LECTURE NOTES FOR THE COURSE "CONVERGENCES IN PROBABILITY" GIVEN AT THE SAINT PETERSBURG STATE UNIVERSITY

Contents

Date: April 22, 2021.

1. Introduction to the basic concepts of convergence in probability theory

During this class, we consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, that is:

- Ω is a non-empty set, the set of outcomes,
- $\mathcal{F}$ (the event space) is a σ-algebra: $\emptyset \in \mathcal{F}$; if $A \in \mathcal{F}$, then $\Omega \setminus A \in \mathcal{F}$ and $\mathcal{F}$ is stable by countable unions: if $A_n \in \mathcal{F}$, $n \in \mathbb{N}$, then $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$,
- $\mathbb{P}$ is a probability measure, that is, $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$ and if $(A_n)_{n\in\mathbb{N}}$ is a sequence of pairwise disjoint elements of $\mathcal{F}$, then $\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$.

Definition 1.1. *A random variable is a function $X\colon \Omega \to \mathbb{R}$ such that for each Borel set B of the real line, $X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in B\}$ belongs to $\mathcal{F}$ (equivalently, we require that $X^{-1}((-\infty, x]) \in \mathcal{F}$ for each $x \in \mathbb{R}$).*

General goal, that will be made more precise later: given a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ and functions $f_n\colon \mathbb{R}^n \to \mathbb{R}$, we would like to understand as best as possible the asymptotic behavior of the sequence $(f_n(X_1, \dots, X_n))$. E.g., $f_n(X_1, \dots, X_n) = \sum_{i=1}^n X_i$.

1.1. Almost sure convergence.

Definition 1.2. *We say that a sequence of random variables $(X_n)_{n\geqslant 1}$ converges almost surely to a random variable X if there exists $\Omega' \in \mathcal{F}$ such that $\mathbb{P}(\Omega') = 1$ and for each $\omega \in \Omega'$, $X_n(\omega) \to X(\omega)$ as $n \to \infty$.*

Example 1.3. Let X_n be such that $X_n(\omega) = 1$ if $\omega \in A_n$, $A_n \in \mathcal{F}$ and $X_n(\omega) = 0$ otherwise. If $\mathbb{P}\left(\bigcup_{k\geqslant n} A_k\right) \to 0$, then $X_n \to 0$ almost surely.

Theorem 1.4 (Egoroff)**.** *Let $(X_n)_{n\geqslant 1}$ be a sequence of random variables which converges almost surely to X. For each positive δ, there exists a set $A_\delta \in \mathcal{F}$ such that and $\mathbb{P}(\Omega \setminus A_\delta) \leqslant \delta$ and $\sup_{\omega\in A_\delta} |X_n(\omega) - X(\omega)| \to 0$.*

In other words, Ω can be splitted into two sets: on the first one, uniform convergence takes place, while the second one can have a measure as small as we wish.

Example 1.5. Let $\Omega = (0, 1)$ endowed with Lebesgue measure and let $X_n(\omega) = \omega^n$. One can take $A_\delta = (0, 1 - \delta)$.

1.2. Convergence in probability.

Definition 1.6. *We say that the sequence of random variables $(X_n)_{n\geqslant 1}$ converges in probability to X if for each positive ε,*

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0. \tag{1.2.1}$$

Example 1.7. Let X_n be such that $X_n(\omega) = 1$ if $\omega \in A_n$, $A_n \in \mathcal{F}$ and $X_n(\omega) = 0$ otherwise. $X_n \to 0$ in probability if and only if $\mathbb{P}(A_n) \to 0$.

1.3. Link between the almost sure convergence and convergence in probability.

Theorem 1.8 (First Borel-Cantelli lemma)**.** *Let $(A_n)_{n\geqslant 1}$ be a sequence of events such that $\sum_{n\geqslant 1} \mathbb{P}(A_n) < \infty$. Then the event $\limsup_{n\to\infty} A_n = \bigcap_{N\geqslant 1} \bigcup_{n\geqslant N} A_n$ has probability zero.*

Proof. For each N_0, $\mathbb{P}\left(\bigcap_{N\geqslant 1}\bigcup_{n\geqslant N} A_n\right)\leqslant \mathbb{P}\left(\bigcup_{n\geqslant N_0} A_n\right)\leqslant \sum_{n\geqslant N_0}\mathbb{P}\left(A_n\right)$. □

A natural question is whether the converse is true, that is, whether $\mathbb{P}\left(\limsup_{n\to\infty} A_n\right)=0$ implies $\sum_{n\geqslant 1}\mathbb{P}\left(A_n\right)<\infty$. This is in general not the case. For example, let $(\Omega,\mathcal{F},\mathbb{P}) = \left([0,1],\mathcal{B}\left([0,1]\right),\lambda\right)$, where λ denotes the Lebesgue measure and let $A_n=(0,1/n)$. Then $\limsup_{n\to\infty} A_n=\emptyset$ but $\mathbb{P}\left(A_n\right)=1/n$ and $\sum_{n\geqslant 1}\mathbb{P}\left(A_n\right)=\infty$.

Nevertheless, a partial converse can be made under the assumption of independence.

Definition 1.9. *We say that a sequence of events $\left(A_n\right)_{n\geqslant 1}$ is independent if for each finite subset I of $\mathbb{N}$, $\mathbb{P}\left(\bigcap_{i\in I} A_i\right)=\prod_{i\in I}\mathbb{P}\left(A_i\right)$.*

Theorem 1.10 (Second Borel-Cantelli lemma)**.** *Let $\left(A_n\right)_{n\geqslant 1}$ be an independent sequence of events. Suppose that $\sum_{n\geqslant 1}\mathbb{P}\left(A_n\right)=\infty$. Then $\mathbb{P}\left(\limsup_{n\to\infty} A_n\right)=1$.*

Suppose that $X_n\to X$ in probability. We know that $\lim_{n\to\infty}\mathbb{P}\left(\left|X_n-X\right|>\varepsilon\right)=0$ For a fixed k, let $\varepsilon=2^{-k}$. We can find an integer n_k such that $\mathbb{P}\left(\left|X_{n_k}-X\right|>2^{-k}\right)\leqslant 2^{-k}$.

It is possible to construct inductively an increasing sequence of integers $\left(n_k\right)_{k\geqslant 1}$ such that for each k, $\mathbb{P}\left(\left|X_{n_k}-X\right|>2^{-k}\right)\leqslant 2^{-k}$.

By the first Borel-Cantelli lemma, $\limsup_{k\to\infty}\left\{\left|X_{n_k}-X\right|>2^{-k}\right\}$ has probability zero hence there is $\Omega'\in\mathcal{F}$ having probability one such that for each $\omega\in\Omega'$, there exists an integer $k\left(\omega\right)$ for which $k\geqslant k\left(\omega\right)$ implies $\left|X_{n_k}\left(\omega\right)-X\left(\omega\right)\right|\leqslant 2^{-k}$.

We showed the following: if $X_n\to X$ in probability, there is a subsequence $\left(X_{n_k}\right)_{k\geqslant 1}$ which converges almost surely to X.

Moreover, we can show (we will do during the seminar) that $X_n\to X$ almost surely if and only if $\sup_{k\geqslant n}\left|X_k-X\right|\to 0$ in probability. Hence almost sure convergence implies convergence in probability.

The converse is not true: let X_n be an independent sequence of random variables where X_n takes the value 1 with probability $1/n$ and 0 with probablity $1-1/n$. Then $X_n\to 0$ in probability but not almost surely, by the second Borel-Cantelli lemma.

1.4. **Convergence in $\mathbb{L}^p$.** A random variable S is said simple if its range is finite. Such a random variable can be expressed as $S=\sum_{i=1}^{N} a_i\mathbf{1}\left(A_i\right)$ for some $N\geqslant 1$, $a_i\in\mathbb{R}$ and $A_i\in\mathcal{F}$ are pairwise disjoint and $\mathbf{1}\left(A_i\right)$ denotes the indicator function of the set A.

For such randon variables, $\mathbb{E}\left[S\right]$ is defined as $\sum_{i=1}^{N} a_i\mathbb{P}\left(A_i\right)$.

For a non-negative random variable X, the expectation of X is defined as

$$\mathbb{E}\left[X\right]=\sup\left\{\mathbb{E}\left[S\right], S \text{ is simple and } 0\leqslant S\leqslant X\right\}. \tag{1.4.1}$$

For a not necessarily non-negative random variable, we decompose it as $X=X^+-X^-$, where $X^+=\max\left\{X,0\right\}$ and $X^-=\min\left\{X,0\right\}$.

Note that a consequence of this is that

$$\lim_{\delta\to 0}\sup_{A\in\mathcal{F},\mathbb{P}(A)\leqslant\delta}\mathbb{E}\left[\left|X\right|\mathbf{1}\left(A\right)\right]=0. \tag{AC}$$

Theorem 1.11 (Dominated convergence theorem under convergence in probability)**.** *Let $\left(X_n\right)_{n\geqslant 1}$ be a sequence of random variables which converges in probability to X. Suppose that $\mathbb{E}\left[\sup_{n\geqslant 1}\left|X_n\right|\right]<\infty$. Then $\mathbb{E}\left[\left|X_n-X\right|\right]\to 0$.*

Proof. Suppose not. Then there is a positive ε_0 and an increasing sequence of integers $(n_k)_{k\geqslant 1}$ such that $\mathbb{E}\left[|X_{n_k}-X|\right]\geqslant\varepsilon_0$. We extract from $(|X_{n_k}-X|)_{k\geqslant 1}$ a subsequence $\left(\left|X_{n_{k_\ell}}-X\right|\right)_{\ell\geqslant 1}$ which converges almost surely to 0. By Egoroff's theorem, we can pick A_δ such that $\sup_{\omega\in A_\delta}\left|X_{n_{k_\ell}}(\omega)-X(\omega)\right|\to 0$ and $\mathbb{P}(\Omega\setminus A_\delta)\leqslant\delta$. Therefore,

$$\begin{aligned}\varepsilon_0\leqslant\mathbb{E}\left[\left|X_{n_{k_\ell}}-X\right|\right]&\leqslant\sup_{\omega\in A_\delta}\left|X_{n_{k_\ell}}(\omega)-X(\omega)\right|+\mathbb{E}\left[\left|X_{n_{k_\ell}}-X\right|\mathbf{1}(\Omega\setminus A_\delta)\right]\\&\leqslant\sup_{\omega\in A_\delta}\left|X_{n_{k_\ell}}(\omega)-X(\omega)\right|+2\sup_{A\in\mathcal{F},\mathbb{P}(A)\leqslant\delta}\mathbb{E}\left[\sup_{n\geqslant 1}|X_n|\mathbf{1}(A)\right]\end{aligned}$$

and (AC) gives a contradiction. □

The space $\mathbb{L}^p$, $1\leqslant p<\infty$, consists of the equivalence class of random variables X such that $\mathbb{E}\left[|X|^p\right]<\infty$.

Definition 1.12. *We say that the sequence of random variables $(X_n)_{n\geqslant 1}$ converges in $\mathbb{L}^p$ to X, $1\leqslant p<\infty$ if*

$$\lim_{n\to\infty}\mathbb{E}\left[|X_n-X|^p\right]=0. \tag{1.4.2}$$

Example 1.13. If $X_n=\mathbf{1}(A_n)$, then $X_n\to X$ in $\mathbb{L}^p$ if and only if $\mathbb{P}(A_n)\to 0$.

Proposition 1.14. *If $(X_n)_{n\geqslant 1}$ converges in $\mathbb{L}^p$ to X and in probability to Y, then $X=Y$ a.s..*

1.5. **Comparison of convergence in $\mathbb{L}^p$ with almost sure convergence and convergence in probability.** Let us first compare convergence in probability with convergence in $\mathbb{L}^p$.

Suppose that $(X_n)_{n\geqslant 1}$ converges in $\mathbb{L}^p$ to X for some $1\leqslant p<\infty$. Integrating the pointwise inequality

$$\varepsilon^p\mathbf{1}\left\{|X_n-X|>\varepsilon\right\}\leqslant|X_n-X|^p,$$

we derive that $X_n\to X$ in probability.

Therefore, convergence in $\mathbb{L}^p$ implies convergence in probability.

The converse is not true, not even if $X_n,X\in\mathbb{L}^p$. Indeed, take $\Omega=(0,1)$ endowed with the Borel σ-algebra and Lebesgue measure. Let $X_n=n^{1/p}\mathbf{1}\left(\left(0,\frac{1}{n}\right)\right)$. Then $X_n\to 0$ in probability but not in $\mathbb{L}^p$.

Then let us compare almost sure and convergence in $\mathbb{L}^p$. Since convergence in probability does not imply convergence in $\mathbb{L}^p$, it is clear that almost sure convergence cannot imply convergence in $\mathbb{L}^p$.

Moreover, convergence in $\mathbb{L}^p$ does not imply almost sure convergence: let X_n be an independent sequence of random variables where X_n takes the value 1 with probability $1/n$ and 0 with probablity $1-1/n$. Then $X_n\to 0$ in probability but not almost surely, by the second Borel-Cantelli lemma.

How to go from convergence in probability to convergence in $\mathbb{L}^p$?

Suppose that $(X_n)_{n\geqslant 1}$ converges in probability to X. We would like to show the convergence in $\mathbb{L}^p$ for some $1\leqslant p<\infty$.

- First we have to assume that $X_n, X \in \mathbb{L}^p$, otherwise it is hopeless.
- Note that we should in particular have $\sup_{n \geqslant 1} \mathbb{E}\left[|X_n|^p\right] < \infty$.

Recall that for a non-negative random variable Y,

$$\mathbb{E}[Y] = \int_0^\infty \mathbb{P}(Y > t)\,\mathrm{d}t. \tag{1.5.1}$$

Therefore,

$$\mathbb{E}\left[|X_n - X|^p\right] = \int_0^\infty \mathbb{P}\left(|X_n - X| > t^{1/p}\right)\mathrm{d}t. \tag{1.5.2}$$

Note that for each fixed R, $\int_0^R \mathbb{P}\left(|X_n - X| > t^{1/p}\right)\mathrm{d}t \leqslant \varepsilon + R\mathbb{P}\left(|X_n - X| > \varepsilon^{1/p}\right)$ hence

$$\limsup_{n\to\infty} \mathbb{E}\left[|X_n - X|^p\right] = \limsup_{n\to\infty} \int_R^\infty \mathbb{P}\left(|X_n - X| > t^{1/p}\right)\mathrm{d}t \tag{1.5.3}$$

1.6. **Uniform integrability.** For a random variable X,

$$\int_R^\infty \mathbb{P}(|X| > t)\,\mathrm{d}t \leqslant R\mathbb{P}(X > R) + \int_R^\infty \mathbb{P}(|X| > t)\,\mathrm{d}t \leqslant \mathbb{E}\left[|X|\,\mathbf{1}\{|X| > R\}\right], \tag{1.6.1}$$

where $\mathbf{1}(A)(\omega)$ is a random variable taking the value 1 if $\omega \in A$ and 0 otherwise.

Therefore, if a sequence $(X_n)_{n\geqslant 1}$ converges in probability to X, we need that

$$\lim_{R\to\infty} \limsup_{n\to\infty} \mathbb{E}\left[|X_n - X|^p\,\mathbf{1}\{|X_n - X| > R\}\right] = 0. \tag{1.6.2}$$

Definition 1.15 (Uniform integrability)**.** *A sequence of random variables $(X_n)_{n\geqslant 1}$ is uniformly integrable (UI) if*

$$\lim_{R\to\infty} \sup_{n\geqslant 1} \mathbb{E}\left[|X_n|\,\mathbf{1}\{|X_n| > R\}\right] = 0. \tag{UI}$$

Let us mention the following properties of uniform integrability

(1) If $X_n = X \in \mathbb{L}^1$, then $(X_n)_{n\geqslant 1}$ is UI.
(2) More generally, if $X_n \to X$ in $\mathbb{L}^1$, then $(X_n)_{n\geqslant 1}$ is UI. Indeed,

$$\mathbb{E}\left[|X_n|\,\mathbf{1}\{|X_n| > R\}\right] \leqslant \mathbb{E}\left[|X_n - X|\right] + \mathbb{E}\left[|X|\,\mathbf{1}\{|X_n| > R\}\right] \tag{1.6.3}$$

(3) If $\sup_{n\geqslant 1} |X_n|$ is integrable, then $(X_n)_{n\geqslant 1}$ is UI.
(4) More generally, if there exists $p > 1$ such that $\sup_{n\geqslant 1} \mathbb{E}\left[|X_n|^p\right] < \infty$, then $(X_n)_{n\geqslant 1}$ is UI.

Let us give an equivalent characterization of uniform integrability, which may be more tractable in some cases.

Proposition 1.16. *A sequence $(X_n)_{n\geqslant 1}$ is UI if and only if the following two conditions are satisfied:*

(1) $\sup_{n\geqslant 1} \mathbb{E}\left[|X_n|\right] < \infty$ *and*
(2) for each positive ε, there exists $\delta > 0$ such that if $A \in \mathcal{F}$ satisfies $\mathbb{P}(A) \leqslant \delta$, then $\sup_{n\geqslant 1} \mathbb{E}\left[|X_n|\,\mathbf{1}(A)\right] < \varepsilon$.

Note that the second condition does not imply the first one (in the case where $\mathbb{P}$ is a Dirac mass for example).

Proof of Proposition 1.16. $\Rightarrow$: taking R such that $\sup_{n\geqslant 1}\mathbb{E}\left[|X_n|\mathbf{1}\left\{|X_n|>R\right\}\right]\leqslant 1$, we get $\mathbb{E}\left[|X_n|\right]\leqslant 1+R$. Moreover, for any $A\in\mathcal{F}$,

$$\begin{aligned}\mathbb{E}\left[|X_n|\mathbf{1}(A)\right]&=\mathbb{E}\left[|X_n|\mathbf{1}\left\{|X_n|>R\right\}\mathbf{1}(A)\right]+\mathbb{E}\left[|X_n|\mathbf{1}\left\{|X_n|\leqslant R\right\}\mathbf{1}(A)\right]\\&\leqslant \mathbb{E}\left[|X_n|\mathbf{1}\left\{|X_n|>R\right\}\right]+R\mathbb{P}(A).\end{aligned}$$

For a fixed $\varepsilon>0$, choose R such that $\sup_{n\geqslant 1}\mathbb{E}\left[|X_n|\mathbf{1}\left\{|X_n|>R\right\}\right]<\varepsilon/2$ and take $\delta=\varepsilon/(2R)$.

$\Leftarrow$: Let $\varepsilon>0$ and let δ be such that if $A\in\mathcal{F}$ satisfies $\mathbb{P}(A)\leqslant\delta$, then $\sup_{n\geqslant 1}\mathbb{E}\left[|X_n|\mathbf{1}(A)\right]<\varepsilon$.

We first show that there exists an R such that $\sup_{n\geqslant 1}\mathbb{P}(|X_n|>R)\leqslant\delta$. Indeed, by Markov's inequality,

$$\sup_{n\geqslant 1}\mathbb{P}(|X_n|>R)\leqslant\frac{1}{R}\sup_{n\geqslant 1}\mathbb{E}\left[|X_n|\right]$$

and we can take $R>\sup_{n\geqslant 1}\mathbb{E}\left[|X_n|\right]/\delta$.

We then get that for each fixed n (with $A=\left\{|X_n|>R\right\}$) that $\mathbb{E}\left[|X_n|\mathbf{1}\left\{|X_n|>R\right\}\right]\leqslant\varepsilon$. □

Theorem 1.17 (Dominated convergence theorem under convergence in probability and UI)**.** *Let $(X_n)_{n\geqslant 1}$ be a UI sequence of random variables which converges in probability to X. Then $\mathbb{E}\left[|X_n-X|\right]\to 0$.*

Proof. Suppose not. Then there is a positive ε_0 and an increasing sequence of integers $(n_k)_{k\geqslant 1}$ such that $\mathbb{E}\left[|X_{n_k}-X|\right]\geqslant\varepsilon_0$. We extract from $(|X_{n_k}-X|)_{k\geqslant 1}$ a subsequence $\left(\left|X_{n_{k_\ell}}-X\right|\right)_{\ell\geqslant 1}$ which converges almost surely to 0. By Egoroff's theorem, we can pick A_δ such that $\sup_{\omega\in A_\delta}\left|X_{n_{k_\ell}}(\omega)-X(\omega)\right|\to 0$ and $\mathbb{P}(\Omega\setminus A_\delta)\leqslant\delta$. Therefore,

$$\begin{aligned}\varepsilon_0\leqslant\mathbb{E}\left[\left|X_{n_{k_\ell}}-X\right|\right]&\leqslant\sup_{\omega\in A_\delta}\left|X_{n_{k_\ell}}(\omega)-X(\omega)\right|+\mathbb{E}\left[\left|X_{n_{k_\ell}}-X\right|\mathbf{1}(\Omega\setminus A_\delta)\right]\\&\leqslant\sup_{\omega\in A_\delta}\left|X_{n_{k_\ell}}(\omega)-X(\omega)\right|+2\sup_{n\geqslant 1}\mathbb{E}\left[|X_n|\mathbf{1}(\Omega\setminus A_\delta)\right]\qquad(1.6.4)\end{aligned}$$

and the previous Proposition applied with $\varepsilon=\varepsilon_0/2$ gives a contradiction. □

1.7. **A necessary and sufficient condition for uniform integrability.**

Theorem 1.18 (De la Vallée-Poussin)**.** *A sequence of random variables $(X_n)_{n\geqslant 1}$ is uniformly integrable if and only if there exists a convex non-decreasing function $\Phi\colon[0,\infty)\to[0,\infty)$ such that $\lim_{x\to\infty}\Phi(x)/x=\infty$ and $\sup_{n\geqslant 1}\mathbb{E}\left[\Phi(|X_n|)\right]<\infty$.*

Example 1.19.
(1) $\Phi(x)=x^p$ for $p>1$.
(2) $\Phi(x)=x(\ln(1+x))^\alpha$, $\alpha>0$.

Proof. $\Leftarrow$ For each positive R and $n\geqslant 1$,

$$|X_n|\mathbf{1}\left\{|X_n|>R\right\}=\frac{|X_n|}{\Phi(|X_n|)}\Phi(|X_n|)\mathbf{1}\left\{|X_n|>R\right\}\leqslant\sup_{x\geqslant R}\frac{x}{\Phi(x)}\Phi(|X_n|)$$

$\Rightarrow$ Define a sequence of real numbers $(R_k)_{k\geqslant 1}$ such that for each $n,k\geqslant 1$, $\mathbb{E}\left[|X_n|\mathbf{1}\left\{|X_n|>R_k\right\}\right]\leqslant 2^{-k}$ and $R_{k+1}\geqslant R_k+1$. This is possible by definition of uniform integrability: let R_1 be such that $\mathbb{E}\left[|X_n| \mathbf{1}\left\{|X_n| > R_1\right\}\right] \leqslant 1/2$. Suppose that $R_1, \dots, R_k$ have been constructed in such a way that $R_{i+1} \geqslant R_i + 1$ for each $1 \leqslant i \leqslant k$. Since $\sup_{n \geqslant 1} \mathbb{E}\left[|X_n| \mathbf{1}\left\{|X_n| > R\right\}\right] \to 0$ as R going to infinity, there exists some t_0 such that $\sup_{n \geqslant 1} \mathbb{E}\left[|X_n| \mathbf{1}\left\{|X_n| > R\right\}\right] \leqslant 2^{-k-1}$ for each $R > t_0$. Choose $R_{k+1} = \max\left\{R_k + 1, t_0\right\}$.

We then define

$$\Phi\left(x\right) = \sum_{k \geqslant 1} \max\left\{0, x - R_k\right\}. \tag{1.7.1}$$

Note that this function is well-defined: if $x \geqslant 0$, then $x \in [R_{k_0}, R_{k_0+1})$ for some k_0 hence the terms for $k > k_0 + 1$ vanishes. Moreover, Φ is a series of convex functions hence convex, and non-decreasing.

For each K,

$$\liminf_{x \to \infty} \frac{\Phi\left(x\right)}{x} \geqslant \liminf_{x \to \infty} \sum_{k=1}^{K} \max\left\{0, 1 - \frac{R_k}{x}\right\} = K$$

hence $\lim_{x \to \infty} \Phi\left(x\right)/x = \infty$.

Moreover,

$$\mathbb{E}\left[\max\left\{0, |X_n| - R_k\right\}\right] \leqslant \mathbb{E}\left[|X_n| \mathbf{1}\left\{|X_n| > R_k\right\}\right] \leqslant 2^{-k}$$

hence $\mathbb{E}\left[\Phi\left(|X_n|\right)\right] \leqslant \sum_{k \geqslant 1} 2^{-k} = 1$. □

1.8. **Uniform integrability and convergence in general.** Given a uniformly integrable sequence $\left(X_n\right)_{n \geqslant 1}$, nothing can be asserted about its convergence: for example, if $X_{2n} = X$ and $X_{2n+1} = Y \neq X$, with X and Y integrable, the sequence $\left(X_n\right)_{n \geqslant 1}$ does not converge in any sense, but there are convergent subsequences.

If $\left(X_n\right)_{n \geqslant 1}$ is bounded in $\mathbb{L}^p$ for some $p > 1$, that is, $\sup_{n \geqslant 1} \mathbb{E}\left[|X_n|^p\right] < \infty$, it is possible to extract a subsequence which converges weakly in $\mathbb{L}^p$ that is, there exists an increasing sequence of integers $\left(n_k\right)_{k \geqslant 1}$ and a random variable $X \in \mathbb{L}^p$ such that for each $Y \in \mathbb{L}^q$, $\mathbb{E}\left[X_{n_k} Y\right] \to \mathbb{E}\left[XY\right]$ as $k \to \infty$, where $1/p + 1/q = 1$.

Question: what happens if we only have uniform integrability?

Theorem 1.20 (Dunford-Pettis)**.** *Let $\left(X_n\right)_{n \geqslant 1}$ be a sequence of random variables on a probability space $\left(\Omega, \mathcal{F}, \mathbb{P}\right)$ which is bounded in $\mathbb{L}^1$. The following are equivalent:*

(1) the sequence $\left(X_n\right)_{n \geqslant 1}$ is uniformly integrable;
(2) for each subsequence of $\left(X_n\right)_{n \geqslant 1}$ there exists a further subsequence $\left(X_{n_k}\right)_{k \geqslant 1}$ and a random variable $X \in \mathbb{L}^1$ such that for each $A \in \mathcal{F}$, $\mathbb{E}\left[X_{n_k} \mathbf{1}\left(A\right)\right] \to \mathbb{E}\left[X \mathbf{1}\left(A\right)\right]$ as $k \to \infty$.

An illustrative example: let $X_n = n \mathbf{1}\left(\left(0, 1/n\right)\right)$, where $\Omega = \left(0, 1\right)$, $\mathcal{F}$ is the Borel σ-algebra and $\mathbb{P}$ the Lebesgue measure. The sequence $\left(X_n\right)_{n \geqslant 1}$ is bounded in $\mathbb{L}^1$ but is not uniformly integrable. If for a subsequence $\left(X_{n_k}\right)_{k \geqslant 1}$, $\mathbb{E}\left[X_{n_k} \mathbf{1}\left(A\right)\right] \to \mathbb{E}\left[X \mathbf{1}\left(A\right)\right]$ as $k \to \infty$, $A \in \mathcal{F}$, then $X = 0$ (take $A = (a, b)$ for $0 < a < b < 1$) but with $A = \left(0, 1\right)$, $\mathbb{E}\left[X_{n_k} \mathbf{1}\left(A\right)\right] = 1$.

Proof of Theorem 1.20. $\Rightarrow$: assume without loss of generality that $X_n \geqslant 0$; if not, write $X_n = X_n' - X_n''$ with $X_n' = \max\left\{0, X_n\right\}, X_n'' \geqslant 0$, find a weakly convergent subsequence for X_n' and extract a further subsequence for X_n''.

For each integer $\ell \geqslant 1$, define the random variable $X_{n,\ell} := X_n \mathbf{1}\left\{X_n \leqslant \ell\right\}$. Since $\sup_{n\geqslant 1} \mathbb{E}\left[X_{n,1}^2\right] < \infty$, there exists an increasing sequence of integers $(n_{k,1})_{k\geqslant 1}$ and a random variable $Y_1 \in \mathbb{L}^2$ such that for each $A \in \mathcal{F}$, $\mathbb{E}\left[X_{n_{k,1},1}\mathbf{1}\left(A\right)\right] \to \mathbb{E}\left[Y_1\mathbf{1}\left(A\right)\right]$.

Since $\sup_{n\geqslant 1} \mathbb{E}\left[X_{n,2}^2\right] < \infty$, there exists an increasing sequence of integers $(n_{k,2})_{k\geqslant 1}$ and a random variable $Y_2 \in \mathbb{L}^2$ such that for each $A \in \mathcal{F}$, $\mathbb{E}\left[X_{n_{k,2},2}\mathbf{1}\left(A\right)\right] \to \mathbb{E}\left[Y_2\mathbf{1}\left(A\right)\right]$ and $\{n_{k,2}, k \geqslant 1\} \subset \{n_{k,1}, k \geqslant 1\}$.

Continuing this process, we get increasing sequences of integers $(n_{k,\ell})_{k\geqslant 1}$ such that

$$I_\ell := \{n_{k,\ell}, k \geqslant 1\} \subset \{n_{k,\ell-1}, k \geqslant 1\}, \ell \geqslant 2,$$

and random variables Y_ℓ, $\ell \geqslant 1$ such that for each ℓ and $A \in \mathcal{F}$, $\mathbb{E}\left[X_{n_{k,\ell},\ell}\mathbf{1}\left(A\right)\right] \to \mathbb{E}\left[Y_\ell\mathbf{1}\left(A\right)\right]$.

Let n_ℓ be the ℓ-th element of I_ℓ. Then $n_{\ell+1} > n_\ell$ and for each $\ell \geqslant 1$ and $A \in \mathcal{F}$,

$$\lim_{k\to\infty} \mathbb{E}\left[X_{n_k,\ell}\mathbf{1}\left(A\right)\right] = \mathbb{E}\left[Y_\ell\mathbf{1}\left(A\right)\right].$$

We found an increasing sequence of integers $(n_k)_{k\geqslant 1}$ and random variables Y_ℓ such that for each $\ell \geqslant 1$

$$\lim_{k\to\infty} \mathbb{E}\left[X_{n_k}\mathbf{1}\left\{X_{n_k} \leqslant \ell\right\}\mathbf{1}\left(A\right)\right] = \mathbb{E}\left[Y_\ell\mathbf{1}\left(A\right)\right].$$

Lemma 1.21. *The sequence $(Y_\ell)_{\ell\geqslant 1}$ converges in $\mathbb{L}^1$ to some random variable X.*

Proof. First notice that for each $A \in \mathcal{F}$,

$$\mathbb{E}\left[\left(Y_\ell - Y_{\ell-1}\right)\mathbf{1}\left(A\right)\right] = \lim_{k\to\infty} \mathbb{E}\left[X_n\mathbf{1}\left\{\ell - 1 < X_n \leqslant \ell\right\}\mathbf{1}\left(A\right)\right] \geqslant 0$$

hence $Y_\ell \geqslant Y_{\ell-1}$ a.s. and $Y_\ell \to X$ a.s. for some random variable X. By Fatou's lemma,

$$\mathbb{E}\left[X\right] \leqslant \liminf_{\ell\to\infty} \mathbb{E}\left[Y_\ell\right] \leqslant \sup_{n\geqslant 1} \mathbb{E}\left[X_n\right].$$

□

To conclude the proof of the direction $\Rightarrow$, let $\varepsilon > 0$ be fixed and ℓ such that $\mathbb{E}\left[X - Y_\ell\right] < \varepsilon$ and $\sup_{n\geqslant 1} \mathbb{E}\left[X_n\mathbf{1}\left\{X_n > \ell\right\}\right] < \varepsilon$ (by UI). For $A \in \mathcal{F}$,

$$\begin{aligned}\left|\mathbb{E}\left[X_{n_k}\mathbf{1}\left(A\right)\right] - \mathbb{E}\left[X\mathbf{1}\left(A\right)\right]\right| &\leqslant \left|\mathbb{E}\left[X_{n_k}\mathbf{1}\left(A\right)\right] - \mathbb{E}\left[X_{n_k,\ell}\mathbf{1}\left(A\right)\right]\right| \\ &+ \left|\mathbb{E}\left[X_{n_k,\ell}\mathbf{1}\left(A\right)\right] - \mathbb{E}\left[Y_\ell\mathbf{1}\left(A\right)\right]\right| + \left|\mathbb{E}\left[Y_\ell\mathbf{1}\left(A\right)\right] - \mathbb{E}\left[X\mathbf{1}\left(A\right)\right]\right|. \quad (1.8.1)\end{aligned}$$

From the estimates:

- $\left|\mathbb{E}\left[X_{n_k}\mathbf{1}\left(A\right)\right] - \mathbb{E}\left[X_{n_k,\ell}\mathbf{1}\left(A\right)\right]\right| \leqslant \mathbb{E}\left[X_{n_k}\mathbf{1}\left\{X_{n_k} > \ell\right\}\right] \leqslant \sup_{n\geqslant 1} \mathbb{E}\left[X_n\mathbf{1}\left\{X_n > \ell\right\}\right]$;
- $\left|\mathbb{E}\left[Y_\ell\mathbf{1}\left(A\right)\right] - \mathbb{E}\left[X\mathbf{1}\left(A\right)\right]\right| < \varepsilon$,

we derive that for each fixed ℓ,

$$\limsup_{k\to\infty} \left|\mathbb{E}\left[X_{n_k}\mathbf{1}\left(A\right)\right] - \mathbb{E}\left[X\mathbf{1}\left(A\right)\right]\right| \leqslant \sup_{n\geqslant 1} \mathbb{E}\left[X_n\mathbf{1}\left\{X_n > \ell\right\}\right] + \varepsilon \leqslant 2\varepsilon, \quad (1.8.2)$$

hence $\mathbb{E}\left[X_{n_k}\mathbf{1}\left(A\right)\right] \to \mathbb{E}\left[X\mathbf{1}\left(A\right)\right]$ for each $A \in \mathcal{F}$.

$\Leftarrow$ We actually have to prove that if $(X_n)_{n\geqslant 1}$ is a sequence which is bounded in $\mathbb{L}^1$ and such that there exists an integrable random variable X satisfying $\mathbb{E}\left[X_n\mathbf{1}\left(A\right)\right] \to \mathbb{E}\left[X\mathbf{1}\left(A\right)\right]$ for each $A \in \mathcal{F}$, then $(X_n)_{n\geqslant 1}$ is uniformly integrable. We can assume without loss of generality that $X = 0$.

Define the pseudo-metric on $\mathcal{F}$ by $\rho(A,B)=\mathbb{P}(A\Delta B)=\mathbb{E}\left[\left|\mathbf{1}\left(A\right)-\mathbf{1}\left(B\right)\right|\right]$, where $A\Delta B=(A\cup B)\setminus(A\cap B)$.

It is sufficient to show that for each $\varepsilon>0$, there is $\delta>0$ such that if $\mathbb{P}(A)<\delta$ then $\left|\mathbb{E}\left[X_n\mathbf{1}\left(A\right)\right]\right|<\varepsilon$. Indeed, decomposing a set A as $A^+\cup A^-$, where $A^-:=A\cap\{X_n\leqslant 0\}$ and $A^+:=A\cap\{X_n>0\}$, we can see that $\mathbb{E}\left[\left|X_n\right|\mathbf{1}\left(A\right)\right]<2\varepsilon$ whenever $\mathbb{P}(A)<\delta$.

For a fixed $\varepsilon>0$, we define

$$F_N:=\bigcap_{n\geqslant N}\left\{A\in\mathcal{F},\left|\mathbb{E}\left[X_n\mathbf{1}\left(A\right)\right]\right|\leqslant\varepsilon\right\}.$$

Each F_N is closed (since for a fixed n, the map $A\mapsto\left|\mathbb{E}\left[X_n\mathbf{1}\left(A\right)\right]\right|$ is continuous) and $\bigcup_N F_N=\mathcal{F}$, hence by Baire's theorem, there is N_0, r_0 and $A_0\in\mathcal{F}$ such that $B_\rho(A_0,r_0):=\{A\in\mathcal{F},\rho\left(A_0,A\right)<r_0\}\subset F_{N_0}$. Let B such that $\mathbb{P}(B)<r_0$. Since $\mathbb{P}(A_0\Delta(A_0\cup B))<r_0$, $\mathbb{P}(A_0\Delta(A_0\cap B^c))<r_0$ and

$$\mathbb{E}\left[X_n\mathbf{1}\left(B\right)\right]=\mathbb{E}\left[X_n\mathbf{1}\left(A_0\cup B\right)\right]-\mathbb{E}\left[X_n\mathbf{1}\left(A_0\cap B^c\right)\right],$$

we have $\left|\mathbb{E}\left[X_n\mathbf{1}\left(B\right)\right]\right|\leqslant 2\varepsilon$ whenever $n\geqslant N_0$ and $\mathbb{P}(B)<r_0$.

For $1\leqslant n\leqslant N_0-1$, there exists $\delta_n>0$ such that if $\mathbb{P}\left(B\right)<\delta_n$, then $\left|\mathbb{E}\left[X_n\mathbf{1}\left(B\right)\right]\right|\leqslant 2\varepsilon$ hence we can take $\delta=\min\left\{r_0,\delta_1,\dots,\delta_{N_0-1}\right\}$. □

1.9. **Convergence in distribution.**

Definition 1.22. *We say that the sequence of real-valued random variables $\left(X_n\right)_{n\geqslant 1}$ converges in distribution to X if for each continuous and bounded function $f\colon\mathbb{R}\to\mathbb{R}$, the convergence $\lim_{n\to\infty}\mathbb{E}\left[f\left(X_n\right)\right]=\mathbb{E}\left[f\left(X\right)\right]$ takes place.*

Proposition 1.23. *A sequence $\left(X_n\right)_{n\geqslant 1}$ converges in distribution to X if and only if for each continuity point x of the cumulative distribution function of X, $\lim_{n\to\infty}\mathbb{P}\left(X_n\leqslant x\right)=\mathbb{P}\left(X\leqslant x\right)$.*

Let us recall some properties of the convergence in distribution.

Proposition 1.24. *If a sequence $\left(X_n\right)_{n\geqslant 1}$ converges in probability to X, then it converges in distribution to X.*

If $X_n\to c$ in distribution, where c is a constant, then $X_n\to c$ in probability.

Proposition 1.25 (Continuous mapping theorem)**.** *If $\left(X_n\right)_{n\geqslant 1}$ converges in distribution to X and $g\colon\mathbb{R}\to\mathbb{R}$ is continuous, then $\left(g\left(X_n\right)\right)_{n\geqslant 1}$ converges in distribution to $g\left(X\right)$.*

1.10. **Convergence in distribution and uniform integrability.** Suppose that $\mathbb{E}\left[\left|X_n\right|\right]<\infty$ for each n and $X_n\to X$ in distribution. What can be said about the convergence of $\left(\mathbb{E}\left[X_n\right]\right)_{n\geqslant 1}$? Certainly we have to assume that $\sup_{n\geqslant 1}\mathbb{E}\left[\left|X_n\right|\right]<\infty$.

Proposition 1.26. *Let $\left(X_n\right)_{n\geqslant 1}$ be a uniformly integrable sequence which converges in distribution to some random variable X. Then $\mathbb{E}\left[X_n\right]\to\mathbb{E}\left[X\right]$.*

Example 1.27. Let $\left(\xi_i\right)_{i\geqslant 1}$ be an i.i.d. sequence, where ξ_1 is centered and has unit variance. Define

$$X_n=\frac{1}{\sqrt{n}}\sum_{i=1}^n\xi_i.$$

Then $\left(X_n^2\right)_{n\geqslant 1}$ is uniformly integrable (see seminar) hence for each $1\leqslant p\leqslant 2$, $\mathbb{E}\left[\left|X_n\right|^p\right]\to\mathbb{E}\left[\left|N\right|^p\right]$, where N is standard normal.

Proof of the Proposition 1.26.

Lemma 1.28. *A sequence $(X_n)_{n\geqslant 1}$ is uniformly integrable if and only if*

$$\lim_{R\to\infty}\sup_{n\geqslant 1}\int_R^\infty \mathbb{P}\left(|X_n|>t\right)dt=0. \tag{UI}$$

Proof. Observe that for a non-negative random variable X,

$$\begin{aligned}\mathbb{E}\left[X\mathbf{1}\left\{X>R\right\}\right]&=\int_0^\infty \mathbb{P}\left(\left\{X>t\right\}\cap\left\{X>R\right\}\right)dt\\ &=R\mathbb{P}\left(X>R\right)+\int_R^\infty \mathbb{P}\left(X>t\right)dt. \end{aligned}\tag{1.10.1}$$

Suppose that $(X_n)_{n\geqslant 1}$ is uniformly integrable. Then

$$\sup_{n\geqslant 1}\mathbb{E}\left[|X_n|\,\mathbf{1}\left\{|X_n|>R\right\}\right]\geqslant\sup_{n\geqslant 1}\int_R^\infty \mathbb{P}\left(|X_n|>t\right)dt$$

and (UI) holds. ∎

Suppose that (UI) holds. Then observe that $\sup_{n\geqslant 1}R\mathbb{P}\left(|X_n|>R\right)\to 0$, because

$$\int_R^{2R}\mathbb{P}\left(|X_n|>t\right)dt\geqslant R\mathbb{P}\left(|X_n|>2R\right).$$

Now we can conclude the proof of the Proposition: first assume that $X_n\geqslant 0$. We have

$$\left|\mathbb{E}\left[X_n\right]-\mathbb{E}\left[X\right]\right|=\left|\int_0^\infty\left(\mathbb{P}\left(X_n>t\right)-\mathbb{P}\left(X>t\right)\right)dt\right|.$$

We apply the dominated convergence theorem on $(0,R)$ endowed with the Borel σ-algebra and Lebesgue measure and the function $f_n\colon t\mapsto\mathbb{P}\left(X_n>t\right)-\mathbb{P}\left(X>t\right)$. Note that $f_n(t)\to 0$, except on a set which is at most countable (hence of measure 0). Therefore,

$$\limsup_{n\to\infty}\left|\mathbb{E}\left[X_n\right]-\mathbb{E}\left[X\right]\right|\leqslant\sup_{k\geqslant 1}\int_R^\infty\mathbb{P}\left(X_k>t\right)dt+\int_R^\infty\mathbb{P}\left(X>t\right)dt$$

and uniform integrability allows to conclude.

We have shown that if $(X_n)_{n\geqslant 1}$ is a non-negative uniformly integrable sequence which converges in distribution to X, then $\mathbb{E}\left[X_n\right]\to\mathbb{E}\left[X\right]$.

If X_n is not necessary non-negative, write $X_n=\max\left\{X_n,0\right\}-\max\left\{0,-X_n\right\}$; apply the continuous mapping theorem to see that $\left(\max\left\{X_n,0\right\}\right)_{n\geqslant 1}$ converges in distribution to $\max\left\{X,0\right\}$ and $\left(\max\left\{-X_n,0\right\}\right)_{n\geqslant 1}$ converges in distribution to $\max\left\{-X,0\right\}$.

Moreover, the sequence $\left(\max\left\{X_n,0\right\}\right)_{n\geqslant 1}$ and $\left(\max\left\{-X_n,0\right\}\right)_{n\geqslant 1}$ are uniformly integrable hence by the result for non-negative random variables

$$\begin{aligned}\mathbb{E}\left[X_n\right]&=\mathbb{E}\left[\max\left\{X_n,0\right\}\right]-\mathbb{E}\left[\max\left\{-X_n,0\right\}\right]\\ &\to\mathbb{E}\left[\max\left\{X,0\right\}\right]-\mathbb{E}\left[\max\left\{-X,0\right\}\right]\\ &=\mathbb{E}\left[X\right].\end{aligned}$$

∎

Alternative proof of Proposition 1.26. Consider $X_n^R := \max\{-R, \min\{X_n, R\}\}$ and $X^R := \max\{-R, \min\{X, R\}\}$. By definition of the convergence in distribution, $\mathbb{E}\left[X_n^R\right] \to \mathbb{E}\left[X^R\right]$ for each R.

Then

$$\left|\mathbb{E}\left[X_n\right] - \mathbb{E}\left[X\right]\right| \leqslant \left|\mathbb{E}\left[X_n^R\right] - \mathbb{E}\left[X^R\right]\right| + \left|\mathbb{E}\left[X_n - X_n^R\right]\right| + \left|\mathbb{E}\left[X\right] - \mathbb{E}\left[X^R\right]\right|.$$

The term $\left|\mathbb{E}\left[X_n^R\right] - \mathbb{E}\left[X^R\right]\right|$ can be controlled by the uniform integrability assumption and $\left|\mathbb{E}\left[X\right] - \mathbb{E}\left[X^R\right]\right| \to 0$ by dominated convergence.

□

Uniform integrability seems to be convient to establish convergence of moments and is easier to establish in general than the latter. One can wonder whether one can deal with such a problem without uniform integrability. The next proposition breaks some hopes.

Proposition 1.29. *Let $(X_n)_{n\geqslant 1}$ be a sequence of non-negative integrable random variables which converges in distribution to some random variable X. Suppose that $\mathbb{E}\left[X_n\right] \to \mathbb{E}\left[X\right]$. Then $(X_n)_{n\geqslant 1}$ is uniformly integrable.*

Proof. Consider $X_n^R := \min\{X_n, R\}$ and $X^R := \min\{X, R\}$. By definition of the convergence in distribution, $\mathbb{E}\left[X_n^R\right] \to \mathbb{E}\left[X^R\right]$ for each R.

Since

$$\left|\mathbb{E}\left[X_n - X_n^R\right]\right| \leqslant \left|\mathbb{E}\left[X_n\right] - \mathbb{E}\left[X\right]\right| + \left|\mathbb{E}\left[X\right] - \mathbb{E}\left[X^R\right]\right| + \left|\mathbb{E}\left[X_n^R\right] - \mathbb{E}\left[X^R\right]\right|,$$

we get that

$$\lim_{R\to\infty} \limsup_{n\to\infty} \left|\mathbb{E}\left[X_n - X_n^R\right]\right| = 0,$$

which guarantees uniform integrability. □

What can be said without uniform integrability?

Proposition 1.30. *Let $(X_n)_{n\geqslant 1}$ be a sequence of random variables which converges in distribution to X. Then*

$$\mathbb{E}\left[|X|\right] \leqslant \liminf_{n\to\infty} \mathbb{E}\left[|X_n|\right]. \tag{1.10.2}$$

We do not assume that X_n is integrable for each n. Even if each X_n is integrable, X may not be integrable (see seminar).

Proof. Write

$$\mathbb{E}\left[|X|\right] = \int_0^\infty \mathbb{P}\left(|X| > t\right) dt = \int_0^\infty \liminf_{n\to\infty} \mathbb{P}\left(|X_n| > t\right) dt, \tag{1.10.3}$$

since $\mathbb{P}\left(|X_n| > t\right) \to \mathbb{P}\left(|X| > t\right)$ for Leb-almost every t. Then use Fatou's lemma. □

1.11. **Tightness and convergence in distribution.**

Definition 1.31. *We say that the sequence of random variables $(X_n)_{n\geqslant 1}$ is tight if*

$$\lim_{R\to\infty} \sup_{n\geqslant 1} \mathbb{P}\left(|X_n| > R\right) = 0. \tag{1.11.1}$$

Proposition 1.32. *Let $(X_n)_{n\geqslant 1}$ be a sequence of random variables. Suppose that $X_n \to X$ in distribution. Then $(X_n)_{n\geqslant 1}$ is tight.*

Proof. Let $\varepsilon > 0$ and let R_0 be such that $-R_0$ and R_0 are continuity points of the cumulative distribution function of X such that $\mathbb{P}\left(-R_0 < X \leqslant R_0\right) > 1 - \varepsilon/2$. Since $X_n \to X$ in distribution, there exists n_0 such that for each $n \geqslant n_0$,

$$\left|\mathbb{P}\left(-R_0 < X_n \leqslant R_0\right) - \mathbb{P}\left(-R_0 < X \leqslant R_0\right)\right| \leqslant \varepsilon/2.$$

Therefore, $\sup_{n \geqslant n_0} \mathbb{P}\left(\left|X_n\right| > R_0\right) \leqslant \varepsilon$.

For each $1 \leqslant n \leqslant n_0 - 1$, there exists R_n such that $\mathbb{P}\left(\left|X_n\right| > R_n\right) \leqslant \varepsilon$. Therefore, for $R \geqslant \max\left\{R_0, R_1, \dots, R_{n_0-1}\right\}$, $\sup_{n \geqslant 1} \mathbb{P}\left(\left|X_n\right| > R\right) \leqslant \varepsilon$. □

1.12. **Convergence of probability measures.**

Definition 1.33. *We say that a sequence of probability measures $\left(\mu_n\right)_{n \geqslant 1}$ on the real line converges in distribution to a probability measure μ if for each continuous and bounded function $f\colon \mathbb{R} \to \mathbb{R}$,*

$$\int f(x)\, d\mu_n(x) \to \int f(x)\, d\mu(x).$$

Definition 1.34. *We say that a sequence of probability measures $\left(\mu_n\right)_{n \geqslant 1}$ is tight if*

$$\lim_{R \to \infty} \sup_{n \geqslant 1} \mu_n\left(\mathbb{R} \setminus [-R, R]\right) = 0.$$

Tightness implies convergence in distribution of a subsequence.

Theorem 1.35 (Prokhorov)**.** *Let $\left(\mu_n\right)_{n \geqslant 1}$ be a tight sequence of probability measures. Then there exists a subsequence $\left(\mu_{n_k}\right)_{k \geqslant 1}$ and a probability measure μ such that $\left(\mu_{n_k}\right)_{k \geqslant 1}$ converges in distribution to μ.*

In order words, tightness plays the same role for the convergence in distribution as boundedness plays for convergence of sequences of real numbers. We will see a proof in the context of probability measures on metric spaces.

Representation of a limiting probability measure as a random variable.

Theorem 1.36. *Let $\left(X_n\right)_{n \geqslant 1}$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mu_n\colon B \mapsto \mathbb{P}\left(X_n \in B\right)$. Suppose that $\left(\mu_n\right)_{n \geqslant 1}$ converges in distribution to a probability measure μ defined on $\mathbb{R}$. Then there exists a random variable $X\colon \Omega \to \mathbb{R}$ such that $\mu(B) = \mathbb{P}(X \in B)$ for each Borel set B.*

This is a non-trivial result and beyond the scope of this course. If you are interested, here is a link.

2. Convergence in distribution in metric spaces

2.1. **Motivation.** Given an independent identically distributed sequence of random variables $\left(X_i\right)_{i \geqslant 1}$ and $t \in [0, 1]$, we define

$$W_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} X_i, \tag{2.1.1}$$

where for $x \in \mathbb{R}$, $[x]$ denotes the only integer such that $[x] \leqslant x < [x] + 1$.

If X_1 is centered and $\mathbb{E}\left[X_1^2\right] = 1$, then for each $t \in [0, 1]$, $W_n(t) \to \sqrt{t} N$, where N has a standard normal distribution.

However, we would like to have information on the maximum of partial sums, that is, $N^{-1/2}\max_{1\leqslant n\leqslant N}\left|\sum_{i=1}^{N}X_i\right| = \sup_{t\in[0,1]}\left|W_N\left(t\right)\right|$ and use $\left\|W_N\right\|_\infty \to \left\|W\right\|_\infty$, where W is some process.

We view W_n as an element of a function space endowed with a norm or a metric.

Question: how to define and establish the convergence in distribution in metric spaces?

2.2. **Recall of some properties of metric spaces.** Recall that a metric space (S,d) consists of non-empty set S and a map $d\colon S\times S\to[0,\infty)$ such that $d\left(x,y\right)=0$ iff $x=y$, $d\left(x,y\right)=d\left(y,x\right)\leqslant d\left(x,a\right)+d\left(y,a\right)$ for each $a,x,y\in S$.

Definition 2.1. *We say that the metric space (S,d) is separable if there exists a sequence $\left(x_n\right)_{n\geqslant 1}$ of elements of S which dense in S, that is, for each $x\in S$ and each positive ε, there exists $n\geqslant 1$ such that $d\left(x,x_n\right)<\varepsilon$.*

Example 2.2. The space $C[0,1]$ of the continuous functions on the unit interval endowed with the metric $d\left(f,g\right)=\sup_{t\in[0,1]}\left|f\left(t\right)-g\left(t\right)\right|$ is separable, since the set of polynomials with rational coefficients is dense.

Recall that a sequence $\left(x_n\right)_{n\geqslant 1}$ is Cauchy if $\lim_{\min\{m,n\}\to\infty}d\left(x_n,x_m\right)=0$ and that a metric space (S,d) is complete if each Cauchy sequence $\left(x_n\right)_{n\geqslant 1}$ is convergent.

2.3. **Probability measure on metric spaces, definition.** We consider a separable complete metric space (S,d). Denote by $\mathcal{B}\left(S\right)$ the Borel-σ-algebra, that is, the smallest (for the inclusion) σ-algebra containing the open sets for the topology induced by the metric d.

We will consider probability measures on S, that is, maps $\mathbb{P}\colon B\in\mathcal{B}\left(S\right)\mapsto\mathbb{P}\left(B\right)\in[0,1]$ satisfying $\mathbb{P}\left(S\right)=1$ and σ-additivity.

Definition 2.3. *Let $\left(\mathbb{P}_n\right)_{n\geqslant 1}$ be a sequence of probability measures on a separable complete metric space (S,d). We say that $\left(\mathbb{P}_n\right)_{n\geqslant 1}$ converges weakly to the probability measure $\mathbb{P}$ if for each continuous and bounded function $f\colon S\to\mathbb{R}$, $\int_S f\left(x\right)d\mathbb{P}_n\left(x\right)\to\int_S f\left(x\right)d\mathbb{P}\left(x\right)$. We write $\mathbb{P}_n\Rightarrow\mathbb{P}$.*

In general, this is not easy to check directly because we do not have a "nice" characterization of the continuous bounded functions.

2.4. **Convergence of probability measures: characteristic functionals (1).** Recall that $\mathbb{P}_n\Rightarrow\mathbb{P}$ if for each continuous and bounded function $f\colon S\to\mathbb{R}$, $\int_S f\left(x\right)d\mathbb{P}_n\left(x\right)\to\int_S f\left(x\right)d\mathbb{P}\left(x\right)$.

In general, we do not have at our disposal an equivalent of cumulative distribution functions. One could try to use an analogue of charateristic functions.

If $(X,\left\|\cdot\right\|)$ is a Banach space, we define the charateristic functional of a probability measure $\mathbb{P}$ on X by

$$\widehat{\mathbb{P}}\left(\ell\right):=\int_X\exp\left(i\ell\left(x\right)\right)d\mathbb{P}\left(x\right),\ell\in X',\tag{2.4.1}$$

where X' is the set of all linear continuous maps from X to $\mathbb{R}$.

Notice that if $\mathbb{P}_n\Rightarrow\mathbb{P}$, then for each $\ell\in X'$, $\widehat{\mathbb{P}_n}\left(\ell\right)\to\widehat{\mathbb{P}}\left(\ell\right)$, because the map $x\mapsto\exp\left(i\ell\left(x\right)\right)$ is continuous and bounded.

However, if $X = \ell^2$ (space of real-valued square summable sequences), e_n is the n-th vector of the canonical basis and $\mathbb{P}_n$ is the Dirac mass at e_n, we have $\widehat{\mathbb{P}_n}(\ell) \to 1$ for each $\ell \in X'$ but no weak convergence.

Recall that

$$\widehat{\mathbb{P}}(\ell) := \int_X \exp(i\ell(x))\, d\mathbb{P}(x), \ell \in X'. \tag{2.4.2}$$

Characteristic functional do not play the same role as in the $\mathbb{R}^d$-valued case as convergence of $\widehat{\mathbb{P}}(\ell)$ for each ℓ does not guarantee weak convergence.

2.5. **Portmanteau theorem.**

Theorem 2.4. *Let $(\mathbb{P}_n)_{n\geqslant 1}$ be a sequence of probability measures on a metric space (S,d) and let $\mathbb{P}$ be a probability measure on S. The following statements are equivalent:*

(1) $\mathbb{P}_n \Rightarrow \mathbb{P}$.
(2) For each uniformly continuous and bounded function $f\colon S \to \mathbb{R}$, $\int_S f(x)\, d\mathbb{P}_n(x) \to \int_S f(x)\, d\mathbb{P}(x)$.
(3) For all closed set $F \subset S$, $\limsup_{n\to\infty} \mathbb{P}_n(F) \leqslant \mathbb{P}(F)$.
(4) For all open set $O \subset S$, $\liminf_{n\to\infty} \mathbb{P}_n(O) \geqslant \mathbb{P}(O)$.
(5) For each subset A of S such that $\mathbb{P}\left(\overline{A} \setminus \mathring{A}\right) = 0$, $\mathbb{P}_n(A) \to \mathbb{P}(A)$.

Note that implication 1. $\Rightarrow$ 2. holds by definition; 3. $\Rightarrow$ 4. follows by taking complement.

Proof that 2. $\Rightarrow$ 3. in Theorem 2.4. We assume that for each uniformly continuous and bounded function $f\colon S \to \mathbb{R}$, $\int_S f(x)\, d\mathbb{P}_n(x) \to \int_S f(x)\, d\mathbb{P}(x)$. We have to prove that for all closed set $F \subset S$, $\limsup_{n\to\infty} \mathbb{P}_n(F) \leqslant \mathbb{P}(F)$.

Let F be a closed set and consider for each integer k,

$$f_{F,k}(x) = \max\{0, 1 - kd(x,F)\}, \tag{2.5.1}$$

where $d(x,A) = \inf\{d(x,a), a \in A\}$.

Then $f_{F,k}$ is a uniformly continuous and bounded function. Moreover, if $x \in F$, then $d(x,F) = 0$ hence $\mathbf{1}(F) \leqslant f_{F,k}$.

Integrating with respect to $\mathbb{P}_n$ gives $\mathbb{P}_n(F) \leqslant \int f_{F,k}(x)\, d\mathbb{P}_n(x)$ and taking the lim sup gives

$$\limsup_{n\to\infty} \mathbb{P}_n(F) \leqslant \int f_{F,k}(x)\, d\mathbb{P}(x) \leqslant \mathbb{P}(G_k),$$

where $G_k = \{x \in X, d(x,F) < 1/k\}$. We conclude by noticing that $\mathbb{P}(G_k) \to \mathbb{P}(F)$, as $\bigcap_{k\geqslant 1} G_k = \{x \in X, d(x,F) = 0\} = F$. □

As mentioned before, 3. and 4. are equivalent, by taking complements.

Proof that 4. $\Rightarrow$ 5. in Theorem 2.4. Observe that

$$\liminf_{n\to\infty} \mathbb{P}_n\left(\mathring{A}\right) \leqslant \liminf_{n\to\infty} \mathbb{P}_n(A) \leqslant \limsup_{n\to\infty} \mathbb{P}_n(A) \leqslant \limsup_{n\to\infty} \mathbb{P}_n\left(\overline{A}\right).$$

Therefore,

$$\mathbb{P}\left(\mathring{A}\right) \leqslant \liminf_{n\to\infty} \mathbb{P}_n(A) \leqslant \limsup_{n\to\infty} \mathbb{P}_n(A) \leqslant \mathbb{P}\left(\overline{A}\right).$$

Since $\mathbb{P}\left(\overline{A} \setminus \mathring{A}\right) = 0$, it follows that $\mathbb{P}\left(\mathring{A}\right) = \mathbb{P}\left(\overline{A}\right) = \mathbb{P}(A)$ hence $\mathbb{P}_n(A) \to \mathbb{P}(A)$.

□

Proof that 5. ⇒ 1. in Theorem 2.4. Let f be a continuous and bounded function. Assume first that for some $M > 0$, $0 \leqslant f(x) \leqslant M$ for each $x \in S$. Then

$$\int_S f(x)\, d\mathbb{P}_n(x) = \int_0^M \mathbb{P}_n(\{x \mid f(x) > t\})\, dt.$$

For each t, the set $A_t := \{x \mid f(x) > t\}$ is open (since f is continuous) and $\overline{A_t} \subset \{x \mid f(x) \geqslant t\}$. Therefore, the set of the t such that $\mathbb{P}\left(\overline{A_t} \setminus \mathring{A}_t\right) > 0$ is contained in the set of the t such that $\mathbb{P}(\{x \in S \mid f(x) = t\}) > 0$, which is at most countable. Therefore, by the dominated convergence theorem applied on $(0, M)$ endowed with the Lebesgue measure,

$$\int_S f(x)\, d\mathbb{P}_n(x) \to \int_S f(x)\, d\mathbb{P}(x).$$

For the general case, write $f = \max\{f, 0\} - \max\{-f, 0\}$. □

2.6. **Tightness.**

Definition 2.5. *Let (S, d) be a metric space. The sequence of Borel probability measures (on S) $(\mathbb{P}_n)_{n \geqslant 1}$ is tight if for each positive $\varepsilon > 0$, there exists a compact set $K = K(\varepsilon)$ such that $\sup_{n \geqslant 1} \mathbb{P}_n(S \setminus K) \leqslant \varepsilon$.*

When $S = \mathbb{R}$ and $d(x, y) = |x - y|$, this is equivalent to the definition of the previous lecture, as a compact set is contained in a bounded interval.

If $\mathbb{P}_n = \mathbb{P}$ for each n, is $(\mathbb{P}_n)_{n \geqslant 1}$ tight?

How is tightness related to weak convergence?

Theorem 2.6 (Tightness of a single probability measure)**.** *Let (S, d) be a separable complete metric space and let $\mathbb{P}$ be a Borel probability measure on S. For each positive ε, there exists a compact set K such that $\mathbb{P}(S \setminus K) \leqslant \varepsilon$.*

Proof. By separability, there exists a sequence $(x_i)_{i \geqslant 1}$ of elements of S such that for each $k \geqslant 1$, $S = \bigcup_{i \geqslant 1} B(x_i, 1/k)$, where $B(x, r) = \{y \in S \mid d(x, y) < r\}$. Let $\varepsilon > 0$ be fixed. For each $k \geqslant 1$, there exists N_k such that $\mathbb{P}\left(\bigcup_{i=1}^{N_k} \overline{B(x_i, 1/k)}\right) > 1 - \varepsilon 2^{-k}$, as $\bigcup_{i=1}^{N} B(x_i, 1/k) \uparrow S$. Let

$$K = \bigcap_{k \geqslant 1} \bigcup_{i=1}^{N_k} \overline{B(x_i, 1/k)}.$$

Then K is compact, as it is a relatively compact closed set in a complete metric space. Moreover,

$$\begin{aligned}
\mathbb{P}(S \setminus K) &= \mathbb{P}\left(\bigcup_{k \geqslant 1} \bigcap_{i=1}^{N_k} \left(S \setminus \overline{B(x_i, 1/k)}\right)\right) \\
&\leqslant \sum_{k \geqslant 1} \mathbb{P}\left(\bigcap_{i=1}^{N_k} \left(S \setminus \overline{B(x_i, 1/k)}\right)\right) \\
&\leqslant \sum_{k \geqslant 1} \varepsilon 2^{-k} \\
&= \varepsilon
\end{aligned}$$

hence $\mathbb{P}(K) \geqslant 1 - \varepsilon$. □

2.7. **Prokhorov theorem.**

Theorem 2.7. *Let (S,d) be a metric space.*

(1) *Assume that (S,d) is separable and complete. If $(\mathbb{P}_n)_{n\geqslant 1}$ is a sequence of probability measures such that every subsequence has a weakly convergent subsequence, then $(\mathbb{P}_n)_{n\geqslant 1}$ is tight.*
(2) *If a sequence of probability measures $(\mathbb{P}_n)_{n\geqslant 1}$ is tight, then it admits a weakly convergent subsequence.*

Proof of 1.

Lemma 2.8. *Let $(O_i)_{i\geqslant 1}$ be a sequence of open sets of S such that $S=\bigcup_{i\geqslant 1} O_i$. Then for each positive ε, there exists $k\geqslant 1$ such that for each $n\geqslant 1$, $\mathbb{P}_n\left(\bigcup_{i=1}^k O_i\right)>1-\varepsilon$.*

Proof. Suppose not. Then there exists an ε_0 such that for each k, there exists $n_k\geqslant 1$ such that $\mathbb{P}_{n_k}\left(\bigcup_{i=1}^k O_i\right)\leqslant 1-\varepsilon_0$. Extract from $(\mathbb{P}_{n_k})_{k\geqslant 1}$ a weakly convergent subsequence $\left(\mathbb{P}_{n_{k_\ell}}\right)_{\ell\geqslant 1}$ to some $\mathbb{P}$. Then apply item 4. of portmanteau theorem with $O=\bigcup_{i=1}^K O_i$ for a fixed K, we get that

$$\mathbb{P}\left(\bigcup_{i=1}^K O_i\right)\leqslant \liminf_{\ell\to\infty}\mathbb{P}_{n_{k_\ell}}\left(\bigcup_{i=1}^K O_i\right).$$

As K is fixed, $K\leqslant k_\ell$ for ℓ large enough hence $\bigcup_{i=1}^K O_i\subset\bigcup_{i=1}^{k_\ell} O_i$ and we get that

$$\mathbb{P}\left(\bigcup_{i=1}^K O_i\right)\leqslant 1-\varepsilon_0.$$

As K is arbitrary, we would get that $\mathbb{P}(S)\leqslant 1-\varepsilon_0$, a contradiction. □

Let $\varepsilon>0$ be fixed. Take a dense countable set $\{x_i, i\geqslant 1\}$. For each m, apply the lemma to $O_i=B(x_i,1/m)$ and $\varepsilon'=\varepsilon 2^{-m}$. We get an integer k_m such that for each $n\geqslant 1$, $\mathbb{P}_n\left(\bigcup_{i=1}^{k_m} B(x_i,1/m)\right)>1-\varepsilon 2^{-m}$. Let

$$K:=\bigcap_{m\geqslant 1}\bigcup_{i=1}^{k_m}\overline{B(x_i,1/m)}.$$

Then K is compact (as S is complete)

$$\mathbb{P}_n(S\setminus K)\leqslant\sum_{m\geqslant 1}\varepsilon 2^{-m}\leqslant\varepsilon.$$

□

Proof of 2 of Theorem 2.7 . We will first give a proof in the particular case where (S,d) is compact. Let $(\mathbb{P}_n)_{n\geqslant 1}$ be a tight collection of probability measures on S.

It is known that $C(S)$, the space of continuous functions from S to $\mathbb{R}$, endowed with the metric $\rho(f,g):=\sup_{t\in[0,1]}|f(t)-g(t)|$, is separable.

Let $(f_j)_{j\geqslant 1}$ be a dense sequence of $C(S)$. Using a diagonal extraction process, it possible to find a subsequence $(\mathbb{P}_{n_k})_{k\geqslant 1}$ such that for each $j\geqslant 1$, $\left(\int_S f_j(x)\,d\mathbb{P}_{n_k}(x)\right)_{k\geqslant 1}$ converges. By density, we derive that for each $f \in C(S)$, $\left(\int_S f(x)\, d\mathbb{P}_{n_k}(x)\right)_{k \geqslant 1}$ is Cauchy hence convergent to some $\ell(f)$.

The map $\ell \colon f \in C(S) \mapsto \ell(f)$ is linear, continuous and bounded. By Riesz theorem, it can be represented as the integral with respect to a probability measure $\mathbb{P}$, which is the weak limit of $(\mathbb{P}_{n_k})_{k \geqslant 1}$.

Proposition 2.9. *Let (S,d) be a separable metric space. There exists a compact metric space (Y,ρ) and a map $T\colon S \to Y$ such that $T \colon S \to T(S)$ is an homeomorphism.*

Proof. Let $Y = [0,1]^{\mathbb{N}} = \left\{(y_i)_{i \geqslant 1}, 0 \leqslant y_i \leqslant 1\right\}$ and

$$\rho\left((y_i)_{i \geqslant 1}, (y_i')_{i \geqslant 1}\right) = \sum_{i \geqslant 1} 2^{-i} \left|y_i - y_i'\right|.$$

It is easy to check that ρ is well-defined and is a metric. With this metric, Y is compact as for each k, Y can be covered by finitely many balls of radius $\leqslant 2^{-k}$ (take as center points of the form $\left(i_1 2^{-k}, \dots, i_k 2^{-k}, 0, 0, \dots\right)$, $0 \leqslant i_1, \dots, i_k \leqslant 2^k$).

Since S is separable, there exists a sequence $(a_i)_{i \geqslant 1}$ which is dense in S. Define for $i \geqslant 1$ the map $f_i \colon S \to [0,1]$ by

$$f_i(x) = \min\{1, d(x, a_i)\}.$$

Since for each $i \geqslant 1$, the map $x \mapsto d(x, a_i)$ is continuous, so is a_i.

For $x \in S$, we define the map $T \colon S \to Y$ by

$$T(x) := (f_i(x))_{i \geqslant 1}.$$

The map T is continuous, since for each k,

$$\rho(T(x), T(x')) \leqslant 2^{-k} + \sum_{i=1}^{k} 2^{-i} d(x, x').$$

We will then show that T is injective and then a homeomorphism.

Lemma 2.10. *Let $C \subset S$ be a closed set and $x \in S \setminus C$. Then there exist a positive ε_0 and $i \geqslant 1$ such that $f_i(x) \leqslant \varepsilon_0/3$ and for each $y \in C$, $f_i(y) \geqslant 2\varepsilon_0/3$.*

Proof. Let $\varepsilon_0 := \min\{1, d(x, C)\}$, where $d(x, C) = \inf\{d(x,y), y \in C\}$. Since C is closed and $x \notin C$, $\varepsilon_0 > 0$. Moreover, by definition, $\varepsilon_0 \leqslant 1$. Let $i \geqslant 1$ be such that $d(x, a_i) \leqslant \varepsilon_0/3$. Then

$$f_i(x) = \min\{1, d(x, a_i)\} \leqslant \min\{1, \varepsilon_0/3\} = \varepsilon_0/3.$$

Let $y \in C$. First observe that by the reversed triangular inequality,

$$f_i(y) = \min\{1, d(y, a_i)\} \geqslant \min\{1, d(y, x) - d(x, a_i)\}$$

Since $y \in C$, $d(y, x) \geqslant d(x, C)$ and we derive that

$$f_i(y) \geqslant \min\{1, d(x, C) - d(x, a_i)\} \geqslant \min\{1, d(x, C) - \varepsilon_0/3\}$$

Since $d(x, C) \geqslant \min\{1, d(x, C)\} = \varepsilon_0$, we get $f_i(y) \geqslant 2\varepsilon_0/3$. ◻

The map T is injective: let $x \neq y$ and apply the Lemma with $C = \{y\}$: we get $\varepsilon_0 > 0$ and $i \geqslant 1$ such that $f_i(x) \leqslant \varepsilon_0/3$ and $f_i(y) \geqslant 2\varepsilon_0/3$. Therefore, $f_i(x) \neq f_i(y)$ hence $T(x) \neq T(y)$.

As a consequence, $T\colon S \to T(S)$ is a bijection.

We have already seen that T is continuous. In order to see that the inverse map is continuous, it suffices to prove that if $(x_n)_{n\geqslant 1}$ is a sequence of elements of S such that $T(x_n) \to T(x)$, then $d(x_n, x) \to 0$. We will prove the contrapositive: if $d(x_n, x)$ does not converge to 0, there exists a positive δ and a subsequence $(x_{n_k})_{k\geqslant 1}$ such that $d(x, x_{n_k}) > \delta$ for each k. Apply the lemma to $C = \overline{\{x_{n_k}, k \geqslant 1\}}$, we get that there exist $\varepsilon_0 > 0$ and i such that for each $k \geqslant 1$, $f_i(x) \leqslant \varepsilon_0/3$ and $2\varepsilon_0/3 \leqslant f_i(x_{n_k})$. In particular, $f_i(x_{n_k})$ does not converge to $f_i(x)$ hence $T(x_{n_k})$ does not converge to $T(x)$.

This ends the proof of the Proposition. □

Let (Y, ρ) be the compact metric space obtained in the previous proposition and $T\colon S \to Y$ be a map such that $T\colon S \to T(S)$ is a homeomorphism. Let $(\mathbb{P}_n)_{n\geqslant 1}$ be a tight sequence of probability measures on S. Define

$$\mu_n(B) := \mathbb{P}_n\left(T^{-1}B\right), B \in \mathcal{B}(Y).$$

[note that $T^{-1}B$ is a Borel subset of S, hence this definition makes sense].

Then $(\mu_n)_{n\geqslant 1}$ is a sequence of probability measures on the compact set Y. By Prokhorov theorem (in compact case), we can extract a subsequence $(\mu_{n_k})_{k\geqslant 1}$ which converges weakly to some probability measure μ on Y.

Let $Y_0 := T(S)$.

Lemma 2.11. *There exists a set $Y_1 \in \mathcal{B}(Y)$ such that $Y_1 \subset Y_0$ and $\mu(Y_1) = 1$.*

Proof. Since $(\mathbb{P}_n)_{n\geqslant 1}$ is tight, for each $j \geqslant 1$, there exists a compact set $K_j \subset S$ such that $\mathbb{P}_n(S \setminus K_j) \geqslant 1 - 1/j$. Since T is continuous, $T(K_j)$ is a compact subset of Y hence closed in Y. By portmanteau theorem,

$$\mu(T(K_j)) \geqslant \limsup_{k\to\infty} \mu_{n_k}(T(K_j))$$

and by definition of μ_{n_k}, $\mu_{n_k}(T(K_j)) = \mathbb{P}_{n_k}(K_j) \geqslant 1 - 1/j$ hence $\mu(T(K_j)) \geqslant 1 - 1/j$. Let $Y_1 = \bigcup_{j\geqslant 1} T(K_j) \subset Y_0$; then $Y_1 \in \mathcal{B}(Y)$ and $\mu(Y_1) = 1$. □

Define $\mu_0(A) := \mu(A \cap Y_0)$ for $A \in \mathcal{B}(Y_0)$ and $\mathbb{P}(A) = \mu_0(T(A))$ for $A \in \mathcal{B}(S)$. Then $\mathbb{P}$ is a probability measure. Let us show that $\mathbb{P}_{n_k} \Rightarrow \mathbb{P}$. To do so, let F be a closed subset of S. Then $T(F)$ is closed in $T(X) = Y_0$ which means that there exists a closed subset C of Y such that $T(F) = C \cap Y_0$. Observe that $F = T^{-1}C$, as $T(F)$ contains no point of the complement of Y_0. Therefore,

$$\limsup_{k\to\infty} \mathbb{P}_{n_k}(F) = \limsup_{k\to\infty} \mathbb{P}_{n_k}\left(T^{-1}C\right).$$

By definition of μ_{n_k}, this becomes

$$\limsup_{k\to\infty} \mathbb{P}_{n_k}(F) = \limsup_{k\to\infty} \mu_{n_k}(C)$$

and since $\mu_{n_k} \to \mu$ weakly,

$$\limsup_{k\to\infty} \mathbb{P}_{n_k}(F) \leqslant \mu(C).$$

Now, with Y_1 like in the previous lemma (that is, $Y_1 \subset Y_0$ and $\mu(Y_1) = 1$),

$$\mu(C) = \mu(C \cap Y_1) + \mu(C \cap (Y \setminus Y_1)).$$

Since μ is concentrated on Y_1, $\mu(C \cap (Y \setminus Y_1)) = 0$. Moreover, $\mu(C \cap Y_1) = \mu(C \cap Y_0) = \mu(T(F)) = \mathbb{P}(F)$.

We showed that $\limsup_{k\to\infty} \mathbb{P}_{n_k}(F) \leqslant \mathbb{P}(F)$ for each closed set F hence $\mathbb{P}_{n_k} \Rightarrow \mathbb{P}$. □

How to prove weak convergence in metric spaces? Let (S, d) be a separable complete metric space. Suppose that we want to prove that a sequence of probability measure $(\mathbb{P}_n)_{n\geqslant 1}$ converges weakly.

We have seen that tightness is a necessary condition, like in the real valued case. We also have to prove that the potential limits of subsequences are the same.

Therefore, we have to look for a sufficient condition which guarantees that two probability measures coincide, that is, reducing the checking of $\mathbb{P}(B) = \mathbb{P}'(B)$ for each Borel set B to a small class of sets.

Regularity of Borel measures on metric spaces

Theorem 2.12. *Let $\mathbb{P}$ be a probability measure on the Borel subsets of a metric space (S, d). For each Borel set A and each positive ε, there exists a closed set F and an open set O such that $F \subset A \subset O$ and $\mathbb{P}(O \setminus F) < \varepsilon$.*

Proof. Let $\mathcal{A} := \{A \subset S \mid \forall \varepsilon > 0, \exists F$ closed and O open s.t.$F \subset A \subset O$ and $\mathbb{P}(O \setminus F) < \varepsilon\}$.

We will show that $\mathcal{A}$ is a σ-algebra containing the open sets.

$\emptyset \in \mathcal{A}$ because we can choose $F = O = \emptyset$.

Let $A \in \mathcal{A}$. For a fixed positive ε, let F closed and O open such that $\mathbb{P}(O \setminus F) < \varepsilon$ and $F \subset A \subset O$. Let $F' := S \setminus O$ and $O' = S \setminus F$. Then $\mathbb{P}(O' \setminus F') < \varepsilon$ and $F' \subset S \setminus A \subset O'$. Therefore, $S \setminus A \in \mathcal{A}$.

Let A_k, $k \geqslant 1$ be elements of $\mathcal{A}$. We first show that $\bigcup_{k=1}^n A_k \in \mathcal{A}$ for each n. Indeed, for each $k \geqslant 1$, let F_k closed and O_k open such that $F_k \subset A_k \subset O_k$ and $\mathbb{P}(O_k \setminus F_k) < 2^{-k}\varepsilon$. Let $F := \bigcup_{k=1}^n F_k$ and $O := \bigcap_{k=1}^n O_k$. Then F is closed, O is open and $F \subset \bigcup_{k=1}^n A_k \subset O$. Moreover,

$$\mathbb{P}(F \setminus O) \leqslant \sum_{k=1}^n \mathbb{P}(F_k \setminus O) \leqslant \sum_{k=1}^n \mathbb{P}(F_k \setminus O_k) \leqslant \varepsilon \sum_{k=1}^n 2^{-k} < \varepsilon.$$

In order to show that $\bigcup_{k\geqslant 1} A_k \in \mathcal{A}$, we take n such that $\mathbb{P}\left(\bigcup_{k\geqslant 1} A_k \setminus (\bigcup_{k=1}^n A_k)\right) < \varepsilon/2$. Let F_1 and F_2 be closed and such that $\mathbb{P}(\bigcup_{k=1}^n A_k \setminus F_1) < \varepsilon/2$ and $\mathbb{P}((S \setminus \bigcup_{k=1}^n A_k) \setminus F_2) < \varepsilon/2$. Take $F = F_1$ and $O = S \setminus F_2$.

This shows that $\mathcal{A}$ is a σ-algebra.

It remains to check that $\mathcal{A}$ contains the open sets, or equivalently, the closed sets.

If A is closed, we can choose $F = A$. Let $O_k := \{x \in S \mid d(x, F) < 1/k\}$, where $d(x, F) = \inf\{d(x, y), y \in F\}$. Then O_k is open, $O_{k+1} \subset O_k$ and $\bigcap_{k\geqslant 1} O_k = \{x \in S \mid d(x, F) = 0\}$ which is equal to F, as F is closed.

Therefore, $\mathbb{P}(O_k \setminus F) \to 0$. □

Why "regularity"? Because $\mathbb{P}(A) = \sup\{\mathbb{P}(F), F$ closed $, F \subset A\}$ and $\mathbb{P}(A) = \inf\{\mathbb{P}(O), O$ open $, A \subset O\}$.

Corollary 2.13. *Let $\mathbb{P}$ and $\mathbb{P}'$ be two probability measures on a metric space (S, d). Suppose that for each open set O, $\mathbb{P}(O) = \mathbb{P}'(O)$. Then $\mathbb{P}(B) = \mathbb{P}'(B)$ for each Borel set B.*

2.8. **Case of the space $C[0,1]$.** In general, it is hard to check that $\mathbb{P}(O) = \mathbb{P}'(O)$ for each open set O.

Let us see how to treat the particular case $O = B(f_0, r)$, that is, where O is an open ball. Saying that $f \in O$ is equivalent to say that there exists k_0 such that for each $t \in [0,1] \cap \mathbb{Q}$, $|f(t) - f_0(t)| \leqslant r - 1/k_0$.

Therefore, knowing the probability of sets of the form

$$\left\{x \in C[0,1] \mid (x(t_1), \dots, x(t_d)) \in B, B \in \mathcal{B}\left(\mathbb{R}^d\right), d \geqslant 1, 0 \leqslant t_1, \dots, t_d \leqslant 1\right\} \tag{2.8.1}$$

is enough to determine the probability of a ball.

Let $(X_n)_{n \geqslant 1}$ be a sequence of random elements of $C[0,1]$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, that is, for each $\omega \in \Omega$, the map $t \in [0,1] \mapsto X_n(\omega, t)$ is continuous and for each $B \in \mathcal{B}(C[0,1])$, $\{\omega, t \mapsto X_n(\omega, t)\} \in \mathcal{F}$.

Definition 2.14 (Finite dimensional distributions). *The law of X_n, denoted by $\mathbb{P}_n$, is defined as $\mathbb{P}_n(B) = \mathbb{P}\{\omega \mid (t \mapsto X_n(\omega, t)) \in B\}$. We say that $X_n \Rightarrow X$ weakly in $C[0,1]$ if $\mathbb{P}_n \Rightarrow \mathbb{P}$, where $\mathbb{P}(B) = \mathbb{P}\{\omega \mid (t \mapsto X(\omega, t)) \in B\}$.*

The collection of the finite dimensional distributions of X_n is the collection of the distributions of the vectors $(X_n(t_i))_{i=1}^d$ where $d \geqslant 1$ and $t_1, \dots, t_d \in [0,1]$.

Example 2.15. Suppose that $X_n(\omega, t) = x_n(t)$, where $x_n \colon [0,1] \to \mathbb{R}$ is a continuous (deterministic) function.

Then the law of $(X_n(t_i))_{i=1}^d$ is the Dirac mass at the point $(x_n(t_i))_{i=1}^d$ of $\mathbb{R}^d$.

Theorem 2.16 (Finite dimensional distributions characterize distributions). *Let X and Y be two random elements of $C[0,1]$ having the same finite dimensional distributions, that is, if $d \geqslant 1$ and $t_1, \dots, t_d \in [0,1]$, then the vectors $(X(t_i))_{i=1}^d$ and $(Y(t_i))_{i=1}^d$ have the same distribution. Then X and Y have the same distribution, that is, for each $B \in \mathcal{B}(C[0,1])$, $\mathbb{P}\{X \in B\} = \mathbb{P}\{Y \in B\}$.*

Proof. By the regularity of measures, it suffices to prove that $\mathbb{P}\{X \in O\} = \mathbb{P}\{Y \in O\}$ for each open set O. Since $C[0,1]$ is separable, each open set can be written as a countable union of open balls. Using $\mathbb{P}(O_n) \to \mathbb{P}(O)$ if $O_n \uparrow O$, it suffices to prove $\mathbb{P}\{X \in O\} = \mathbb{P}\{Y \in O\}$ when O is a finite union of open balls.

Let $O := \bigcup_{i=1}^N B(f_i, r_i)$, where $f_i \in C[0,1]$ and $r_i > 0$. Let $(q_k)_{k \geqslant 1}$ be an enumeration of the rationals of $[0,1]$. Since

$$B(f_i, r_i) = \bigcup_{j \geqslant 1} \bigcap_{k \geqslant 1} \left\{x \in C[0,1], |f_i(q_k) - x(q_k)| \leqslant r_i - j^{-1}\right\},$$

O can be expressed as $\bigcap_{K \geqslant 1} F_K$, where

$$F_K = \bigcup_{i=1}^N \bigcup_{j \geqslant 1} \bigcap_{k=1}^K \left\{x \in C[0,1], |f_i(q_k) - x(q_k)| \leqslant r_i - j^{-1}\right\}.$$

Define the Borel subset B_K of $\mathbb{R}^K$ by

$$B_K := \left\{ (v_k)_{k=1}^K \in \mathbb{R}^K, \exists 1 \leqslant i \leqslant N \mid |f_i(q_k) - v_k| < r_i \right\}.$$

Then $\{X \in F_K\} = \left\{ (X(q_k))_{k=1}^K \in B_K \right\}$ and $\mathbb{P}\{X \in F_K\} = \mathbb{P}\{Y \in F_K\}$ follows from equality of the finite dimensional distributions. □

We thus have the following strategy to prove the weak convergence in $C[0,1]$.

Theorem 2.17. *Let $(X_n)_{n\geqslant 1}$ be a sequence of random elements of $C[0,1]$. Suppose that $(X_n)_{n\geqslant 1}$ is tight and that there exists a random element X of $C[0,1]$ such that for each $d \geqslant 1$ and $t_1, \dots, t_d \in [0,1]$, the sequence of vectors $\left((X_n(t_i))_{i=1}^d\right)_{n\geqslant 1}$ converges in distribution to $(X(t_i))_{i=1}^d$. Then $X_n \Rightarrow X$ weakly in $C[0,1]$.*

Proof. Suppose not. Then there exists a continuous and bounded function $f\colon C[0,1] \to \mathbb{R}$, an increasing sequence of integers $(n_k)_{k\geqslant 1}$ and $\varepsilon_0 > 0$ such that for each k,

$$|\mathbb{E}[f(X_{n_k})] - \mathbb{E}[f(X)]| > \varepsilon_0.$$

By Prokhorov theorem, extract from $(X_{n_k})_{k\geqslant 1}$ a weakly convergent subsequence to some Y. We thus get $|\mathbb{E}[f(Y)] - \mathbb{E}[f(X)]| > \varepsilon_0$. But the finite dimensional distributions of $(X_{n_k})_{k\geqslant 1}$ converge to those of X, then by the previous theorem, X and Y have the same distribution and we get a contradiction. □

How to prove tightness in $C[0,1]$?

Proving weak convergence in $C[0,1]$ reduces to show the convergence of finite dimensional distributions (which is easier, as we only deal with vectors) and tightness.

Recall (Arzelà-Ascoli) that a set $K \subset C[0,1]$ is compact if and only if it is closed, $\sup_{x\in K}|x(0)| < \infty$ and

$$\lim_{\delta\to 0}\sup_{x\in K}\omega(x,\delta) = 0,$$

where $\omega(x,\delta) := \sup\{|x(t) - x(s)|, s,t \in [0,1], |t-s| < \delta\}$.

Question: is there a simple characterization of tightness of a sequence of random elements of $C[0,1]$?

Here we present a tightness criterion.

Theorem 2.18. *A sequence of random elements of $C[0,1]$ is tight if and only if the following two conditions are satisfied:*

(1) the sequence $(X_n(0))_{n\geqslant 1}$ is tight in $\mathbb{R}$ and
(2) for each positive ε,

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\{\sup\{|X_n(t) - X_n(s)|, s,t \in [0,1], |t-s| < \delta\} > \varepsilon\} = 0. \tag{2.8.2}$$

Proof of Theorem 2.18. $\Rightarrow$ We assume that $(X_n)_{n\geqslant 1}$ is tight in $C[0,1]$. For each positive ε, there exists a compact set $K(\varepsilon) \subset C[0,1]$ such that $\sup_{n\geqslant 1}\mathbb{P}\{X_n \notin K(\varepsilon)\} < \varepsilon$. By Arzelà-Ascoli theorem, $\sup_{x\in K(\varepsilon)}|x(0)| \leqslant C(\varepsilon)$ and for some $\delta(\varepsilon)$,

$$\sup_{x\in K}\omega(x,\delta(\varepsilon)) < \varepsilon, \tag{2.8.3}$$

Since $\mathbb{P}\left\{\left|X_n(0)\right| > C(\varepsilon)\right\} \leqslant \mathbb{P}\left\{X_n \notin K(\varepsilon)\right\} < \varepsilon$, it follows that $\sup_{n\geqslant 1} \mathbb{P}\left\{\left|X_n(0)\right| > C(\varepsilon)\right\} < \varepsilon$ hence $\left(X_n(0)\right)_{n\geqslant 1}$ is tight in $\mathbb{R}$.

Let $\varepsilon > 0$ and let $\delta(\varepsilon)$ like in (2.8.3). Then

$$\left\{\sup\left\{\left|X_n(t) - X_n(s)\right|, s, t \in [0,1], |t-s| < \delta(\varepsilon)\right\} > \varepsilon\right\} \subset \left\{X_n \notin K(\varepsilon)\right\}$$

hence

$$g(\varepsilon) := \lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}\left\{\sup\left\{\left|X_n(t) - X_n(s)\right|, s, t \in [0,1], |t-s| < \delta\right\} > \varepsilon\right\} \leqslant \varepsilon. \tag{2.8.4}$$

We got $g(\varepsilon) \leqslant \varepsilon$. If ε_0 is fixed, for $\varepsilon < \varepsilon_0$, $g(\varepsilon_0) \leqslant g(\varepsilon) \leqslant \varepsilon$ hence $g(\varepsilon) = 0$.

$\Leftarrow$ Let ε be fixed; we have to find a compact set K such that $\sup_{n\geqslant 1} \mathbb{P}\left\{X_n \notin K\right\} < \varepsilon$. Since $\left(X_n(0)\right)_{n\geqslant 1}$ is tight in $\mathbb{R}$, there exists C such that $\sup_{n\geqslant 1} \mathbb{P}\left\{\left|X_n(0)\right| > C\right\} < \varepsilon/2$.

The second condition applied with $1/j$ for each fixed $j \geqslant 1$ gives δ_j such that

$$\limsup_{n\to\infty} \mathbb{P}\left\{\sup\left\{\left|X_n(t) - X_n(s)\right|, s, t \in [0,1], |t-s| < \delta_j\right\} > 1/j\right\} < \varepsilon 2^{-j}.$$

That is, there exists N_j such that

$$\sup_{n\geqslant N_j} \mathbb{P}\left\{\sup\left\{\left|X_n(t) - X_n(s)\right|, s, t \in [0,1], |t-s| < \delta_j\right\} > 1/j\right\} < \varepsilon 2^{-j}.$$

For each $n \leqslant N_j$, $\{X_n\}$ is tight in $C[0,1]$ (since this space is separable and complete) hence we can find $\delta_{n,j}$ such that

$$\sup_{n\leqslant N_j} \mathbb{P}\left\{\sup\left\{\left|X_n(t) - X_n(s)\right|, s, t \in [0,1], |t-s| < \delta_{n,j}\right\} > 1/j\right\} < \varepsilon 2^{-j}.$$

Therefore, taking $\delta'_j := \min\left\{\delta_j, \delta_{1,j}, \dots, \delta_{N_j,j}\right\}$

$$\sup_{n\geqslant 1} \mathbb{P}\left\{\sup\left\{\left|X_n(t) - X_n(s)\right|, s, t \in [0,1], |t-s| < \delta'_j\right\} > 1/j\right\} < \varepsilon 2^{-j}.$$

We got $\sup_{n\geqslant 1} \mathbb{P}\left\{\left|X_n(0)\right| > C\right\} < \varepsilon/2$ and

$$\sup_{n\geqslant 1} \mathbb{P}\left\{\sup\left\{\left|X_n(t) - X_n(s)\right|, s, t \in [0,1], |t-s| < \delta'_j > 1/j\right\}\right\} < \varepsilon 2^{-j}.$$

Let $K \subset C[0,1]$ be defined by

$$K := \left\{x : |x(0)| \leqslant C\right\} \cap \bigcap_{j\geqslant 1} \left\{x : \sup\left\{\left|x(t) - x(s)\right|, s, t \in [0,1], |t-s| < \delta'_j\right\} \leqslant 1/j\right\}. \tag{2.8.5}$$

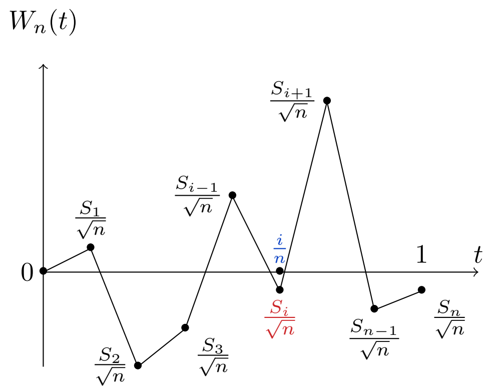
Then K satisfies the wanted requirements. □

3. Invariance principle

3.1. **Partial sum process.** We would like to study the asymptotic behavior of partial sums of a sequence of random variables $\left(X_i\right)_{i\geqslant 1}$, that is, see how $f_n\left(\sum_{i=1}^{k}, 1 \leqslant k \leqslant n\right)$, $f_n \colon \mathbb{R}^n \to \mathbb{R}$ behaves.

We would like to find a sequence of random continous functions W_n which contains all the information of the partial sums, that is, knowning $W_n(t)$ for each $0 \leqslant t \leqslant 1$ allows to derive the value of S_k for each $1 \leqslant k \leqslant n$.

FIGURE 1. The map $t \mapsto W_n(t)$



We can try

$$W_n(t) = \sum_{i=1}^{\lfloor nt \rfloor} X_i, 0 \leqslant t \leqslant 1,$$

where for $x \in \mathbb{R}$, $\lfloor x \rfloor \leqslant x < \lfloor x \rfloor + 1$.

But this function is not continuous, as there are jumps at k/n, $k \in \{1, \ldots, n\}$.

Instead, we define

$$W_n(t) := \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{\lfloor nt \rfloor} X_i + (nt - \lfloor nt \rfloor) X_{\lfloor nt \rfloor + 1} \right), \quad 0 < t \leqslant 1,$$

and $W_n(0) = 0$. Then

- for each $k \in \{0, \ldots, n-1\}$, the map $t \mapsto W_n(t)$ is piecewise linear on $[(k-1)/n, k/n)$. Indeed, for $t \in [(k-1)/n, k/n)$, $\lfloor nt \rfloor = k-1$ and $\sqrt{n} W_n(t) = \sum_{i=1}^{k-1} X_i + (nt - k - +1) X_k$.
- As $t \to k/n^-$, $W_n(t) \to \sum_{i=1}^{k} X_i / \sqrt{n}$ hence $t \mapsto W_n(t)$ is continuous.
- Actually, the map $t \mapsto W_n(t)$ is Hölder continous of exponent α for each $0 < \alpha \leqslant 1$.
- For $k \in \{1, \ldots, n\}$, $W_n(k/n) = \sum_{i=1}^{k} = S_k / \sqrt{n}$.

3.2. **Statement.** Let

$$W_n(t) := \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{\lfloor nt \rfloor} X_i + (nt - \lfloor nt \rfloor) X_{\lfloor nt \rfloor + 1} \right), \quad 0 < t \leqslant 1,$$

We would like to study the convergence of $(W_n)_{n \geqslant 1}$ in the space $S := C[0, 1]$ endowed with the uniform metric $\rho(f, g) = \sup_{0 \leqslant t \leqslant 1} |f(t) - g(t)|$. As a starting point, let us do the i.i.d. case.

One needs to look at the convergence of finite dimensional distributions and show tightness.

Definition 3.1 (Brownian motion). *We call a process $(B_t)_{0 \leqslant t \leqslant 1}$ a standard Brownian motion if it is a centered Gaussian process such that $\operatorname{Cov}(B_s, B_t) = \min\{s, t\}$ and for almost every ω, the map $t \mapsto B_t(\omega)$ is continuous.*

We will see in the next theorem that such a process does exist. Actually, the map $t \mapsto B_t(\omega)$ is almost surely Hölder continuous with exponent α for each $0 < \alpha < 1/2$.

Note that $\operatorname{Cov}(B_s, B_t) = \min\{s,t\}$ implies that for $0 = t_0 < t_1 < \cdots < t_d \leqslant 1$, the family $\left(B_{t_i} - B_{t_{i-1}}\right)_{i=1}^d$ is independent. Indeed, this family forms a Gaussian vector and one can check that $\operatorname{Cov}\left(B_{t_i} - B_{t_{i-1}}, B_{t_j} - B_{t_{j-1}}\right) = 0$ for $i \neq j$.

Theorem 3.2 (Donsker, 1951)**.** *Let $(X_i)_{i\geqslant 1}$ be a centered i.i.d. sequence of random variables such that $\mathbb{E}\left[X_1^2\right] = 1$. Let W_n be defined by*

$$W_n(t) := \frac{1}{\sqrt{n}}\left(\sum_{i=1}^{\lfloor nt \rfloor} X_i + (nt - \lfloor nt \rfloor) X_{\lfloor nt \rfloor + 1}\right), \quad 0 < t \leqslant 1,$$

and $W_n(0) = 0$. Then $(W_n)_{n\geqslant 1}$ converges in distribution in $C[0,1]$ to a standard Brownian motion.

This theorem is usually called functional central limit theorem. We can also meet the terminolog "invariance principle", as the limiting process does not depend on the distribution of X_1. We have a standard Brownian motion provided that X_1 is centered and $\mathbb{E}\left[X_1^2\right] = 1$.

3.3. **Convergence of the finite dimensional distributions.** Let

$$W_n'(t) := \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt \rfloor} X_i +, \quad 0 \leqslant t \leqslant 1.$$

We have seen during the seminar that the finite dimensional distributions of $(W_n')_{n\geqslant 1}$ converge to those of a standard Brownian motion. Therefore, it suffices to show that for each $0 \leqslant t \leqslant 1$,

$$\frac{1}{\sqrt{n}}(nt - \lfloor nt \rfloor) X_{\lfloor nt \rfloor + 1} \to 0 \text{ in probability.}$$

Since $0 \leqslant nt - \lfloor nt \rfloor \leqslant 1$, we get

$$\mathbb{P}\left(\frac{1}{\sqrt{n}}(nt - \lfloor nt \rfloor)\left|X_{\lfloor nt \rfloor + 1}\right| > \varepsilon\right) \leqslant \mathbb{P}\left(\frac{1}{\sqrt{n}}\left|X_{\lfloor nt \rfloor + 1}\right| > \varepsilon\right).$$

Moreover, $\left|X_{\lfloor nt \rfloor + 1}\right|$ has the same distribution as $|X_1|$ hence

$$\mathbb{P}\left(\frac{1}{\sqrt{n}}(nt - \lfloor nt \rfloor)\left|X_{\lfloor nt \rfloor + 1}\right| > \varepsilon\right) \leqslant \mathbb{P}\left(|X_1| > \varepsilon\sqrt{n}\right) \to 0.$$

3.4. **Tightness of partial sum process, a sufficient condition.** In the special case of partial sum process defined by

$$W_n(t) := \frac{1}{\sqrt{n}}\left(\sum_{i=1}^{\lfloor nt \rfloor} X_i + (nt - \lfloor nt \rfloor) X_{\lfloor nt \rfloor + 1}\right), \quad 0 < t \leqslant 1,$$

$W_n(0) = 0$, one can give a tightness sufficient condition in terms of the partials sums $S_k := \sum_{i=1}^k X_i$.

Proposition 3.3. *Suppose that the sequence* $(X_i)_{i\geqslant 1}$ *is such that the family*

$$\left(\frac{1}{n}\max_{1\leqslant k\leqslant n}\left(S_{k+i}-S_i\right)^2\right)_{n\geqslant 1, i\geqslant 0}$$

is uniformly integrable. Then $(W_n)_{n\geqslant 1}$ *is tight in* $C[0,1]$.

Proof. We use the tightness criterion given at page 2 of Lecture 8. Since $W_n(0)=0$, it suffices to check that for each positive ε,

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\left\{\sup\left\{\left|W_n(t)-W_n(s)\right|, s,t\in[0,1], |t-s|<\delta\right\}>\varepsilon\right\}=0.$$

We first give an upper bound of $\sup\left\{\left|W_n(t)-W_n(s)\right|, s,t\in[0,1],|t-s|<\delta\right\}$ for a fixed δ and $n>1/\delta$. Define the interval $I_k:=[k\delta,(k+1)\delta)\cap[0,1]$, $0\leqslant k\leqslant\lfloor 1/\delta\rfloor$ and $J_k:=[k\delta,(k+1)\delta]\cap[0,1]$. If $s,t\in[0,1]$ are such that $|t-s|<\delta$, then s belongs to some I_k and $t\in I_j$ where $|k-j|\leqslant 1$ hence

$$\begin{aligned}\sup&\left\{\left|W_n(t)-W_n(s)\right|, s,t\in[0,1],|t-s|<\delta\right\}\\&\leqslant\max_{0\leqslant k\leqslant\lfloor 1/\delta\rfloor}\max\left\{\sup_{s,t\in J_k}\left|W_n(s)-W_n(t)\right|+\sup_{s\in J_k}\sup_{t\in J_{k+1}}\left|W_n(s)-W_n(t)\right|\right\}.\end{aligned}\tag{3.4.1}$$

Note that

$$\begin{aligned}\max_{0\leqslant k\leqslant\lfloor 1/\delta\rfloor}&\sup_{s\in J_k}\sup_{t\in J_{k+1}}\left|W_n(s)-W_n(t)\right|\\&\leqslant\max_{0\leqslant k\leqslant\lfloor 1/\delta\rfloor}\sup_{s\in J_k}\left|W_n(s)-W_n((k+1)\delta)\right|\\&\qquad+\max_{0\leqslant k\leqslant\lfloor 1/\delta\rfloor}\sup_{t\in J_{k+1}}\left|W_n((k+1)\delta)-W_n(t)\right|\end{aligned}\tag{3.4.2}$$

and these two terms are both smaller than

$$\max_{0\leqslant k\leqslant\lfloor 1/\delta\rfloor+1}\sup_{s,t\in J_k}\left|W_n(s)-W_n(t)\right|.$$

We thus got

$$\begin{aligned}\sup&\left\{\left|W_n(t)-W_n(s)\right|, s,t\in[0,1],|t-s|<\delta\right\}\\&\leqslant 4\max_{0\leqslant k\leqslant\lfloor 1/\delta\rfloor+1}\sup_{t\in J_k}\left|W_n(t)-W_n(k\delta)\right|.\end{aligned}\tag{3.4.3}$$

We have to control $\sup_{k\delta\leqslant t\leqslant(k+1)\delta}\left|W_n(t)-W_n(k\delta)\right|$ in terms of partials sums. Let t be such that $k\delta\leqslant t\leqslant(k+1)\delta$. Then $i/n\leqslant t<(i+1)/n$ for some i and $j/n\leqslant k\delta<(j+1)/n$ for some j, namely, $j=\lfloor nk\delta\rfloor$. Let us first use the bound

$$\begin{aligned}&\left|W_n(t)-W_n(k\delta)\right|\\&\qquad\leqslant\left|W_n(t)-W_n(i/n)\right|+\left|W_n(i/n)-W_n(j/n)\right|+\left|W_n(j/n)-W_n(k\delta)\right|\end{aligned}$$

For the term $\left|W_n(t)-W_n(i/n)\right|$, we use the fact that W_n is affine on $[i/n,(i+1)/n)$ and the slope is X_{i+1} to get that

$$\left|W_n(t)-W_n(i/n)\right|\leqslant\frac{1}{\sqrt{n}}\max_{1\leqslant\ell\leqslant n+1}\left|X_\ell\right|.$$

By the same argument,

$$|W_n(j/n) - W_n(k\delta)| \leqslant \frac{1}{\sqrt{n}} \max_{1\leqslant \ell\leqslant n+1} |X_\ell| .$$

Since $W_n(i/n) = S_i$ and $W_n(j/n) = S_j$, we finally get that

$$|W_n(t) - W_n(k\delta)| \leqslant \frac{2}{\sqrt{n}} \max_{1\leqslant \ell\leqslant n+1} |X_\ell| + \frac{1}{\sqrt{n}} \left|S_i - S_{\lfloor nk\delta\rfloor}\right| .$$

Finally, $i - j \leqslant n\delta$ hence we got that for each n such that $n > 1/\delta$,

$$\begin{aligned}\sup\{|W_n(t) - W_n(s)|, s,t\in[0,1], |t-s|<\delta\} \\ \leqslant \frac{8}{\sqrt{n}} \max_{1\leqslant \ell\leqslant n+1} |X_\ell| + \frac{4}{\sqrt{n}} \max_{0\leqslant k\leqslant \lfloor 1/\delta\rfloor} \max_{\lfloor nk\delta\rfloor+1\leqslant i\leqslant \lfloor nk\delta\rfloor+\lfloor n\delta\rfloor} \left|S_i - S_{\lfloor nk\delta\rfloor}\right| . \quad (3.4.4)\end{aligned}$$

Therefore, it suffices to show that $\frac{1}{\sqrt{n}} \max_{1\leqslant \ell\leqslant n+1} |X_\ell| \to 0$ in probability and for each positive ε,

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}\left(\max_{0\leqslant k\leqslant \lfloor 1/\delta\rfloor} \max_{\lfloor nk\delta\rfloor+1\leqslant i\leqslant \lfloor nk\delta\rfloor+\lfloor n\delta\rfloor} \left|S_i - S_{\lfloor nk\delta\rfloor}\right| > \varepsilon\sqrt{n}\right) = 0.$$

The prove that $\frac{1}{\sqrt{n}} \max_{1\leqslant \ell\leqslant n+1} |X_\ell| \to 0$ in probability will be done during the seminar. In order to prove that

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}\left(\max_{0\leqslant k\leqslant \lfloor 1/\delta\rfloor} \max_{\lfloor nk\delta\rfloor+1\leqslant i\leqslant \lfloor nk\delta\rfloor+\lfloor n\delta\rfloor} \left|S_i - S_{\lfloor nk\delta\rfloor}\right| > \varepsilon\sqrt{n}\right) = 0,$$

we start from the union bound

$$\begin{aligned}\mathbb{P}\left(\max_{0\leqslant k\leqslant \lfloor 1/\delta\rfloor} \max_{\lfloor nk\delta\rfloor+1\leqslant i\leqslant \lfloor nk\delta\rfloor+\lfloor n\delta\rfloor} \left|S_i - S_{\lfloor nk\delta\rfloor}\right| > \varepsilon\sqrt{n}\right) \\ \leqslant \sum_{0\leqslant k\leqslant \lfloor 1/\delta\rfloor} \mathbb{P}\left(\max_{\lfloor nk\delta\rfloor+1\leqslant i\leqslant \lfloor nk\delta\rfloor+\lfloor n\delta\rfloor} \left|S_i - S_{\lfloor nk\delta\rfloor}\right| > \varepsilon\sqrt{n}\right). \quad (3.4.5)\end{aligned}$$

We are reduced to prove that

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \sum_{0\leqslant k\leqslant \lfloor 1/\delta\rfloor} \mathbb{P}\left(\max_{\lfloor nk\delta\rfloor+1\leqslant i\leqslant \lfloor nk\delta\rfloor+\lfloor n\delta\rfloor} \left|S_i - S_{\lfloor nk\delta\rfloor}\right| > \varepsilon\sqrt{n}\right) = 0.$$

Let $M_{j,i}$ denote the random variable $\frac{1}{j} \max_{1\leqslant k\leqslant j} (S_{k+i} - S_i)^2$. Recall that $(M_{j,i})_{i,j\geqslant 1}$ is assumed to be uniformly integrable. We have to prove that

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \sum_{0\leqslant k\leqslant \lfloor 1/\delta\rfloor} \mathbb{P}\left(M_{\lfloor n\delta\rfloor, \lfloor nk\delta\rfloor} > \left(\varepsilon \frac{\sqrt{n}}{\sqrt{\lfloor n\delta\rfloor}}\right)^2\right) = 0.$$

By Markov's inequality,

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\sum_{0\leqslant k\leqslant\lfloor 1/\delta\rfloor}\mathbb{P}\left(M_{\lfloor n\delta\rfloor,\lfloor nk\delta\rfloor}>\left(\varepsilon\frac{\sqrt{n}}{\sqrt{\lfloor n\delta\rfloor}}\right)^2\right)$$
$$\leqslant\lim_{\delta\to 0}\limsup_{n\to\infty}\left(\varepsilon\frac{\sqrt{n}}{\sqrt{\lfloor n\delta\rfloor}}\right)^{-2}$$
$$\sum_{0\leqslant k\leqslant\lfloor 1/\delta\rfloor}\mathbb{E}\left[M_{\lfloor n\delta\rfloor,\lfloor nk\delta\rfloor}\mathbf{1}\left\{M_{\lfloor n\delta\rfloor,\lfloor nk\delta\rfloor}>\left(\varepsilon\frac{\sqrt{n}}{\sqrt{\lfloor n\delta\rfloor}}\right)^2\right\}\right].\quad(3.4.6)$$

As a consequence,

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\sum_{0\leqslant k\leqslant\lfloor 1/\delta\rfloor}\mathbb{P}\left(M_{\lfloor n\delta\rfloor,\lfloor nk\delta\rfloor}>\left(\varepsilon\frac{\sqrt{n}}{\sqrt{\lfloor n\delta\rfloor}}\right)^2\right)$$
$$\leqslant\lim_{\delta\to 0}\limsup_{n\to\infty}\varepsilon^{-2}\delta\left(\lfloor 1/\delta\rfloor+1\right)\sup_{j,i}\mathbb{E}\left[M_{j,i}\mathbf{1}\left\{M_{j,i}>\left(\varepsilon\frac{\sqrt{n}}{\sqrt{\lfloor n\delta\rfloor}}\right)^2\right\}\right].\quad(3.4.7)$$

For each fixed δ, there exists n_0 such that for $n\geqslant n_0$,

$$\frac{\sqrt{n}}{\sqrt{\lfloor n\delta\rfloor}}\geqslant\frac{1}{2\sqrt{\delta}}.$$

Therefore,

$$\limsup_{n\to\infty}\sup_{j,i}\mathbb{E}\left[M_{j,i}\mathbf{1}\left\{M_{j,i}>\left(\varepsilon\frac{\sqrt{n}}{\sqrt{\lfloor n\delta\rfloor}}\right)^2\right\}\right]$$
$$\leqslant\sup_{j,i}\mathbb{E}\left[M_{j,i}\mathbf{1}\left\{M_{j,i}>\left(\frac{\varepsilon}{2\sqrt{\delta}}\right)^2\right\}\right]\quad(3.4.8)$$

and we get that

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\sum_{0\leqslant k\leqslant\lfloor 1/\delta\rfloor}\mathbb{P}\left(M_{\lfloor n\delta\rfloor,\lfloor nk\delta\rfloor}>\left(\varepsilon\frac{\sqrt{n}}{\sqrt{\lfloor n\delta\rfloor}}\right)^2\right)$$
$$\leqslant\varepsilon^{-2}\lim_{\delta\to 0}\sup_{j,i}\mathbb{E}\left[M_{j,i}\mathbf{1}\left\{M_{j,i}>\left(\frac{\varepsilon}{2\sqrt{\delta}}\right)^2\right\}\right]=0.\quad(3.4.9)$$

□

3.5. **Uniform integrability of partial sums.** It remains to check the following.

Proposition 3.4. *Let* $(X_i)_{i\geqslant 1}$ *be an i.i.d. centered sequence such that* $\mathbb{E}\left[X_1^2\right]=1$ *and* $S_k=\sum_{i=1}^k X_i$. *Then the family*

$$\left\{\frac{1}{n}\max_{1\leqslant k\leqslant n}S_k^2,n\geqslant 1\right\}$$

is uniformly integrable.

Note that this will give immediately uniform integrability of

$$\left(\frac{1}{n}\max_{1\leqslant k\leqslant n}\left(S_{k+i}-S_i\right)^2\right)_{n\geqslant 1, i\geqslant 0}.$$

Indeed, for each fixed i,

$$\mathbb{E}\left[\frac{1}{n}\max_{1\leqslant k\leqslant n}\left(S_{k+i}-S_i\right)^2\mathbf{1}\left\{\frac{1}{n}\max_{1\leqslant k\leqslant n}\left(S_{k+i}-S_i\right)^2>R\right\}\right] = \mathbb{E}\left[\frac{1}{n}\max_{1\leqslant k\leqslant n}\left(S_k\right)^2\mathbf{1}\left\{\frac{1}{n}\max_{1\leqslant k\leqslant n}\left(S_k\right)^2>R\right\}\right]. \quad (3.5.1)$$

Equality

$$\mathbb{E}\left[\frac{1}{n}\max_{1\leqslant k\leqslant n}\left(S_{k+i}-S_i\right)^2\mathbf{1}\left\{\frac{1}{n}\max_{1\leqslant k\leqslant n}\left(S_{k+i}-S_i\right)^2>R\right\}\right] = \mathbb{E}\left[\frac{1}{n}\max_{1\leqslant k\leqslant n}\left(S_k\right)^2\mathbf{1}\left\{\frac{1}{n}\max_{1\leqslant k\leqslant n}\left(S_k\right)^2>R\right\}\right]. \quad (3.5.2)$$

follows from the fact that for each i, the vectors $(X_{i+1},\dots,X_{n+i})$ and $(X_1,\dots,X_n)$ have the same distribution (their characteristic function is $(t_1,\dots,t_n)\mapsto\prod_{j=1}^n\mathbb{E}\left[\exp\left(it_jX_j\right)\right]$).

During the seminar, we have see uniform integrability of $\left\{\frac{1}{n}S_n^2, n\geqslant 1\right\}$.

We would like to use a similar proof, but we have to handle maximas of partial sums.

Define for a fixed integer m the random variables

$$X_{i,m} := X_i\mathbf{1}\left\{\left|X_i\right|\leqslant m\right\}-\mathbb{E}\left[X_i\mathbf{1}\left\{\left|X_i\right|\leqslant m\right\}\right],$$
$$X'_{i,m} := X_i\mathbf{1}\left\{\left|X_i\right|> m\right\}-\mathbb{E}\left[X_i\mathbf{1}\left\{\left|X_i\right|> m\right\}\right].$$

Note that since X_i is centered, $X_i = X_{i,m}+X'_{i,m}$. By exercise 15 (sheet 1) appliyed with $Y_n^{(m)} := \frac{1}{n}\max_{1\leqslant k\leqslant n}\left(\sum_{i=1}^k X_{i,m}\right)^2$, it suffices to prove that

(1) for each fixed m, the sequence $\left(Y_n^{(m)}\right)_{n\geqslant 1}$ is uniformly integrable and
(2)
$$\lim_{m\to\infty}\sup_{n\geqslant 1}\mathbb{E}\left[\frac{1}{n}\max_{1\leqslant k\leqslant n}\left(\sum_{i=1}^k X'_{i,m}\right)^2\right]=0. \quad (3.5.3)$$

For items 1, we will control the moment of order 4. Hence in both cases, we are forced to find good upper bounds for

$$\mathbb{E}\left[\left(\frac{1}{n}\max_{1\leqslant k\leqslant n}\left(\sum_{i=1}^k X_i\right)^2\right)^p\right], p\in\{1,2\},$$

where $(X_i)_{i\geqslant 1}$ is i.i.d. and centered. Note that for $p=1$,

$$\mathbb{E}\left[\frac{1}{n}\left(\sum_{i=1}^n X_i\right)^2\right]=\mathbb{E}\left[X_1^2\right]$$

and for $p=2$,

$$\mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)^4\right]=n\mathbb{E}\left[X_1^4\right]+3n\left(n-1\right)\left(\mathbb{E}\left[X_1^2\right]\right)^2.$$

This follows by an expansion of $\left(\sum_{i=1}^n X_i\right)^4 = \sum_{i_1,i_2,i_3,i_4=1}^n X_{i_1}X_{i_2}X_{i_3}X_{i_4}$.

Then after having taken the expectation,

$$\mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)^4\right] = \sum_{i_1,i_2,i_3,i_4=1}^n \mathbb{E}\left[X_{i_1}X_{i_2}X_{i_3}X_{i_4}\right].$$

The only cases where $\mathbb{E}\left[X_{i_1}X_{i_2}X_{i_3}X_{i_4}\right]$ is not zero is when all the indices are equal or two of them have a value j_1 and the other two a value $j_2 \neq j_1$.

- In the first case, we have n possibilities ($i_1 = i_2 = i_3 = i_4 = i \in \{1, \dots, n\}$).
- In the second case, we have $\binom{n}{2}\binom{4}{2} = 3n(n-1)$ possibilities.

We admit for the moment the Doob's inequality:

Proposition 3.5. *Let $p > 1$ and let $(X_i)_{i\geqslant 1}$ be an i.i.d. centered sequence such that $\mathbb{E}\left[|X_1|^p\right] < \infty$. Let $S_k := \sum_{i=1}^k X_i$. Then*

$$\mathbb{E}\left[\max_{1\leqslant k\leqslant n} |S_k|^p\right] \leqslant \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[|S_n|^p\right].$$

Let us see how Doob's inequality allows to conclude. Recall that $Y_n^{(m)} := \frac{1}{n}\max_{1\leqslant k\leqslant n}\left(\sum_{i=1}^k X_{i,m}\right)^2$ and $X_{i,m} := X_i\mathbf{1}\{|X_i| \leqslant m\} - \mathbb{E}\left[X_i\mathbf{1}\{|X_i| \leqslant m\}\right],$. We show that for each fixed m, $\sup_{n\geqslant 1}\mathbb{E}\left[\left(Y_n^{(m)}\right)^2\right] < \infty$. Applying Doob's inequality with $p = 4$ gives

$$\begin{aligned}\mathbb{E}\left[\left(Y_n^{(m)}\right)^2\right] &\leqslant \left(\frac{4}{3}\right)^4 \mathbb{E}\left[\frac{1}{n^2}\left(\sum_{i=1}^n X_{i,m}\right)^4\right] \\ &\leqslant \left(\frac{4}{3}\right)^4 \frac{1}{n^2}\left(\left(n\mathbb{E}\left[X_{1,m}\right]^4\right) + 3n(n-1)\left(\mathbb{E}\left[X_{1,m}^2\right]\right)^2\right) \leqslant C_m. \end{aligned} \tag{3.5.4}$$

In order to check that

$$\lim_{m\to\infty}\sup_{n\geqslant 1}\mathbb{E}\left[\frac{1}{n}\max_{1\leqslant k\leqslant n}\left(\sum_{i=1}^k X'_{i,m}\right)^2\right] = 0. \tag{3.5.5}$$

where

$$X'_{i,m} := X_i\mathbf{1}\{|X_i| > m\} - \mathbb{E}\left[X_i\mathbf{1}\{|X_i| > m\}\right],$$

we apply Doob's inequality with $p = 2$ in order to derive that

$$\begin{aligned}\mathbb{E}\left[\frac{1}{n}\max_{1\leqslant k\leqslant n}\left(\sum_{i=1}^k X'_{i,m}\right)^2\right] &\leqslant 4\mathbb{E}\left[\frac{1}{n}\left(\sum_{i=1}^n X'_{i,m}\right)^2\right] \\ &= 4\mathbb{E}\left[\left(X'_{1,m}\right)^2\right] \\ &\leqslant 4\mathbb{E}\left[X_1^2\mathbf{1}\{|X_1| > m\}\right].\end{aligned}$$

The only remaining thing to prove in order to finish the proof of Donsker's theorem is the

Proposition 3.6 (Doob's inequality)**.** *Let $p>1$ and let $(X_i)_{i\geqslant 1}$ be an i.i.d. centered sequence such that $\mathbb{E}\left[|X_1|^p\right]<\infty$. Let $S_k:=\sum_{i=1}^k X_i$. Then*

$$\mathbb{E}\left[\max_{1\leqslant k\leqslant n}|S_k|^p\right]\leqslant\left(\frac{p}{p-1}\right)^p\mathbb{E}\left[|S_n|^p\right].$$

We start by the following:

Lemma 3.7. *With the notations of the Proposition, for each $x>0$, the inequality*

$$x\mathbb{P}\left(\max_{1\leqslant k\leqslant n}|S_k|>x\right)\leqslant\mathbb{E}\left[|S_n|\,\mathbf{1}\left\{\max_{1\leqslant k\leqslant n}|S_k|>x\right\}\right]. \tag{3.5.6}$$

Let $M_j:=\max_{1\leqslant k\leqslant j}|S_k|$. Define the events

$$A_j=\{|S_j|>x\}\cap\{M_{j-1}\leqslant x\}\,,j\geqslant 2,A_1=\{|S_1|>x\}\,.$$

The collection $(A_j)_{j=1}^n$ is pairwise disjoint and $\bigcup_{j=1}^n A_j=\{\max_{1\leqslant k\leqslant n}|S_k|>x\}$. Therefore,

$$x\mathbb{P}\left(\max_{1\leqslant k\leqslant n}|S_k|>x\right)=\sum_{j=1}^n x\mathbb{P}\left(A_j\right).$$

By Markov's inequality, $x\mathbb{P}\left(A_j\right)\leqslant\mathbb{E}\left[|S_j|\,\mathbf{1}\left(A_j\right)\right]$ hence

$$x\mathbb{P}\left(\max_{1\leqslant k\leqslant n}|S_k|>x\right)\leqslant\sum_{j=1}^n\mathbb{E}\left[|S_j|\,\mathbf{1}\left(A_j\right)\right].$$

Suppose that we prove that for each j,

$$\mathbb{E}\left[|S_j|\,\mathbf{1}\left(A_j\right)\right]\leqslant\mathbb{E}\left[|S_n|\,\mathbf{1}\left(A_j\right)\right]. \tag{3.5.7}$$

Then we would get that

$$x\mathbb{P}\left(\max_{1\leqslant k\leqslant n}|S_k|>x\right)\leqslant\sum_{j=1}^n\mathbb{E}\left[|S_n|\,\mathbf{1}\left(A_j\right)\right]$$

and using again the fact that $(A_j)_{j=1}^n$ is pairwise disjoint and $\bigcup_{j=1}^n A_j=\{\max_{1\leqslant k\leqslant n}|S_k|>x\}$ would give the wanted inequality.

Let us prove (3.5.7). Denote

$$B_j:=\left\{(x_1,\dots,x_j)\in\mathbb{R}^j,\left|\sum_{i=1}^j x_i\right|>x,\max_{1\leqslant k\leqslant j-1}\left|\sum_{i=1}^k x_i\right|\leqslant x\right\}.$$

Then, by independence and Fubini's theorem,

$$\mathbb{E}\left[|S_n|\,\mathbf{1}\left(A_j\right)\right]=\int_{\mathbb{R}^j}\int_{\mathbb{R}^{n-j}}\left|\sum_{i=1}^n x_i\right|\mathbf{1}_{B_j}\left(x_1,\dots,x_j\right)d\mathbb{P}_{X_1}\left(x_1\right)\dots d\mathbb{P}_{X_n}\left(x_n\right).$$

By Jensen's inequality,

$$\int_{\mathbb{R}^{n-j}}\left|\sum_{i=1}^n x_i\right|\mathbf{1}_{B_j}\left(x_1,\dots,x_j\right)d\mathbb{P}_{X_{j+1}}\left(x_{j+1}\right)\dots d\mathbb{P}_{X_n}\left(x_n\right)$$
$$\geqslant\left|\int_{\mathbb{R}^{n-j}}\sum_{i=1}^n x_i\mathbf{1}_{B_j}\left(x_1,\dots,x_j\right)d\mathbb{P}_{X_{j+1}}\left(x_{j+1}\right)\dots d\mathbb{P}_{X_n}\left(x_n\right)\right| \tag{3.5.8}$$

For each $i \geqslant j+1$,

$$\int_{\mathbb{R}^{n-j}} x_i \mathbf{1}_{B_j}(x_1,\dots,x_j)\, d\mathbb{P}_{X_{j+1}}(x_{j+1})\dots d\mathbb{P}_{X_n}(x_n)$$
$$= \mathbf{1}_{B_j}(x_1,\dots,x_j)\int_{\mathbb{R}} x_i d\mathbb{P}_{X_i}(x_i) = \mathbf{1}_{B_j}(x_1,\dots,x_j)\,\mathbb{E}[X_i] = 0, \quad (3.5.9)$$

We thus got

$$\mathbb{E}[|S_n|\,\mathbf{1}(A_j)] \geqslant \int_{\mathbb{R}^j}\left|\sum_{i=1}^{j} x_i\right| \mathbf{1}_{B_j}(x_1,\dots,x_j)\, d\mathbb{P}_{X_1}(x_1)\dots d\mathbb{P}_{X_j}(x_j)$$
$$= \mathbb{E}[|S_j|\,\mathbf{1}(A_j)]. \quad (3.5.10)$$

From the lemma,

$$\int_0^\infty px^{p-1}\mathbb{P}\left(\max_{1\leqslant k\leqslant n}|S_k|>x\right)dx \leqslant \int_0^\infty px^{p-2}\mathbb{E}\left[|S_n|\,\mathbf{1}\left\{\max_{1\leqslant k\leqslant n}|S_k|>x\right\}\right]dx \quad (3.5.11)$$

By Fubini-Tonnelli theorem,

$$\mathbb{E}\left[\max_{1\leqslant k\leqslant n}|S_k|^p\right] = p\mathbb{E}\left[|S_n|\int_0^\infty x^{p-2}\mathbf{1}\left\{\max_{1\leqslant k\leqslant n}|S_k|>x\right\}dx\right]$$

Note that

$$p\int_0^\infty x^{p-2}\mathbf{1}\left\{\max_{1\leqslant k\leqslant n}|S_k|>x\right\}dx = \frac{p}{p-1}\max_{1\leqslant k\leqslant n}|S_k|^{p-1}.$$

Letting $M_n := \max_{1\leqslant k\leqslant n}|S_k|$ and $\|X\|_q = \left(\mathbb{E}[|X|^q]\right)^{1/q}$, we thus got

$$\|M_n\|_p^p \leqslant \frac{p}{p-1}\left\|S_n M_n^{p-1}\right\|_1.$$

By Hölder's inequality,

$$\|S_n M_n\|_1 \leqslant \|S_n\|_p \left\|M_n^{p-1}\right\|_{p/(p-1)} = \|S_n\|_p \|M_n\|_p^{p-1},$$

hence

$$\|M_n\|_p^p \leqslant \frac{p}{p-1}\|S_n\|_p \|M_n\|_p^{p-1},$$

from which Doob's inequality directly follows.

This ends the proof of Donsker's theorem.

4. Martingales

4.1. Conditional expectation.

Definition 4.1. *Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. The conditional expectation of an integrable random variable X, denoted by $\mathbb{E}[X \mid \mathcal{G}]$ is the unique (up to almost sure equality) random variable Y which is $\mathcal{G}$-measurable and such that for each $G\in\mathcal{G}$, $\mathbb{E}[X\mathbf{1}(G)] = \mathbb{E}[Y\mathbf{1}(G)]$.*

When $X\in\mathbb{L}^2$, $\mathbb{E}[X \mid \mathcal{G}]$ is the orthogonal projection of X on the subspace $\mathbb{L}^2(\mathcal{G})$ of $\mathbb{L}^2(\mathcal{F})$.

Therefore, $\mathbb{E}[X \mid \mathcal{G}]$ can be seen as the "closest" random variable of X with the constraint of being $\mathcal{G}$-measurable.

Let us recall some propertites of conditional expectation.

Proposition 4.2. *Let X and Y be two random variables such that X, Y and XY are integrable. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. If X is $\mathcal{G}$-measurable, then $\mathbb{E}[XY \mid \mathcal{G}] = X\mathbb{E}[Y \mid \mathcal{G}]$.*

Proposition 4.3. *Let X be an integrable random variable and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Suppose that X is independent of $\mathcal{G}$. Then $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$.*

Proposition 4.4. *Let X be an integrable random variable and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Let $\varphi \colon \mathbb{R} \to \mathbb{R}$ be a convex function. Then the following inequality holds almost surely: $\varphi(\mathbb{E}[X \mid \mathcal{G}]) \leqslant \mathbb{E}[\varphi(X) \mid \mathcal{G}]$.*

4.2. **Definition of martingales.**

Definition 4.5 (Martingales)**.** *A sequence of sub-σ-algebras $(\mathcal{F}_k)_{k\geqslant 0}$ is a filtration if the inclusion $\mathcal{F}_k \subset \mathcal{F}_{k+1}$ holds for all $k \geqslant 0$.*

Definition 4.6. *A sequence of random variables $(S_n)_{n\geqslant 1}$ is a martingale with respect to the filtration $(\mathcal{F}_k)_{k\geqslant 0}$ if for each $n \geqslant 1$, S_n is integrable, $\mathcal{F}_n$-measurable and $\mathbb{E}[S_n \mid \mathcal{F}_{n-1}] = S_{n-1}$.*

Example 4.7. Let $(X_i)_{i\geqslant 1}$ be an independent sequence where $\mathbb{E}[|X_i|] < \infty$, $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_k = \sigma(X_1, \ldots, X_k)$. Let $S_n := \sum_{i=1}^n X_i$. If each X_i is centered, then $(S_n)_{n\geqslant 1}$ is a martingale with respect to the filtration $(\mathcal{F}_k)_{k\geqslant 0}$.

Theorem 4.8 (Law of large numbers for martingales)**.** *Let $(S_n)_{n\geqslant 1}$ be a martingale with respect to the filtration $(\mathcal{F}_k)_{k\geqslant 0}$. Let $X_n = S_n - S_{n-1}$ for $n \geqslant 2$ and $X_1 = S_1$. Let $1 \leqslant p < 2$. Suppose that for each n, X_n has the same distribution as X_1 and that $\mathbb{E}[|X_1|^p]$ is finite. Then*

$$\frac{1}{n^{1/p}} S_n \to 0 \text{ almost surely.}$$