Math 212 Lecture 16

Wiener measure and Brownian motion.

The Big Path Space.

We begin by constructing Wiener measure following a paper by Nelson, *Journal of Mathematical Physics* **5** (1964) 332-343.

Let $\dot{\mathbf{R}}^{\mathbf{n}}$ denote the one point compactification of $\mathbf{R}^n$. Let

$$\Omega := \prod_{0 \leq t < \infty} \dot{\mathbf{R}}^{\mathbf{n}} \tag{1}$$

be the product of copies of $\dot{\mathbf{R}}^{\mathbf{n}}$, one for each non-negative t. This is an uncountable product, and so a huge space, but by Tychonoff's theorem, it is compact and Hausdorff. We can think of a point ω of Ω as being a function from $\mathbf{R}_+$ to $\dot{\mathbf{R}}^{\mathbf{n}}$, i.e. as as a "curve" with no restrictions whatsoever.

Let F be a continuous function on the m-fold product:

$$F : \prod^{m} \dot{\mathbf{R}}^{\mathbf{n}} \to \mathbf{R},$$

and let $t_1 \leq t_2 \leq \cdots \leq t_m$ be fixed "times". Define

$$\phi = \phi_{F;t_1,\ldots,t_m} : \ \Omega \to \mathbf{R}$$

by

$$\phi(\omega) := F(\omega(t_1), \ldots, \omega(t_m)).$$

Finite functions.

$$F : \prod_{i=1}^{m} \dot{\mathbf{R}}^{\mathbf{n}} \to \mathbf{R},$$

and let $t_1 \leq t_2 \leq \cdots \leq t_m$ be fixed "times". Define

$$\phi = \phi_{F;t_1,\ldots,t_m} : \ \Omega \to \mathbf{R}$$

by

$$\phi(\omega) := F(\omega(t_1), \ldots, \omega(t_m)).$$

We can call such a function a **finite** function since its value at ω depends only on the values of ω at finitely many points. The set of such functions satisfies our abstract axioms for a space on which we can define integration. Furthermore, the set of such functions is an algebra containing 1 and which separates points, so is dense in $C(\Omega)$ by the Stone-Weierstrass theorem. Let us call the space of such functions $C_{fin}(\Omega)$.

If we define an integral I on $C_{fin}(\Omega)$ then, by the Stone-Weierstrass theorem it extends to $C(\Omega)$ and therefore, by the Riesz representation theorem, gives us a regular Borel measure on Ω.

For each $x \in \mathbf{R}^n$ we are going to define such an integral, I_x by

$$I_x(\phi) =$$

$$\int \cdots \int F(x_1, x_2, \ldots, x_m) p(x, x_1; t_1) p(x_1, x_2; t_2 - t_1) \cdots p(x_{m-1}, x_m, t_m - t_{m-1}) dx_1 \ldots dx_m$$

when $\phi = \phi_{F, t_1, \ldots, t_m}$ where

$$p(x, y; t) = \frac{1}{(2\pi t)^{n/2}} e^{-(x-y)^2/2t} \tag{2}$$

(with $p(x, \infty) = 0$) and all integrations are over $\dot{\mathbf{R}}^{\mathbf{n}}$. In order to check that this is well defined, we have to verify that if F does not depend on a given x_i then we get the same answer if we define ϕ in terms of the corresponding function of the remaining $m - 1$ variables. This amounts to the computation

$$\int p(x, y; s) p(y, z, t) dy = p(x, z; s + t).$$

To prove:

$$\int p(x, y; s)p(y, z, t)dy = p(x, z; s + t).$$

If $n = 1$ this is the computation

$$\frac{1}{2\pi t}\int_{\mathbf{R}} e^{-(x-y)^2/2s}e^{-(y-z)^2 2t}dy = \frac{1}{2\pi(s+t)}e^{(x-z)^2/2(s+t)}.$$

If we make the change of variables $u = x - y$ this becomes

$$n_t \star n_s = n_{t+s}$$

where

$$n_r(x) := \frac{1}{\sqrt{r}}e^{-x^2/2r}.$$

In terms of our "scaling operator" S_a given by $S_a f(x) = f(ax)$ we can write

$$n_r = r^{-\frac{1}{2}}S_{r^{-\frac{1}{2}}}n$$

where n is the unit Gaussian $n(x) = e^{-x^2/2}$.

To prove:

$$n_t \star n_s = n_{t+s}$$

where

$$n_r(x) := \frac{1}{\sqrt{r}} e^{-x^2/2r}.$$

In terms of our "scaling operator" S_a given by $S_a f(x) = f(ax)$ we can write

$$n_r = r^{-\frac{1}{2}} S_{r^{-\frac{1}{2}}} n$$

where n is the unit Gaussian $n(x) = e^{-x^2/2}$. Now the Fourier transform takes convolution into multiplication, satisfies

$$(S_a f)\hat{} = (1/a) S_{1/a} \hat{f},$$

and takes the unit Gaussian into the unit Gaussian. Thus upon Fourier transform, the equation $n_t \star n_s = n_{t+s}$ becomes the obvious fact that

$$e^{-s\xi^2/2} e^{-t\xi^2/2} = e^{-(s+t)\xi^2/2}.$$

The same proof (or an iterated version of the one dimensional result) applies in n-dimensions.

Probabilistic interpretation.

So, for each $x \in \mathbf{R}^n$ we have defined a measure on Ω. We denote the measure corresponding to I_x by pr_x. It is a probability measure in the sense that $\mathrm{pr}_x(\Omega) = 1$.

The intuitive idea behind the definition of pr_x is that it assigns probability

$$\mathrm{pr}_x(E) :=$$

$$\int_{E_1} \cdots \int_{E_m} p(x, x_1; t_1) p(x_1, x_2; t_2 - t_1) \cdots p(x_{m-1}, x_m, t_m - t_{m-1}) dx_1 \ldots dx_m$$

to the set of all paths ω which start at x and pass through the set E_1 at time t_1, the set E_2 at time t_2 etc. and we have denoted this set of paths by E.

The heat equation.

We pause to reflect upon the computation we did in the preceding section. Define the operator T_t on the space $\mathcal{S}$ (or on $\mathcal{S}'$) by

$$(T_t f)(x) = \int_{\mathbf{R}^n} p(x,y,t) f(y) dy. \tag{3}$$

In other words, T_t is the operation of convolution with

$$t^{-n/2} e^{-x^2/2t}.$$

We have verified that

$$T_t \circ T_s = T_{t+s}. \tag{4}$$

Also, we have verified that when we take Fourier transforms,

$$(T_t f)\hat{}(\xi) = e^{-t\xi^2/2} \hat{f}(\xi). \tag{5}$$

If we let $t \to 0$ in this equation we get

$$\lim_{t \to 0} T_t = \text{Identity}. \tag{6}$$

$$T_t \circ T_s = T_{t+s}. \tag{4}$$

$$(T_t f)\hat{}(\xi) = e^{-t\xi^2/2}\hat{f}(\xi). \tag{5}$$

$$\lim_{t \to 0} T_t = \text{Identity}. \tag{6}$$

Using some language we will introduce later, conditions (4) and (6) say that the T_t form a continuous semi-group of operators. If we differentiate (5) with respect to t, and let

$$u(t,x) := (T_t f)(x)$$

we see that u is a solution of the "heat equation"

$$\frac{\partial^2 u}{(\partial t)^2} = \frac{\partial^2 u}{(\partial x^1)^2} + \cdots + \frac{\partial^2 u}{(\partial x^n)^2}$$

with the initial conditions $u(0,x) = f(x)$. In terms of the operator

$$\Delta := -\left(\frac{\partial^2}{(\partial x^1)^2} + \cdots + \frac{\partial^2}{(\partial x^n)^2}\right)$$

we are tempted to write

$$T_t = e^{-t\Delta},$$

in analogy to our study of elliptic operators on compact manifolds.

Paths are continuous with probability one.

The purpose of this subsection is to prove that if we use the measure pr_x, then the set of discontinuous paths has measure zero.

We begin with some technical issues. We recall that the statement that a measure μ is regular means that for any Borel set A

$$\mu(A) = \inf\{\mu(G) : \ A \subset G, \ G \text{ open } \}$$

and for any open set U

$$\mu(U) = \sup\{\mu(K) : \ K \subset U, \ K \text{ compact}\}.$$

This second condition has the following consequence: Suppose that Γ is any collection of open sets which is closed under finite union. If

$$O = \bigcup_{G \in \Gamma} G$$

then

$$\mu(O) = \sup_{G \in \Gamma} \mu(G)$$

since any compact subset of O is covered by finitely many sets belonging to Γ. The importance of this stems from the fact that we can allow Γ to consist of uncountably many open sets, and we will need to impose uncountably many conditions in singling out the space of continuous paths, for example. Indeed, our first task will be to show that the measure pr_x is concentrated on the space of continuous paths in $\mathbf{R}^n$ which do not go to infinity too fast.

We begin with the following computation in one dimension:

$$\mathrm{pr}_0(\{|\omega(t)| > r\}) = 2 \cdot \left(\frac{1}{2\pi t}\right)^{1/2} \int_r^\infty e^{-x^2/2t} dx \le \left(\frac{2}{\pi t}\right)^{1/2} \int_r^\infty \frac{x}{r} e^{-x^2/2t} dx =$$

$$\left(\frac{2}{\pi t}\right)^{1/2} \frac{t}{r} \int_r^\infty \frac{x}{t} e^{-x^2/2t} dx = \left(\frac{2t}{\pi}\right)^{1/2} \frac{e^{-r^2/2t}}{r}.$$

For fixed r this tends to zero (very fast) as $t \to 0$. In n-dimensions $\|y\| > \epsilon$ (in the Euclidean norm) implies that at least one of its coordinates y_i satisfies $|y_i| > \epsilon/\sqrt{n}$ so we find that

$$\mathrm{pr}_x(\{|\omega(t) - x| > \epsilon\}) \le c e^{-\epsilon^2/2nt}$$

for a suitable constant depending only on n. In particular, if we let $\rho(\epsilon, \delta)$ denote the supremum of the above probability over all $0 < t \le \delta$ then

$$\rho(\epsilon, \delta) = o(\delta). \tag{7}$$

Lemma 1 *Let* $0 \le t_1 \le \cdots \le t_m$ *with* $t_m - t_1 \le \delta$. *Let*

$$A := \{\omega |\ |\omega(t_j) - \omega(t_1)| > \epsilon \quad \text{for some } j = 1, \ldots m\}.$$

Then

$$\mathrm{pr}_x(A) \le 2\rho(\frac{1}{2}\epsilon, \delta) \tag{8}$$

independently of the number m *of steps.*

Proof. Let

$$B := \{\omega |\ |\omega(t_1) - \omega(t_m)| > \frac{1}{2}\epsilon\}$$

let

$$C_i := \{\omega |\ |\omega(t_i) - \omega(t_m)| > \frac{1}{2}\epsilon\}$$

and let

$$D_i = \{\omega |\ |\omega(t_1) - \omega(t_i)| > \epsilon \text{ and } |\omega(t_1) - \omega(t_k)| \le \epsilon\ k = 1, \ldots i-1\}.$$

If $\omega \in A$, then $\omega \in D_i$ for some i by the definition of A, by taking i to be the first j that works in the definition of A. If $\omega \notin B$ and $\omega \in D_i$ then $\omega \in C_i$ since it has to move a distance of at least $\frac{1}{2}\epsilon$ to get back from outside the ball of radius ϵ to inside the ball of radius $\frac{1}{2}\epsilon$.

$$B := \{\omega | \ |\omega(t_1) - \omega(t_m)| > \frac{1}{2}\epsilon\} \qquad C_i := \{\omega | \ |\omega(t_i) - \omega(t_m)| > \frac{1}{2}\epsilon\}$$

$$D_i = \{\omega | \ |\omega(t_1) - \omega(t_i)| > \epsilon \text{ and } |\omega(t_1) - \omega(t_k)| \leq \epsilon \ k = 1, \ldots i-1\}.$$

$$D_i := \{\omega | \ |\omega(t_1) - \omega(t_i)| > \epsilon \text{ and } |\omega(t_1) - \omega(t_k)| \leq \epsilon \ k = 1, \ldots i-1\}.$$

If $\omega \in A$, then $\omega \in D_i$ for some i by the definition of A, by taking i to be the first j that works in the definition of A. If $\omega \notin B$ and $\omega \in D_i$ then $\omega \in C_i$ since it has to move a distance of at least $\frac{1}{2}\epsilon$ to get back from outside the ball of radius ϵ to inside the ball of radius $\frac{1}{2}\epsilon$. So we have

$$A \subset B \cup \bigcup_{i=1}^{m} (C_i \cap D_i)$$

and hence

$$\mathrm{pr}_x(A) \leq \mathrm{pr}_x(B) + \sum_{i=1}^{m} \mathrm{pr}_x(C_i \cap D_i). \tag{9}$$

$$\operatorname{pr}_x(A) \le \operatorname{pr}_x(B) + \sum_{i=1} \operatorname{pr}_x(C_i \cap D_i). \tag{9}$$

Now we can estimate $\operatorname{pr}_x(C_i \cap D_i)$ as follows. For ω to belong to this intersection, we must have $\omega \in D_i$ and then the path moves a distance at least $\frac{\epsilon}{2}$ in time $t_n - t_i$ and these two events are independent, so $\operatorname{pr}_x(C_i \cap D_i) \le \rho(\frac{\epsilon}{2}, \delta) \operatorname{pr}_x(D_i)$. Here is this argument in more detail: Let

$$F = \mathbf{1}_{\{(y,z)|\ |y-z|>\frac{1}{2}\epsilon\}}$$

so that

$$\mathbf{1}_{C_i} = \phi_{F,t_i,t_n}.$$

Similarly, let G be the indicator function of the subset of $\dot{\mathbf{R}}^n \times \dot{\mathbf{R}}^n \times \cdots \times \dot{\mathbf{R}}^n$ (i copies) consisting of all points with

$$|x_k - x_1| \le \epsilon, \quad k = 1, \dots, i-1, \quad |x_1 - x_i| > \epsilon$$

so that

$$\mathbf{1}_{D_i} = \phi_{G,t_1,\dots,t_j}.$$

Then

$$\mathrm{pr}_x(C_i \cap D_i) =$$

$$\int \cdots \int p(x, x_1; t_1) \cdots p(x_{i-1}, x_i; t_i - t_{i-1}) F(x_1, \ldots, x_i) G(x_i, x_n) p(x_i, x_n) dx_1 \cdots dx_n.$$

The last integral (with respect to x_n) is $\leq \rho(\frac{1}{2}\epsilon, \delta)$. Thus

$$\mathrm{pr}_x(C_i \cap D_i) \leq \rho(\frac{\epsilon}{2}, \delta)\, \mathrm{pr}_x(D_i).$$

The D_i are disjoint by definition, so

$$\sum \mathrm{pr}_x(D_i) \leq \mathrm{pr}_x(\bigcup D_i) \leq 1.$$

So

$$\mathrm{pr}_x(A) \leq \mathrm{pr}_x(B) + \rho(\frac{1}{2}\epsilon, \delta) \leq 2\rho(\frac{1}{2}\epsilon, \delta).$$

QED

Let

$$E : \{\omega || \omega(t_i) - \omega(t_j)| > 2\epsilon \ \text{ for some } 1 \leq j < k \leq m\}.$$

Then $E \subset A$ since if $|\omega(t_j) - \omega(t_k)| > 2\epsilon$ then either $|\omega(t_1) - \omega(t_j)| > \epsilon$ or $|\omega(t_1) - \omega(t_k)| > \epsilon$ (or both). So

$$\mathrm{pr}_x(E) \leq 2\rho(\frac{1}{2}\epsilon, \delta). \tag{10}$$

Lemma 2 *Let* $0 \leq a < b$ *with* $b - a \leq \delta$. *Let*

$$E(a, b, \epsilon) := \{\omega | \ |\omega(s) - \omega(t)| > 2\epsilon \ \textit{ for some } \ s, t \in [a, b]\}.$$

Then

$$\mathrm{pr}_x(E(a, b.\epsilon)) \leq 2\rho(\frac{1}{2}\epsilon, \delta).$$

Proof. Here is where we are going to use the regularity of the measure. Let S denote a finite subset of $[a, b]$ and and let

$$E(a, b, \epsilon, S) := \{\omega | \ |\omega(s) - \omega(t)| > 2\epsilon \ \text{ for some } \ s, t \in S\}.$$

Then $E(a, b, \epsilon, S)$ is an open set and $\mathrm{pr}_x(E(a, b, \epsilon, S)) < 2\rho(\frac{1}{2}\epsilon, \delta)$ for any S. The union over all S of the $E(a, b, \epsilon, S)$ is $E(a, b, \epsilon)$. The regularity of the measure now implies the lemma. QED

Let k and n be integers, and set

$$\delta := \frac{1}{n}.$$

Let

$$F(k,\epsilon,\delta) := \{\omega|\ |\omega(t)-\omega(s)| > 4\epsilon \ \text{ for some } t,s \in [0,k],\ \text{with } |t-s| < \delta\}.$$

Then we claim that

$$\mathrm{pr}_x(F(k,\epsilon,\delta)) < 2k\frac{\rho(\frac{1}{2}\epsilon,\delta)}{\delta}. \tag{11}$$

Indeed, $[0,k]$ is the union of the $nk = k/\delta$ subintervals $[0,\delta],[\delta,2\delta],\ldots,[k-\delta,\delta]$. If $\omega \in F(k,\epsilon,\delta)$ then $|\omega(s)-\omega(t)| > 4\epsilon$ for some s and t which lie in either the same or in adjacent subintervals. So ω must lie in $E(a,b,\epsilon)$ for one of these subintervals, and there are kn of them. QED

Let $\omega \in \Omega$ be a continuous path in $\mathbf{R}^n$. Restricted to any interval $[0, k]$ it is uniformly continuous. This means that for any $\epsilon > 0$ it belongs to the complement of the set $F(k, \epsilon, \delta)$ for some δ. We can let $\epsilon = 1/p$ for some integer p. Let $\mathcal{C}$ denote the set of continuous paths from $[0, \infty)$ to $\mathbf{R}^n$. Then

$$\mathcal{C} = \bigcap_k \bigcap_\epsilon \bigcup_\delta F(k, \epsilon, \delta)^c$$

so the complement $\mathcal{C}^c$ of the set of continuous paths is

$$\bigcup_k \bigcup_\epsilon \bigcap_\delta F(k, \epsilon, \delta),$$

a countable union of sets of measure zero since

$$\mathrm{pr}_x \left(\bigcap_\delta F(k, \epsilon, \delta) \right) \leq \lim_{\delta \to 0} 2k\rho(\frac{1}{2}\epsilon, \delta)/\delta = 0.$$

We have thus proved a famous theorem of Wiener:

Theorem 1 [Wiener.] *The measure* pr_x *is concentrated on the space of continuous paths, i.e.* $\mathrm{pr}_x(\mathcal{C}) = 1$. *In particular, there is a probability measure on the space of continuous paths starting at the origin which assigns probability*

$$\mathrm{pr}_0(E) =$$

$$\int_{E_1} \cdots \int_{E_m} p(0, x_1; t_1) p(x_1, x_2; t_2 - t_1) \cdots p(x_{m-1}, x_m, t_m - t_{m-1}) dx_1 \ldots dx_m$$

to the set of all paths ω *which start at* 0 *and pass through the set* E_1 *at time* t_1, *the set* E_2 *at time* t_2 *etc. and we have denoted this set of paths by* E.

Norbert Wiener

Born: 26 Nov 1894 in Columbia, Missouri, USA
Died: 18 March 1964 in Stockholm, Sweden

Embedding in $\mathcal{S}'$.

For convenience in notation let me now specialize to the case $n = 1$. Let

$$\mathcal{W} \subset \mathcal{C}$$

consist of those paths ω with $\omega(0) = 0$ and

$$\int_0^\infty (1+t)^{-2} w(t) dt < \infty.$$

Proposition 1 **[Stroock]** *The Wiener measure* pr_0 *is concentrated on* $\mathcal{W}$.

Indeed, we let $E(|\omega(t)|)$ denote the expectation of the function $|\omega(t)|$ of ω with respect to Wiener measure, so

$$E(|\omega(t)|) = \frac{1}{\sqrt{2\pi t}} \int_{\mathbf{R}} |x| e^{-x^2/2t} dx = \frac{1}{\sqrt{2\pi t}} \cdot t \int_0^\infty \frac{x}{t} e^{-x^2/t} dx = Ct^{1/2}.$$

Thus, by Fubini,

$$E\left(\int_0^\infty (1+t)^{-2} |w(t)| dt \right) = \int_0^\infty (1+t)^{-2} E(|w(t)|) < \infty.$$

Hence the set of ω with $\int_0^\infty (1+t)^{-2} |w(t)| dt = \infty$ must have measure zero. QED

Now each element of $\mathcal{W}$ defines a tempered distribution, i.e. an element of $\mathcal{S}'$ according to the rule

$$\langle \omega, \phi \rangle = \int_0^\infty \omega(t)\phi(t)dt. \tag{12}$$

We claim that this map from $\mathcal{W}$ to $\mathcal{S}'$ is measurable and hence

the Wiener measure pushes forward to give a measure on $\mathcal{S}'$.

the Wiener measure pushes forward to give a measure on $\mathcal{S}'$.

To see this, let us first put a different topology (of uniform convergence) on $\mathcal{W}$. In other words, for each $\omega \in \mathcal{W}$ let $U_\epsilon(\omega)$ consist of all ω_1 such that

$$\sup_{t \geq 0} |\omega_1(t) - \omega(t)| < \epsilon,$$

and take these to form a basis for a topology on $\mathcal{W}$. Since we put the weak topology on $\mathcal{S}'$ it is clear that the map (12) is continuous relative to this new topology. So it will be sufficient to show that each set $U_\epsilon(\omega)$ is of the form $A \cap \mathcal{W}$ where A is in $\mathcal{B}(\Omega)$, the Borel field associated to the (product) topology on Ω.

So first consider the subsets $V_{n,\epsilon}(\omega)$ of $\mathcal{W}$ consisting of all $\omega_1 \in \mathcal{W}$ such that

$$\sup_{t \geq 0} |\omega_1(t) - \omega(t)| \leq \epsilon - \frac{1}{n}.$$

Clearly

$$U_\epsilon(\omega) = \bigcup_n V_{n,\epsilon}(\omega),$$

a countable union, so it is enough to show that each $V_{n,\epsilon}(\omega)$ is of the form $A_n \cap \mathcal{W}$ where $A_n \in \mathcal{B}(X)$. Now by the definition of the topology on Ω, if r is any real number, the set

$$A_{n,r} := \{\omega_1 | \ |\omega_1(r) - \omega(r)| \leq \epsilon - \frac{1}{n}$$

is closed.

the set

$$A_{n,r} := \{\omega_1 | \ |\omega_1(r) - \omega(r)| \leq \epsilon - \frac{1}{n}$$

is closed. So if we let r range over the non-negative rational numbers $\mathbb{Q}_+$, then

$$A_n = \bigcap_{r \in \mathbb{Q}_+} A_{n,r}$$

belongs to $\mathcal{B}(\Omega)$. But if ω_1 is continuous, then if $\omega_1 \in A_n$ then $\sup_{t \in \mathbb{R}_+} |\omega_1(t) - \omega(t)| \leq \epsilon - \frac{1}{n}$, and so

$$A_n \cap \mathcal{W} = V_{n,\epsilon}(\omega)$$

as was to be proved.

Stochastic processes and generalized stochastic processes.

In the standard probability literature a **stochastic process** is defined as follows: one is given an index set T and for each $t \in T$ one has a random variable $X(t)$. More precisely, one has some probability triple $(\Omega, \mathcal{F}, P)$ and for each $t \in T$ a real valued measurable function on $(\Omega, \mathcal{F})$. So a stochastic process $\mathbf{X}$ is just a collection $\mathbf{X} = \{X(t), t \in T\}$ of random variables. Usually $T = \mathbb{Z}$ or $\mathbb{Z}_+$ in which case we call $\mathbf{X}$ a **discrete time random process** or $T = \mathbb{R}$ or $\mathbb{R}_+$ in which case we call $\mathbf{X}$ a **continuous time random process**. Thus the word *process* means that we are thinking of T as representing the set of all times.

Brownian motion.

A realization of $\mathbf{X}$, that is the set $X(t)(\omega)$ for some $\omega \in \Omega$ is called a **sample path**. If T is one of the above choices, then $\mathbf{X}$ is said to have **independent increments** if for all $t_0 < t_1 < t_2 < \cdots < t_n$ the random variables

$$X(t_1) - X(t_0),\ X(t_1) - X(t_1), \ldots, X(t_n) - X(t_{n-1})$$

are independent.

For example, consider Wiener measure, and let $X(t)(\omega) = \omega(t)$ (say for $n = 1$). This is a continuous time stochastic process with independent increments which is known as (one dimensional) **Brownian motion**. The idea (due to Einstein) is that one has a small (visible) particle which, in any interval of time is subject to many random bombardments in either direction by small invisible particles (say molecules) so that the central limit theorem applies to tell us that the change in the position of the particle is Gaussian with mean zero and variance equal to the length of the interval.

Suppose that $\mathbf{X}$ is a continuous time random variable with the property that for almost all ω, the sample path $X(t)(\omega)$ is continuous. Let ϕ be a continuous function of compact support. Then the Riemann approximating sums to the integral

$$\int_T X(t)(\omega)\phi(t)dt$$

will converge for almost all ω and hence we get a random variable

$$\langle \mathbf{X}, \phi \rangle$$

where

$$\langle \mathbf{X}, \phi \rangle(\omega) = \int_T X(t)(\omega)\phi(t)dt,$$

the right hand side being defined (almost everywhere) as the limit of the Riemann approximating sums.

The same will be true if ϕ vanishes rapidly at infinity and the sample paths satisfy (a.e.) a slow growth condition such as given by Proposition 1 in addition to being continuous a.e.

The notation $\langle \mathbf{X}, \phi \rangle$ is justified since $\langle \mathbf{X}, \phi \rangle$ clearly depends linearly on ϕ.

The notation $\langle \mathbf{X}, \phi \rangle$ is justified since $\langle \mathbf{X}, \phi \rangle$ clearly depends linearly on ϕ.

But now we can make the following definition due to Gelfand. We may restrict ϕ further by requiring that ϕ belong to $\mathcal{D}$ or $\mathcal{S}$. We then consider a rule Z which assigns to each such ϕ a random variable which we might denote by $Z(\phi)$ or $\langle Z, \phi \rangle$ and which depends linearly on ϕ and satisfies appropriate continuity conditions. Such an object is called a **generalized random process**. The idea is that (just as in the case of generalized functions) we may not be able to evaluated $Z(t)$ at a given time t, but may be able to evaluate a "smeared out version" $Z(\phi)$.

Computational goal.

The purpose of the next few sections is to do the following computation: We wish to show that for the case Brownian motion, $\langle \mathbf{X}, \phi \rangle$ is a Gaussian random variable with mean zero and with variance

$$\int_0^\infty \int_0^\infty \min(s,t)\phi(s)\phi(t)dsdt.$$

First we need some results about Gaussian random variables.

Generalities about expectation and variance.

Let V be a vector space (say over the reals and finite dimensional). Let X be a V-valued random variable. That is, we have some measure space $(M, \mathcal{F}, \mu)$ (which will be fixed and hidden in this section) where μ is a probability measure on M, and $X : M \to V$ is a measurable function. If X is integrable, then

$$E(X) := \int X d\mu$$

is called the **expectation** of X and is an element of V.

The function $X \otimes X$ is a $V \otimes V$ valued function, and if it is integrable, then

$$\mathrm{Var}(X) = E(X \otimes X) - E(X) \otimes E(X) = E\left(X - E(X)\right) \otimes ((X - E(X))$$

is called the **variance** of X and is an element of $V \otimes V$. It is by its definition a symmetric tensor, and so can be thought of as a quadratic form on V^*.

Functoriality.

If $A : V \to W$ is a linear map, then AX is a W valued random variable, and

$$E(AX) = AE(X), \qquad \mathrm{Var}(AX) = (A \otimes A)\,\mathrm{Var}(X) \tag{12}$$

assuming that $E(X)$ and $\mathrm{Var}(X)$ exist. We can also write this last equation as

$$\mathrm{Var}(AX)(\eta) = \mathrm{Var}(X)(A^*\eta), \qquad \eta \in W^* \tag{13}$$

if we think of the variance as quadratic function on the dual space.

The characteristic function.

The function on V^* given by

$$\xi \mapsto E(e^{i\xi \cdot X})$$

is called the **characteristic function** associated to X and is denoted by ϕ_X. Here we have used the notation $\xi \cdot v$ to denote the value of $\xi \in V^*$ on $v \in V$. It is a version of the Fourier transform (with the conventions used by the probabilists). More precisely, let $X_*\mu$ denote the push forward of the measure μ by the map X, so that $X_*\mu$ is a probability measure on V. Then ϕ_X is the Fourier transform of this measure except that there are no powers of 2π in front of the integral and a plus rather than a minus sign is before the i in the exponent. These are the conventions of the probabilists. What is important for us is the fact that the Fourier transform determines the measure, i.e. ϕ_X determines $X_*\mu$. The probabilists would say that the *law* of the random variable (meaning $X_*\mu$) is determined by its characteristic function.

$$\mathrm{Var}(AX)(\eta) = \mathrm{Var}(X)(A^*\eta), \quad \eta \in W^* \tag{13}$$

To get a feeling for (13) consider the case where $A = \xi$ is a linear map from V to $\mathbf{R}$. Then $\mathrm{Var}(X)(\xi) = \mathrm{Var}(\xi \cdot X)$ is the usual variance of the scalar valued random variable $\xi \cdot X$. Thus we see that $\mathrm{Var}(X)(\xi) \geq 0$, so $\mathrm{Var}(X)$ is non-negative definite symmetric bilinear form on V^*. The variance of a scalar valued random variable vanishes if and only if it is a constant. Thus $\mathrm{Var}(X)$ is positive definite unless X is concentrated on hyperplane.

Suppose that $A : V \to W$ is an isomorphism, and that $X_*\mu$ is absolutely continuous with respect to Lebesgue measure, so

$$X_*\mu = \rho dv$$

where ρ is some function on V (called the probability density of X). Then $(AX)_*\mu$ is absolutely continuous with respect to Lebesgue measure on W and its density σ is given by

$$\sigma(w) = \rho(A^{-1}w)|\det A|^{-1} \tag{14}$$

as follows from the change of variables formula for multiple integrals.

Gaussian measures and their variances.

Let d be a positive integer. We say that N is a **unit** (d-dimensional) **Gaussian random variable** if N is a random variable with values in $\mathbf{R}^d$ with density

$$(2\pi)^{-d/2}e^{-(x_1^2\cdots+x_d^2)/2}.$$

It is clear that $E(N)=0$ and, since

$$(2\pi)^{-d/2}\int x_i x_j e^{-(x_1^2\cdots+x_d^2)/2}dx=\delta_{ij},$$

that

$$\mathrm{Var}(N)=\sum_i \delta_i\otimes\delta_i \tag{15}$$

where $\delta_1,\ldots,\delta_d$ is the standard basis of $\mathbf{R}^d$. We will sometimes denote this tensor by I_d. In general we have the identification $V\otimes V$ with $\mathrm{Hom}(V^*,V)$, so we can think of the $\mathrm{Var}(X)$ as an element of $\mathrm{Hom}(V^*,V)$ if X is a V-valued random variable. If we identify $\mathbf{R}^d$ with its dual space using the standard basis, then I_d can be thought of as the identity matrix.

A V-valued random variable X is called **Gaussian** if (it is equal in law to a random variable of the form)

$$AN + a$$

where

$$A : \mathbf{R}^d \to V$$

is a linear map, where $a \in V$, and where N is a unit Gaussian random variable. Clearly

$$E(X) = a,$$

$$\mathrm{Var}(X) = (A \otimes A)(I_d)$$

$$\mathrm{Var}(X) = (A \otimes A)(I_d)$$

or, put another way,

$$\mathrm{Var}(X)(\xi) = I_d(A^*\xi)$$

and hence

$$\phi_X(\xi) = \phi_N(A^*\xi)e^{i\xi\cdot a} = e^{-\frac{1}{2}I_d(A^*\xi)}e^{i\xi\cdot a}$$

or

$$\phi_X(\xi) = e^{-\mathrm{Var}(X)(\xi)/2 + i\xi\cdot E(X)}. \tag{17}$$

It is a bit of a nuisance to carry along the $E(X)$ in all the computations, so we shall restrict ourselves to **centered Gaussian** random variables meaning that $E(X) = 0$. Thus for a centered Gaussian random variable we have

$$\phi_X(\xi) = e^{-\operatorname{Var}(X)(\xi)/2}. \tag{18}$$

Conversely, suppose that X is a V valued random variable whose characteristic function is of the form

$$\phi_X(\xi) = e^{-Q(\xi)/2},$$

where Q is a quadratic form. Since $|\phi_X(\xi)| \leq 1$ we see that Q must be non-negative definite. Suppose that we have chosen a basis of V so that V is identified with $\mathbf{R}^q$ where $q = \dim\ V$. By the principal axis theorem we can always find an orthogonal transformation (c_{ij}) which brings Q to diagonal form. In other words, if we set

$$\eta_j := \sum_i c_{ij}\xi_i$$

then

$$Q(\xi) = \sum_j \lambda_j \eta_j^2.$$

The λ_j are all non-negative since Q is non-negative definite. So if we set

$$a_{ij} := \lambda_j^{\frac{1}{2}} c_{ij}, \text{ and } A = (a_{ij})$$

we find that $Q(\xi) = I_q(A^*\xi)$. Hence X has the same characteristic function as a Gaussian random variable hence must be Gaussian.

As a corollary to this argument we see that

A random variable X is centered Gaussian if and only if $\xi \cdot X$ is a real valued Gaussian random variable with mean zero for each $\xi \in V^*$.

The variance of a Gaussian with density.

$$\sigma(w) = \rho(A^{-1}w)|\det A|^{-1} \tag{14}$$

In our definition of a centered Gaussian random variable we were careful not to demand that the map A be an isomorphism. For example, if A were the zero map then we would end up with the δ function (at the origin for centered Gaussians) which (for reasons of passing to the limit) we want to consider as a Gaussian random variable.

But suppose that A *is* an isomorphism. Then by (14), X will have a density which is proportional to

$$e^{-S(v)/2}$$

where S is the quadratic form on V given by

$$S(v) = J_d(A^{-1}v)$$

and J_d is the unit quadratic form on $\mathbf{R}^d$:

$$J_d(x) = x_1^2 \cdots + x_d^2$$

or, in terms of the basis $\{\delta_i^*\}$ of the dual space to $\mathbf{R}^d$,

$$J_d = \sum_i \delta_i^* \otimes \delta_i^*.$$

$$J_d = \sum_i \delta_i^* \otimes \delta_i^*.$$

Here $J_d \in (\mathbf{R}^d)^* \otimes (\mathbf{R}^d)^* = \mathrm{Hom}(\mathbf{R}^d, (\mathbf{R}^d)^*)$. It is the inverse of the map I_d. We can regard S as belonging to $\mathrm{Hom}(V, V^*)$ while we also regard $\mathrm{Var}(X) = (A \otimes A) \circ I_d$ as an element of $\mathrm{Hom}(V^*, V)$. I claim that $\mathrm{Var}(X)$ and S are inverses to one another. Indeed, dropping the subscript d which is fixed in this computation, $\mathrm{Var}(X)(\xi, \eta) = I(A^*\xi, A^*\eta) = \eta \cdot (A \circ I \circ A^*)$ when thought of as a bilinear form on $V^* \otimes V^*$, and hence

$$\mathrm{Var}(X) = A \circ I \circ A^*$$

when thought of as an element of $\mathrm{Hom}(V^*, V)$. Similarly thinking of S as a bilinear form on V we have $S(v, w) = J(A^{-1}v, A^{-1}w) = J(A^{-1}v) \cdot A^{-1}w$ so

$$S = A^{-1*} \circ J \circ A^{-1}$$

when S is thought of as an element of $\mathrm{Hom}(V, V^*)$. Since I and J are inverses of one another, the two above displayed expressions for S and $\mathrm{Var}(X)$ show that these are inverses on one another.

A computational tool.

This has the following very important computational consequence:

Suppose we are given a random variable X with (whose law has) a density proportional to $e^{-S(v)/2}$ where S is a quadratic form which is given as a "matrix" $S = (S_{ij})$ in terms of a basis of V^*. Then $\mathrm{Var}(X)$ is given by S^{-1} in terms of the dual basis of V.

The variance of Brownian motion.

For example, consider the two dimensional vector space with coordinates (x_1, x_2) and probability density proportional to

$$\exp -\frac{1}{2}\left(\frac{x_1^2}{s} + \frac{(x_2 - x_1)^2}{t-s}\right)$$

where $0 < s < t$. This corresponds to the matrix

$$\begin{pmatrix} \frac{t}{s(t-s)} & -\frac{1}{t-s} \\ -\frac{1}{t-s} & \frac{1}{t-s} \end{pmatrix} = \frac{1}{t-s}\begin{pmatrix} \frac{t}{s} & -1 \\ -1 & 1 \end{pmatrix}$$

whose inverse is

$$\begin{pmatrix} s & s \\ s & t \end{pmatrix}$$

which thus gives the variance.

So, if we let

$$B(s,t) := \min(s,t) \tag{19}$$

we can write the above variance as

$$\begin{pmatrix} B(s,s) & B(s,t) \\ B(t,s) & B(t,t) \end{pmatrix}.$$

Now suppose that we have picked some finite set of times $0 < s_1 < \cdots < s_n$ and we consider the corresponding Gaussian measure given by our formula for Brownian motion on a one-dimensional space for a path starting at the origin and passing successively through the points x_1 at time s_1, x_2 at time s_2 etc. We can compute the variance of this Gaussian to be

$$(B(s_i, s_j))$$

since the projection onto any coordinate plane (i.e. restricting to two values s_i and s_j) must have the variance given above.

Let $\phi \in \mathcal{S}$. We can think of ϕ as a (continuous) linear function on $\mathcal{S}'$. For convenience let us consider the real spaces $\mathcal{S}$ and $\mathcal{S}'$, so ϕ is a real valued linear function on $\mathcal{S}'$. Applied to Stroock's version of Brownian motion which is a probability measure living on $\mathcal{S}'$ we see that ϕ gives a real valued random variable. Recall that this was given by integrating $\phi \cdot \omega$ where ω is a continuous path of slow growth, and then integrating over Wiener measure on paths.

This is the limit of the Gaussian random variables given by the Riemann approximating sums

$$\frac{1}{n}(\phi(s_1)x_1 + \cdots + \phi(s_{n^2})x_{n^2})$$

where $s_k = k/n$, $k = 1, \ldots, n^2$, and $(x_1, \ldots, x_{n^2})$ is an n^2 dimensional centered Gaussian random variable whose variance is $(\min(s_i, s_j))$.

This is the limit of the Gaussian random variables given by the Riemann approximating sums

$$\frac{1}{n}(\phi(s_1)x_1 + \cdots + \phi(s_{n^2})x_{n^2})$$

where $s_k = k/n$, $k = 1, \ldots, n^2$, and $(x_1, \ldots, x_{n^2})$ is an n^2 dimensional centered Gaussian random variable whose variance is $(\min(s_i, s_j))$. Hence this Riemann approximating sum is a one dimensional centered Gaussian random variable whose variance is $\dfrac{1}{n^2}\displaystyle\sum_{i,j} \min(s_i, s_j)\phi(s_i)\phi(s_j)$.

Passing to the limit we see that integrating $\phi \cdot \omega$ defines a real valued centered Gaussian random variable whose variance is

$$\int_0^\infty \int_0^\infty \min(s,t)\phi(s)\phi(t)dsdt = 2\int_0^\infty \int_{0\le s\le t} s\phi(s)\phi(t)dsdt, \tag{21}$$

Let us say that a probability measure μ on $\mathcal{S}'$ is a **centered Gaussian process** if every $\phi \in \mathcal{S}$, thought of as a function on the probability space $(\mathcal{S}', \mu)$ is a real valued centered Gaussian random variable; in other words $\phi_*(\mu)$ is a centered Gaussian probability measure on the real line. If we denote this process by Z, then we may write $Z(\phi)$ for the random variable given by ϕ. We clearly have $Z(a\phi + b\psi) = aZ(\phi) + bZ(\psi)$ in the sense of addition of random variables, and so we may think of Z as a rule which assigns, in a linear fashion, random variables to elements of $\mathcal{S}$. With some slight modification (we, following Stroock, are using $\mathcal{S}$ instead of $\mathcal{D}$ as our space of test functions) this notion was introduced by Gelfand some fifty years ago. (See Gelfand and Vilenkin, *Generalized Functions* volume IV.)

If we have generalized random process Z as above, we can consider its derivative in the sense of generalized functions, i.e.

$$\dot{Z}(\phi) := Z(-\dot{\phi}).$$

The derivative of Brownian motion is white noise.

To see how this derivative works, let us consider what happens for Brownian motion. Let ω be a continuous path of slow growth, and set

$$\omega_h(t) := \frac{1}{h}(\omega(t+h) - \omega(t)).$$

The paths ω are not differentiable (with probability one) so this limit does not exist as a function. But the limit does exist as a generalized function, assigning the value

$$\int_0^\infty -\dot{\phi}(t)\omega(t)dt$$

to ϕ. Now if $s < t$ the random variables $\omega(t+h) - \omega(t)$ and $\omega(s+h) - \omega(s)$ are independent of one another when $h < t - s$ since Brownian motion has independent increments. Hence we expect that this limiting process be independent at all points in some generalized sense. (No actual, as opposed to generalized, process can have this property. We will see more of this point in a moment when we compute the variance of $\dot{Z}$.)

$$\int_0^\infty \int_0^\infty \min(s,t)\phi(s)\phi(t)dsdt = 2\int\int_{0\le s\le t} s\phi(s)\phi(t)dsdt. \tag{20}$$

$$\dot{Z}(\phi) := Z(-\dot{\phi}).$$

In any event, $\dot{Z}(\phi)$ is a centered Gaussian random variable whose variance is given (according to (20)) by

$$2\int_0^\infty \left(\int_0^t s\dot{\phi}(s)ds\right)\dot{\phi}(t)dt.$$

We can integrate the inner integral by parts to obtain

$$\int_0^t s\dot{\phi}(s)ds = t\phi(t) - \int_0^t \phi(s)ds.$$

In any event, $Z(\phi)$ is a centered Gaussian random variable whose variance is given (according to (20)) by

$$2\int_0^\infty \left(\int_0^t s\dot{\phi}(s)ds\right)\dot{\phi}(t)dt.$$

$$\int_0^t s\dot{\phi}(s)ds = t\phi(t) - \int_0^t \phi(s)ds.$$

Integration by parts now yields

$$\int_0^\infty t\phi(t)\dot{\phi}(t)dt = -\frac{1}{2}\int_0^\infty \phi(t)^2dt$$

and

$$-\int_0^\infty \left(\int_0^t \phi(s)ds\right)\dot{\phi}(t)dt = \int_0^\infty \phi(t)^2dt.$$

We conclude that the variance of $\dot{Z}(\phi)$ is given by

$$\int_0^\infty \phi(t)^2 dt$$

which we can write as

$$\int_0^\infty \int_0^\infty \delta(s-t)\phi(s)\phi(t)dsdt.$$

Notice that now the "covariance function" is the generalized function $\delta(s-t)$. The generalized process (extended to the whole line) with this covariance is called white noise because it is a Gaussian process which is stationary under translations in time and its covariance "function" is $\delta(s-t)$, signifying independent variation at all times, and the Fourier transform of the delta function is a constant, i.e. assigns equal weight to all frequencies.