$f : S \to \mathbb{R}$

$f(X_n)$ random variable

for every continuous real function f. Note that the limit measure is necessarily a probability measure.

Let X_n and X be random elements in S, defined possibly on different probability spaces. The sequence (X_n) *converges in distribution to* X if the corresponding induced measures converge weakly. Equivalently,

$$\mathbf{E}f(X_n) \to \mathbf{E}f(X)$$

$\omega : [0,\infty) \to \mathbb{R}$

ω is cont.

for all continuous and bounded f.

(S, ϱ)

$(S, \mathcal{B}(S))$

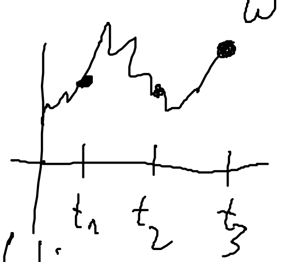
Assume that $X_n \to X$ in distribution. For any $0 \le t_1 < \ldots < t_k$ consider the projection $\pi_{t_1,\ldots,t_k} : C[0,\infty) \to \mathbb{R}^k$

$$\pi_{t_1,\ldots,t_k}(\omega) = (\omega(t_1), \ldots, \omega(t_k)) .$$

$(\omega(t_1), \omega(t_2), \omega(t_3))$

This is clearly continuous. For a continuous bounded function $f : \mathbb{R}^k \to \mathbb{R}$ the composite function $f(\pi_{t_1,\ldots,t_k})$ is bounded and continuous. Therefore, by the definition of convergence in distribution

$$\mathbf{E}f(\pi_{t_1,\ldots,t_k}(X_n)) \to \mathbf{E}f(\pi_{t_1,\ldots,t_k}(X))$$



that is

$$\mathbf{E}f(\overbrace{X_n(t_1), \ldots, X_n(t_k)}^{k \text{ dimensional random vector}}) \to \mathbf{E}f(X(t_1), \ldots, X(t_k)).$$

That is, for any $0 \le t_1 < \ldots < t_k$

convergence in distribution in $\mathbb{R}^k$

$$(X_n(t_1), \ldots, X_n(t_k)) \xrightarrow{\mathcal{D}} (X(t_1), \ldots, X(t_k)).$$

This means that the finite dimensional distributions converge.

We proved the following.

Proposition 8. *If (X_n) converges in distribution to X then all finite dimensional distributions converge.*

$\forall k \in \mathbb{N}$

$\forall\, 0 \le t_1 < \ldots < t_k < \infty$

...

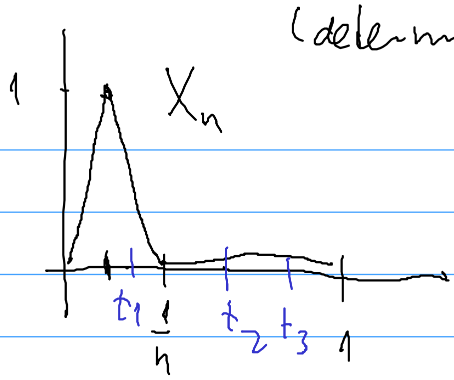
The converse is not true in general.

Example 8. Let

$$X_n(t) = nt\mathbf{I}_{[0,(2n)^{-1}]}(t) + (1 - nt)\mathbf{I}_{((2n)^{-1},n^{-1}]}(t), \quad t \ge 0.$$

Then all finite dimensional distributions converge to the corresponding finite dimensional distributions of $X \equiv 0$. However, convergence as a process does not hold.

(determin)



as $n \to \infty$

$$(X_n(t_1), \dots, X_n(t_k)) \xrightarrow{\mathcal{D}} (0, 0, \dots, 0)$$

$$f : C[0,1] \to \mathbb{R} \qquad f(\omega) = \max_{t \in [0,1]} \omega(t)$$

No convergence in distribution in $C[0,1]$.

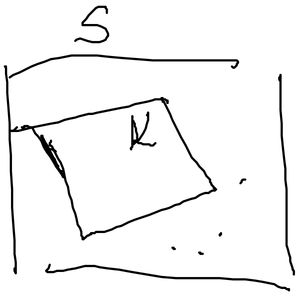
Convergence in distribution in $C[0,\infty)$

$$\Updownarrow$$

can handle this part → Convergence of finite dimensional distribution

+ ? something extra

"tightness"



tightness ~ „the mass does not go away"

fences

In what follows we try to understand what goes wrong in the example above, and state a converse of the Proposition above.

A family of probability measures Π on $(S, \mathcal{B}(S))$ is *tight* if for every $\varepsilon > 0$ there exists a compact set $K \subset S$ such that $P(K) \geq 1 - \varepsilon$ for all $P \in \Pi$. The family Π is *relatively compact* if each sequence of elements from Π contains a convergent subsequence. A family of random elements is tight (relatively compact) if the family of induced measures is tight (relatively compact).

"weak convergence"

Theorem 15 (Prohorov)**.** *Let Π be a family of probability measures on a complete separable metric space S. Then Π is tight if and only if it is relatively compact.*

= Bolzano–Weierstrass thm

The modulus of continuity plays an important role in characterization of tightness on C. Fix $T > 0$ and $\delta > 0$, and let $\omega \in C[0, \infty)$. The modulus of continuity on $[0, T]$

$$m^T(\omega, \delta) = \max \left\{ |\omega(s) - \omega(t)| : \ |s - t| \leq \delta, 0 \leq s, t \leq T \right\}.$$

Exercise 20. Show that m^T is continuous in $\omega \in C[0, \infty)$ under the metric ρ, is nondecreasing in δ, and $\lim_{\delta \downarrow 0} m^T(\omega, \delta) = 0$ for each $\omega \in C[0, T]$.

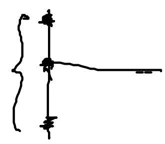
continuous functions are uniformly continuous on compacts

Theorem 16 (Arselà–Ascoli)**.** *A set $A \subset C[0, \infty)$ has compact closure if and only if the following two conditions hold:*

(i) $\sup_{\omega \in A} |\omega(0)| < \infty$*;*

(ii) for every $T > 0$

uniformly in A



$$\lim_{\delta \downarrow 0} \sup_{\omega \in A} m^T(\omega, \delta) = 0.$$

Now we can characterize tightness of probability measures.

{thm:tightness-C}

Theorem 17. *A sequence (P_n) of probability measures on $(C[0, \infty), \mathcal{B})$ is tight if and only if the following two conditions hold:*

(i) $\lim_{\lambda \uparrow \infty} \sup_{n \geq 1} P_n(\omega : \ |\omega(0)| > \lambda) = 0$*;*

(ii) for all $T > 0$ and $\varepsilon > 0$

$$\lim_{\delta \downarrow 0} \sup_{n \geq 1} P_n(\omega : \ m^T(\omega, \delta) > \varepsilon) = 0.$$

{thm:conv-spaceC}

Theorem 18. *Let (X_n) be a tight sequence of continuous processes such that its finite dimensional distributions converge. Then the sequence of induced*

$\mathbb{R}$

Bolzano - Weierstrass

bounded sequence — convergent subsequence

$\updownarrow$

tightness

measures (P_n) converge weakly to a measure P such that the coordinate mapping $W_t(\omega) = \omega_t$ on $C[0,\infty)$ satisfies

$$(X_n(t_1), \ldots, X_n(t_k)) \xrightarrow{\mathcal{D}} (W(t_1), \ldots, W(t_k)),$$

for any $0 \leq t_1 < \ldots < t_k < \infty$, $k \geq 1$.

Proof. Tightness is the same as relative compactness. Therefore, every subsequence contains a further convergent subsequence. Because of the convergence of finite dimensional distributions any two limit measure has the same finite dimensional distributions. But finite dimensional distributions determine the measure. □

3.3 Donsker theorem

Let $\xi, \xi_1, \xi_2, \ldots$ be iid random variables with $\mathbf{E}\xi = 0$, $\mathbf{E}\xi^2 = \sigma^2 \in (0,\infty)$, and let $S_n = \sum_{i=1}^n \xi_i$ denote the partial sum. Define the continuous time process $(Y_t)_{t\geq 0}$ as

$$Y_t = S_{\lfloor t \rfloor} + (t - \lfloor t \rfloor)\xi_{\lfloor t \rfloor + 1},$$

$\lfloor \cdot \rfloor$ integer part

where $\rfloor \cdot \rfloor$ stands for the usual integer part. For $n \in \mathbb{N}$ define the scaled process

$$X_t^{(n)} = \frac{1}{\sigma\sqrt{n}} Y_{nt}, \quad t \geq 0. \longrightarrow P_n$$

Then $X_t^{(n)} - X_s^{(n)}$ for $s, t \in \mathbb{N}/n$ is independent of $\sigma(\xi_1, \ldots, \xi_{sn})$, and by the CLT its distribution tends to $N(0, t-s)$.

Theorem 19 (Invariance principle of Donsker). *Let P_n denote the measure on $(C[0,\infty), \mathcal{B}(C[0,\infty)))$ induced by $X^{(n)}$. Then P_n converges weakly to a measure $P_\star$. Under $P_\star$ the coordinate mapping $W_t(\omega) = \omega(t)$, $\omega \in C[0,\infty)$ is SBM.*

Proof. According to Theorem 18 we have to show that $(X^{(n)})$ is tight and the finite dimensional distributions converge to those of a SBM.

To prove tightness we have to show that the conditions of Theorem 17 hold for P_n. This can be done by proving some maximal inequalities. We skip this part.

We prove the convergence of finite dimensional distributions. Fix $d \in \mathbb{N}$ and $0 \leq t_1 < \ldots < t_d < \infty$. We have to show that

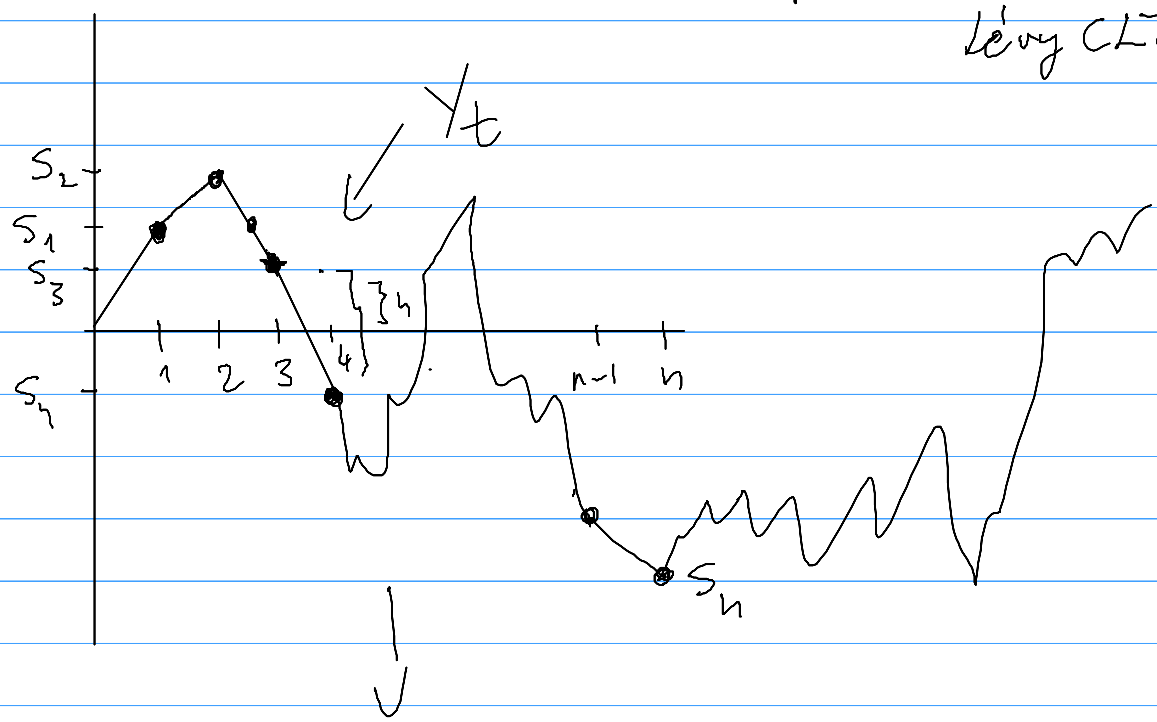
$$\left(X_{t_1}^{(n)}, \ldots, X_{t_d}^{(n)}\right) \xrightarrow{\mathcal{D}} \underbrace{(W_{t_1}, \ldots, W_{t_d})}.$$

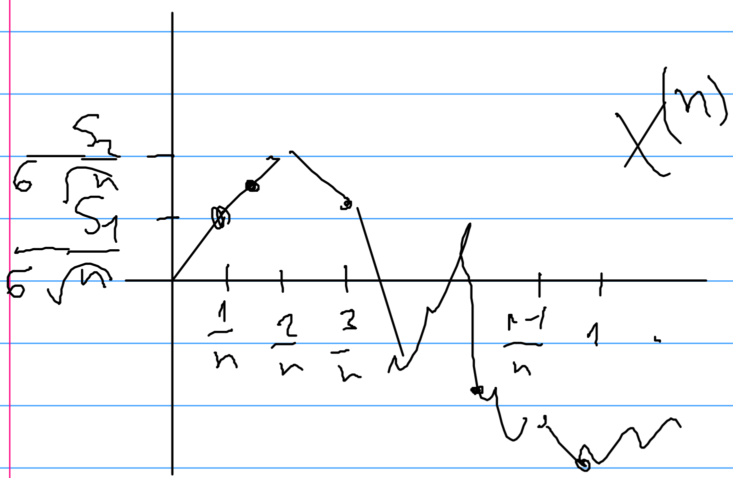
d-dimensional random vector

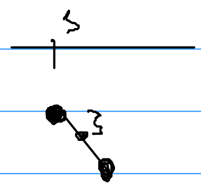
$\xi, \xi_1, \xi_2, \dots$ iid $\quad E\xi = 0, \quad E\xi^2 = \sigma^2.$

$S_n = \xi_1 + \dots + \xi_n \qquad \dfrac{S_n}{\sqrt{n}\,\sigma} \xrightarrow{\mathcal{D}} N(0,1)$

Lévy CLT







$$X_s^{(n)} = \frac{S_{\lfloor sn \rfloor}}{\sqrt{n}\sigma} + \frac{sn - \lfloor sn \rfloor}{\sqrt{n}\sigma} \cdot \xi_{\lfloor sn \rfloor + 1} \quad \xrightarrow{\mathcal{D}/\mathbf{P}} 0$$

To ease notation let $d = 2$ and $(t_1, t_2) = (s, t)$. We want to show that

$$(X_s^{(n)}, X_t^{(n)}) \xrightarrow{\mathcal{D}} (W_s, W_t)\,.$$

By the definition of $X^{(n)}$

$$\left\| (X_s^{(n)}, X_t^{(n)}) - \frac{1}{\sigma\sqrt{n}}(S_{\lfloor sn \rfloor}, S_{\lfloor tn \rfloor}) \right\| \xrightarrow{\mathbf{P}} 0,$$

Ex. 22.

therefore it is enough to show that

$$\frac{1}{\sigma\sqrt{n}}(S_{\lfloor sn \rfloor}, S_{\lfloor tn \rfloor}) \xrightarrow{\mathcal{D}} (W_s, W_t)\,.$$

By Lévy's CLT

independent

$$\frac{1}{\sigma\sqrt{n}}(S_{\lfloor sn \rfloor}, S_{\lfloor tn \rfloor} - S_{\lfloor sn \rfloor}) \xrightarrow{\mathcal{D}} (\sqrt{s}Z, \sqrt{t-s}Z'),$$

Ex. 21.

where Z, Z' are independent $N(0, 1)$. Therefore

$$\frac{1}{\sigma\sqrt{n}}(S_{\lfloor sn \rfloor}, S_{\lfloor tn \rfloor}) \xrightarrow{\mathcal{D}} (\sqrt{s}Z, \sqrt{s}Z + \sqrt{t-s}Z') \stackrel{\mathcal{D}}{=} (W_s, W_t),$$

as claimed. □

In the proof above we used the following simple statements.

HW

Exercise 21. Let (X_n) be a sequence of random elements in the metric space (S_1, ρ_1) converging in distribution to X. Let $\varphi : S_1 \to S_2$ be continuous, where (S_2, ρ_2) is another metric space. Show that $\varphi(X_n)$ converges in distribution to $\varphi(X)$.

H.W.

Exercise 22. Let (X_n), (Y_n) be random elements in the separable metric space (S, ρ) defined on the same probability space. Show that if X_n converges in distribution to X and $\rho(X_n, Y_n) \to 0$ in probability then Y_n converges in distribution to X.

As a consequence of Donsker's invariance principle we obtain limit result for the path of random walks. Let us restrict to the interval $[0, 1]$ and consider the space $C[0, 1]$ with the supremum norm. Consider the continuous functional

$$f : C[0, 1] \to \mathbb{R};\ \omega \mapsto \max_{t \in [0,1]} \omega(t).$$

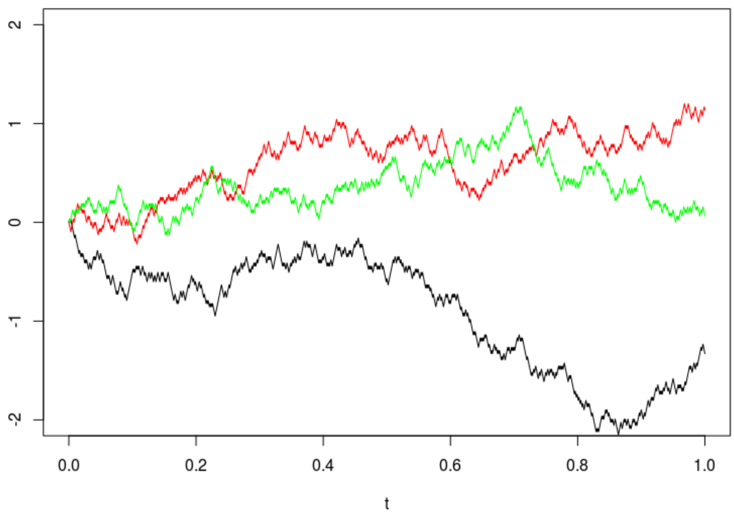


Figure 1: Simulation of 3 independent SBM

Since $X^{(n)} \to W$ in distribution (in $C[0,1]$) we have that $f(X^{(n)}) \to f(W)$ in distribution (in $\mathbb{R}$!). That is

$$\mathbf{P}(\max_{t\in[0,1]} X_t^{(n)} \leq x) \to \mathbf{P}(\max_{t\in[0,1]} W_t \leq x),$$

for each $x \in \mathbb{R}$ (well, only for continuity point of the limit, but it is continuous). By the definition of $X^{(n)}$ we can rewrite the RHS to get

$$\mathbf{P}\left(\max_{k\leq n} S_k \leq \sqrt{n}\sigma x\right) \to \mathbf{P}(\max_{t\in[0,1]} W_t \leq x).$$

Next we determine the LHS. Using the reflection principle

$$\begin{aligned}
&\mathbf{P}\left(\max_{t\in[0,1]} W_t > x\right) \\
&= \mathbf{P}\left(\max_{t\in[0,1]} W_t > x,\ W_1 > x\right) + \mathbf{P}\left(\max_{t\in[0,1]} W_t > x,\ W_1 < x\right) \\
&= 2\mathbf{P}\left(\max_{t\in[0,1]} W_t > x,\ W_1 > x\right) \\
&= 2\mathbf{P}\left(W_1 > x\right) = 2\left(1 - \Phi(x)\right).
\end{aligned}$$

$\max_{t\leq s} W_t = M_s \quad (M_t)$

$$X_n \xrightarrow{\mathcal{D}} W \quad \text{in } C[0,1]$$

by def.: $\quad E\, h(X_n) \to E\, h(X)$

for all $h : C[0,1] \to \mathbb{R}$ cont. & bounded

$\Rightarrow$

$$f(w) = \sup\{ w_t : t \in [0,1] \}$$

$$f(X_n) \xrightarrow{\mathcal{D}} f(W) \quad \text{in } \mathbb{R})$$

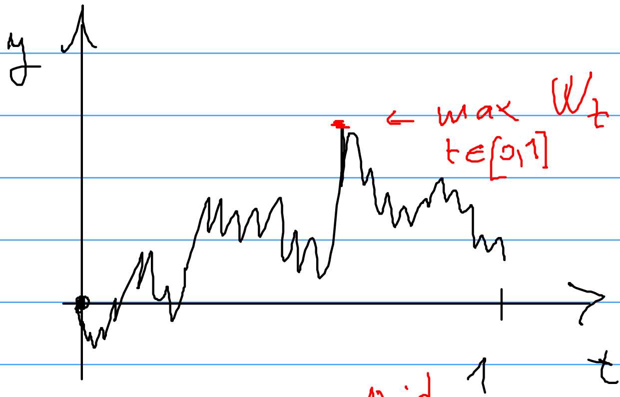
usual random variables

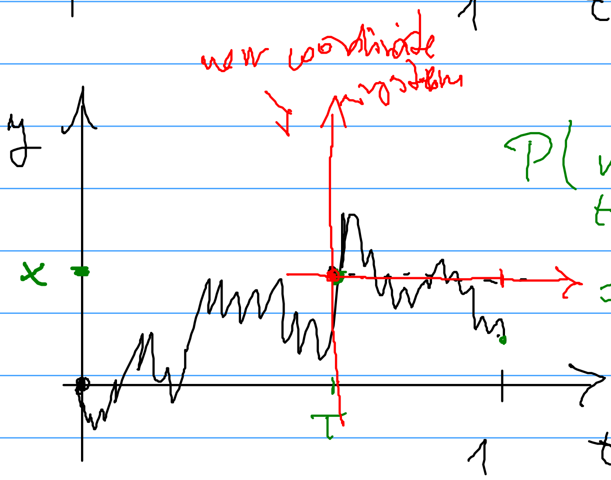
$$E\, g(f(X_n)) \to E\, g(f(W))$$

for all cont. bounded $g : \mathbb{R} \to \mathbb{R}$

Choose: $h = g \circ f$.

$$P\left(\max_{t\in[0,1]} W_t > x\right) =$$





$$P\left(\max_{t\in[0,1]} W_t > x\right) =$$

$$= P(\max > x, W_1 > x)$$

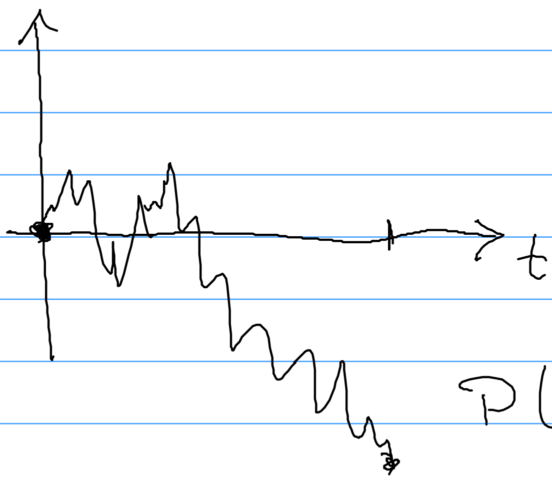
$$+ P(\max > x, W_1 \le x) =$$

$$= 2\cdot P(\max > x, W_1 > x) \overset{(*)}{=}$$

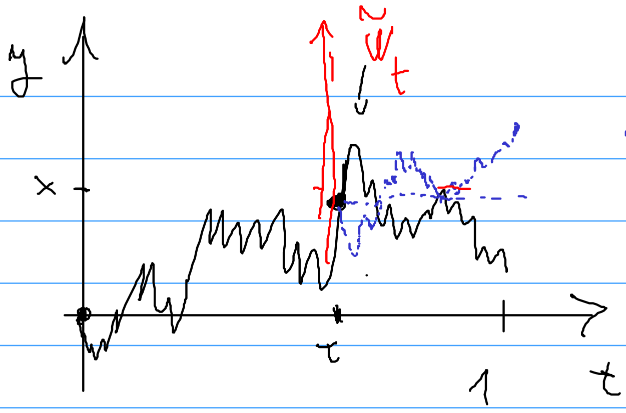
$$P\left(W_1 > x \;\middle|\; \max_{t\in[0,1]} W_t > x\right) \overset{(*)}{=} P(W_1 - W_T > 0)$$

$$= \frac{1}{2}$$

$$\boxed{\begin{aligned} &= 2\cdot P(W_1 > x) \\ &= 2\cdot\left(1 - \Phi(x)\right) \end{aligned}}$$

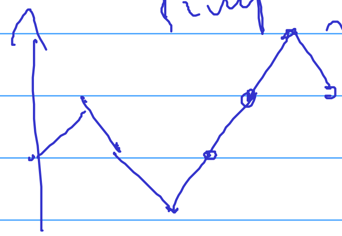


$$P(W_1 > 0) = \frac{1}{2}$$



reflection principle

simple symm. r.w.



$$W_{s+t} - W_s = \tilde{W}_t \qquad s \text{ fixed}$$

Summarizing

$$\lim_{n\to\infty} \mathbf{P}\left(\max_{k\le n} S_k \le \sqrt{n}\sigma x\right) = 2\Phi(x) - 1.$$

3.4 Markov property

Assume that we have a SBM (W_t) and we know everything up to time s. Conditioned on that information, what is the distribution of W_t, $t > s$?

Formally, $(W_t, \mathcal{F}_t)$ is a SBM, and we are interested in the conditional probabilities

indep. past up to s.

$$\mathbf{P}(W_t \in A|\mathcal{F}_s).$$

Since $W_t = W_s + W_t - W_s$, where W_s is $\mathcal{F}_s$-measurable and $W_t - W_s$ is independent of $\mathcal{F}_s$, we obtain

only thing that matters is the value W_s.

$$\mathbf{P}(W_t \in A|\mathcal{F}_s) = \mathbf{P}(W_t \in A|W_s) = \mathbf{P}_{W_s}(W_{t-s} \in A),$$

where $\mathbf{P}_x(W_u \in A) = \mathbf{P}(W_u \in A|W_0 = x)$, that is under $\mathbf{P}_x$ W is a SBM starting at x. That is knowing the whole past up to s gives no more information than knowing only W_s. This is the Markov property.

To make the previous argument formal we need the following.

Exercise 23. Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space, $\mathcal{G} \subset \mathcal{A}$ a sub-σ-algebra, X, Y random variables such that X is independent of $\mathcal{G}$ and Y is $\mathcal{G}$-measurable. Then

$$\mathbf{P}(X + Y \in A|\mathcal{G}) = \mathbf{P}(X + Y \in A|Y) \quad \mathbf{P} - \text{a.s.}$$

$X+Y \in A$

and

$$\mathbf{P}(X + Y \in A|Y = y) = \mathbf{P}(X + y \in A) \quad \mathbf{P}Y^{-1} - \text{a.s.}$$

$\downarrow$

$(X, Y) \in B \times C$

For the latter note that for some $\sigma(Y)/\mathcal{B}(\mathbb{R})$-measurable h

$$\mathbf{P}(X + Y \in A|Y) = h(Y).$$

$\to$ λ-system + extension work.

So the latter statement claims that $h(y) = \mathbf{P}(X + y \in A)$ a.s. with respect to the induced measure $\mathbf{P}Y^{-1}$.

A (d-dimensional) adapted process (X_t) is *Markov process with initial distribution* μ if

(i) $\mathbf{P}(X_0 \in A) = \mu(A)$;

(ii) $\mathbf{P}(X_{t+s} \in A|\mathcal{F}_s) = \mathbf{P}(X_{t+s} \in A|X_s)$, for all A and $t, s > 0$.