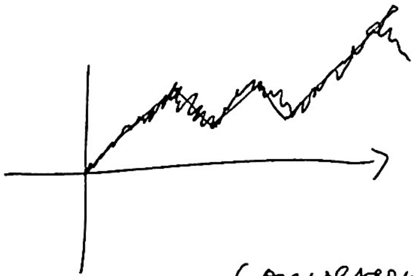
Lecture 2 : Invariance Principle



convergence of RW to BM.

We will discuss convergence of metric-space-valued random variables.

I. Weak Convergence

<u>Def</u> $\mathcal{X}$: topological space (metric space)

$C_b(\mathcal{X}) = \{$ bounded continuous functions $\}$

$X_n : \Omega \to \mathcal{X}$ $\quad X_n \rightharpoonup X$ (convergence in law) means

$\forall f \in C_b(\mathcal{X})$ $\quad \mathbb{E}[f(X_n)] \longrightarrow \mathbb{E}[f(X)] \iff \mu_n(f) \to \mu(f).$

For convergence of curves, take $\mathcal{X} = C([0,T], \mathbb{R})$

Recall convergence a.s.

$\Downarrow$

convergence in $\mathbb{P}$

$\Downarrow$

convergence in law.

<u>Prop</u> (Portmanteau).

The following conditions are equivalent.

① $\mu_n \rightharpoonup \mu$.

② $\mu_n(f) \to \mu(f)$ $\quad \forall f$ C_b unif. continuous

③ $\forall F$ closed in $\mathcal{X}$ $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$

④ $\forall O$ open in $\mathcal{X}$ $\liminf_{n\to\infty} \mu_n(O) \geq \mu(O)$.

⑤ $\forall A$ measurable, $\mu(\partial A) = 0$, $\lim_{n\to+\infty} \mu_n(A) = \mu(A)$.

Pf ① $\Rightarrow$ ② Trivial

③ $\cancel{\Leftarrow}\Rightarrow$ ④ :

$$\liminf_{n\to\infty} \mu_n(O)$$
$$= 1 - \limsup_{n\to\infty} \mu_n(O^c)$$
$$\geq 1 - \mu(O^c)$$
$$= \mu(O)$$

④ $\Rightarrow$ ③ Exactly the same.

③ + ④ $\Rightarrow$ ⑤.

$$\mu(\bar{A}) \geq \limsup_{n\to\infty} \mu_n(\bar{A}) \geq \limsup_{n\to\infty} \mu_n(A) \geq \liminf_{n\to\infty} \mu_n(A)$$
$$\geq \liminf_{n\to\infty} \mu_n(A^\circ) \geq \mu(A^\circ)$$

Since $\mu(\partial A) = 0$,

$$\mu(\bar{A}) = \mu(A^\circ) = \mu(A) = \limsup \mu_n(A) = \liminf \mu_n(A)$$

$$\mu(A) = \lim_{n\to+\infty} \mu_n(A).$$

Standard example for strict inequality :

$$\mu_n = \delta_{\frac{1}{n}} \to \delta_0 \qquad O = (0,1)$$

② $\Rightarrow$ ③

Let $F = 1_F$ ~~..~~ . We want to approximate F by C_b, unif. cont. functions. Let $f_\delta(x) = \frac{1}{\delta}(\delta - d(x,F))_+ = (1 - \delta^{-1} d(x,F))_+$

f_δ is obviously bounded and continuous.

$$|f_\delta(x) - f_\delta(y)| \leq \delta^{-1} |d(x,F) - d(y,F)|$$
$$\leq \delta^{-1} d(x,y).$$

So f_δ is Lipschitz $\Rightarrow$ uniformly continuous. So $\lim_{n\to\infty} \mu_n(f_\delta) = \mu(f_\delta)$.

~~$\leq \mu(F) + \varepsilon$ for~~

$\mu(f_\delta) \leq \mu(F) + \varepsilon$ for $\delta \searrow 0$ by dominated convergence

$\lim\limits_{n\to\infty} \mu_n(f_\delta) \geq \limsup\limits_{n\to+\infty} \mu_n(F)$ since $f_\delta \geq F$.

Hence, $\limsup\limits_{n\to+\infty} \mu_n(F) \leq \mu(F) + \varepsilon \quad \forall \varepsilon > 0$.

So ③ is obtained.

⑤ $\Longrightarrow$ ①

Suppose f is positive, $0 \leq f \leq M$.

$$\mu_n(f) = \int_0^M \mu_n(f > t)\, dt \quad \text{(Standard corollary of Fubini)}$$

$$\downarrow$$

$$= \int_0^M \mu(f > t)\, dt \quad \text{by dominated convergence,}$$

(Dominated by 1) contingent on $\mu_n(f > t) \longrightarrow \mu(f > t)$.
How to get this convergence. To apply ⑤, we need $\partial\{f > t\}$ to have zero μ-measure.

Define RV $X \sim \mu$. Then $\mu(\{f > t\}) = \mathbb{P}[\underbrace{f(X)}_{=:Y} > t]$ $= 1 - F_Y(t)$. Recall that a CDF has at most countably many points of discontinuity.

So it is clear that $\mu_n(f > t) \to \mu(f > t)$ except countably many points. ▨

> **Def** (Tightness) $(\mu_n)_{n\in\mathbb{N}}$ family of probability measures on $\mathcal{X}$, they are tight iff $\forall \varepsilon > 0$ $\exists K_\varepsilon$ compact in $\mathcal{X}$ s.t.
> $\mu_n(K_\varepsilon) \geq 1 - \varepsilon, \quad \forall n \in \mathbb{N}$.

Def $\mathcal{X}$ is a Polish space iff. $\exists$ metric d on $\mathcal{X}$ s.t.

① $(\mathcal{X}, d)$ is complete

② $(\mathcal{X}, d)$ is separable

Let $\mathcal{M}_1(\mathcal{X}) = \{$ Prob measures on $\mathcal{X} \}$

Thm If $\mathcal{X}$ is Polish, then $\mathcal{M}_1(\mathcal{X})$ is also Polish.
In particular, if $\mathcal{X}$ is compact, then $\mathcal{M}_1(\mathcal{X})$ is also compact.

What is the metric on $\mathcal{M}_1(\mathcal{X})$? Levy distance. We have ~~talked~~ about other ~~measures~~ distances on measures, such as total variation and Wasserstein distance. These generate stronger topologies. Levy distance generates exactly the weak topology. Generally, this distance is difficult to work with.

e.g. $\mathcal{X}$ Polish. X_n is $\mathcal{X}$-valued. $\hat{\mu} = \frac{1}{n}\sum_{k=1}^{n} X_k \in \mathcal{M}_1(\mathcal{X})$
This is a random measure, with $\mathrm{Law}(\hat{\mu}) \in \mathcal{M}_1(\mathcal{M}_1(\mathcal{X}))$.

Thm (Prokhrov). $\mathcal{X}$ Polish, $(\mu_n)_{n\in\mathbb{N}} \subseteq \mathcal{M}_1(\mathcal{X})$, then they are relatively compact $\Longleftrightarrow$ they are tight.

Relative compact: their limit (subsequential) is not necessarily in this family, but ~~exits~~ exists in ~~$(\mu_n)_{n\in\mathbb{N}}$~~ $\mathcal{M}_1(\mathcal{X})$ by Polish property.

Rf: $\mathcal{X} = \mathbb{R}$, $(X_n)_{n\in\mathbb{N}}$ from F_{X_n}, we can construct

$F \nearrow$ (Helly's selection ~~pro~~ principle, diagonal argument).

We can find the ~~dist~~ "CDF" of the limit. But is it necessarily a CDF? No

Exa $X_n = n$. Then $\mu_n = \delta_n$. So this mass keeps drifting to the right. So $F = 0$. (vague convergence).

Clearly, one cannot reconstruct a random variable from ~~this~~. However, the tightness assumption forbids such a loss of mass phenomenon.

$$F(K) \overset{\sim}{\cancel{=}} \lim_{n\to+\infty} F_{X_n}(K) \geq 1-\varepsilon.$$

So this can give us a CDF and we can reconstruct a RV for this (standard).

Thm (Banach–Alaoglu).
Let ~~B~~ V be a normed vector space, V^* its analytic ~~dual~~ dual. Then put on it a weak-$*$ topology. $f_n \xrightarrow{*} f \in V^*$ if $\forall x \in V$, $f_n(x) \longrightarrow f(x)$
On V^*, the unit ball is compact under ~~*~~ the weak-$*$ topology.

How does this functional analytic theorem relates to probability? Our weak topology on $\mathcal{M}_1(\mathcal{X})$ is really a weak-$*$ topology.
Exer Let $V = C(\mathcal{X}, \|\cdot\|_{L^\infty})$

$$V^* = \underbrace{C^*(\mathcal{X}, \|\cdot\|_{L^\infty})}_{=\{\text{bdd lin functionals}\}} \supsetneq \mathcal{M}_1(\mathcal{X})$$

So this C^* ~~is isomorphic to~~ consists of signed measures on $\mathcal{X}$. Is ~~this space~~ compact? The space V^* is weak-$*$-compact. But $\mathcal{M}_1(\mathcal{X})$ this doesn't mean that $\mathcal{M}_1(\mathcal{X})$ is compact, ~~even though~~ it is just a subset. It would be if it is a <u>closed</u> subset. But is it closed? Not necessarily, not if the ambient space $\mathcal{X}$ is not compact.

When $\mathcal{X}$ is compact, then $C^*(\mathcal{X}, \|\cdot\|_\infty) = \{$ finite signed measures $\}$ and it is true that $\mathcal{M}_1(\mathcal{X})$ is compact.

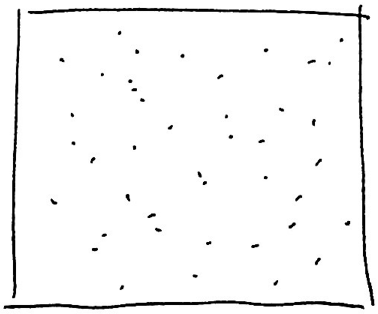
So this suggests that the compactness of $\mathcal{X}$ ~~is truly nec~~, and hence that of $\mathcal{M}_1(\mathcal{X})$ is truly necessary for Prokhorov.

<u>Prf</u> (Prokhorov).

($\Longrightarrow$) Suppose that $(\mu_n)_{\mathbb{N}}$ is relatively compact.

- ~~Obs~~ Observation: Let (x_k) dense subset of $\mathcal{X}$. $\forall \varepsilon > 0\ \forall k$, $\exists\, \phi(k) \in \mathbb{N}$ s.t. $\forall \mu_n \in (\mu_n)_{n\in\mathbb{N}}$, we have

$$\mu_n\Big(\bigcup_{m \le \phi(k)} B_{2^{-k-1}}(x_m)\Big) \ge 1 - \frac{\varepsilon}{2^{k+1}}$$



$\mathcal{X}$ If we ε-thicken (x_k), then $\mathcal{X}$ is covered. So we should cover all the probabilities/masses. Right now, we only want to approximate the measure.

- <u>Prf</u> of observation. Suppose not. Then $\exists\, k, \varepsilon$, and a sequence $(\mu_n)_{n\in\mathbb{N}}$ with (take $\phi(k) = k$)

$$\mu_n\Big(\bigcup_{m \le n} B_{2^{-k-1}}(x_m)\Big) < 1 - \frac{\varepsilon}{2^{k+1}}$$

By relative compactness, $\mu_n \to \mu$ for some probability measure μ.

Recall Portmanteau: $\mu\Big(\bigcup_{n \le n} B_{2^{-k-1}}(x_m)\Big) \le \liminf_{n\to} \nu \le 1 - \frac{\varepsilon}{2^{k+1}}$

We let $n \to +\infty$, then $\mu(\mathcal{X}) = \mu\Big(\bigcup^{+\infty} B_{2^{-k-1}}(x_m)\Big) \le 1 - \frac{\varepsilon}{2^{k+1}}$

Now, we want to find the compact ~~set~~ set K_ε for tightness.

$$K_\varepsilon := \bigcap_{k=1}^{+\infty} \bigcup_{m \le \phi(k)} B_{2^{-k-1}}(x_m)$$

$$\mu_n(K_\varepsilon^c) = \mu_n\left(\bigcup_{k=1}^{+\infty} \left(\bigcup_{m\in\phi(k)} B_{2^{-k-1}}(x_m)\right)^c \right)$$

$$\le \sum_{k=1}^{+\infty} 1 - \left(1 - \frac{\varepsilon}{2^{k+1}}\right)$$

$$= \varepsilon$$

So $\mu_n(K_\varepsilon) \ge 1-\varepsilon$.

Why is K_ε compact? Recall Analysis I.

$(\Leftarrow)$ Idea: Study the problem in a compact space ~~$\mathcal{X}$~~. $\mathcal{Y}$.

Lemma $\mathcal{Y} = [0,1]^{\mathbb{N}}$. — compact (Tychonoff)

$\exists$ continuous injection $\Psi : \mathcal{X} \hookrightarrow \cancel{\mathcal{X}}\ \mathcal{Y}$ s.t.

$\mathcal{X}$ homeomorphic to $\Psi(\mathcal{X})$.

If we have a continuous map, then it induces a measure.

Let $\nu \in \mathcal{M}_1(\mathcal{Y})$, $\nu_n = \Psi_* \mu_n$ (pushforward)

~~$\forall A \in \mathcal{B}(\mathcal{Y})$, ν~~

$$\forall f \in C_b(\mathcal{Y}), \quad \int_{\mathcal{Y}} f\, d\nu_n := \int_{\mathcal{X}} (f\circ\Psi)\, d\mu_n.$$

Because $\mathcal{Y}$ is compact, $\mathcal{M}_1(\mathcal{Y})$ is also compact.

$\Rightarrow (\nu_n)_{n\in\mathbb{N}}$ has limit distribution.

So ν is a Probability ~~that~~ measure. Are we done? No, we have not used the tightness ~~cond~~ condition. Where is the problem? ~~ν might~~

ν is a ~~m~~ p-measure on $\mathcal{Y}$, ~~not~~ not necessarily on $\Psi(\mathcal{X})$.
We need tightness to ensure that $\nu(\Psi(\mathcal{X}))=1$.
Then we can find the limit

$$\mu := \Psi^{-1}_{*}(\nu) \quad \text{in the sense}$$

$$\forall g \in C_b(\mathcal{X}), \quad \int_{\mathcal{X}} g \, d\mu = \int_{\mathcal{Y}} (g \circ \Psi)^{-1} \, d\nu .$$

$\forall \varepsilon > 0 \ \exists K_\varepsilon \subseteq \mathcal{X}$ s.t.

$$\sup_{n \in \mathbb{N}} \mu_n(K_\varepsilon) \geq 1-\varepsilon .$$

Then $\nu_n(\Psi(K_\varepsilon)) = \mu_n(K_\varepsilon) \geq 1-\varepsilon$.

Portmanteau $\nu(\mathcal{X}) \geq \nu(\Psi(K_\varepsilon)) = \limsup_{n\to\infty} \nu_n(\Psi(K_\varepsilon)) \geq 1-\varepsilon$. $\forall \varepsilon > 0$.

<u>Pf of Lem</u>

$\mathcal{X}$ separable $\Rightarrow$ have countable dense subset.
Usual metric spaces don't have coordinates. But on a separable metric space, the countable dense subset specifies the point if we know its distance from all of these points.

Let $(x_k)_{k \in \mathbb{N}}$ dense in $\mathcal{X}$.
$\forall x \in \mathcal{X}$ we define $\Psi(x) := (d(x, x_k) \wedge 1)_{k \in \mathbb{N}} \in \mathcal{Y}$.

Continuity of Ψ

On $\mathcal{Y}$, we can place a metric

$$d_{\mathcal{Y}}\left((a_n)_{n\in\mathbb{N}}, (b_n)_{n\in\mathbb{N}}\right) = \sum_{n=0}^{+\infty} \frac{|a_n - b_n|}{2^{n+1}}.$$

$$d_{\mathcal{Y}}\left(\Psi(x), \Psi(x')\right) \le \sum_{k=0}^{+\infty} \frac{\left|d_{\mathcal{X}}(x, x_k) - d_{\mathcal{X}}(x', x_k)\right|}{2^{n+1}}$$

$$\le \sum_{k=0}^{+\infty} \frac{d_{\mathcal{X}}(x, x')}{2^{n+1}}$$

$$\le d_{\mathcal{X}}(x, x')$$

So Ψ is even Lipschitz.

Injectivity

Let $x \neq x'$. Then $\exists\, x_k$ s.t. $d_{\mathcal{X}}(x, x_k) \neq d_{\mathcal{X}}(x', x_k)$ by density.

So ~~Even better,~~ $\Psi : \mathcal{X} \to \Psi(\mathcal{X})$ is a bijection.

~~Part~~ $\mathcal{C}([0,1], \mathbb{R})$.

Prop ~~Ude~~ Under the distance

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

$C([0,1])$ is a Polish space.

Pf Complete — see Analysis I.

Separable — Stone-Weierstrass Theorem.

To use our previous theory, we want to have compact sets. What sets are compact on $C[0,T]$?

<u>Thm</u> (Arzela - Ascoli)

$$\omega(f,\delta) := \sup_{\substack{s,t\in[0,T]\\ |s-t|<\delta}} |f(s)-f(t)|$$

$(f_n)_{n\in\mathbb{N}}$ is relatively compact in $C[0,T]$

$\iff \begin{cases} (f_n(0))_{n\in\mathbb{N}} \text{ is uniformly bounded on } [0,T] \\ \lim\limits_{\delta\to 0} \sup\limits_{n\in\mathbb{N}} \omega(f_n,\delta) = 0. \end{cases}$

<u>Prop</u> (Criteria for Tightness).

$(\mu_n)_{n\in\mathbb{N}}$ is tight on $\mathcal{M}_1(C[0,T])$ iff.

$$\begin{cases} \lim\limits_{A\to+\infty} \sup\limits_{n} \mu_n(f : |f(0)| > A) = 0 & \text{— ①} \\ \forall \varepsilon \ \lim\limits_{\delta\to 0} \sup\limits_{n} \mu_n(f : \omega(f,\delta) > \varepsilon) = 0 & \text{— ②} \end{cases}$$

<u>Pf</u> These conditions correspond to the conditions in Arzela-Ascoli.

From ①: $\forall \varepsilon' \ \exists A_{\varepsilon'}$ s.t.

$$\sup_n \mu_n(f : |f(0)| > A_\varepsilon) \le \frac{\varepsilon'}{2}$$

$\exists \ \delta_{\varepsilon'}$ s.t.

$$\sup_n \mu_n(f : \omega(f,\delta_{\varepsilon'}) > \varepsilon) \le \frac{\varepsilon'}{2}$$

$\Rightarrow \forall n \ \mu_n(f : |f(0)| \le A_{\varepsilon'}, \omega(f,\delta) \le \varepsilon) \ge 1-\varepsilon'.$

Prop (Criterion Kolmogorov - Chentsov).

$(X^{(n)}) \subseteq C[0,T]$ ~~~~ (μ_n) is tight on $\mathcal{M}_1(C[0,T])$ if it satisfies

$X^{(n)}(0)$ is tight in $\mathbb{R}$

$\forall\ \exists\ a \geq 1, b > 1$ s.t. $\forall\ s,t \in [0,T]$, $\forall n$

$$\mathbb{E}\left[\left|X^{(n)}(s) - X^{(n)}(t)\right|^a\right] \leq C|t-s|^b$$

Pf $\mathbb{E}[\omega(X^{(n)}, \delta)] - ?$

Step 1 : ~~Let T~~ D_k a dyadic partition.

Let $T=1$ $D_k = \left\{ \frac{0}{2^k}, \dots, \frac{2^k}{2^k} \right\}$

$t_k := \underbrace{0, t^1, \dots, t^k}_{k-\text{tuple}} \mid \dots$

$$X^{(n)}(t) = X^{(n)}(t_0) + \sum_{\ell=0}^{+\infty} \left(X^{(n)}(t_{\ell+1}) - X^{(n)}(t_\ell)\right)$$

$$= X^{(n)}(t_k) + \sum_{\ell=k}^{+\infty} \left(X^{(n)}(t_{\ell+1}) - X^{(n)}(t_\ell)\right).$$

$\xi_k^{(n)} := \sup_{a \in D_k} \left|X^{(n)}(a+2^{-k}) - X^{(n)}(a)\right|$ Jump in state 2^{-k}

$$X^{(n)}(t) - X^{(n)}(s) = \begin{matrix} X^{(n)}(t_k) \\ -X^{(n)}(s_k) \end{matrix} + \sum_{\ell=k}^{+\infty} \left(X^{(n)}(t_{\ell+1}) - X^{(n)}(t_\ell)\right) + \sum \cdots$$

$\forall\ |t-s| \leq 2^{-k}$

$$\left|X^{(n)}(t) - X^{(n)}(s)\right| \leq 4 \sum_{\ell \geq k} \xi_k^{(n)}$$

$$\Longrightarrow \omega(X^{(n)}, 2^{-k}) \leq 4 \sum_{\ell \geq k} \xi_k^{(n)}.$$

ℓ^p triangle inequality

$$\mathbb{E}^{\frac{1}{a}}\left[(\omega(X^{(n)},2^{-k}))^a\right] \le 4\sum_{\ell\le k}\mathbb{E}^{\frac{1}{a}}\left((\xi_k^{(n)})^a\right)$$

$$\mathbb{E}\left[(\xi_k^{(n)})^a\right] \le \sum_{a\in D_k}\underbrace{\mathbb{E}\left[\left|X^{(n)}(a+2^{-k})-X^{(n)}(a)\right|^a\right]}$$

$$\le 2^k(2^{-k})^b$$

$$\mathbb{E}\left[(\xi_k^{(n)})^a\right] \le 2^{-k\underbrace{(b-1)}_{>0}}$$

$$\mathbb{E}^{\frac{1}{a}}\left[(\omega(X^{(n)},2^{-k}))^a\right] \le 4\sum_{\ell\ge k} 2^{-\frac{\ell(b-1)}{a}}$$

$$\le C\,2^{-\frac{k(b-1)}{a}}$$

$$\Longrightarrow \mathbb{E}\left[(\omega(X^{(n)},2^{-k}))^a\right] \le C\,2^{-k(b-1)}$$

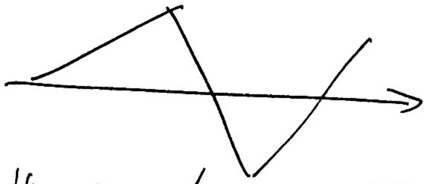
$$\forall\varepsilon>0,\ \lim_{k\to+\infty}\sup_n \mathbb{P}\left[\omega(X^{(n)},2^{-k})>\varepsilon\right]$$

$$\le \lim_{k\to+\infty}\frac{C\,2^{-k(b-1)}}{\varepsilon^a} \quad \text{(Markov inequality)}$$

$$=0.$$

~~Thm (Donsker).~~ Donsker's theorem

$S_n \longrightarrow$ (a_i) is a set of points.

But we treat it as S_n 

We assume existence of 4th moment. — (not necessary)

① $\text{Moment}_4 \Longrightarrow$ tightness

② Finite dim CLT (monotone class)

show that what the trajectory converges to is indeed the Brownian motion.