Chapter 16

Convergence of Random Walks

This lecture examines the convergence of random walks to the Wiener process. This is very important both physically and statistically, and illustrates the utility of the theory of Feller processes.

Section 16.1 finds the semi-group of the Wiener process, shows it satisfies the Feller properties, and finds its generator.

Section 16.2 turns random walks into cadlag processes, and gives a fairly easy proof that they converge on the Wiener process.

16.1 The Wiener Process is Feller

Recall that the Wiener process $W(t)$ is defined by starting at the origin, by independent increments over non-overlapping intervals, by the Gaussian distribution of increments, and by continuity of sample paths (Examples 38 and 78). The process is homogeneous, and the transition kernels are (Section 11.1)

$$\mu_t(w_1, B) = \int_B dw_2 \frac{1}{\sqrt{2\pi t}} e^{-\frac{(w_2-w_1)^2}{2t}} \tag{16.1}$$

$$\frac{d\mu_t(w_1, w_2)}{d\lambda} = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(w_2-w_1)^2}{2t}} \tag{16.2}$$

where the second line gives the density of the transition kernel with respect to Lebesgue measure.

Since the kernels are known, we can write down the corresponding evolution operators:

$$K_t f(w_1) = \int dw_2 f(w_2) \frac{1}{\sqrt{2\pi t}} e^{-\frac{(w_2-w_1)^2}{2t}} \tag{16.3}$$

We saw in Section 11.1 that the kernels have the semi-group property, so the evolution operators do too.

Let's check that $\{K_t\}, t \geq 0$ is a Feller semi-group. The first Feller property is easier to check in its probabilistic form, that, for all t, $y \to x$ implies $W_y(t) \xrightarrow{d} W_x(t)$. The distribution of $W_x(t)$ is just $\mathcal{N}(x, t)$, and it is indeed true that $y \to x$ implies $\mathcal{N}(y, t) \to \mathcal{N}(x, t)$. The second Feller property can be checked in its semi-group form: as $t \to 0$, $\mu_t(w_1, B)$ approaches $\delta(w - w_1)$, so $\lim_{t\to 0} K_t f(x) = f(x)$. Thus, the Wiener process is a Feller process. This implies that it has cadlag sample paths (Theorem 158), but we already knew that, since we know it's continuous. What we did not know was that the Wiener process is not just Markov but strong Markov, which follows from Theorem 159.

It's easier to find the generator of $\{K_t\}, t \geq 0$, it will help to re-write it in an equivalent form, as

$$K_t f(w) = \mathbf{E}\left[f(w + Z\sqrt{t})\right] \quad (16.4)$$

where Z is an independent $\mathcal{N}(0, 1)$ random variable. (You should convince yourself that this is equivalent.) Now let's pick an $f \in C_0$ which is also twice continuously differentiable, i.e., $f \in C_0 \cap C^2$. Look at $K_t f(w) - f(w)$, and apply Taylor's theorem, expanding around w:

$$\begin{aligned}
K_t f(w) - f(w) &= \mathbf{E}\left[f(w + Z\sqrt{t})\right] - f(w) && (16.5)\\
&= \mathbf{E}\left[f(w + Z\sqrt{t}) - f(w)\right] && (16.6)\\
&= \mathbf{E}\left[Z\sqrt{t}f'(w) + \frac{1}{2}tZ^2 f''(w) + R(Z\sqrt{t})\right] && (16.7)\\
&= \sqrt{t}f'(w)\mathbf{E}[Z] + t\frac{f''(w)}{2}\mathbf{E}\left[Z^2\right] + \mathbf{E}\left[R(Z\sqrt{t})\right] && (16.8)\\
\lim_{t\downarrow 0} \frac{K_t f(w) - f(w)}{t} &= \frac{1}{2}f''(w) + \lim_{t\downarrow 0} \frac{\mathbf{E}\left[R(Z\sqrt{t})\right]}{t} && (16.9)
\end{aligned}$$

So, we need to investigate the behavior of the remainder term $R(Z\sqrt{t})$. We know from Taylor's theorem that

$$\begin{aligned}
R(Z\sqrt{t}) &= \frac{tZ^2}{2}\int_0^1 du\ f''(w + uZ\sqrt{t}) - f''(w) && (16.10)\\
& && (16.11)
\end{aligned}$$

Since $f \in C_0 \cap C^2$, we know that $f'' \in C_0$. Therefore, f'' is uniformly continuous, and has a modulus of continuity,

$$m(f'', h) = \sup_{x,y:\ |x-y|\leq h} |f''(x) - f''(y)| \quad (16.12)$$

which goes to 0 as $h \downarrow 0$. Thus

$$\begin{aligned}
\left|R(Z\sqrt{t})\right| &\leq \frac{tZ^2}{2}m(f'', Z\sqrt{t}) && (16.13)\\
\lim_{t\to 0} \frac{\left|R(Z\sqrt{t})\right|}{t} &\leq \lim_{t\to 0} \frac{Z^2 m(f'', Z\sqrt{t})}{2} && (16.14)\\
&= 0 && (16.15)
\end{aligned}$$

Plugging back in to Equation 16.9,

$$Gf(w) = \frac{1}{2}f''(w) + \lim_{t \downarrow 0} \frac{\mathbf{E}\left[R(Z\sqrt{t})\right]}{t} \quad (16.16)$$

$$= \frac{1}{2}f''(w) \quad (16.17)$$

That is, $G = \frac{1}{2}\frac{d^2}{dw^2}$, one half of the Laplacian. We have shown this only for $C_0 \cap C^2$, but this is clearly a linear subspace of C_0, and, since C^2 is dense in C, it is dense in C_0, i.e., this is a core for the generator. Hence the generator is really the extension of $\frac{1}{2}\frac{d^2}{dw^2}$ to the whole of C_0, but this is too cumbersome to repeat all the time, so we just say it's the Laplacian.

16.2 Convergence of Random Walks

Let $X_1, X_2, \ldots$ be a sequence of IID variables with mean 0 and variance 1. The random walk process S_n is then just $\sum_{i=1}^{n} X_i$. It is a discrete-time Markov process, and consequently also a strong Markov process. Imagine each step of the walk takes some time h, and imagine this time interval becoming smaller and smaller. Then, between any two times t_1 and t_2, the number of steps of the random walk will be about $\frac{t_2 - t_1}{h}$, which will go to infinity. The displacement of the random walk between t_1 and t_2 will then be a sum of an increasingly large number of IID random variables, and by the central limit theorem will approach a Gaussian distribution. Moreover, if we look at the interval of time from t_2 to t_3, we will see another Gaussian, but all of the random-walk steps going into it will be independent of those going into our first interval. So, we expect that the random walk will in some sense come to look like the Wiener process, no matter what the exact distribution of the X_1. Let's consider this in more detail.

Define $Y_n(t) = n^{-1/2}\sum_{i=0}^{[nt]} X_i = n^{-1/2} S_{[nt]}$, where $X_0 = 0$ and $[nt]$ is the integer part of the real number nt. You should convince yourself that this is a Markov process, with cadlag sample paths.

We want to consider the limiting distribution of Y_n as $n \to \infty$. First of all, we should convince ourselves that a limit distribution exists. But this is not too hard. For any fixed t, $Y_n(t)$ approaches a Gaussian distribution by the central limit theorem. For any fixed finite collection of times $t_1 \leq t_2 \ldots \leq t_k$, $Y_n(t_1), Y_n(t_2), \ldots Y_n(t_k)$ approaches a limiting distribution if $Y_n(t_1), Y_n(t_2) - Y_n(t_1), \ldots Y_n(t_k) - Y_n(t_{k-1})$ does, but that again will be true by the (multivariate) central limit theorem. Since the limiting finite-dimensional distributions exist, some limiting distribution exists (via Theorem 23). It remains to identify it.

Lemma 173 $Y_n \stackrel{fd}{\to} W$.

PROOF: For all n, $Y_n(0) = 0 = W(0)$. For any $t_2 > t_1$,

$$\mathcal{L}\left(Y_n(t_2) - Y_n(t_1)\right) = \mathcal{L}\left(\frac{1}{\sqrt{n}} \sum_{i=[nt_1]}^{[nt_2]} X_i\right) \quad (16.18)$$

$$\xrightarrow{d} \mathcal{N}(0, t_2 - t_1) \quad (16.19)$$

$$= \mathcal{L}\left(W(t_2) - W(t_1)\right) \quad (16.20)$$

Finally, for any three times $t_1 < t_2 < t_3$, $Y_n(t_3) - Y_n(t_2)$ and $Y_n(t_2) - Y_n(t_1)$ are independent for sufficiently large n, because they become sums of disjoint collections of independent random variables. Thus, the limiting distribution of Y_n starts at the origin and has independent Gaussian increments. Since these properties determine the finite-dimensional distributions of the Wiener process, $Y_n \xrightarrow{fd} W$. □

Theorem 174 $Y_n \xrightarrow{d} W$.

PROOF: By Theorem 165, it is enough to show that $Y_n \xrightarrow{fd} W$, and that any of the properties in Theorem 166 hold. The lemma took care of the finite-dimensional convergence, so we can turn to the second part. A sufficient condition is property (1) inn the latter theorem, that $|Y_n(\tau_n + h_n) - Y_n(\tau_n)| \xrightarrow{P} 0$ for all finite optional times τ_n and any sequence of positive constants $h_n \to 0$.

$$|Y_n(\tau_n + h_n) - Y_n(\tau_n)| = n^{-1/2}\left|S_{[n\tau_n + nh_n]} - S_{[n\tau_n]}\right| \quad (16.21)$$

$$\stackrel{d}{=} n^{-1/2}\left|S_{[nh_n]} - S_0\right| \quad (16.22)$$

$$= n^{-1/2}\left|S_{[nh_n]}\right| \quad (16.23)$$

$$= n^{-1/2}\left|\sum_{i=0}^{[nh_n]} X_i\right| \quad (16.24)$$

To see that this converges in probability to zero, we will appeal to Chebyshev's inequality: if Z_i have common mean 0 and variance σ^2, then, for every positive ϵ,

$$\mathbb{P}\left(\left|\sum_{i=1}^{m} Z_i\right| > \epsilon\right) \leq \frac{m\sigma^2}{\epsilon^2} \quad (16.25)$$

Here we have $Z_i = X_i/\sqrt{n}$, so $\sigma^2 = 1/n$, and $m = [nh_n]$. Thus

$$\mathbb{P}\left(n^{-1/2}\left|S_{[nh_n]}\right| > \epsilon\right) \leq \frac{[nh_n]}{n\epsilon^2} \quad (16.26)$$

As $0 \leq [nh_n]/n \leq h_n$, and $h_n \to 0$, the bounding probability must go to zero for every fixed ϵ. Hence $n^{-1/2}\left|S_{[nh_n]}\right| \xrightarrow{P} 0$. □

Corollary 175 (The Invariance Principle) *Let $X_1, X_2, \ldots$ be IID random variables with mean μ and variance σ^2. Then*

$$Y_n(t) \equiv \frac{1}{\sqrt{n}} \sum_{i=0}^{[nt]} \frac{X_i - \mu}{\sigma} \quad \stackrel{d}{\to} \quad W(t) \tag{16.27}$$

PROOF: $(X_i - \mu)/\sigma$ has mean 0 and variance 1, so Theorem 174 applies. □

This result is called "the invariance principle", because it says that the limiting distribution of the sequences of sums depends only on the mean and variance of the individual terms, and is consequently *invariant* under changes which leave those alone. Both this result and the previous one are known as the "functional central limit theorem", because convergence in distribution is the same as convergence of all bounded continuous *functionals* of the sample path. Another name is "Donsker's Theorem", which is sometimes associated however with the following corollary of Theorem 174.

Corollary 176 (Donsker's Theorem) *Let $Y_n(t)$ and $W(t)$ be as before, but restrict the index set T to the unit interval $[0, 1]$. Let f be any function from $\mathbf{D}([0, 1])$ to $\mathbb{R}$ which is measurable and a.s. continuous at W. Then $f(Y_n) \stackrel{d}{\to} f(W)$.*

PROOF: Exercise. □

This version is especially important for statistical purposes, as we'll see a bit later.

16.3 Exercises

Exercise 16.1 *Go through all the details of Example 138.*

- a *Show that $\mathcal{F}_t^X \subseteq \mathcal{F}_t^W$ for all t, and that $\mathcal{F}^X \subset \mathcal{F}^W$.*
- b *Show that $\tau = \inf_t X(t) = (0, 0)$ is a $\mathcal{F}^X$-optional time, and that it is finite with probability 1.*
- c *Show that X is Markov with respect to both its natural filtration and the natural filtration of the driving Wiener process.*
- d *Show that X is not strongly Markov at τ.*
- e *Which, if any, of the Feller properties does X have?*

Exercise 16.2 *Consider a d-dimensional Wiener process, i.e., an $\mathbb{R}^d$-valued process where each coordinate is an independent Wiener process. Find the generator.*

Exercise 16.3 *Prove Donsker's Theorem (Corollary 176).*

Exercise 16.4 (Diffusion equation) *As mentioned in class, the partial differential equation*

$$\frac{1}{2}\frac{\partial^2 f(x,t)}{\partial x^2} = \frac{\partial f(x,t)}{\partial t}$$

is called the diffusion equation*. From our discussion of initial value problems in Chapter 12 (Corollary 126 and related material), it is clear that the function* $f(x,t)$ *solves the diffusion equation with initial condition* $f(x,0)$ *if and only if* $f(x,t) = K_t f(x,0)$*, where* K_t *is the evolution operator of the Wiener process.*

a *Take* $f(x,0) = (2\pi 10^{-4})^{-1/2} e^{-\frac{x^2}{2\cdot 10^{-4}}}$. $f(x,t)$ *can be found analytically; do so.*

b *Estimate* $f(x,10)$ *over the interval* $[-5,5]$ *stochastically. Use the fact that* $K_t f(x) = \mathbf{E}[f(W(t))|W(0)=x]$*, and that random walks converge on the Wiener process. (Be careful that you scale your random walks the right way!) Give an indication of the error in this estimate.*

c *Can you find an analytical form for* $f(x,t)$ *if* $f(x,0) = \mathbf{1}_{[-0.5,0.5]}(x)$*?*

d *Find* $f(x,10)$*, with the new initial conditions, by numerical integration on the domain* $[-10,10]$*, and compare it to a stochastic estimate.*