Advanced Probability

==
(2nd November, Friday)

Chapter 5. Weak Convergence

5.1. Definitions

Let E be a metric space. Whenever we are talking about a metric space, the σ-algebra is given by the Borel σ-algebra. Write $C_b(E)$ for the set of bounded continuous functions on E.

- Let $(\mu_n : n \in \mathbb{N})$ be a sequence of probability measures and let μ be another probability measure on E. We say that $\mu_n \to \mu$ **weakly** (as $n \to \infty$) if $\mu_n(f) \to \mu(f)$ for all $f \in C_b(\mathbb{R})$.

Theorem 5.1.1) The following are equivalent.

(a) $\mu_n \to \mu$ weakly on E

(b) $\liminf_{n\to\infty} \mu_n(U) \geq \mu(U)$ for all U open

(c) $\limsup_{\mu(F)} \leq \mu(F)$ for all F closed.

(d) $\mu_n(B) \to \mu(B)$ for all $B \in \mathcal{B}$ such that $\mu(\partial B) = 0$.(Boundary is the set of limit points of B that are not contained in B.)

proof) Exercise.

For an example, consider a sequence $(x_n)_n \subset \mathbb{R}$ such that $x_n \to 0$ as $n \to \infty$. We want to have $\delta_{x_n} \to \delta_0$. Indeed, this is true in the weak sense. However, the sequence has $\delta_{x_n}(\{0\}) = 0$ for all n, hence we should have inequality in condition (c).

We have a similar version of the theorem for the real line.

Proposition 5.1.2) Consider the case $E = \mathbb{R}$. TFAE

(a) $\mu_n \to \mu$ weakly for some probability measure μ.

(b) $F_n(x) \to F(x)$ for all $x \in \mathbb{R}$ such that $F(x^-) = F(x)$. (Here, $F(x) = \mu((\infty, x])$ is the **distribution function** of μ.) (Sometimes called convergence of distributions)

(c) There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random variables X_n, X on Ω such that $X_n \sim \mu_n$, $X \sim \mu$ and $X_n \to X$ almost surely.

proof) See probability and measure notes.

5.2. Prohorov's Theorem

When does a sequence of probability measures has a converging subsequence?

Let E be a metric space and $(\mu_n : n \in \mathbb{N})$ be a sequence of probability measures on E.

- We say that $(\mu_n)_n$ is **tight** if for all $\epsilon > 0$, there is a compact set $K \subset E$ such that

$$\mu_n(E \backslash K) \le \epsilon \quad \forall n \in \mathbb{N}$$

For example, the sequence $(\delta_n)_n$ is *not* tight.

Theorem 5.2.1) Let $(\mu_n : n \in \mathbb{N})$ be a sequence of probability measures on a metric space E and suppose that $(\mu_n : n \in \mathbb{N})$ is tight. Then there exists a subsequence $(n_k)_k \subset \mathbb{N}$ and probability measure μ on E such that $\mu_{n_k} \to \mu$ weakly as $k \to \infty$.

This gives a version of weakly sequential compactness of probability measures. We are only going to prove this for $\mathbb{R}$. This theorem is hard to prove in general.(e.g. there is a method using Monge-Kantorovich metric defined for Polish spaces. For this method, see "Topics in Optimal Transport", C.Villani, Ame.Soc.Math. For the general version, see the attached note)

proof for $E = \mathbb{R}$) By a diagonal argument and by passing to a subsequence, it suffices to consider the case where $F_n(x) \to g(x)$ as $n \to \infty$ for all $x \in \mathbb{Q}$ for some $g(x) \in [0,1]$, where F_n is the distribution function of F_n. Now $g : \mathbb{Q} \to [0,1]$ is non-decreasing so g has a non-decreasing extension $G : \mathbb{R} \to [0,1]$, i.e.

$$G(x) = \lim_{q \searrow x, q \in \mathbb{Q}} g(q)$$

which has only countably many discontinuities.(because there should be a rational number in each discontinuity). Now we must have

$$F_n(x) \to G(x) \quad \forall x \text{ s.t. } G \text{is continuous at } x$$

Set $F(x) = G(x^+)$, then F and G have same points of continuity, so $F_n(x) \to F(x)$ for all $x \in \mathbb{R}$.

We are only left to check that $G(x) \to 1$ as $x \to \infty$ using tightness condition.

Since $(\mu_n : n \in \mathbb{N})$ is tight, given $\epsilon > 0$, there exists $R < \infty$ such that $\mu_n(\mathbb{R} \backslash (-R, R)) \le \epsilon$ for all ϵ so $F_n(-R) \le \epsilon$, $F_n(R) \ge 1 - \epsilon$. So

$$\begin{aligned} F(x) &\to 0 \quad \text{as } x \to -\infty \\ F(x) &\to 1 \quad \text{as } x \to \infty \end{aligned}$$

So F is distribution function. So there exists a probability measure μ such that $\mu((-\infty, x]) = F(x)$. Then $\mu_n \to \mu$ by **Prop 5.1.2.**

(End of proof) □

5.3. Weak Convergence and Characteristic Functions

Take $E = \mathbb{R}^d$.

- For a probability measures mu on $\mathbb{R}^d$, define its **characteristic function** $\phi : \mathbb{R}^d \to \mathbb{C}$ by

$$\phi(u) = \int_{\mathbb{R}^d} e^{i\langle u, x\rangle} \mu(dx)$$

Lemma 5.3.1) Fix $d = 1$. For all $\lambda \in (0, \infty)$,

$$\mu(\mathbb{R} \backslash (-\lambda, \lambda)) \le C\lambda \int_0^\lambda (1 - \mathrm{Re}(\phi(u)))du$$

where $C = (1 - \sin(1))^{-1} < \infty$.

proof) Consider for $t \geq 1$. Let $A(t) = t^{-1}\int_0^t (1-\cos v)dv$. Then

$$A(t) \geq A(0) = 1 - \sin(t)$$

(to see this, observe that $A(t)$ is the average of $(1-\cos(v))$ on interval $(0,t)$ and divide the cases $|t| \leq \pi/2$ and $|t| \geq \pi/2$)
So $Ct^{-1}\int_0^t (1-\cos(v))dv \geq 1$. Substitute $v = uy$, $u = v/y$,

$$Ct^{-1}\int_0^{t/y} (1-\cos(uy))ydu \geq 1$$

Put $t/y = 1/\lambda$, $\lambda = y/t$, $t = y/\lambda \geq 1$ to see

$$C\lambda \int_0^{1/\lambda} (1-\cos(uy))du \geq 1$$

whenever $t = y/\lambda \geq 1$(this was the assumption we started with). Now for general $y \in \mathbb{R}$, has

$$C\lambda \int_0^{1/\lambda} (1-\cos(uy))du \geq 1_{|y|\geq\lambda}$$

Now integrate with respect to μ and use Fubini.

$$\begin{aligned}\mu(\mathbb{R}\backslash(-\lambda,\lambda)) &\leq C\lambda \int_{\mathbb{R}}\int_0^{1/\lambda} (1-\cos(uy))du\mu(dy) \\ &= C\lambda \int_0^{1/\lambda}\int_{\mathbb{R}} (1-\cos(uy))du\mu(dy)\end{aligned}$$

(End of proof) □

==

(5th November, Monday)

Theorem 5.3.2) Let μ_n, μ be probability measures on $\mathbb{R}^d$ with characteristic functions ϕ_n, ϕ. Then the following are equivalent

(a) $\mu_n \to \mu$ weakly on $\mathbb{R}^d$.

(b) $\phi_n(u) \to \phi(u)$ for all $u \in \mathbb{R}^d$.

We will prove only for the case $d = 1$.

proof) It is clear that (a) implies (b). Suppose (b) holds. We prove via a 'compactness argument'. We aim to show that the sequence $(\mu_n)_n$ tight, and therefore has a converging subsequence, and show that the converging point is in fact μ.

Note that $\phi(0) = 1$ and ϕ is continuous. Given $\epsilon > 0$, there exists $\lambda < \infty$ such that

$$C\lambda \int_0^{1/\lambda} (1-\mathrm{Re}(\phi(u)))du \leq \epsilon/2$$

with $C = (1-\sin(1))^{-1} < \infty$. By dominated convergence,

$$\int_0^{1/\lambda} (1-\mathrm{Re}(\phi_n(u)))du \xrightarrow{n\to\infty} \int_0^{1/\lambda} (1-\mathrm{Re}(\phi(u)))du$$

so for sufficiently large n, by **Lemma 5.3.1,**,

$$\mu_n(\mathbb{R}\backslash(-\lambda,\lambda)) \leq C\lambda \int_0^{1/\lambda} (1-\mathrm{Re}(\phi_n(u)))du \leq \epsilon$$

Since ϵ was arbitrary, we see that $(\mu_n : n \in \mathbb{N})$ is tight. By Prohorov's theorem, we have a converging subsequence $\mu_{n_k} \to \nu$ for some probability measure ν.

Suppose for a contradiction that $\nu \neq \mu$. Therefore, there exists $\epsilon > 0$, and $f \in C_b(\mathbb{R}^n)$ such that

$$|\mu_{n_k}(f) - \mu(f)| \geq \epsilon \quad \forall k$$

By above argument, we have $\mu_{n_k} \to \nu$. But then, since e^{inx} is a bounded continuous function,

$$\int_{\mathbb{R}} e^{inx} \nu(dx) = \lim_{k \to \infty} \phi_{n_k}(n) = \phi(n)$$

which indicates $\mu = \nu$ by uniqueness of characteristic functions (see PM notes), a contradiction.

(End of proof) $\square$

In fact, the proof of the theorem implies a slightly stronger statement, which is less useful.

Theorem 5.3.3) *(Lévy's continuity theorem for characteristic functions)* Let $(\mu_n : n \in \mathbb{N})$ be a sequence of probability measures on $\mathbb{R}^n$ with characteristic functions ϕ_n. Suppose $\phi_n(u) \to \phi(u)$ for all u for some function ϕ (not necessarily a characteristic function) such that ϕ is continuous at 0. Then ϕ is the characteristic function of some probability measure μ on $\mathbb{R}^d$ and $\mu_n \to \mu$ weakly on $\mathbb{R}^d$.

6. Large Deviations

6.1. Cramérs theorem

Theorem 6.1.1) Let $(X_n : n \in \mathbb{N})$ be a sequence of integrable *i.i.d.* random variables in $\mathbb{R}$. Set $m = \mathbb{E}(X_1)$, $S_n = X_1 + \cdots + X_n$. We know $S_n/n \to \delta_m$ in probability, so if $(m - \epsilon, m + \epsilon) \cap B = \phi$ then $\mathbb{P}(S_n/n \in B) \to 0$ as $n \to \infty$. Then in fact the convergence rate is given by $\sim \exp(-n\alpha(B))$ for some α. To be precise, for all $a \geq m = \mathbb{E}(X_1)$, as $n \to \infty$,

$$\frac{1}{n} \log \mathbb{P}(S_n \geq na) \to -\psi^*(a)$$

where ψ^* is the *Legendre transform* of the *cumulant generating function* $\psi(\lambda) = \log(\mathbb{E}(e^{\lambda X_1}))$, where Legendre transform is given by

$$\psi^*(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x - \psi(\lambda)\}$$

In particular, for n sufficiently large and in case $\psi^*(a) < \infty$, we get

$$-\psi^*(a) - \epsilon \leq \frac{1}{n} \log(\mathbb{P}(S_n \geq a)) \leq -\psi^*(a) + \epsilon$$

and therefore

$$e^{-n(\psi^*(a)+\epsilon)} \leq \mathbb{P}(S_n \geq na) \leq e^{-n(\psi^*(a)-\epsilon)}.$$

Note : ψ is always a convex function, so ψ^* is also a convex function.

Examples :

(i) $X_1 \sim N(0,1)$, then $\mathbb{E}(e^{\lambda X_1}) = e^{\lambda^2/2}$, $\psi(\lambda) = \lambda^2/2$ and $\psi^*(x) = x^2/2$. Hence

$$\frac{1}{n} \log(\mathbb{P}(S_n \geq a)) \to -\frac{a^2}{2} \quad \forall a \geq 0$$

Can check this directly, using the fact that $S_n \sim N(0,n)$ in this case.

(ii) $X_1 \sim \text{Exp}(1)$, then

$$\mathbb{E}(e^{\lambda X_1}) = \int_0^\infty e^{\lambda x} e^{-x} dx = \begin{cases} \infty & \text{if } \lambda \geq 1 \\ \frac{1}{1-\lambda} & \text{if } \lambda < 1 \end{cases}$$

so $\psi(\lambda) = -\log(1-\lambda)$ if $\lambda < 1$ and ∞ otherwise, and $\psi^*(x) = x - 1 - \log(x)$ for $x > 0$. Cramér's theorem implies that

$$\frac{1}{n}\log \mathbb{P}(S_n \geq na) \to -(a - 1 - \log(a)) \quad \forall a \geq 1$$

On the other hand, $\mathrm{Var}(X_1) = 1 < \infty$, so $\frac{S_n - n}{\sqrt{n}} \to N(0,1)$ by CLT. So

$$\mathbb{P}(S_n \geq n + a\sqrt{n}) \to \int_a^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

so Cramér's theorem gives a result of a different flavour from CLT for distributions with bounded variation : while CLT provides a description for distribution near the average, Cramér gives an explanation of tail distribution of S_n.

preparation for proof of Cramér's theorem) Let $\mu(B) = \mathbb{P}(X_1 \in B)$. Exclude the easy case where $\mu = \delta_m$. Define for $\lambda \geq 0$ with $\psi(\lambda) < \infty$, the **tilted distribution** μ_λ by

$$\mu_\lambda(dx) \propto e^{\lambda x}\mu(dx)$$

For $K \geq m = \mathbb{E}(X_1)$, define the conditional distribution by

$$\mu_K(dx|x \leq K) \propto 1_{\{x \leq K\}}\mu(dx)$$

The CGF(cumulant generating function) of μ_K is then given by

$$\psi_K(\lambda) = \log(\mathbb{E}(e^{\lambda X_1}|X_1 \leq K))$$

==

(7th November, Wednesday)

We now start proving the following theorem.

Theorem 6.1.1) Let $(X_n : n \in \mathbb{N})$ be a sequence of integrable *i.i.d.* random variables in $\mathbb{R}$. Set $m = \mathbb{E}(X_1)$, $S_n = X_1 + \cdots + X_n$. Then for all $a \geq m = \mathbb{E}(X_1)$, as $n \to \infty$,

$$\frac{1}{n}\log \mathbb{P}(S_n \geq na) \to -\psi^*(a)$$

where $\psi(\lambda) = \log(\mathbb{E}(e^{\lambda X_1}))$, and $\psi^*(x) = \sup_{\lambda \in \mathbb{R}}\{\lambda x - \psi(\lambda)\}$.

proof) *(Upper bound)* For all $\lambda \geq 0$ and $n \geq 1$

$$\mathbb{P}(S_n \geq na) = \mathbb{P}(e^{\lambda S_n} \geq e^{\lambda na}) \leq e^{-\lambda na}\mathbb{E}(e^{\lambda S_n}) = e^{-(\lambda a - \psi(\lambda))n}$$

so $\frac{1}{n}\log \mathbb{P}(S_n \geq na) \leq -(\lambda a - \psi(\lambda))$ and

$$\frac{1}{n}\log \mathbb{P}(S_n \geq na) \leq -\psi^*(a)$$

(Lower bound) It remains to show the lower bound. That is, we aim to prove

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}(S_n \geq na) \geq -\psi^*(a)$$

Consider first the case where $\mathbb{P}(X_1 \leq a) = 1$. Then

$$\mathbb{E}(e^{\lambda(X_1 - a)}) \xrightarrow{\lambda\to\infty} \mathbb{P}(X_1 = a)$$

Call $p = \mathbb{P}(X_1 = a)$, so $\lambda a - \psi(\lambda) \to -\log(p)$. So in particular,

$$\psi^*(a) \geq -\log(p)$$

Now $\mathbb{P}(S_n \geq na) = p^n$ so

$$\frac{1}{n}\log \mathbb{P}(S_n \geq na) = \log(p) \geq -\psi^*(a)$$

hence we can eliminate the case $\mathbb{P}(X_1 \leq a) = 1$.

Next consider the case $\psi(\lambda) < \infty$ for all $\lambda \geq 0$ and $\mathbb{P}(X_1 > a) > 0$. Fix $\epsilon > 0$ and set $b = a + \epsilon$, $c = a + 2\epsilon$, choosing ϵ small enough so $\mathbb{P}(X_1 > b) > 0$. Then there exists λ such that $\psi'(\lambda) = b$ - where the differentiability and the existence is justified in the following proposition :

Proposition 6.1.2) Suppose X is integrable and not a.s. constant. Then

$$\psi_K(\lambda) = \log \mathbb{E}(e^{\lambda X_1} | X_1 \leq K) < \infty \quad \forall K < \infty$$
$$\textit{and} \quad \psi_K(\lambda) \nearrow \psi(\lambda) \quad \text{as } K \to \infty$$

Moreover in the case $\psi(\lambda) < \infty$ for all $\lambda \geq 0$, ψ has a continuous derivative on $[0, \infty)$ and is C^2 on $(0, \infty)$ with

$$\psi'(\lambda) = \int_{\mathbb{R}} x\mu_\lambda(dx)$$
$$\psi''(\lambda) = \mathrm{Var}(\mu_\lambda) > 0$$

and ψ' is a homeomorphism from $[0, \infty)$ to $[m, \sup(\mathrm{supp}(\mu))$.
proof) (Exercise)

Now we use the idea of tilting the probability measure. Define a new probability measure $\mathbb{P}_\lambda$ by $d\mathbb{P}_\lambda = e^{\lambda S_n - n\psi(\lambda)} d\mathbb{P}$. Then observe that under $\mathbb{P}_\lambda$ the random variables $X_1, \cdots, X_n$ are independent with distributions μ_λ and that $\mathbb{E}_\lambda(X_1) = b$. Consider the event

$$A_n = \{\left|\frac{S_n}{n} - b\right| \leq \epsilon\} = \{(b - \epsilon)n = an \leq S_n \leq (b + \epsilon)n = cn\}$$

By the weak law of large numbers, $\mathbb{P}_\lambda(A_n) \to 1$. So

$$\begin{aligned}\mathbb{P}(S_n \geq na) \geq \mathbb{P}(A_n) &= \mathbb{E}_\lambda\Big(1_{A_n} e^{-\lambda S_n + \psi(\lambda)n}\Big) \\ &\geq e^{-\lambda cn + \psi(\lambda)n} \mathbb{P}_\lambda(A_n)\end{aligned}$$

So

$$\frac{1}{n}\log \mathbb{P}(S_n \geq na) \geq -\lambda c + \psi(\lambda) + \frac{\log(\mathbb{P}_\lambda(A_n))}{n}$$

and

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}(S_n \geq na) \geq -(\lambda c - \psi(\lambda)) \geq -\psi^*(c)$$

Now ψ^* is continuous at a (recall, ψ^* is a Legendre transform of a convex function so is convex, and therefore continuous. Or, see **Lemma 6.1.3**) and $\epsilon > 0$ is arbitrary so the desired lower bound follows on letting $\epsilon \to 0$.

Finally, consider the general case $\mathbb{P}(X_1 > a) > 0$ but allowing $\psi(\lambda) = \infty$ for some $\lambda \geq 0$. For $K > a$, we have $\mathbb{P}(X_1 > a | X_1 \leq K) > 0$ and $\psi_K(\lambda) < \infty$ for all $\lambda \geq 0$. So preceding argument shows

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}_K(S_n > na) \geq -\psi_K^*(a)$$

where $\mathbb{P}_K$ is the probability measure given by

$$d\mathbb{P}_K^{(n)} \propto 1_{\{X_1 \leq K, \cdots, X_n \leq K\}} d\mathbb{P}$$

(To see this, note, under $\mathbb{P}_K$, random variables $X_1, \cdots X_n$ are independent with distribution $\mu(\cdot | x \leq K)$). But

$$\mathbb{P}(S_n \geq na) \geq \mathbb{P}(S_n \geq na | X_1 \leq K, \cdots, X_n \leq K) = \mathbb{P}_K(S_n \geq na)$$

and $\psi_K^*(a) \searrow \psi^*(a)$ as $K \to \infty$ (by **Lemma 6.1.3**) so we see

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}(S_n \geq na) \geq -\psi_K^*(a) \nearrow -\psi^*(a)$$

(End of proof) □

One different way to see that ψ^* is continuous at a is presented in the following lemma.

Lemma 6.1.3) For all $a \geq m$, with $\mathbb{P}(X_1 > 0) > 0$ we have $\psi_K^*(a) \searrow \psi^*(a)$ as $K \to \infty$. Moreover in the case $\psi(\lambda) < \infty$ for all $\lambda \geq 0$, ψ^* is continuous at a and we have $\psi^*(a) = \lambda^* a - \psi(\lambda^*)$ where λ^* is uniquely determined by $\psi'(\lambda^*) = a$.

proof) Consider first the later case where $\psi(\lambda) < \infty$ for $\lambda \geq 0$. Then by **Proposition 6.1.2** wee see that

$$\psi^*(a) = \lambda^* a - \psi(\lambda^*)$$

where $a = \psi'(\lambda^*)$ and ψ^* is continuous at a with $\lambda^* = (\psi')^{-1}(a)$.

For the first part, note that ψ_K^* is non-increasing in K. For K sufficiently large, we have

$$\mathbb{P}(X_1 > a | X_1 \leq K) > 0$$

and $a \geq m \geq m_K$ (where $m_K = \mathbb{E}(X_1 | \leq X_1 \leq K)$) and $\psi_K(\lambda) < \infty$ for all $\lambda \geq 0$, so we may apply the preceding argument to μ_K to see that

$$\psi_K^*(a) = \lambda_K^* a - \psi_K(\lambda_K^*)$$

where $\lambda_K^* \geq 0$ is determined by $\psi_K'(\lambda_K^*) = a$. Now $\psi_K'(\lambda)$ is non-decreasing in K and λ, so $\lambda_K^* \searrow \lambda^*$ for some $\lambda^* \geq 0$. Also $\psi_K'(\lambda) \geq m_K$ for all $\lambda \geq 0$ so

$$\psi_K(\lambda_K^*) \geq \psi_K(\lambda^*) + m_K(\lambda_K^* - \lambda^*)$$

Then

$$\psi_K^*(a) = \lambda_K^* a - \psi_K(\lambda_K^*) \leq \lambda_K^* a - \psi_K(\lambda^*) - m_K(\lambda_K^* - \lambda^*) \to \lambda^* a - \psi(\lambda^*) \leq \psi^*(a)$$

So $\psi_K^*(a) \searrow \psi^*(a)$ as $K \to \infty$ as claimed.

(End of proof) $\square$

===

(9th November, Friday)

7. Borwnian Motion

7.1. Definition

Let $(X_t)_{t\geq 0}$ is a random process in $\mathbb{R}^d$. We say $(X_t)_{t\geq 0}$ is a **Brownian motion** if :

(i) For all $s, t \geq 0$, the random variable $X_{s+t} - X_s$ is Gaussian, of mean 0 and variance tI and is independent of $\mathcal{F}_s^X = \sigma(X_r : r \leq s)$

(ii) for all $\omega \in \Omega$ the map $t \mapsto X_t(\omega) : [0, \infty) \to \mathbb{R}^d$ is *continuous.*

Condition (i) means that, for all $s \geq 0$, $t > 0$, all Borel sets $B \subset \mathbb{R}^d$ and all $A \in \mathcal{F}_s^X$,

$$\mathbb{P}(\{X_{s+t} - X_s \in B\} \cap A) = \mathbb{P}(A) \int_B (2\pi t)^{-\frac{d}{2}} e^{-|y|^2/2t} dy$$

Or, in terms of conditional expectation, (i) is equivalent to : for all $s, t \geq 0$ and all $f \in C_b(\mathbb{R}^d)$,

$$\mathbb{E}(f(X)_{s+t} | \mathcal{F}_s^X) = P_t f(X_x) \quad \text{a.s.}$$

where P_t is the **heat semigroup**, i.e.

$$P_0 f(x) = \int_{\mathbb{R}^d} p(t, x, y) f(y) dy, \quad p(t, x, y) = (2\pi t)^{-\frac{d}{2}} e^{-|y-x|^2/2t}$$

If $X_0 = x$ then we call $(X_t)_{t\geq 0}$ a **Brownian motion starting from** x. In this case, condition (i) is equivalent following property : for all $t_1, \cdots, t_n \geq 0$ with $t_1 < \cdots < t_n$ and all $B \in \mathcal{B}(\mathbb{R}^{dn})$

$$\mathbb{P}((X_{t_1}, \cdots, X_{t_n}) \in \mathcal{B}) = \int_B \prod_{i=1}^n p(s_i, x_{i-1}, x_i) dx_i$$

where $t_0 = 0$, $x_0 = x$, $s_i = t_i - t_{i-1}$.

Given independent Brownian motions $(X_t^1)_{t\geq 0}, \cdots, (X_t^d)_{t\geq 0}$ in $\mathbb{R}$ starting from 0 and given $x = (x^1, \cdots, x^d) \in \mathbb{R}^d$, the process $(x + (X_t^1, \cdots, X_t^d))_{t\geq 0}$ is a Brownian motion in $\mathbb{R}^d$ starting from x and we obtain all Brownian motion starting from x in $\mathbb{R}^d$ in this way.

7.2. Wiener's theorem

Brownian motion was established as a mathematical object only after 1920's.

Let $W_d = C([0, \infty), \mathbb{R}^d)$, and $x_t : W_d \to \mathbb{R}^d$, $x_t(w) = w(t)$ be the coordinate functions. We may endow W_d with σ-algebra $\mathcal{W}_d = \sigma(x_t : t \geq 0)$.

Given a continuous process $(X_t)_{t\geq 0}$ in $\mathbb{R}^d$ on Ω, we can define

$$X : \Omega \to W_d, \quad X(\omega)(t) = X_t(\omega)$$

then X is $\mathcal{W}_d$-measurable so X has a law on $(W_d, \mathcal{W}_d)$.

Theorem 7.2.1.) *(Wiener)* For all $d \geq 1$ and $x \in \mathbb{R}^d$, there exist a unique probability measure μ_x on $(W_d, \mathcal{W}_d)$ such that $(x_t)_{t\geq 0}$ is a Brownian motion in $\mathbb{R}^d$ staring from x. In particular, Brownian motion exists.

proof) Conditions (i) and (ii) determine the finite dimensional distributions of a Brownian motion and hence determine the law of any BM on $(W_d, \mathcal{W}_d)$ (with given starting point - hence such probability measure is unique.

Suppose we have a measure μ_0 on $(W_1, \mathcal{W}_1)$ such that $(x_t)_{t\geq 0} \sim \mathrm{BM}_0$ in $\mathbb{R}$. For $x \in \mathbb{R}$, $(x + x_t)_{t\geq 0} \sim \mathrm{BM}_x$ so could take μ_x as law of this process. Then for $x = (x_1, \cdots, x_d) \in \mathbb{R}^d$, the measure $\mu_{x_1} \otimes \cdots \otimes \mu_{x_d}$ has required properties. So we only have to work in 1 dimension, starting at 0.

Define $\mathbb{D}_N = \{k2^{-N} : k \in \mathbb{Z}^+\}$ and $\mathbb{D} = \bigcup_{N\geq 0} \mathbb{D}_N$. There exists some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a family $(Y_t : t \in \mathbb{D})$ of independent $N(0,1)$ random variables. First define for $t \in \mathbb{D}_0 = \mathbb{Z}^+$,

$$\xi_t = Y_1 + \cdots + Y_t$$

Then $(\xi_n)_{n\in\mathbb{D}_0}$ is Gaussian and $(\xi_{t+1} - \xi_t : t \in \mathbb{D}_0)$ are independent and has distribution $\sim N(0,1)$. We define recursively $(\xi_t)_{t\in \mathsf{D}_N}$ as follows for $t \in \mathbb{D}_{N+1} \backslash \mathbb{D}_N$:

: set $r = t - 2^{-N-1}$, $s = t + 2^{-N-1} \in \mathbb{D}_{\mathbb{N}}$, set $Z_t = 2^{-\frac{N+2}{2}} Y_t$ and define $\xi_t = \frac{\xi_r + \xi_s}{2} + Z_t$.

We will show by induction that for all $N \geq 0$, $(\xi_{t+2^{-N}} - \xi_t : t \in \mathbb{D}_N)$ are independent, $\sim N(0, 2^{-N})$ random variables

: Suppose true for N. Take $t \in \mathbb{D}_{N+1} - \mathbb{D}_N$ and r, s as above. Then

$$\xi_t - \xi_r = \frac{\xi_s - \xi_r}{2} + Z_t, \quad \xi_s - \xi_t = \frac{\xi_s - \xi_r}{2} - Z_t$$
$$\mathrm{Var}\Big(\frac{\xi_s - \xi_r}{2}\Big) = \frac{1}{4} 2^{-N}, \quad \mathrm{Var}(Z_t) = 2^{-N-2}$$

so

$$\mathrm{Var}(\xi_t - \xi_r) = \frac{1}{4} 2^{-N} + 2^{-N-2} = 2^{-N-1} = \mathrm{Var}(\xi_s - \xi_r)$$
$$\mathrm{cov}(\xi_t - \xi_r, \xi_s - \xi_t) = 0$$

Also for any interval $(u, v]$ disjoint from $(r, s]$ with $u, v \in \mathbb{D}_{N+1}$,

$$\mathrm{cov}(\xi_s - \xi_r, \xi_v - \xi_u) = \mathrm{cov}(\xi_s - \xi_t, \xi_v - \xi_u) = 0$$

So the induction proceeds.

==
(12th November, Monday)

proof of Wiener's theorem continues) We constructed $(\xi_t : t \in \mathbb{D})$ such that for all $s, t \geq 0$, $\xi_{s+t} - \xi_s \sim N(0,t)$ and is independent of $\sigma(\xi_r : r \leq s, r \in \mathbb{D})$. Choose $p > 2$ and set $C_p = \mathbb{E}(|\xi_1|^p)$. Then $C_p < \infty$ and $\mathbb{E}(|\xi_t - \xi_s|^p) = C_p|t-s|^{p/2}$ so by Kolmogorov's lemma(**Theorem 4.2.1**), there exists a continuous process $(X_t)_{t\geq 0}$ such that $X_t = \xi_t$ a.s. for all $t \in \mathbb{D}$.

For $s \geq 0$, $t > 0$ and for any $A \in \mathcal{F}_s^X$ there exist sequence (s_n) in $[0,\infty)$, (t_n) in $(0,\infty)$ and $A_0 \in \sigma(\xi_r : r \leq s, r \in \mathbb{D})$. s.t. $s_n \to s$, $t_n \to t$ as $n \to \infty$ and $1_A = 1_{A_0}$ a.s. (to check this, consider $\{A \in \mathcal{F}_s : \text{this holds }\}$). Then for any $f \in C_b(\mathbb{R})$,

$$\begin{aligned}\mathbb{E}(f(X_{s_n+t_n} - X_{s_n})1_A) &= \mathbb{E}(f(\xi_{s_n+t_n} - \xi_{s_n})1_{A_0}) \\ &= \mathbb{P}(A_0)\int_{\mathbb{R}} p(t_0,0,y)f(y)dy\end{aligned}$$

so letting $n \to \infty$ using bounded convergence theorem gives

$$\mathbb{E}(f(X_{t+s} - X_s)1_A) = \mathbb{P}(A)\int_{\mathbb{R}} p(t,0,y)f(y)dy$$

so X is a Brownian motion as required.

(End of proof) □

7.3. Symmetries of Brownian Motion

Proposition 7.3.1) Let $(X_t)_{t\geq 0}$ be a $\mathrm{BM}_0(\mathbb{R}^d)$ and let $\sigma \in (0,\infty)$ and $U \in O(d)$. Then the following processes are also $\mathrm{BM}_0(\mathbb{R}^d)$.

(i) **(Scaling property)** $(\sigma X_{\sigma^{-2}t})_{t\geq 0}$,

(ii) **(Rotation invariance)** $(UX_t)_{t\geq 0}$.

In fact $\mathrm{BM}_0(\mathbb{R}^d)$ is characterized among continuous Gaussian processes by its means and covariances,

$$\mathbb{E}(X_t) = 0, \quad \mathrm{cov}(X_s^i, X_t^j) = \mathbb{E}(X_s^i X_t^j) = \delta_{ij}(s \wedge t)$$

proof) Exercise.

7.4. Brownian Motion in a Given Filtration

Suppose given a filtration $(\mathcal{F}_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$.

- We say that $(X_t)_{t\geq 0}$ is a $(\mathcal{F}_t)_{t\geq 0}$**-BM** if
  (a) X_t is $\mathcal{F}_t$-measurable,
  (b) for all $s, t \geq 0$, the random variable $X_{s+t} - X_s \sim N(0, tI)$ and is independent of $\mathcal{F}_s$,
  (c) for all $\omega \in \Omega$, the map $t \mapsto \Big(X_t(\omega) : [0,\infty) \to \mathbb{R}^d\Big)$ is continuous.

  This implies that X is a BM in the old sense and any process X which is a BM in the old sense is an $(\mathcal{F}_t^X)_{t\geq 0}$-BM.

Proposition 7.4.1) Let $X = (X_t)_{t\geq 0}$ be a $\mathrm{BM}(\mathbb{R}^d)$ and let F be a bounded measurable function on W_d. Define

$$f(x) = \int F(\omega)\mu_x(d\omega), \quad x \in \mathbb{R}^d$$

Then f is measurable on $\mathbb{R}^d$ and $\mathbb{E}(F(X)|\mathcal{F}_0) = f(X_0)$ a.s. (recall, μ_x is the law of BM_x)

proof) (Was an exercise.) First consider F of form $1_{A^{(a,\eta)}}$ where $A^{(a,\underline{\eta})} = \{g \in W_d : g(a) \geq (\eta_1, \cdots, \eta_d)\} \subset W_d$. Then

$$\int A^{(a,\eta)}(\omega)\mu_x(d\omega) = \mathbb{P}(B^{(x)}(a) \geq \eta) = \int_{\eta_1}^{\infty} \cdots \int_{\eta_d}^{\infty} \frac{1}{(2\pi)^{d/2}} e^{-y^2/2} d^d y$$

where $B^{(x)}$ is a Brownian motion starting at x, so this is measurable. Also, we easily see that any finite intersection of sets $A^{(a,\eta)}$ if of form

$$B^{(a_1,\cdots,a_m)}_{(\eta_1,\cdots,\eta_m)} = \{g \in W_d : g(a_1) \geq \eta_1, \cdots, g(a_1 + \cdots + a_m) \geq \eta_m\}$$

with $a_1, \cdots a_m \geq 0$ so

$$\int B^{(a_1,\cdots,a_m)}_{(\eta_1,\cdots,\eta_m)} \mu_x(d\omega) = \int A^{(a_1,\eta_1)}(\omega)\mu_x(d\omega) \prod_{k=2}^{m} \int A^{(a_k,\eta_k - \eta_{k-1})}(\omega)\mu_0(d\omega)$$

is measurable. We also see that $\mathbb{E}(F(X)|\mathcal{F}_0) = f(X_0)$ a.s. for each case $F = 1_{A^{(a,\eta)}}$ or $F = 1_{B^{(a_1,\cdots,a_m)}(\eta_1,\cdots,\eta_m)}$.

Now notice that the sets $(B^{(a_1,\cdots,a_m)}_{(\eta_1,\cdots,\eta_m)} : m \geq 1, a_j \geq 0, \eta_j \in \mathbb{R})$ forms a π-system generating the σ-algebra $\mathcal{W}_d$. Since any limit of measurable functions is measurable, and $\mathbb{E}(F_n(X)|\mathcal{F}_0) = f_n(X_0)\ \forall n$, $F_n \nearrow F$ implies $\mathbb{E}(F(X)|\mathcal{F}_0) = f(X_0)$, we see from monotone class theorem that the statement also holds for general bounded measurable functions F.

(End of proof) □

7.5. Martingales of BM

Theorem 7.5.1) Let $(X_t)_{t\geq 0}$ be an $(\mathcal{F}_t)_{t\geq 0}$-BM in $\mathbb{R}^d$ and let $f \in C_b^2(\mathbb{R}^d)$. Define

$$M_t = f(X_t) - f(X_0) - \int_0^t \frac{1}{2}\Delta f(X_s) ds$$

Then $(M_t)_{t\geq 0}$ is an $(\mathcal{F}_t)_{t\geq 0}$-martingale.

There are two ways for doing this. One is to use stochastic calculus, and the other is to do Markovian-way(which would be made clear soon). You will see the first method in Stochastic Calculus course next term, and we are going to prove this in the second way.

proof) It is clear that $(M_t)_{t\geq 0}$ is adapted and integrable. Consider for now the case $X_0 = x \in \mathbb{R}^d$. Set

$$m(x) = \mathbb{E}(M_t) = \int_{W_d} \left[f(\omega(t)) - f(\omega(0)) - \int_0^t \frac{1}{2}\Delta f(\omega(s)) ds \right] \mu_x(d\omega)$$

Fix $s \in (0, t]$. We have

$$M_t - M_s = f(X_t) - f(X_s) - \int_s^t \frac{1}{2}\Delta f(X_r) dr$$

so using Fubini,

$$\mathbb{E}(M_t - M_s) = \int_{\mathbb{R}^d} p(t,x,y) f(y) dy - \int p(s,x,y) f(y) dy - \int_s^t \int_{\mathbb{R}^d} p(r,x,y) \times \frac{1}{2}\Delta f(y) dy$$

But since p is a heat kernel, has $\frac{d}{dt}p = \frac{1}{2}\Delta p$ so

$$\begin{aligned} \int_{\mathbb{R}^d} p(r,x,y) \times \frac{1}{2}\Delta f(y) dy &= \int_{\mathbb{R}} \frac{1}{2}\Delta_y p(r,x,y) f(y) dy \\ &= \int_{\mathbb{R}} \frac{d}{dt} p(r,x,y) f(y) dy \end{aligned}$$

and

$$\int_s^t \int_{\mathbb{R}^d} p(r,x,y) \times \frac{1}{2}\Delta f(y) dy dr = \int_s^t \int_{\mathbb{R}} \frac{d}{dt} p(r,x,y) f(y) dy dr$$
$$= \int_{\mathbb{R}^d} (p(t,x,y) - p(s,x,y)) dy$$

so $\mathbb{E}(M_t - M_s) = 0$. But $M_s \to 0$ a.s. as $s \to 0$ so by bounded convergence, $\mathbb{E}(M_s) \to 0$. Hence $m(x) = \mathbb{E}(M_t) = 0$ for all x.

Return to the case of general X_0. By **Prop 7.4.1**,

$$\mathbb{E}(M_t|\mathcal{F}_0) = m(X_0) = 0 \quad \text{a.s.}$$

Now for any $s, t \geq 0$,

$$M_{s+t} - M_s = f(X_{s+t}) - f(X_s) - \frac{1}{2}\int_0^s \Delta f(X_{s+r}) dr$$

and $(X_{s+t})_{t\geq 0}$ is an $(\mathcal{F}_{s+t})_{t\geq 0}$-BM. So we see that

$$\mathbb{E}(M_{s+t} - M_s|\mathcal{F}_s) = 0 \quad \text{a.s.}$$

as required.

(End of proof) $\square$

We are in fact apply the theorem in the case where the assumption $f \in C_b^2(\mathbb{R}^d)$ does not hold. We can actually relax this condition using almost the same proof, e.g. any f of polynomial growth rate would work.

Exercise : see how much we can relax $C_b^2(\mathbb{R}^d)$.

-Every step other than the integration by parts $\int_{\mathbb{R}^d} p(r,x,y) \times \frac{1}{2}\Delta f(y) dy = \int_{\mathbb{R}} \frac{1}{2}\Delta_y p(r,x,y) f(y) dy$ works without assumption that $f \in C_b^2(\mathbb{R}^d)$. Hence, it is enough to have $f \in W^{2,1}(\mathbb{R}^d)$(the Sobolev space) with sub-exponential growth rate.

===

(14th November, Wednesday)

Examples : Let $(X_t^i)_{t\geq 0}$, $i = 1, \cdots, n$ be a BM.

- $f(x) = x^i$ then $\Delta f = 0$, so is a martingale.
- $f(x) = x^i x^j$ then $\Delta f = 0$, so is a martingale.
- $(X_t^i X_t^j - \delta_j t)_{t\geq 0}$ is a martingale.
- $\exp(\lambda X_t - \frac{1}{2}\lambda^2 t)_{t\geq 0}$ is a martingale.
- For all harmonic functions f bounded, $(f(X_t))_{t\geq 0}$ is a martingale.

7.6. Strong Markov Property

Strong Markov Property of Brownian Motion would be particularly useful.

Theorem 7.6.1) Let $(X_t)_{t\geq 0}$ be an $(\mathcal{F}_t)_{t\geq 0}$-BM and let T be a stopping time. Then conditioned on $\{T < \infty\}$, $(X_{T+t})_{t\geq 0}$ is an $(\mathcal{F}_{T+t})_{t\geq 0}$-BM.

proof) Clearly on $\{T < \infty\}$, $(X_{T<t})_{t\geq 0}$ is continuous on $[0,\infty)$. Also on $\{T < \infty\}$, X_{T+t} is $\mathcal{F}_{T+t}$-measurable. Fix $s \geq 0$, $t > 0$, $f \in C_b(\mathbb{R}^d)$, $m, n \in \mathbb{N}$, $A \in \mathcal{F}_{T+s}$ with $A \subset \{T \leq m\}$. For $k = 0, 1, \cdots 2^n m$, set $t_k = k2^{-n}$ and $A_k = A \cap \{t_k - 2^{-n} < T \leq t_k\}$. Note $A_k \in \mathcal{F}_{t_k+s}$ and $A_k = A \cap \{T_n = t_k\}$ where $T_n = 2^{-n}\lceil 2^n T\rceil$, so

$$\mathbb{E}(f(X_{T_n+s+t})1_{A_k}) = \mathbb{E}(f(X_{t_k+s+t})1_{A_k})$$
$$= \mathbb{E}(P_t f(X_{t_k+s})1_{A_k}) = \mathbb{E}(P_t f(X_{T_n+s})1_{A_k})$$

Now sum over k and let $n \to \infty$ using bounded convergence to obtain

$$\mathbb{E}(f(X_{T+s+t})1_A) = \mathbb{E}(P_t f(X_{T+s})1_A)$$

But m and A were arbitrary so this implies, on $\{T < \infty\}$,

$$\mathbb{E}(f(X_{T+t+s})|\mathcal{F}_{T+s}) = P_t f(X_{T+s}) \quad \text{a.s.}$$

as required.

(End of proof) □

7.7. Properties of 1-d BM

Proposition 7.7.1) Let $(X_t)_{t\geq 0}$ be a $\mathrm{BM}_0(\mathbb{R})$. Set $T_a = \inf\{t \geq 0 : X_t = a\}$. Then

$$\mathbb{P}(T_a < \infty) = 1 \quad \text{for all } a \in \mathbb{R} \quad \textit{and}$$
$$\mathbb{P}(T_{-a} \leq T_b) = \frac{b}{a+b} \quad \forall a, b \geq 0 \quad \textit{and}$$
$$\mathbb{E}(T_a \wedge T_b) = ab$$

Moreover T_a has a density f_a on $[0, \infty)$ given by

$$f_a(t) = \frac{1}{\sqrt{2\pi t^3}} e^{-a^2/2t} \quad t \geq 0$$

Moreover the following holds almost surely.

(a) $X_t/t \to 0$ as $t \to \infty$.

(b) $\inf_{t\geq 0} X_t = -\infty$, $\sup_{t\geq 0} X_t = \infty$.

(c) for all $s \geq 0$, there exist $t, n \geq s$ such that $X_t < 0 < X_n$.

(d) for all $s > 0$ there exist $t, n \in [0, s)$ such that $X_t < 0 < X_n$.

proof) Exercise.

Theorem 7.7.2) Let $X \sim \mathrm{BM}_0(\mathbb{R})$. Then the following properties hold almost surely :

(a) for all $\alpha < 1/2$, $(X_t)_{t\geq 0}$ is locally Hölder continuous of exponent α.

(b) for all $\alpha > 1/2$ there is no non-trivial interval on which X is Hölder continuous of exponent α.

proof)

(a) For $s, t \geq 0$ with $s < t$, $X_t - X_s \sim (t-s)^{1/2} X_1$ so for all $p < \infty$,

$$\mathbb{E}(|X_t - X_s|^p) = (t-s)^{p/2} C_p$$

where $C_p = \mathbb{E}(|X_1|^p) < \infty$. Given $\alpha < 1/2$ we can find $p < \infty$ so that $\alpha + \frac{1}{p} < \frac{1}{2}$. Then by Kolmogorov's lemma there exist $K \in L^p(\mathbb{P})$ such that

$$|X_t - X_s| \leq K|t-s|^\alpha \quad \text{for all } s, t \in [0, 1]$$

So by scaling of BM, $(X_t)_{t\geq 0}$ is locally Hölder continuous of exponent α a.s.

Finally consider $\alpha_n \nearrow 1/2$, $\alpha_n < 1/2$ to see $(X_t)_{t\geq 0}$ is Hölder continuous of exponent α a.s. for all $\alpha < 1/2$.

(b) Let $m, n \in \mathbb{N}$ with $m \geq n$. For $s, t \in \mathbb{D}_n$, define

$$[X]^m_{s,t} = \sum_\tau (X_{\tau+2^{-m}} - X_\tau)^2$$

where we sum over all $\tau \in \mathbb{D}_n$ with $\tau \in (s, t]$. The random variables $(X_{\tau+2^{-m}} - X_\tau)^2$ are independent of mean 2^{-m} and variance 2^{-2m+1} (Here we need $\mathbb{E}(X_1^4) = 3$, $\mathrm{Var}(X_1^2) = 2$). So $\mathbb{E}([X]^m_{s,t}) = 2^m(t-s)2^{-m} = t-s$ and $\mathrm{Var}([X]_{s,t}) = 2^m(t-s)2^{-2m+1}$. So $[X]^m_{s,t} \to t-s$ a.s.

Suppose X is Hölder α on $[s, t]$ then for some constant K,

$$|X_{\tau+2^{-m}} - X_\tau|^2 \leq K^2(2^{-m})^{2\alpha}$$

so $[X]^m_{s,t} \leq 2^m(t-s)K^2 2^{-2m\alpha} \to 0$ as $m \to \infty$. So $(X_t)_{t\geq 0}$ is not Hölder α on $[s, t]$, which contradicts with our earlier result.

(End of proof) □