Lecture 13

Separating classes

Often probabilist study random quantities X which, instead of mapping into $\mathbb{R}$, map into more complicated spaces.

E.g. X = random tree

X = random continuous function

X = random finite set

$\vdots$

In this case it doesn't make sense to talk about $E(X) = \int_\Omega x\, dP$ since our definition of integration requires X maps into $\bar{\mathbb{R}}$.

Moreover, characterizing the distribution of X, i.e. PX^{-1}, using densities is difficult especially if there is no canonical measure on the range space of X like Lebesgue measure to compare densities w.r.t.

However, it is often the case that the range space of X has a natural metric space associated with it.

$$\underset{\mathcal{F},P}{\Omega'} \xrightarrow{X(\omega)} \underset{\mathcal{B}(\Omega)}{\Omega} \leftarrow \text{metric space}$$

Since $\mathcal{B}(\Omega)$ is the Borel σ-field we have available to us the notion of continuity for functions mapping $\Omega \to \mathbb{R}$

$$\underset{\mathcal{F},P}{\Omega'} \xrightarrow{X(\omega)} \underset{\mathcal{B}(\Omega)}{\Omega} \xrightarrow{f(\omega)} \mathbb{R} \Rightarrow$$

$f(X)$ is a r.v. and $Ef(X)$ is defined.

Now to study properties of X a general approach is to study how $Ef(X)$ behaves for "nice" functions f, such as those which are continuous and bdd.

This will be a general approach used in many of the future lectures.

In this lecture we will use this approach for characterizing the distribution of X: i.e. find some nice function classes Γ s.t.

$$Ef(X) = Ef(Y),\ \forall f \in \Gamma \implies X \overset{\mathcal{D}}{=} Y.$$

or $\int_\Omega f\,dP = \int_\Omega f\,dQ,\ \forall f \in \Gamma \implies P = Q$ on $\mathcal{B}(\Omega)$

There is basically one important theorem in this lecture. The immediate utility of this theorem will be to establish the following result:

> If X and Y are d-dimensional random vectors such that $E e^{iX\cdot k} = E e^{iY\cdot k}$ $\forall k \in \mathbb{R}^d$ then $X \overset{\mathcal{D}}{=} Y$.
>
> The characteristic functions for X and Y.

However the main use of this result will be for establishing the existence of the Wiener measure (also called the Wiener process or Brownian motion).

Wiener Measure

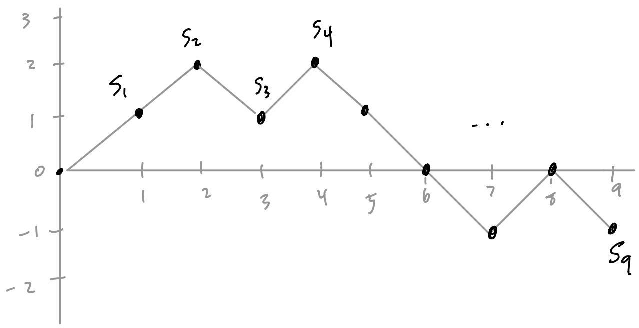
The Wiener measure is probably one of the most important objects in advanced probability.

I find it interesting in that you need some "high powered" theory to rigiourisly show its existance, yet you can easily give a non-rigourous description of its properties.

Recall in lecture 6 we discussed the 1-d random walk $S_0, S_1, \dots$ defined by

$$S_n = \sum_{k=1}^{n} R_k, \quad S_0 = 0$$

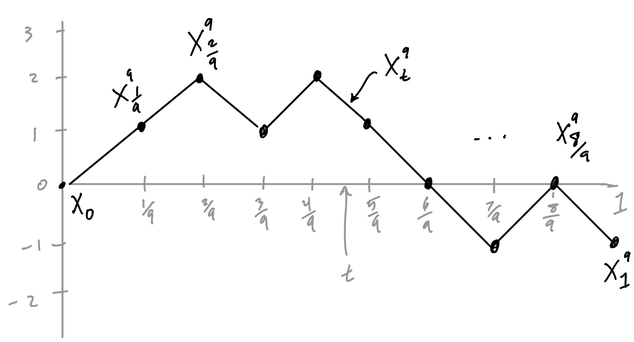
where $R_1, R_2, \dots$ are independent Rademacher r.v.s s.t. $P(R_k = 1) = P(R_k = -1) = 1/2$.



Notice that $E(S_n) = \sum_{k=1}^{n} E(R_k) = 0$

$$\underbrace{\mathrm{var}(S_n)}_{E(S_n - E(S_n))^2} = \sum_{k=1}^{n} \underbrace{\mathrm{var}(R_k)}_{E(R_k^2) = 1} = n$$

Now for each n re-scale the x-axis by $\frac{1}{n}$ define X_t^n on $[0,1]$ by linear interp of $X_{i/n}^n = S_i$



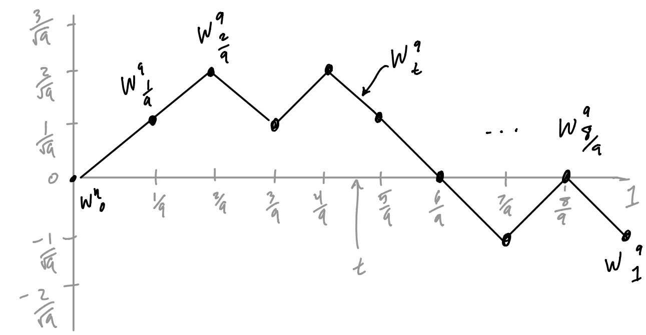
Notice X_t^n is continuous on $[0,1]$, i.e. $X_t^n \in C[0,1]$, and

$$E(X_t^n) = 0 \qquad \mathrm{var}(X_{\frac{i}{n}}^n) = n.$$

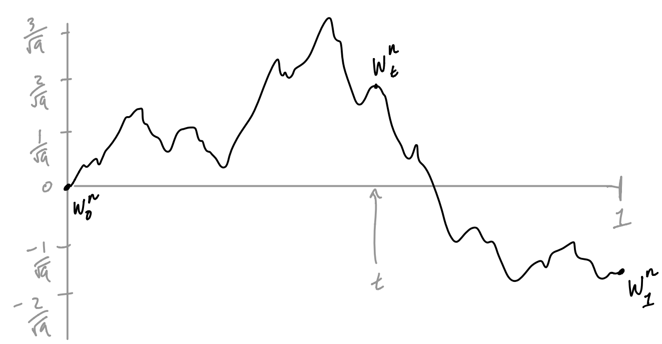
If we want a limit we need to re-scale X_t^n so the variability explode... try re-scaling so $\mathrm{var}(cX_1^n) = 1$.

So re-scale the y-axis by $\frac{1}{\sqrt{n}}$ and

define $W_t^n := \frac{1}{\sqrt{n}} X_t^n$



For large n, W_t^n looks like this:



Clearly $W_t^n \in C[0,1]$ so, in some sense, we constructed n measures on $C[0,1]$ with sup-norm metric used for $\mathcal{B}(C[0,1])$.

<u>Question 1</u>: Why is

$$(\Omega, \mathcal{F}) \xrightarrow{W^n} (C[0,1], \mathcal{B}(C[0,1]))$$

measurable?

This is easy to see by writing

$$W_t^n(\omega) = W_t^n(R_1(\omega), \dots, R_n(\omega))$$

$$\begin{tikzcd}
\Omega \arrow[r, "{(R_1,\dots,R_n)}\ ⓜ"] \arrow[rr, bend right, "W_t^n\ ⓜ"'] & \mathbb{R}^n \arrow[r, "\substack{\text{continuous} \\ \text{w.r.t. sup-norm}}"] & C[0,1]
\end{tikzcd}$$

$\mathcal{F}, P$ $\quad$ $\mathcal{B}(\mathbb{R}^n)$ $\quad$ $\mathcal{B}(C[0,1])$

Question 2: Does there exist a limit

$$W_t \overset{?}{=} \lim_{n\to} W_t^n$$

Definitely not to a fixed function in $C[0,1]$ since $var(W_t^n) = 1$, $\forall n$.

However we will later show that W_t^n does have a random limit W_t (in distribution):

$$(*) \qquad Ef(W_\cdot^n) \overset{n\to\infty}{\rightarrow} Ef(W_\cdot)$$

for all continuous and bdd $f: C[0,1] \to \mathbb{R}$.

This W_t is Brownian motion on $[0,1]$.

To show this we will show

(i) $\{W_\cdot^n\}_{n\in\mathbb{N}}$ is tight

(ii) For any fixed $t_1, \dots, t_m$

$$E\, f(W_{t_1}^n, \dots, W_{t_m}^n) \longrightarrow Ef(W_{t_1}, \dots, W_{t_m})$$

$\forall$ continuous and bdd $f: \mathbb{R}^m \to \mathbb{R}$.

The fact that (i) & (ii) are sufficient uses the main theorem in this lecture.

Question 3: To show (ii) how do we know what $Ef(W_{t_1}, \dots, W_{t_m})$ is?

Since $E(W_t^n) = 0$ the CLT will imply

$$(W_{t_1}, \dots, W_{t_m}) \overset{\mathcal{D}}{=} N_m(0, \Sigma)$$

Also note that for integers $j < i < q < p$

$$W_{\frac{i}{n}}^n = \frac{1}{\sqrt{n}} S_i$$

$$W_{i/n}^n - W_{j/n}^n = \frac{1}{\sqrt{n}}(S_i - S_j) = \frac{1}{\sqrt{n}} \sum_{k=j+1}^{i} R_k$$

$W_{i/n}^n - W_{j/n}^n$ indep of $W_{p/n}^n - W_{q/n}^n$

So that

$$var(W_{i/n}^n) = \frac{1}{n} var(S_i) = \frac{i}{n}$$

$$var(W_{i/n}^n - W_{j/n}^n) = \frac{1}{n} var(S_i - S_j) = \frac{i-j}{n}$$

$$cov(W_{i/n}^n - W_{j/n}^n,\ W_{p/n}^n - W_{q/n}^n) = 0$$

This suggest that

$$E(W_t) = 0$$

$$var(W_t) = t$$

$$var(W_t - W_s) = |t-s|$$

note: $var(W_t) = t$ is a consequence of this.

$W_t - W_s$ indep of $W_x - W_y$ when $0 \le t < s < x < y \le 1$.

$\therefore\ cov(W_t, W_s) = \frac{1}{2}(|t| + |s| - |t-s|)$

which is enough to determine:

$$Ef(W_{t_1}, \dots, W_{t_m})$$

Some definitions

Definition:
If Ω is a complete and separable metric space then Ω is called a Polish space.

In some sense Polish spaces are the most general spaces we will need to work with in the remainder of the class. One can generalize results to non-polish spaces but I don't think the extra complexity is worth it.

One nice thing about Polish spaces is that
$$\mathcal{B}(\Omega) = \sigma\langle \text{open balls}\rangle$$
and any probability measure on $(\Omega, \mathcal{B}(\Omega))$ is automatically a Radon measure and therefore satisfies

$$P(A) = \sup\{P(K) : \text{compact } K \subset A\}$$

proof is similar to what was done in HW 2.

Note: I'll often write "Let $(\Omega, \mathcal{B}(\Omega))$ be Polish..." which means Ω is a polish space with metric d and $\mathcal{B}(\Omega)$ is the Borel σ-field w.r.t d.

It will also be useful to define notation for some common function spaces when Ω is a metric space.

$C(\Omega) := \{\text{continuous maps } f:\Omega \to \mathbb{R}\}$

$C_b(\Omega) := \{\text{bdd and continuous maps } f:\Omega\to\mathbb{R}\}$

$C_c(\Omega) := \{\text{compactly supported continuous maps } f:\Omega\to\mathbb{R}\}$

$\text{Lip}(\Omega) := \{f:\Omega\to\mathbb{R} \text{ s.t. } |f(x)-f(y)| \le c\, d(x,y) \text{ for some } c\ge 0 \text{ \& all } x,y\in\Omega\}$

$C^k(\mathbb{R}^d) := \{k\text{-times } \underbrace{\text{differentiable}}_{\text{continuously}} \text{ maps } f:\mathbb{R}^d\to\mathbb{R}\}$

Definition:
Let $(\Omega, \mathcal{B}(\Omega))$ be Polish and $\Gamma \subset \underset{\uparrow \text{ quasi-integrable}}{\mathcal{Q}}(\Omega, \mathcal{B}(\Omega))$

- Γ is a separating class for $(\Omega, \mathcal{B}(\Omega))$ if
$$\int_\Omega f\,dP = \int_\Omega f\,dQ,\ \forall f\in\Gamma \implies P = Q$$
for all probability measures P, Q on $(\Omega, \mathcal{B}(\Omega))$.

- If $\mathcal{P}$ is a collection of probability measures on $(\Omega, \mathcal{B}(\Omega))$, Γ separates $\mathcal{P}$ if
$$\int_\Omega f\,dP = \int_\Omega f\,dQ,\ \forall f\in\Gamma \implies P = Q$$
$\forall P, Q\in\mathcal{P}$

- If $\mathcal{P}$ is a collection of random variables mapping into $(\Omega, \mathcal{B}(\Omega))$, Γ separates $\mathcal{P}$ if
$$Ef(X) = Ef(Y),\ \forall f\in\Gamma \implies X \overset{\mathcal{D}}{=} Y$$
$\forall X, Y\in\mathcal{P}$.

Before we prove the main result of this section lets briefly discuss convolutions.

Convolution

working with functions $f \in C(\mathbb{R}^d)$ can be annoying since they are not regular enough to do thing like Taylor approximations, etc...

To get access to these tools we often smooth f by convolving with a member of $C_c^\infty(\mathbb{R}^d)$.

Let $\varphi \in C_c^\infty(\mathbb{R}^d)$ s.t. φ has support in $[-1,1]^d$ and $\int_{\mathbb{R}^d} \varphi(x)dx = 1$. Define

$$\varphi_\varepsilon(x) = \frac{1}{\varepsilon^d}\varphi\left(\frac{x}{\varepsilon}\right)$$ also satisfies $\int \varphi_\varepsilon(x)dx = 1$

$$\varphi_\varepsilon * f(x) = \int_{\mathbb{R}^d} \varphi_\varepsilon(x-y)f(y)\,dy.$$

Now $\varphi_\varepsilon * f$ are very nice in that:

- $\varphi_\varepsilon * f \in C^\infty(\mathbb{R}^d)$
- $\varphi_\varepsilon * f \in C_c^\infty(\mathbb{R}^d)$ if $f \in C_c(\mathbb{R}^d)$
- $\partial^\alpha(\varphi_\varepsilon * f) = (\partial^\alpha \varphi_\varepsilon) * f \quad \forall$ multi index α

Moreover, as $\varepsilon \to 0$, $\varphi_\varepsilon * f$ approximates f in a way that works well with integrals:

- $\varphi_\varepsilon * f \to f$ uniformly on compacts
- $\varphi_\varepsilon * f \to f$ uniformly if $f \in C_c(\mathbb{R}^d)$

Convolutions also have an important probabilistic interpretation:
If $\varphi(x) \geq 0$ then φ is the density of some random vector Z.

In this case

$$\varphi * f(x) = \mathbb{E} f(x - Z)$$
$$\varphi_\varepsilon * f(x) = \mathbb{E} f(x - \varepsilon Z)$$

and if $f(x)$ is also a density for some r.v. X indep of Z and defined on the same probability space as Z then

$\varphi * f$ is the density of $X + Z$

So, if you want to study a r.v. X which is not very well behaved you can often study $X + \varepsilon Z$ for a very nice Z, indep of X, & limit $\varepsilon \to 0$.

Remark: One can use probability theory to construct a $\varphi \in C_c^\infty([-1,1])$ which is a probability density. In a previous lecture you showed

$$\tilde{U} = \sum_{k=1}^\infty 2^{-k} X_k \sim \mathrm{Unif}(0,1).$$ 0 or 1 w.p. $\frac{1}{2}$ each

Now let $U_1, U_2, \ldots$ be independent copies of $U \sim \mathrm{Unif}([-1,1])$ & define

$$Z = \sum_{k=1}^\infty 2^{-k} U_k$$

The density φ of Z is the infinite convolution of the densities of $2^{-k}U_k$ & satisfies $\varphi \in C_c^\infty([-1,1])$ since $|Z| \leq 1$

these have compactly supported densities

11

Main theorem for separating classes

we can already construct a bunch of separating classes using indicators of generating sets. Indeed suppose $\mathcal{P} \subset \mathcal{B}(\Omega)$ is a π-system s.t.

$$\mathcal{B}(\Omega) = \sigma\langle \mathcal{P} \rangle.$$

By π-uniquness of prob. measures we have that for any probabilities P & Q on $(\Omega, \mathcal{B}(\Omega))$:

$$P(A) = Q(A),\ \forall \in \mathcal{P} \Rightarrow P = Q$$

$$\therefore \int_\Omega I_A \, dP = \int_\Omega I_A \, dQ,\ \forall A \in \mathcal{P} \Rightarrow P = Q$$

$\therefore$ $\Gamma = \{ I_A : A \in \mathcal{P} \}$ is a separating class for $(\Omega, \mathcal{B}(\Omega))$.

e.g. since $\mathcal{B}(\Omega) = \sigma\langle \underbrace{\text{closed sets}}_{\text{forms a } \pi\text{-system}} \rangle$

$\Gamma := \{ I_C : \text{closed } C \subset \Omega \}$ is a separating class.

The key, however, is to work with a separating class that is either

"small" $\Rightarrow$ don't need to check $\int_\Omega f\,dP = \int_\Omega f\,dQ$ for too many $f \in \Gamma$

12

or

"regular" $\Rightarrow$ the functions $f \in \Gamma$ are well behaved and easier to control.

The following result illustrates both

Theorem 1

Suppose $(\Omega, \mathcal{B}(\Omega))$ is Polish.
Then each of the following function classes is a separating class over $(\Omega, \mathcal{B}(\Omega))$

(i) $Lip(\Omega) \cap C_b(\Omega)$, always

(ii) $C_c(\Omega)$, if Ω is locally compact

i.e. $\forall x \in \Omega$ $\exists$ open $\mathcal{U}$ s.t. $x \in \mathcal{U}$ & $\bar{\mathcal{U}}$ is compact

(iii) $C_c^\infty(\Omega)$, if $\Omega = \mathbb{R}^d$.

(iv) $\{ e^{ix \cdot k} \text{ s.t. } k \in \mathbb{R}^d \}$, if $\Omega = \mathbb{R}^d$

(v) $\{ \underbrace{x_1^{p_1} x_2^{p_2} \cdots x_n^{p_n}}_{\text{monomials}} \text{ s.t. } p_1, \dots p_n \in \mathbb{N} \}$, if Ω is a compact subset of $\mathbb{R}^d$.

Proof:

Suppose P and Q are two probability measures on $(\Omega, \mathcal{B}(\Omega))$ s.t.

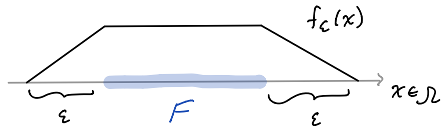
$$(*) \quad \int_\Omega f\,dP = \int_\Omega f\,dQ \quad \forall f \in \Gamma$$

For (i) Suppose $\Gamma := Lip_1(\Omega) \cap C_b(\Omega)$.

If we show $P(F) = Q(F)$, $\forall$ closed $F \subset \Omega$, we are done (by π-uniqness).

Let F be closed & $\varepsilon > 0$.

Define $f_\varepsilon(x) := \left(1 - \frac{d(x,F)}{\varepsilon}\right)^+$



Notice that f_ε is bdd and Lipschitz continuous since

$$|f_\varepsilon(x) - f_\varepsilon(y)| \le \left| \frac{d(x,F)}{\varepsilon} - \frac{d(y,F)}{\varepsilon} \right|$$

since $|(1-z)^+ - (1-w)^+| \le |z-w|$ for $w, z \ge 0$

$$\le \frac{d(x,y)}{\varepsilon}$$ left as an exercise.

$\therefore f_\varepsilon \in \Gamma, \ \forall \varepsilon > 0$ (*)

Moreover if $F^\varepsilon := \{y : d(y,F) < \varepsilon\}$ then

$$I_F(x) \le f_\varepsilon(x) \le I_{F^\varepsilon}(x)$$

since $x \in F \Rightarrow f_\varepsilon(x) = 1$ since $x \notin F^\varepsilon \Rightarrow f_\varepsilon(x) = 0$

$$\therefore P(F) \le \int_\Omega f_\varepsilon \, dP = \int_\Omega f_\varepsilon \, dQ \le Q(F^\varepsilon)$$

since $f_\varepsilon \in \Gamma$ by (*)

Since F is closed $\underbrace{F^\varepsilon \downarrow F}$ as $\varepsilon \to 0$

because $x \in \bigcap_{\varepsilon \in \mathbb{Q}} F^\varepsilon \Rightarrow d(x,F) = 0 \Rightarrow x \in F$

$$\therefore P(F) \le \lim_{\varepsilon \downarrow 0} Q(F^\varepsilon) = Q(F)$$

Similarly one obtains $Q(F) \le P(F)$

$\therefore P(F) = Q(F) \ \forall$ closed $F \subset \Omega$

as was to be shown

QED

For (ii)

Suppose $\Gamma := C_c(\Omega)$ and Ω is locally compact.

Since Ω is Polish and P, Q are probabilities

(*) $P(A) = \sup\{P(K) : \text{compact } K \subset A\}$

(**) $Q(A) = \sup\{Q(K) : \text{compact } K \subset A\}$

$\therefore$ we just need to show $P = Q$ on compacts.

Let K be compact and again write

$$f_\varepsilon(x) := \left(1 - \frac{d(x,K)}{\varepsilon}\right)^+.$$

The same proof as in (i) shows

$$P(K) = Q(K).$$

$\therefore P = Q$ by (*) & (**)

Now we just show $f_\varepsilon \in \Gamma$, i.e. that

$$\operatorname{supp} f_\varepsilon := \{x : d(x,K) < \varepsilon\} =: K^\varepsilon$$

is contained in a compact set.

Start by writing

$$K \subset \bigcup_{x \in K} U_x$$

where U_x is an open neighborhood of x s.t. $\overline{U_x}$ is compact... using our assumption for (ii).

Since K is compact $\exists x_1, \dots, x_n$ s.t.

$$K \subset \bigcup_{i=1}^n U_{x_i} =: \underbrace{U}_{\text{open}}$$

Now for all sufficiently small $\varepsilon < 0$

$$\underbrace{K \subset K^\varepsilon \subset U}_{} \subset \underbrace{\bigcup_{i=1}^n \overline{U}_{x_i}}_{\text{Compact}}$$

For suppose $\forall \varepsilon > 0$ $K \subset K^\varepsilon \not\subset U$.

Now $\exists x_n \in K^{\frac{1}{n}} \cap U^c$ s.t.

$x_n \to x \in K \cap U^c$ but this is $\emptyset$

QED

For (iii)

Suppose $\Omega = \mathbb{R}^d$ & $\Gamma = C_c^\infty(\mathbb{R}^d)$.

Now we just need to show

$$\int_\Omega f\,dP = \int_\Omega f\,dQ \quad \forall f \in C_c(\mathbb{R}^d)$$

which will imply $P = Q$ by (ii)

Here is where we use convolutions.
Fix $f \in C_c(\mathbb{R}^d)$ and simply notice that when $\varphi_\varepsilon \in C_c^\infty(\mathbb{R}^d)$ as in previous section then

$$\int_\Omega f\,dP = \lim_{\varepsilon \to 0} \int_\Omega \varphi_\varepsilon * f\,dP \quad \text{by uniform approx since } f \in C_c(\mathbb{R}^d)$$

$$= \lim_{\varepsilon \to 0} \int_\Omega \varphi_\varepsilon * f\,dQ \quad \text{by assumption}$$

$$= \int_\Omega f\,dQ \quad \text{again by uniform}$$

QED

For (iv) and (v)

we use the Stone-Weierstrass thm:

> If $f \in C([-N,N]^d)$ then f is uniformly approximated by polynomials and finite linear combinations of $\sin(x\cdot k)$ & $\cos(x\cdot k)$ for $k \in \frac{2\pi}{2N}\mathbb{Z}^d$.

For (iv) we can use the technique in (iii) and just show that $\forall f \in C_c(\mathbb{R}^d)$
$\exists f_1, f_2, \ldots \in$ finite linear span of $\sin(x\cdot k), \cos(x\cdot k)$

s.t.

$$\int_{\mathbb{R}^d} f_n\,dP \xrightarrow{n\to\infty} \int_{\mathbb{R}^d} f\,dP$$

$$(*) \qquad \int_{\mathbb{R}^d} f_n\,dQ \xrightarrow{n\to\infty} \int_{\mathbb{R}^d} f\,dQ$$

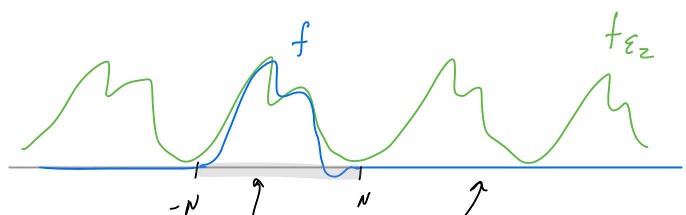
Note $\forall \varepsilon_1 > 0 \ \exists N > 0$ s.t. (N depends on ε_1)

support of $f \subset [-N,N]^d$.
$P(\mathbb{R}^d \setminus [-N,N]^d) < \varepsilon_1$
$Q(\mathbb{R}^d \setminus [-N,N]^d) < \varepsilon_1$

Also note $\forall \varepsilon_2 > 0 \ \exists f_{\varepsilon_2} \in$ finite span $\{\sin(x\cdot k), \cos(x\cdot k)\}$ ($k \in \frac{2\pi}{2N}\mathbb{Z}^d$)

s.t. $\sup_{x \in [-N,N]^d} |f_{\varepsilon_2}(x) - f(x)| < \varepsilon_2$

since f_{ε_2} is $2N$-periodic we have



In here $|f - f_{\varepsilon_2}| < \varepsilon_2$

out here $|f_{\varepsilon_2}| \le \|f\|_\infty + \varepsilon_2$

$\therefore$

$$\left| \int_{\mathbb{R}^d} f_{\varepsilon_2}\,dP - \int_{\mathbb{R}^d} f\,dP \right|$$

$$\le \int_{[-N,N]^d} |f_{\varepsilon_2} - f|\,dP + \int_{\mathbb{R}^d \setminus [-N,N]^d} |f_{\varepsilon_2}|\,dP$$

$$\le \varepsilon_2 P([-N,N]^d) + (\|f\|_\infty + \varepsilon_2)\varepsilon_1$$

$$\le \underbrace{\varepsilon_2 + (\|f\|_\infty + \varepsilon_2)\varepsilon_1}_{\to 0 \text{ as } \varepsilon_1, \varepsilon_2 \to 0}$$

Similar approximation holds for Q.

This gives $(*)$ as was to be shown.

Finally for the mononialis $m(v)$ the same proof works except that the term

$$\underbrace{\int_{\mathbb{R}^d \setminus [-N,N]^d} |f_{\varepsilon_2}| dP}_{\text{... but if } P \text{ and } Q \text{ have compact support, this term is zero for large enough } N.}$$ could be infinite $\forall N$ & ε_2

QED

Application: Multivariate charateristic functions

In Lecture 12 we defined the characteristic function for a univariate r.v. X

$$\phi_X(k) := E e^{ikX} \quad \text{for } k \in \mathbb{R}.$$

The advantage for studying properties of ϕ_X vrs PX^{-1} is that ϕ_X is a regular function $\mathbb{R} \to \mathbb{C}$ whereas working with PX^{-1} is hard since it maps $\mathcal{B}(\mathbb{R}) \to \mathbb{R}$.

The characteristic function for a random vector X

$$\underset{\mathcal{F}, P}{\Omega} \xrightarrow{X(\omega)} \mathbb{R}^d$$

is similarly defined

$$\phi_X(k) = E\left(e^{ik\cdot X}\right) \quad \forall k \in \mathbb{R}^d$$

The following theorem shows that $\phi_X(\cdot)$ encodes all the information about the randomness in X.

Theorem 2: Let X & Y be random vectors taking values in $\mathbb{R}^d$ with characteristic functions ϕ_X & ϕ_Y. If $\phi_X(k) = \phi_Y(k)$ $\forall k \in \mathbb{R}^d$ then

$$X \overset{\mathcal{D}}{=} Y.$$

Proof:

$\phi_X(k) = \phi_Y(k)$ $\forall k \in \mathbb{R}^d$ implies

$$\begin{aligned} E\cos(k\cdot X) &= \text{Real}\, \phi_X(k) \\ &= \text{Real}\, \phi_Y(k) \\ &= E\cos(k\cdot Y) \end{aligned}$$

and

$$\begin{aligned} E\sin(k\cdot X) &= \text{Imag}\, \phi_X(k) \\ &= \text{Imag}\, \phi_Y(k) \\ &= E\sin(k\cdot Y) \end{aligned}$$

$\therefore$ $Ef(X) = Ef(Y)$ $\forall f \in \Gamma$ where Γ is the separating class over $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ given in Theorem 1.

$\therefore$ $X \overset{\mathcal{D}}{=} Y$. QED.

This also gives the following corollary

Corollary 1:

Let X & Y be random vectors taking values in $\mathbb{R}^d$. Then

$$X \overset{\mathcal{D}}{=} Y \iff k\cdot X \overset{\mathcal{D}}{=} k\cdot Y \quad \forall k \in \mathbb{R}^d.$$

Application: Moments characterizing X

Also in lecture 12 we saw that the moment generating function

$$M_X(t) := Ee^{tX}$$

can be used to recover $\phi_X(t)$ when $M_X(t)$ is finite on a non-empty open interval. Combined with Theorem 2 we get the following corollary

Corollary 2: If X is a r.v. then $M_X(t)$ characterizes the distribution of X whenever $M_X(t)$ is finite on a non-empty open interval.

We also noticed a slightly weaker result that the moments of X, i.e. EX^n, can be used to recover $\phi_X(t)$ only when $M_X(t)$ is finite on a non-empty open interval containing 0.

We get an illustration of this and an extension to bdd multivariate r.v.s via separating classes

Theorem 3: Let $X=(X_1,\dots,X_d)$ and $Y=(Y_1,\dots,Y_d)$ be real random variables which are bounded. Then

$$X \overset{\mathcal{D}}{=} Y \iff E(X_1^{\alpha_1}\cdots X_d^{\alpha_d}) = E(Y_1^{\alpha_1}\cdots Y_d^{\alpha_d}) \quad \forall\, \alpha_1,\dots,\alpha_d \in \{0,1,2,\dots\}$$

Proof: Similar to Proof of Theorem 2 using $\Gamma := \{\text{monomials}\}$ from Theorem 1.

Finite Dimensional Distributions of a stochastic Process on $[0,1]$

Let $(Z_t(\omega))_{t\in[0,1]}$ be a collection of random variables indexed by $t\in[0,1]$ all defined on a probability space $(\Omega,\mathcal{F},P)$.

For a fixed t here is the picture

$$\Omega_{\mathcal{F},P} \xrightarrow{Z_t(\omega)} \mathbb{R}_{\mathcal{B}(\mathbb{R})}$$

Suppose, in addition, if we fix $\omega\in\Omega$ then $Z_t(\omega)$ is continuous in t.

$$\therefore \quad \Omega_{\mathcal{F},P} \xrightarrow{(Z_t)_{t\in[0,1]}} C[0,1]$$

Question:

Is $(Z_t)_{t\in[0,1]}$ (m) $\mathcal{F}/\mathcal{B}(C[0,1])$ with the sup-norm metric on $C[0,1]$?

To answer this first note that $C[0,1]$ is separable and complete w.r.t. See Billingsley.

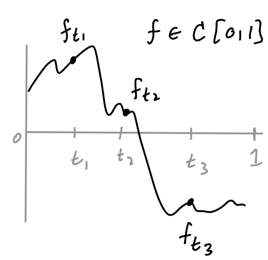
$$d(f,g) := \sup_{t\in[0,1]} |f_t - g_t|$$

$\therefore$ $(C[0,1], \mathcal{B}(C[0,1]))$ is Polish

Second consider the finite coordinate projections

$$\pi_{t_1,\dots,t_n}(f) := (f_{t_1},\dots,f_{t_n})$$

defined on $C[0,1]$.



$$\xmapsto{\pi_{t_1,t_2,t_3}} (f_{t_1},f_{t_2},f_{t_3}) \in \mathbb{R}^3$$

Since we are working with sup-norm its clear that $\pi_{t_1,\dots,t_n}$ is a continuous map from $C[0,1]$ to $\mathbb{R}^n$, hence (m)

$\therefore \; \pi_{t_1,\dots,t_n}^{-1}(A) \in \mathcal{B}(C[0,1]), \; \forall A \in \mathcal{B}(\mathbb{R}^n)$

Theorem (finite projections of $C[0,1]$)

$$\Pi := \{\pi_{t_1,\dots,t_n}^{-1}(A) \mid A \in \mathcal{B}(\mathbb{R}^n),\ t_1,\dots,t_n \in [0,1]\ \&\ n \in \mathbb{N}\} \subset \mathcal{B}(C[0,1])$$

forms a π-system

$$\sigma\langle \Pi \rangle = \mathcal{B}(C[0,1]).$$

Proof:

The π-system property follows since

$$\pi_{t_1,\dots,t_n}^{-1}(A) \cap \pi_{s_1,\dots,s_m}^{-1}(B) = \left\{ f : \begin{matrix} (f_{t_1},\dots,f_{t_n}) \in A \\ \& \\ (f_{s_1},\dots,f_{s_m}) \in B \end{matrix} \right\}$$

$$= \pi_{t_1,\dots,t_n,s_1,\dots,s_m}^{-1}(\underbrace{A \times B}_{\in \mathcal{B}(\mathbb{R}^{n+m})})$$

if coordinates overlap this becomes $(A \times \mathbb{R}^i) \cap (B \times \mathbb{R}^j)$

We already showed $\Pi \subset \mathcal{B}(C[0,1])$ by continuity.

$\therefore \; \sigma\langle \Pi \rangle \subset \mathcal{B}(C[0,1])$.

To show $\mathcal{B}(C[0,1]) \subset \sigma\langle \Pi \rangle$ it will be sufficient to show the open balls

$$B_\varepsilon(f) \in \sigma\langle \Pi \rangle.$$

Indeed

$$B_\varepsilon(f) = \{g \in C[0,1] : \sup_{t \in [0,1]} |f_t - g_t| < \varepsilon\}$$

$$= \bigcap_{t \in [0,1] \cap \mathbb{Q}} \underbrace{\{g \in C[0,1] : |g_t - f_t| < \varepsilon\}}_{\pi_t^{-1}((f_t - \varepsilon, f_t + \varepsilon))}$$

$$\in \sigma\langle \Pi \rangle$$

QED

Therefore to check Z is (m) just check

$$Z^{-1}(\pi_{t_1,\dots,t_n}^{-1}(A)) \in \mathcal{F} \quad \forall A \in \mathcal{B}(\mathbb{R}^n).$$

by generators are enough.

Notice that

(*) $$Z^{-1}(\pi_{t_1,\dots,t_n}^{-1}(A)) = \{ \underbrace{(Z_{t_1},\dots,Z_{t_n})}_{\text{is (m) since each } Z_{t_i} \text{ is (m)}} \in A \} \in \mathcal{F}.$$

$\therefore$ the map

$$\omega \in \Omega \longmapsto (Z_t(\omega))_{t \in [0,1]} \in C[0,1]$$

is indeed measurable. ← Answers the question

Moreover the induced distribution of $(Z_t)_{t \in [0,1]}$ on $(C[0,1], \mathcal{B}(C[0,1]))$ is characterized by the probabilities of sets of the form (*) by π-uniqueness.

This proves the following theorem.

Theorem (finite dimensional distributions)

If $(Z_t)_{t \in [0,1]}$ and $(W_t)_{t \in [0,1]}$ are two continuous stochastic processes s.t.

(*) $$(Z_{t_1}, Z_{t_2}, \dots, Z_{t_n}) \overset{\mathcal{D}}{=} (W_{t_1}, W_{t_2}, \dots, W_{t_n})$$

$\forall n > 0,\ t_1, \dots, t_n \in [0,1]$ then Z and W induce the same probability measures on $(C[0,1], \mathcal{B}(C[0,1]))$.

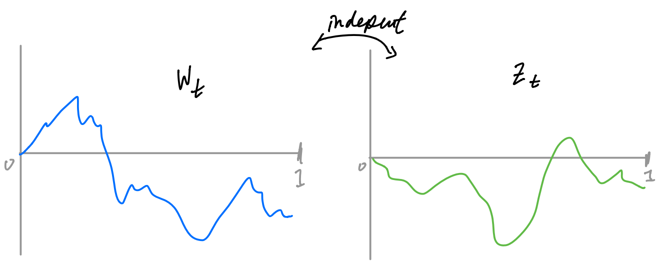
Remark: The collection of all distributions corresponding to $(Z_{t_1},\dots,Z_{t_n})$ $\forall n > 0$ & $\forall t_1,\dots,t_n \in [0,1]$ is called the finite dimensional distributions of $(Z_t)_{t \in [0,1]}$

(f.d.d for short).

Example:

Suppose you proved the Wiener process W_t, discussed in the introduction, exists as a scaled limit of random walks.

By using two independent sequences of Radamacher r.v.s you could then generate two independent Wiener processes



$E(W_t) = 0$

$\mathrm{cov}(W_t, W_s) = s \wedge t$

independent increments

f.d.d of W_t are Gaussian

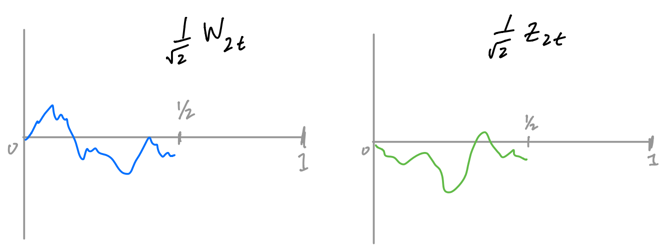
$E(Z_t) = 0$

$\mathrm{cov}(Z_t, Z_s) = s \wedge t$

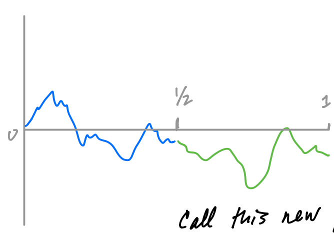
independent increments

f.d.d. of Z_t are Gaussian

Now scale Z_t & W_t as follows



Then link them at the endpoints



Call this new process $(B_t)_{t \in [0,1]}$

I now claim B_t is another Weiner process with probabilities identical to W_t or Z_t.

For example

$$P\Big(\sup_{t\in[0,1]} B_t \geq 10\Big) = P\Big(\sup_{t\in[0,1]} W_t \geq 10\Big)$$

$$P\Big(\int_0^1 \sin^2(B_t)\,dt < 0.25\Big) = P\Big(\int_0^1 \sin^2(Z_t)\,dt < 0.25\Big)$$

$\vdots$

etc.

To prove this all need to show is that the f.d.d match, i.e. that

$$(*) \quad (B_{t_1}, \ldots, B_{t_n}) \overset{\mathcal{D}}{=} (W_{t_1}, \ldots, W_{t_n})$$

$\forall t_1, \ldots, t_n \in [0,1]$.

I'll leave this for you to verify as an exercise.