Probability Theory 0

作者：徐知南而北游
时间：2025 年 7 月 7 日

目录

Chapter 1

Probability Space and Random Variable

§ 1.1 Probability space

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ contains three elements:

- The space Ω : this is a non-empty set. It can be viewed as the set of all possible outcomes.
- The σ-field $\mathcal{F}$: this can be viewed as a collection of all the events.
- The probability measure $\mathbb{P}$: this is a function from $\mathcal{F}$ to $[0, 1]$. It gives a probability to each event.

【Definition 1.1】 Suppose $\mathcal{F}$ is a non-empty collection of subsets of Ω.

- It is a field, if it is closed under complementation and closed under union:

$$A \in \mathcal{F} \Longrightarrow A^c \in \mathcal{F}$$
$$A_1, A_2 \in \mathcal{F} \Longrightarrow A_1 \cup A_2 \in \mathcal{F}$$

- It is a monotone class if

$$A_j \in \mathcal{F}, A_j \subset A_{j+1}, 1 \leqslant j < \infty \Longrightarrow \cup_j A_j \in \mathcal{F}.$$
$$A_j \in \mathcal{F}, A_j \supset A_{j+1}, 1 \leqslant j < \infty \Longrightarrow \cap_j A_j \in \mathcal{F}.$$

- It is a σ-field if it is closed under complementation and closed under countable union:

$$A \in \mathcal{F} \Longrightarrow A^c \in \mathcal{F}.$$
$$A_j \in \mathcal{F}, 1 \leqslant j < \infty \Longrightarrow \cup_j A_j \in \mathcal{F}.$$

【Lemma 1.2】 A field is a σ-field $\iff$ it is a monotone class.

Proof.

1. ($\Longrightarrow$)
2. ($\Longleftarrow$)

$\square$

【Definition 1.3】 Given any collection $\mathcal{C}$ of sets, the σ-field (resp. monotone class) generated by $\mathcal{C}$ is the intersection of all σ-fields (resp. monotone classes) containing $\mathcal{C}$.

Chapter 2

Convergence Concepts

§ 2.1 Convergence: almost sure, in probability,and in L^p

【Definition 2.1】(Almost sure convergence (a.s.)) The sequence of random variables $\{X_n\}$ converges a.s. to the random variable X if there exists a null set $\mathcal{N}$ such that

$$\lim_n X_n(\omega) = X(\omega), \quad \forall \omega \in \Omega \backslash \mathcal{N} \tag{2.1}$$

【Definition 2.2】(Convergence in probability) The sequence $\{X_n\}$ converges in probability to the random variable X if, for every $\epsilon > 0$, we have

$$\lim_n \mathbb{P}\left[|X_n - X| > \epsilon\right] = 0 \tag{2.2}$$

【Definition 2.3】(Convergence in L^p) Assume $p \geqslant 1$. The sequence $\{X_n\}$ converges in L^p to the random variable X if $X_n \in L^p, X \in L^p$ and

$$\lim_n \mathbb{E}\left[|X_n - X|^p\right] = 0 \tag{2.3}$$

【Lemma 2.4】 The sequence $\{X_n\}$ converges a.s. to $X \iff$, for any $\epsilon > 0$, we have

$$\lim_{m\to\infty} \mathbb{P}\left[|X_n - X| \leqslant \epsilon, \quad \forall n \geqslant m\right] = 1 \tag{2.4}$$

Proof.

1. (2.1) $\implies$ (2.4)

2. (2.1) $\impliedby$ (2.4)

□

【Proposition 2.5】 Assume $p > 0$. Then $X_n \to X$ in $L^p \implies X_n \to X$ in probability.

Proof. Assume $X_n \to X$ in L^p. For any $\epsilon > 0$, we have

$$\mathbb{P}\left[|X_n - X| > \epsilon\right] \leqslant \epsilon^{-p}\mathbb{E}\left[|X_n - X|^p\right] \to 0 \tag{2.5}$$

□

【Remark 2.6】 (2.5) is also called Markov inequality

§2.2 Borel Cantelli lemma

【Definition 2.7】 Let $\{E_n\}$ be a sequence of subsets in $\mathcal{F}$. Define

$$\limsup_n E_n = \bigcap_{m=1}^{\infty} \bigcup_{n \geqslant m} E_n, \quad \liminf_n E_n = \bigcup_{m=1}^{\infty} \bigcap_{n \geqslant m} \mathbb{E}_n$$

Note that

$$\liminf_n E_n = \left(\limsup_n E_n^c\right)^c$$

【Theorem 2.8】(Borel Cantelli lemma) .

- For arbitrary sequence $\{E_n\}$, we have

$$\sum_n \mathbb{P}\left[E_n\right] < \infty \Longrightarrow \mathbb{P}\left[E_n \text{ i.o. }\right] = 0$$

- If the events $\{E_n\}$ are independent, we have

$$\sum_n \mathbb{P}\left[E_n\right] = \infty \Longrightarrow \mathbb{P}\left[E_n \text{ i.o. }\right] = 1$$

Proof.

□

Chapter 3

Law of Large Numbers

§ 3.1 simple theorems

【Theorem 3.1】 If the X_j 's are uncorrelated and their second moments have a common bound, then

$$\frac{S_n - \mathbb{E}(S_n)}{n} \to 0 \quad \text{in} L^2 \tag{3.1}$$

【Remark 3.2】 hence also in pr.

Proof. By Chebyshev inequality □

【Theorem 3.3】 If the X_j 's are uncorrelated and their second moments have a common bound, then (3.1) holds almost surely

Proof. Without loss of generality we may suppose that $\mathscr{E}(X_j) = 0$ for each j, so that the X_j 's are orthogonal. then

$$\sigma^2(S_n) = \sum_{j=1}^{n} \sigma^2(X_j) \implies \mathbb{E}\left(S_n^2\right) \le Mn,$$

where M is a bound for the second moments. It follows by Chebyshev's inequality that for each $\epsilon > 0$ we have

$$\sum_n \mathbb{P}\left\{|S_{n^2}| > n^2\epsilon\right\} = \sum_n \frac{M}{n^2\epsilon^2} < \infty$$

Hence by Borel-Cantelli lemma we have

$$\mathbb{P}\left\{|S_{n^2}| > n^2\varepsilon \text{ i.o. }\right\} = 0 \quad \forall\varepsilon \iff \frac{S_{n^2}}{n^2} \to 0 \quad \text{ a.e.}$$

for each $n \ge 1$:

$$D_n = \max_{n^2 \le k < (n+1)^2} |S_k - S_{n^2}|$$

Then we have

$$\mathbb{E}\left(D_n^2\right) \le 2n\mathbb{E}\left(\left|S_{(n+1)^2} - S_{n^2}\right|^2\right) = 2n \sum_{j=n^2+1}^{(n+1)^2} \sigma^2(X_j) \le 4n^2 M$$

By Chebyshev's inequality

$$\mathbb{P}\left\{D_n > n^2\epsilon\right\} \le \frac{4M}{\epsilon^2 n^2}$$

It follows as before that

$$\frac{D_n}{n^2} \to 0 \quad \text{ a.e.}$$

Then we have

$$\frac{|S_k|}{k} \le \frac{|S_{n^2}| + D_n}{n^2}$$

which implies (3.1) □

§ 3.2 Weak law of large numbers

【Definition 3.4】 Two sequences of random variables $\{X_n\}$ and $\{Y_n\}$ are equivalent if

$$\sum_n \mathbb{P}[X_n \neq Y_n] < \infty$$

【Theorem 3.5】 If $\{X_n\}$ and $\{Y_n\}$ are equivalent, then

$$\sum_n (X_n - Y_n) \quad \text{converges almost surely}$$

Furthermore if $a_n \uparrow \infty$, then

$$\frac{1}{a_n}\sum_{j=1}^{n}(X_j - Y_j) \to 0 \quad \text{almost surely}$$

Proof. By the Borel-Cantelli lemma, X_n, Y_n are equivalent implies that

$$\mathbb{P}\{X_n \neq Y_n \text{ i.o. }\} = 0$$

This means that there exists a null set N with the following property: if $\omega \in \Omega \backslash \mathrm{N}$, then there exists $n_0(\omega)$ such that

$$n \geq n_0(\omega) \implies X_n(\omega) = Y_n(\omega)$$

Thus for such an ω, the two numerical sequences $\{X_n(\omega)\}$ and $\{Y_n(\omega)\}$ differ only in a finite number of terms (how many depending on ω). In other words, the series

$$\sum_n (X_n(\omega) - Y_n(\omega))$$

consists of zeros from a certain point on. Both assertions of the theorem are trivial consequences of this fact. □

【Theorem 3.6】 Let $\{X_n\}$ be i.i.d. with finite mean m. Define $S_n = \sum_{j=1}^{n} X_j$, then we have

$$\frac{S_n}{n} \to m, \quad \text{in probability.}$$

Proof. Denote by μ the common law of X_n 's, and suppose $Z \sim \mu$. Since $Z \in L^1$, we have

$$\sum_n \mathbb{P}[|Z| \geqslant n] < \infty$$

We introduce random variables Y_n 's by truncating X_n 's:

$$Y_n = X_n \mathbb{1}_{\{|X_n| \leqslant n\}}$$

Then

$$\sum_n \mathbb{P}[X_n \neq Y_n] = \sum_n \mathbb{P}[|X_n| > n] = \sum_n \mathbb{P}[|Z| \geqslant n] < \infty$$

Hence $\{Y_n\}$ and $\{X_n\}$ are equivalent. Define $T_n = \sum_{j=1}^{n} Y_j$. If we prove

$$\frac{T_n - \mathbb{E}[T_n]}{n} \to 0 \quad \text{in probability}$$

then the conclusion follows, because $\mathbb{E}[T_n]/n \to m$ as $n \to \infty$. For any $\epsilon > 0$,

$$\mathbb{P}[|T_n - \mathbb{E}[T_n]| \geqslant n\epsilon] \leqslant \frac{\operatorname{var}(T_n)}{n^2\epsilon^2}$$

It suffices to show $\operatorname{var}(T_n) = o(n^2)$. Let us calculate $\operatorname{var}(T_n)$.

$$\operatorname{var}(T_n) = \sum_{j=1}^{n} \operatorname{var}(Y_j) \leqslant \sum_{j=1}^{n} \mathbb{E}\left[Y_j^2\right] = \sum_{j=1}^{n} \mathbb{E}\left[Z^2 \mathbb{1}_{\{|Z| \leqslant j\}}\right]$$

The most naive estimate is the following:

$$\operatorname{var}(T_n) \leqslant \sum_{j=1}^{n} \mathbb{E}\left[Z^2 \mathbb{1}_{\{|Z| \leqslant j\}}\right] \leqslant \sum_{j=1}^{n} j^2 = O\left(n^3\right)$$

The less naive estimate is the following:

$$\operatorname{var}(T_n) \leqslant \sum_{j=1}^{n} \mathbb{E}\left[Z^2 \mathbb{1}_{\{|Z| \leqslant j\}}\right] \leqslant \sum_{j=1}^{n} j \mathbb{E}\left[|Z| \mathbb{1}_{\{|Z| \leqslant j\}}\right] \leqslant \sum_{j=1}^{n} j \mathbb{E}[|Z|] = O\left(n^2\right)$$

But we desire a control of $o\left(n^2\right)$. To improve it, let $\{a_n\}$ be a sequence of integers such that $1 \leqslant a_n \leqslant n$, $a_n \to \infty$, but $a_n = o(n)$. Then we have

$$\begin{aligned}
\operatorname{var}(T_n) &\leqslant \sum_{j=1}^{n} \mathbb{E}\left[Z^2 \mathbb{1}_{\{|Z| \leqslant j\}}\right] = \sum_{j \leqslant a_n} + \sum_{a_n < j \leqslant n} \\
&= \sum_{j \leqslant a_n} \mathbb{E}\left[Z^2 \mathbb{1}_{\{|Z| \leqslant j\}}\right] + \sum_{a_n < j \leqslant n} \mathbb{E}\left[Z^2 \mathbb{1}_{\{|Z| \leqslant a_n\}}\right] + \sum_{a_n < j \leqslant n} \mathbb{E}\left[Z^2 \mathbb{1}_{\{a_n < |Z| \leqslant j\}}\right] \\
&\leqslant \sum_{j \leqslant a_n} a_n \mathbb{E}[|Z|] + \sum_{a_n < j \leqslant n} a_n \mathbb{E}[|Z|] + \sum_{a_n < j \leqslant n} n \mathbb{E}\left[|Z| \mathbb{1}_{\{a_n < |Z| \leqslant j\}}\right] \\
&\leqslant O(1) n a_n + O(1) n^2 \mathbb{E}\left[|Z| \mathbb{1}_{\{|Z| > a_n\}}\right]
\end{aligned}$$

The first term is $n a_n = o\left(n^2\right)$ because $a_n = o(n)$; the second term is also $o\left(n^2\right)$ because $\mathbb{E}\left[|Z| \mathbb{1}_{\{|Z| > a_n\}}\right] \to 0$ since $a_n \to \infty$. Therefore, we have $\operatorname{var}(T_n) = o\left(n^2\right)$ as desired. □

【Example 3.7】 Let $\{X_n\}$ be i.i.d. with the common law given by

$$\mathbb{P}[Z = n] = \mathbb{P}[Z = -n] = \frac{c}{n^2 \log n}, \quad n = 2, 3, \ldots,$$

where c is a normalizing constant. Define $S_n = \sum_{j=1}^{n} X_j$. It is clear that $\mathbb{E}[|Z|] = \infty$, but we have

$$\frac{S_n}{n} \to 0, \quad \text{in probability}$$

Proof.

□

§ 3.3 Three series theorem

【Lemma 3.8】 Let $\{X_n\}$ be independent random variables such that $\mathbb{E}[X_n] = 0$ and $\mathbb{E}\left[X_n^2\right] < \infty$. Define $S_n = \sum_{j=1}^{n} X_j$. Then we have

$$\mathbb{P}\left[\max_{1 \leqslant j \leqslant n} |S_j| \geqslant \epsilon\right] \leqslant \frac{\mathbb{E}\left[S_n^2\right]}{\epsilon^2}$$

Proof. Fix $\epsilon > 0$ and define

$$\Lambda = \left\{\max_{1\leqslant j\leqslant n} |S_j| \geqslant \epsilon\right\}$$

Define $T = \min\{j : |S_j| \geqslant \epsilon\}$ to be the first time that $|S_j|$ exceeds ϵ, and define $\Lambda_k = \{T = k\}$:

$$\Lambda_k = \left\{\max_{1\leqslant j\leqslant k-1} |S_j| < \epsilon, |S_k| \geqslant \epsilon\right\}$$

Note that Λ_k 's are disjoint and $\Lambda = \sqcup_{k=1}^n \Lambda_k$. We have

$$\mathbb{E}\left[S_n^2 \mathbb{1}_\Lambda\right] = \sum_{k=1}^n \mathbb{E}\left[S_n^2 \mathbb{1}_{\Lambda_k}\right] = \sum_{k=1}^{\mathbb{E}} \left[S_k^2 \mathbb{1}_{\Lambda_k} + 2S_k\left(S_n - S_k\right)\mathbb{1}_{\Lambda_k} + \left(S_n - S_k\right)^2 \mathbb{1}_{\Lambda_k}\right]$$

Note that $S_k \mathbb{1}_{\Lambda_k}$ and $S_n - S_k$ are independent, thus

$$\mathbb{E}\left[S_k\left(S_n - S_k\right)\mathbb{1}_{\Lambda_k}\right] = \mathbb{E}\left[S_k \mathbb{1}_{\Lambda_k}\right]\mathbb{E}\left[\left(S_n - S_k\right)\right] = 0$$

Therefore,

$$\mathbb{E}\left[S_n^2 \mathbb{1}_\Lambda\right] = \sum_{k=1}^n \mathbb{E}\left[S_k^2 \mathbb{1}_{\Lambda_k} + \left(S_n - S_k\right)^2 \mathbb{1}_{\Lambda_k}\right] \geqslant \sum_{k=1}^n \mathbb{E}\left[S_k^2 \mathbb{1}_{\Lambda_k}\right] \geqslant \sum_{k=1}^n \epsilon^2 \mathbb{P}\left[\Lambda_k\right] = \epsilon^2 \mathbb{P}[\Lambda].$$

Thus we have $\mathbb{P}[\Lambda] \leqslant \mathbb{E}\left[S_n^2\right]/\epsilon^2$, as desired □

【Lemma 3.9】 Let $\{X_n\}$ be independent random variables which are bounded: there exists a constant A such that $|X_n| \leqslant A$ almost surely for all n. Define $S_n = \sum_{j=1}^n X_j$. Then we have

$$\mathbb{P}\left[\max_{1\leqslant j\leqslant n} |S_j| \leqslant B\right] \leqslant \frac{(2B + A)^2}{\operatorname{var}\left(S_n\right)}.$$

Proof. Define $T = \min\{j : |S_j| > B\}$ to be the first time that $|S_j|$ exceeds B. Then we have

$$\{T > k\} = \left\{\max_{1\leqslant j\leqslant k} |S_j| \leqslant B\right\}, \quad \{T = k\} = \left\{\max_{1\leqslant j\leqslant k-1} |S_j| \leqslant B, |S_k| > B\right\}$$

We need to give a upper bound for $\mathbb{P}[T > n]\operatorname{var}\left(S_n\right)$. Let us consider the expectation and the variance of S_k on $\{T > k\}$:

$$a_k := \mathbb{E}\left[S_k \mathbb{1}_{\{T>k\}}\right]/\mathbb{P}[T > k], \quad \mathbb{E}\left[\left(S_k - a_k\right)^2 \mathbb{1}_{\{T>k\}}\right].$$

It is clear that $|a_k| \leqslant B$. We write

$$\mathbb{E}\left[\left(S_{k+1} - a_{k+1}\right)^2 \mathbb{1}_{\{T>k+1\}}\right] = \mathbb{E}\left[\left(S_{k+1} - a_{k+1}\right)^2 \mathbb{1}_{\{T>k\}}\right] - \mathbb{E}\left[\left(S_{k+1} - a_{k+1}\right)^2 \mathbb{1}_{\{T=k+1\}}\right].$$

For the first term,

$$\begin{aligned}
&\mathbb{E}\left[\left(S_{k+1} - a_{k+1}\right)^2 \mathbb{1}_{\{T>k\}}\right] \\
&= \mathbb{E}\left[\left(S_k - a_k + X_{k+1} - a_{k+1} + a_k\right)^2 \mathbb{1}_{\{T>k\}}\right] \\
&= \mathbb{E}\left[\left(S_k - a_k\right)^2 \mathbb{1}_{\{T>k\}}\right] + \mathbb{E}\left[\left(X_{k+1} - a_{k+1} + a_k\right)^2 \mathbb{1}_{\{T>k\}}\right] \\
&= \mathbb{E}\left[\left(S_k - a_k\right)^2 \mathbb{1}_{\{T>k\}}\right] + \mathbb{E}\left[\left(X_{k+1} - a_{k+1} + a_k\right)^2\right]\mathbb{P}[T > k] \\
&\geqslant \mathbb{E}\left[\left(S_k - a_k\right)^2 \mathbb{1}_{\{T>k\}}\right] + \operatorname{var}\left(X_{k+1}\right)\mathbb{P}[T > k].
\end{aligned}$$

The last $\geqslant$ uses the fact $\mathbb{E}\left[(X - c)^2\right] \geqslant \operatorname{var}(X)$ for all $c \in \mathbb{R}$.
For the second term,

$$\mathbb{E}\left[\left(S_{k+1} - a_{k+1}\right)^2 \mathbb{1}_{\{T=k+1\}}\right] = \mathbb{E}\left[\left(S_k + X_{k+1} - a_{k+1}\right)^2 \mathbb{1}_{\{T=k+1\}}\right].$$

Note that, $|S_k| \leqslant B$ on $\{T = k+1\}$, and $|X_{k+1}| \leqslant A$, and $|a_{k+1}| \leqslant B$. Thus, for the second term,

$$\mathbb{E}\left[(S_{k+1} - a_{k+1})^2 \mathbb{1}_{\{T=k+1\}}\right] \leqslant (2B+A)^2 \mathbb{P}[T = k+1].$$

Combining the two estimates, we have

$$\mathbb{E}\left[(S_{k+1} - a_{k+1})^2 \mathbb{1}_{\{T>k+1\}}\right] \geqslant \mathbb{E}\left[(S_k - a_k)^2 \mathbb{1}_{\{T>k\}}\right] + \operatorname{var}(X_{k+1})\, \mathbb{P}[T > k] - (2B+A)^2 \mathbb{P}[T = k+1].$$

Summing over k, we have

$$\begin{aligned}\mathbb{E}\Big[(S_n - a_n)^2 \mathbb{1}_{\{T>n\}}\Big] &\geqslant \mathbb{E}\left[(X_1 - a_1)^2 \mathbb{1}_{\{T>1\}}\right] + \sum_{k=1}^{n-1} \operatorname{var}(X_{k+1})\, \mathbb{P}[T > k] - (2B+A)^2 \mathbb{P}[2 \leqslant T \leqslant n] \\ &\geqslant \mathbb{E}\left[(X_1 - a_1)^2 \mathbb{1}_{\{T>1\}}\right] + (\operatorname{var}(S_n) - \operatorname{var}(X_1))\, \mathbb{P}[T > n] - (2B+A)^2 \mathbb{P}[2 \leqslant T \leqslant n]\end{aligned}$$

Thus

$$\operatorname{var}(S_n)\, \mathbb{P}[T > n] \leqslant \mathbb{E}\left[(S_n - a_n)^2 \mathbb{1}_{\{T>n\}}\right] + \operatorname{var}(X_1)\, \mathbb{P}[T > n] + (2B+A)^2 \mathbb{P}[2 \leqslant T \leqslant n].$$

Note that

$$\mathbb{E}\left[(S_n - a_n)^2 \mathbb{1}_{\{T>n\}}\right] = \mathbb{E}\left[S_n^2 \mathbb{1}_{\{T>n\}}\right] - a_n^2 \mathbb{P}[T > n] \leqslant B^2 \mathbb{P}[T > n].$$

Thus

$$\operatorname{var}(S_n)\, \mathbb{P}[T > n] \leqslant B^2 \mathbb{P}[T > n] + A^2 \mathbb{P}[T > n] + (2B+A)^2 \mathbb{P}[2 \leqslant T \leqslant n] \leqslant (2B+A)^2$$

□

【**Theorem 3.10**】 Let $\{X_n\}$ be independent random variables and define the truncation for a fixed constant $A > 0$:

$$Y_n = X_n \mathbb{1}_{\{|X_n| \leqslant A\}}$$

Then the series $\sum_n X_n$ converges almost surely $\iff$ the following three series all converge:

$$\sum_n \mathbb{P}\left[|X_n| > A\right], \quad \sum_n \mathbb{E}\left[Y_n\right], \quad \sum_n \operatorname{var}(Y_n).$$

Proof. Suppose the three series all converge.

$$\text{the first series converges} \implies \sum_n \mathbb{P}\left[X_n \neq Y_n\right] < \infty \iff \{X_n\}, \{Y_n\} \quad \text{are equivalent}$$

it suffices to show that $\sum_n Y_n$ converges almost surely.

$$\text{the second series converges} \implies \left(\sum_n Y_n \quad \text{converges a.s.} \iff \sum_n (Y_n - \mathbb{E}[Y_n]) \quad \text{converges a.s.} \right)$$

Let us consider the tail of this series

$$T(n,m) := \sum_{j=n}^{m} (Y_j - \mathbb{E}[Y_j])$$

We need to show that, almost surely, the oscillation

$$W_n := \max_{\ell \geqslant k \geqslant n} |T(k,\ell)|$$

is small when n is large. Fix $\epsilon > 0$, by Lemma 3.8, we have

$$\mathbb{P}\left[\max_{n \leqslant j \leqslant m} |T(n,j)| \geqslant \epsilon/2\right] \leqslant 4\epsilon^{-2} \sum_{j=n}^{m} \operatorname{var}(Y_j)$$

Let $m \to \infty$, we have

$$\mathbb{P}\left[\max_{j \geqslant n} |T(n,j)| \geqslant \epsilon/2\right] \leqslant 4\epsilon^{-2} \sum_{j \geqslant n} \text{var}(Y_j)$$

For $\ell \geqslant k \geqslant n$, we have

$$T(k,\ell) = T(n,\ell) - T(n,k)$$

Thus

$$\mathbb{P}[W_n \geqslant \epsilon] = \mathbb{P}\left[\max_{\ell \geqslant k \geqslant n} |T(k,\ell)| \geqslant \epsilon\right] \leqslant \mathbb{P}\left[\max_{j \geqslant n} |T(n,j)| \geqslant \epsilon/2\right] \leqslant 4\epsilon^{-2} \sum_{j \geqslant n} \text{var}(Y_j)$$

Since the third series converges, we have

$$\lim_n \mathbb{P}[W_n \geqslant \epsilon] = 0.$$

Since the sequence of events $\{W_n \geqslant \epsilon\}$ is decreasing in n, we have

$$\mathbb{P}\left[\lim_n W_n \geqslant \epsilon\right] = 0$$

Let $\epsilon \to 0$, we have

$$\mathbb{P}\left[\lim_n W_n = 0\right] = 1$$

This implies the almost sure convergence.

Suppose $\sum_n X_n$ converges almost surely, then we have

$$\mathbb{P}[|X_n| > A \text{ i.o. }] = 0$$

Then Borel Cantelli lemma guarantees the convergence of the first series. As a consequence, the sequences $\{Y_n\}$ and $\{X_n\}$ are equivalent, hence $\sum_n Y_n$ converges almost surely as well. By Lemma 3.9, we have

$$\mathbb{P}\left[\max_{n \leqslant k \leqslant m} \left|\sum_{j=n}^{k} Y_j\right| \leqslant 1\right] \leqslant \frac{(A+2)^2}{\sum_{j=n}^{m} \text{var}(Y_j)}$$

If the third series diverges, then the right hand-side will go to zero as $m \to \infty$. Hence the tail of $\sum_n Y_n$ almost surely would not be bounded by one, so the series could not converge. This confirms the convergence of the third series. By the proof of direction of $\Longleftarrow$, the convergence of the third series implies the convergence of $\sum_n (Y_n - \mathbb{E}[Y_n])$. Combining with the convergence of $\sum_n Y_n$, we have the convergence of the second series. □

§3.4 Strong law of large numbers

【Lemma 3.11】(Kronecker's lemma) Let $\{x_n\}$ be a sequence of real numbers, $\{a_n\}$ be a sequence of numbers such that $0 < a_n \uparrow \infty$. Then

$$\sum_n \frac{x_n}{a_n} \text{ converges} \Longrightarrow \frac{1}{a_n} \sum_{j=1}^{n} x_j \to 0$$

【Theorem 3.12】 Let $\{X_n\}$ be i.i.d. Define $S_n = \sum_{j=1}^{n} X_j$. Then we have

$$\mathbb{E}[|X_1|] < \infty \Longrightarrow \frac{S_n}{n} \to \mathbb{E}[X_1], \text{ almost surely;}$$

$$\mathbb{E}[|X_1|] = \infty \Longrightarrow \limsup_n \frac{|S_n|}{n} = \infty, \text{ almost surely.}$$