms have the form

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

where $m > n$. Since

$$W_{\frac{m+1}{N}T} - W_{\frac{m}{N}T}$$

is independent of

$$g\left(\frac{n}{N}T\right)g\left(\frac{m}{N}T\right)\left(W_{\frac{n+1}{N}T} - W_{\frac{n}{N}T}\right),$$

we have that

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

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

Returning to the N other terms, we see that since $W_{\frac{n+1}{N}T} - W_{\frac{n}{N}T}$ is stochastically independent of $g\left(\frac{n}{N}T\right)$, it follows that

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

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

where the second equation follows because

$$Eg\left(W_{\frac{n+1}{N}T} - W_{\frac{n}{N}T}\right)^2 = \frac{T}{N}.$$

Therefore

$$\left\|\sum_{n=0}^{N-1} g\left(\frac{n}{N}T\right)\left(W_{\frac{n+1}{N}T} - W_{\frac{n}{N}T}\right)\right\|_2^2 = E\left[\sum_{n=0}^{N-1} g\left(\frac{n}{N}T\right)\left(W_{\frac{n+1}{N}T} - W_{\frac{n}{N}T}\right)\right]^2 = \frac{T}{N}\sum_{n=0}^{N-1} Eg\left(\frac{n}{N}T\right)^2. \qquad (14.4)$$

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 $E(g(t)^2)$ from 0 to T. That is,

$$\lim_{N \to \infty} ||V_N||_2^2 = \lim_{N \to \infty} \frac{T}{N}\sum_{n=0}^{N-1} Eg\left(\frac{n}{N}T\right)^2 = \int_0^T E[g(t)^2]\,dt, \qquad (14.5)$$

where V_N is defined by equation 14.3.

I now show that V_N is a Cauchy sequence with respect to $||.||_2$ and do so by imitating the proof of theorem 5.15 in the 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 5.15 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 5.15 depended on the assertion of lemma 5.14 that a continuous function is uniformly continuous on a compact set. The lemma applied to functions from a subset of R^N to a subset of R^K. We need to apply the same assertion to a continuous function from $[0, \infty)$ to $L_2(S, \mathbb{S}, P)$. It is easy to check that the proof of lemma 5.14 applies in this context. Since $g : [0, \infty) \to L_2(S, \mathbb{S}, P)$ is continuous with respect to the norm $||.||_2$ on $L_2(S, \mathbb{S}, P)$, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that if $|t - r| < \delta$, then $||g(t) - g(r)||_2^2 < \varepsilon^2/4T$.

I continue with the imitation of the proof of theorem 5.15. Let K be an integer so large that $K > T/\delta$ and let N and M be integers both exceeding K. I show that $||V_N - V_M||_2 < \varepsilon$. The numbers nT/N and mT/M, for n = 0, 1, ... , N–1 and m = 0, 1, ... , M–1 form Q distinct numbers, where $Q < N + M$. Let these numbers be $t_0, t_1, \ldots, t_Q$, where

$0 = t_0 < t_1 < .... < t_Q = T$. Let

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

It is sufficient to show that $||V_N - \mathcal{I}||_2 < \varepsilon/2$ and $||V_M - \mathcal{I}||_2 < \varepsilon/2$. In order to prove these statements, it is enough to prove that $||\mathcal{I} - V_N||_2 < \varepsilon/2$, if $N > T/\delta$. For n = 0, 1, ... , N − 1, let $nT/N = t_{q(n)}$, so that $q(n) < q(n+1)$, for all n. I hope it is clear that

$$V_N = \sum_{n=0}^{N-1} g\left(\frac{n}{N}T\right)\left[W_{\frac{n+1}{N}T} - W_{\frac{n}{N}T}\right] = \sum_{q=0}^{Q-1} \overline{g}(t_q)\left[W_{t_{q+1}} - W_{t_q}\right],$$

where $\overline{g}(t_q) = g\left(\frac{n}{N}T\right)$, for n such that $q(n) \le q < q(n+1)$. Therefore

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

By a close analogue of equation 14.4,

$$||V_N - \mathcal{I}||_2^2 = \sum_{q=0}^{Q-1}(t_{q+1} - t_q)\,||\overline{g}(t_q) - g(t_q)||_2^2. \tag{14.6}$$

Because $nT/N = t_{q(n)} \le t_q < t_{q(n+1)} = (n+1)T/N$ and $T/N \le \delta$, it follows that

$$||\overline{g}(t_q) - g(t_q)||_2^2 = ||g(nT/N) - g(t_q)||_2^2 < \varepsilon^2/4T.$$

Therefore

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

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

The stochastic integral $\int_0^T g(t)\,dW_t$ is defined to be $\lim_{N\to\infty} V_N$, where the limit is in $L_2(S, \mathcal{S}, P)$. Notice that $\int_0^T g(t)\,dW_t$ is itself a stochastic process, since it is family of random variables indexed by T. In fact, it takes values in $L_2(S, \mathbb{S}, P)$. It is important to remember that $\int_0^T g(t)\, dW_t$ is a random variable. An easy way to recall this is to note that $\int_0^T dW_t = W_t(s)$.

Recall that we could not define

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

for each s, because $W_t(s)$ 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 $L_2(S, \mathbb{S}, P)$.