概率论第 8-10 次习题课

宋语轩

2022 秋, USTC

1 随机变量列的收敛与极限定理

注. 本章节中任何引理, 定理, 命题 (不包括推论和例题等) 均可以在考试中直接使用, 除非题目已做说明及题目特别要求证明他们.

1.1 基本工具与技术

命题 1.1 (矩不等式). 我们有

(1) **Hölder 不等式**: $p, q > 1, \frac{1}{p} + \frac{1}{q} = 1$,

$$\mathbb{E}[|XY|] \leq (\mathbb{E}[|X|^p])^{\frac{1}{p}} (\mathbb{E}[|Y|^q])^{\frac{1}{q}}$$

(2) **Minkowski 不等式**:

$$(\mathbb{E}[|X+Y|^p])^{\frac{1}{p}} \leq (\mathbb{E}[|X|^p])^{\frac{1}{p}} + (\mathbb{E}[|Y|^p])^{\frac{1}{p}}$$

(3) **Markov 不等式**:

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[|X|]}{a}, a > 0$$

(4) **Chebyshev 不等式**:

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{1}{a^2} Var(X)$$

(5) 设 $s > 0$, 若 $\mathbb{E}[|X|^s] < +\infty$, 则对 $r \in [0, s]$, 均有 $\mathbb{E}[|X|^r] < +\infty$. 因此 $X_n \xrightarrow{s} X \implies X_n \xrightarrow{r} X$.

(6) **C_r 不等式**:

$$\mathbb{E}|X_1 + \cdots + X_n|^r \leqslant C_{n,r}(\mathbb{E}[|X_1|^r] + \cdots \mathbb{E}[|X_n|^r]).$$

其中

$$C_{n,r} = \begin{cases} 1, & 0 < r < 1, \\ n^{r-1}, & r \geqslant 1. \end{cases}$$

(7) **Lyapunov 不等式**: 对 $\forall 0 < r < s$, 有

$$(\mathbb{E}[|X|^r])^{\frac{1}{r}} \leqslant (\mathbb{E}[|X|^s])^{\frac{1}{s}}.$$

[0]个人主页: http://home.ustc.edu.cn/~zyx240014/.

发现错误欢迎联系: zyx240014@mail.ustc.edu.cn.

证明. (5) 注意到

$$|X|^r = |X|^r I_{\{|X|\leqslant 1\}} + |X|^r I_{\{|X|>1\}} \leqslant 1 + |X|^s,$$

两边取期望即得.
(6) 当 $0<r\leqslant 1$ 时, 利用 x^r 的凹性, 有

$$|X_1+\cdots+X_n|^r \leqslant |X_1|^r+\cdots+|X_n|^r.$$

当 $r\geqslant 1$ 时, 利用 x^r 的凸性, 有

$$\left(\frac{|X_1|+\cdots+|X_n|}{n}\right)^r \leqslant \frac{|X_1|^r+\cdots+|X_n|^r}{n}$$

再利用 $|X_1+\cdots+X_n| \leqslant |X_1|+\cdots+|X_n|$ 即可. 综上, 有 $|X_1+\cdots+X_n|^r \leqslant C_{n,r}(|X_1|^r+\cdots+|X_n|^r)$.
最后对两边同时取期望即得.
(7) 利用 Hölder 不等式, 并取 $Y=1, p=\dfrac{s}{r}$, 用 $|X|^r$ 代替 $|X|$ 即可. □

注. Marvkov 不等式 (及其推广) 是在尾概率估计中一个重要的不等式, 为随机变量偏离某些值的概率给出了上界. 更一般地, 设 g 是 $\mathbb{R}$ 上的非负 Borel 可测函数, 我们有

(1) 若 g 为偶函数, 且在 $\mathbb{R}^+$ 上单调递增 (指不严格, 下同), 则对 $\forall a\geqslant 0$, 有

$$\mathbb{P}(|X|\geqslant a) = \mathbb{P}(g(|X|)\geqslant g(a)) \leqslant \frac{\mathbb{E}[g(|X|)]}{g(a)}.$$

(2) 若 g 在 $\mathbb{R}$ 上单调递增, 则对 $\forall a\geqslant 0$, 有

$$\mathbb{P}(X\geqslant a) = \mathbb{P}(g(X)\geqslant g(a)) \leqslant \frac{\mathbb{E}[g(X)]}{g(a)}.$$

下面两个简单的特例非常值得关注, 在尾概率估计中非常常见. 对 $a>0$, 有

$$\mathbb{P}(|X|\geqslant a) \leqslant \frac{\mathbb{E}[|X|^r]}{a^r}, r>0, \qquad \mathbb{P}(X>a)\leqslant \mathrm{e}^{-ta}\mathbb{E}[\mathrm{e}^{tX}].$$

对于前者, 越高阶矩得到有关矩的信息越多, 也在证明随机变量的收敛时放缩更易施展手脚, 减少限制. 后者估计便于利用矩母函数的性质, 由此导出一些集中不等式 (超出本课程范围).

例 1.1 (**单边 Chebyshev 不等式**)**.** 设 $Var(X)=\sigma^2\in(0,+\infty)$, 则对 $\forall\lambda>0$, 有

$$\mathbb{P}(X-\mathbb{E}[X]\geqslant\lambda) \leqslant \frac{\sigma^2}{\sigma^2+\lambda^2}.$$

证明. 记 $Y=X-\mathbb{E}[X]$, 设 $u>0$, 则有

$$\mathbb{P}(X-\mathbb{E}[X]\geqslant\lambda) = \mathbb{P}(Y+u\geqslant\lambda+u) \leqslant \mathbb{P}((Y+u)^2\geqslant(\lambda+u)^2) \leqslant \frac{\mathbb{E}[(Y+u)^2]}{(\lambda+u)^2} = \frac{\sigma^2+u^2}{(\lambda+u)^2} \xlongequal{u=\frac{\sigma^2}{\lambda}} \frac{\sigma^2}{\sigma^2+\lambda^2}.$$

其中最后一个不等式是一个简单的函数最值问题, 最大值当 $u=\dfrac{\sigma^2}{\lambda}$ 时取等. □

注. 对称地, 有

$$\mathbb{P}(X-\mathbb{E}[X]\leqslant-\lambda) \leqslant \frac{\sigma^2}{\sigma^2+\lambda^2}.$$

例 1.2. 设 $X_1,\cdots X_n$ 独立同分布, $\mathbb{P}(X_1=1)=\mathbb{P}(X_1=-1)=\frac{1}{2}$, 记 $S_n=\sum_{k=1}^{n}X_i$, 证明: 对 $\forall a>0$, 均有

$$\mathbb{P}(S_n>a)\leqslant \mathrm{e}^{-\frac{a^2}{2n}}.$$

证明. 我们有

$$\mathbb{P}(S_n>a)\leqslant \mathrm{e}^{-ta}\mathbb{E}[\mathrm{e}^{tS_n}]=\mathrm{e}^{-ta}(\mathbb{E}[\mathrm{e}^{tX_1}])^n=\mathrm{e}^{-ta}\left(\frac{\mathrm{e}^t+\mathrm{e}^{-t}}{2}\right)^n\leqslant \mathrm{e}^{-ta+\frac{n}{2}t^2}=\mathrm{e}^{\frac{n}{2}(t-\frac{a}{n})^2-\frac{a^2}{2n}}\xlongequal{t=\frac{a}{n}}\mathrm{e}^{-\frac{a^2}{2n}}.$$

其中

$$\frac{\mathrm{e}^t+\mathrm{e}^{-t}}{2}=\frac{1}{2}\left(\sum_{k=0}^{+\infty}\frac{t^k}{k!}+\sum_{k=0}^{+\infty}(-1)^k\frac{t^k}{k!}\right)=\sum_{k=0}^{+\infty}\frac{t^{2k}}{(2k)!}\leqslant\sum_{k=0}^{+\infty}\frac{t^{2k}}{2^k k!}=\mathrm{e}^{\frac{t^2}{2}}.$$

□

注. $\frac{\mathrm{e}^t+\mathrm{e}^{-t}}{2}\leqslant \mathrm{e}^{\frac{t^2}{2}}$ 是常用的不等式放缩, 值得一记. 类似地还有 $\frac{\mathrm{e}^t-\mathrm{e}^{-t}}{2t}\leqslant \mathrm{e}^{\frac{t^2}{2}}$.

下面是概率测度下证明随机变量列收敛及一些放缩中常用的拆项小技巧, 请读者自行验证:

命题 1.2. 我们有

$$\mathbb{P}(|X+Y|\leqslant a+b)\geqslant \mathbb{P}(|X|\leqslant a,|Y|\leqslant b),\quad \mathbb{P}(|X+Y|\geqslant a+b)\leqslant \mathbb{P}(|X|\geqslant a)+\mathbb{P}(|Y|\geqslant b).$$

注. 独立复制的副产品: 设 X' 是 X 的独立复制 (X,X' i.i.d), 对 $\forall x,a$, 均有

$$\mathbb{P}(|X-X'|>x)=\mathbb{P}(|(X-a)-(X'-a)|>x)\leqslant \mathbb{P}(|X-a|\geqslant\frac{x}{2})+\mathbb{P}(|X'-a|\geqslant\frac{x}{2})=2\mathbb{P}(|X-a|\geqslant\frac{x}{2}).$$

回顾一下 Borel-Cantelli 引理:

定理 1.1 (**Borel-Cantelli 引理**)**.** 记 $\{A_n,\ \text{i.o.}\}=\limsup_{n\to\infty}A_n=\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty}A_m$(infinitely often)

(1) 当 $\sum_{n=1}^{\infty}\mathbb{P}(A_n)<\infty$ 时有, $\mathbb{P}(A_n,\ \text{i.o.})=0$

(2) 假设 $\{A_n\}$ 相互独立, 当 $\sum_{n=1}^{\infty}\mathbb{P}(A_n)=\infty$ 时有 $\mathbb{P}(A_n,\ \text{i.o.})=1$

随机变量的截尾

对于随机变量 X, 主要有以下三类结尾方法:

$$X_1=XI_{\{|X|\leqslant M\}}=\begin{cases}X, & |X|\leqslant M,\\ 0, & |X|>M.\end{cases},\quad X_2=\begin{cases}X, & |X|\leqslant M,\\ M, & X>M,\\ -M, & X<-M.\end{cases},\quad X_3=\begin{cases}X, & X\leqslant M,\\ M, & X>M.\end{cases}.$$

注. 截尾法常常用于证明有关随机变量列的收敛中, 运用了转化的思想. 对于随机变量列 $\{X_n\}$, 我们常常令 $M\uparrow+\infty$ 或者取 $M=n$ (更一般取 $M=k_n$, $\{k_n\}$ 递增) 再令 $n\to+\infty$ 来处理. 这样结尾的好处是截尾后的随机变量具有有界性, 同时利用了随机变量列的单调递增性, 其中 X_2,X_3 两个随机变量本身就是单调递增 (X_2 中大于 0 递增小于 0 递减) 的函数, 递增性更好. 一方面, 有界性的保证大大可以施展手脚 (比如很多定理、命题的适用条件都有有界性的要求, 以及随机变量的有界性能保证其任意阶矩的存在性, 从而便于矩不等式的施展); 另一方面, 递增性或是关于 $X-X_i$ 的性质保证了目标的转化条件 (之前做过的一道作业题: $F_n(x)\to F(x)$ 逐点收敛).

我们下举一例有关截尾的例子, 这也是强大数律的证明中一个重要的转化步骤. 后续我们还会看到一些用到截尾的例子, 可见其精妙之处.

例 1.3. 设非负随机变量 X_i 独立同分布且 $\mathbb{E}[X_1] < \infty$, 令 $Y_n = X_n I_{\{|X_n| \leqslant n\}}$, 则有

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - Y_k) \xrightarrow{\text{a.s.}} 0.$$

从而

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\text{a.s.}} a \Longleftrightarrow \frac{1}{n}\sum_{k=1}^{n} Y_k \xrightarrow{\text{a.s.}} a$$

证明. 对 $X \geqslant 0$, 利用 (见**第 5 次习题课讲义引理 2.1**)

$$\sum_{m=1}^{\infty}\mathbb{P}(X \geq m) \leq \mathbb{E}[X] \leq 1 + \sum_{m=1}^{\infty}\mathbb{P}(X \geq m),$$

有

$$\sum_{n=1}^{+\infty}\mathbb{P}(Y_n \neq X_n) = \sum_{n=1}^{+\infty}\mathbb{P}(X_n \geq n) = \sum_{n=1}^{+\infty}\mathbb{P}(X_1 \geq n) \leq \mathbb{E}[X_1] < \infty$$

利用 Borel-Cantelli 引理, $\mathbb{P}(X_n \neq Y_n \text{ i.o.}) = 0$, 即几乎处处 $\{X_n \neq Y_n\}$ 只发生有限次, 从而

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i) \xrightarrow{\text{a.s.}} 0.$$

□

注. 通过截尾, 证明原随机变量列的收敛转化成证明截尾后的随机变量列的收敛.

补充：独立复制与对称化方法

这个方法在之前证明特征函数的一道例题中 (参见**第 6 次习题课讲义例 2.2**) 也出现过, 在概率中是非常重要的技术, 考虑到内容上限就不作过多补充, 之后的一些概率类课程还会专门用到这个方法, 现在只以两个例子给大家自行体会.

1 设 $f, g : \mathbb{R} \to \mathbb{R}$ 且单调递增, 证明, 对任意随机变量 X, 均有

$$\mathbb{E}[f(X)g(X)] \geqslant \mathbb{E}[f(X)]\mathbb{E}[g(X)].$$

证明. 设 X' 是 X 的独立复制, 则有

$$(f(X) - f(X'))(g(X) - g(X')) \geqslant 0.$$

因此

$$0 \leqslant \mathbb{E}[(f(X) - f(X'))(g(X) - g(X'))] = 2\mathbb{E}[f(X)g(X)] - 2\mathbb{E}[f(X)]\mathbb{E}[g(X)].$$

□

回忆一下, 我们称随机变量 X 是对称的, 若 $X \overset{d}{=} -X$.

2 设随机变量 $e_1, \cdots e_n$ 独立同分布, 且 $\mathbb{E}[e_i] = 0$, 证明: 对任意 $h_1, \cdots h_n \in \mathbb{R}$, 均有

$$\mathbb{E}\left|\sum_{i=1}^{n} h_i e_i\right| \leqslant 2\mathbb{E}\left|\sum_{i=1}^{n} |h_i| e_i\right|$$

证明. 若 X_i 对称, 则

$$h_i X_i \stackrel{d}{=} |h_i| X_i \implies (h_1 X_1, \cdots, h_n X_n) \stackrel{d}{=} (|h_1| X_1, \cdots, |h_n| X_n) \implies \sum_{i=1}^{n} h_i X_i \stackrel{d}{=} \sum_{i=1}^{n} |h_i| X_i.$$

设 $\{e_i'\}$ 是 $\{e_i\}$ 的一个独立复制, 则 $e_i - e_i'$ 对称. 故有

$$\begin{aligned}
\mathbb{E}\left|\sum_{i=1}^{n} h_i e_i\right| &\leqslant \mathbb{E}\left|\sum_{i=1}^{n} h_i (e_i - \mathbb{E}[e_i'|e_1, \cdots, e_n])\right| = \mathbb{E}\left|\mathbb{E}(\sum_{i=1}^{n} h_i [(e_i - e_i')|e_1, \cdots, e_n])\right| \leqslant \mathbb{E}\left|\sum_{i=1}^{n} h_i (e_i - e_i')\right| \\
&= \mathbb{E}\left|\sum_{i=1}^{n} h_i (e_i - e_i')\right| \leqslant \mathbb{E}\left|\sum_{i=1}^{n} |h_i| e_i\right| + \mathbb{E}\left|\sum_{i=1}^{n} |h_i| e_i'\right| \\
&= 2\mathbb{E}\left|\sum_{i=1}^{n} |h_i| e_i\right|.
\end{aligned}$$

□

1.2 浅谈四种收敛

回顾随机变量列四种收敛的定义:

定义 1.1. 设 $X, X_1, X_2, \cdots, X_n$ 为 $(\Omega, \mathcal{F}, \mathbb{P})$ 上随机变量.

(1) **几乎处处 (以概率 1 收敛)** :

$$\mathbb{P}\left(\{\omega \in \Omega : X_n(\omega) \to X(\omega), n \to \infty\}\right) = 1$$

记为 $X_n \xrightarrow{\text{a.s.}} X$.

(2) **r 阶收敛** $(r > 0)$:$\mathbb{E}[|X_n|^r] < \infty, \forall n$ 且

$$\mathbb{E}[|X_n - X|^r] \to 0, n \to \infty$$

记为 $X_n \xrightarrow{r} X$, 特别 $r = 1$ 时为 平均收敛, $r = 2$ 时为 均方收敛.

(3) 依概率收敛:

$$\forall \epsilon > 0, P(|X_n - X| > \epsilon) \to 0, n \to \infty$$

记为 $X_n \xrightarrow{P} X$.

(4) 依分布收敛:

$$F_{X_n} \xrightarrow{w} F_X$$

记为 $X_n \xrightarrow{D} X$.

注. 只有依分布收敛会“忘记”样本空间 (因为分布函数会“忘记”样本空间), 其他类型的收敛都要考虑样本空间. 因而 $X_n \to X, Y_n \to Y \nRightarrow \mathbb{P}(X = Y) = 1$ 及 $X_n + Y_n \xrightarrow{\text{D}} X + Y$.

证明依概率收敛:

(1) 估计 $\mathbb{P}(|X_n - X| > \varepsilon)$, 可能要利用**命题 1.1** 中的矩不等式或其他方法.

(2) 弱大数律 (注意条件是 i.i.d 且一阶矩存在).

(3) 若收敛极限是常数 c, 利用 $X_n \xrightarrow{D} c \Longleftrightarrow X_n \xrightarrow{P} c$, 转化成证明依分布收敛.

⋮

⋮

例 1.4. 有一列零均值随机变量 $X_1, \cdots X_n$, $\exists c > 0$, 使得对 $\forall n \in \mathbb{N}^*, Var(X_n) < c$. 记 $S_n = \sum_{k=1}^{n} X_k$, 若

$$\mathrm{Cov}(X_i, X_j) \to 0 \quad as\ |i-j| \to +\infty,$$

证明: $\frac{S_n}{n} \xrightarrow{P} 0$.

证明. 对 $\forall \varepsilon > 0$, 由 Chebyshev 不等式, 有

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) \leqslant \frac{\mathbb{E}[S_n^2]}{\varepsilon^2 n^2} = \frac{1}{\varepsilon^2 n^2}\sum_{k=1}^{n}\mathrm{Var}(X_k) + \frac{2}{\varepsilon^2 n^2}\sum_{1\leqslant i<j\leqslant n}\mathrm{Cov}(X_i, X_j) \leqslant \frac{c}{\varepsilon^2 n} + \frac{2}{\varepsilon^2 n^2}\sum_{1\leqslant i<j\leqslant n}\mathrm{Cov}(X_i, X_j).$$

对 $\forall \delta > 0$, 由题意知, $\exists N \in \mathbb{N}^*$, 使得 $|\mathrm{Cov}(X_i, X_j)| \leqslant \delta, |i-j| > N$. 又因为 $\mathrm{Cov}(X_i, X_j) < c$, 因此

$$\sum_{1\leqslant i<j\leqslant n}\mathrm{Cov}(X_i, X_j) = \sum_{\substack{1\leqslant i<j\leqslant n\\ |i-j|\leqslant N}}\mathrm{Cov}(X_i, X_j) + \sum_{\substack{1\leqslant i<j\leqslant n\\ |i-j|> N}}\mathrm{Cov}(X_i, X_j) \leqslant nNc + n^2\delta.$$

故有

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) \leqslant \frac{c}{\varepsilon^2 n} + \frac{2Nc}{\varepsilon^2 n} + 2\frac{\delta}{\varepsilon^2}.$$

令 $n \to +\infty$, 再令 $\delta \to 0$, 则有 $\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) = 0$, 即 $\frac{S_n}{n} \xrightarrow{P} 0$.

□

证明几乎处处收敛:

(1) 利用定义转化: 存在 $\Omega_0 \subset \Omega$, 满足 $\mathbb{P}(\Omega_0) = 1$, 且对 $\forall \omega \in \Omega_0$, 有

$$X_n(\omega) \xrightarrow{n\to+\infty} X(\omega),\ n \to +\infty.$$

固定了 ω 后 $X_n(\omega), X(\omega)$ 变为常数, 于是转化成常数列极限问题: $a_n \to a$. 例如:

- $X_n \xrightarrow{a.s.} X, Y_n \xrightarrow{a.s.} Y \implies X_n + Y_n \xrightarrow{a.s.} X + Y, X_nY_n \xrightarrow{a.s.} XY$.
- $X_n \xrightarrow{a.s.} X \implies \frac{1}{n}\sum_{k=1}^{n} X_n \xrightarrow{a.s.} X$. (思考: 换成依概率收敛是否成立?)
- **Cauchy 列**: $X_n \xrightarrow{a.s.} X \Longleftrightarrow \sup_{k\geqslant n}|X_k - X| \xrightarrow{a.s.} 0 \Longleftrightarrow \sup_{k\geqslant n}|X_k - X_n| \xrightarrow{a.s.} 0 \Longleftrightarrow \sup_{k,m\geqslant n}|X_k - X_m| \xrightarrow{a.s.} 0$.
- $\frac{X_n}{n} \xrightarrow{a.s.} 0 \Longleftrightarrow \frac{\max_{1\leqslant k\leqslant n}|X_k|}{n} \xrightarrow{a.s.} 0$.

(2) 运用 Borel-Cantelli 引理解决 (可能会结合**子序列**, **截尾**, 独立复制与对称化等方法):

$$\forall \varepsilon > 0, \sum_{n=1}^{+\infty} \mathbb{P}(|X_n - X| > \varepsilon) < \infty \Longrightarrow \forall \varepsilon > 0, \mathbb{P}(|X_n - X| > \varepsilon, \text{i.o.}) = 0 \Longleftrightarrow X_n \xrightarrow{\text{a.s.}} X.$$

若 $\{X_n - X\}$ 相互独立, 则为等价关系. 最后进行 $\mathbb{P}(|X_n - X| > \varepsilon)$ 的估计, 可能要借助各种矩不等式或其他方法解决, 用 Markov 不等式时对矩选取合适的阶数, 例子见作业题 **Grimmett7.11.6**.

(3) 强大数律 (注意条件是 i.i.d 且一阶矩存在), 例子见作业题 **Grimmett7.5.1**.

(4) * 见后, 例如转化成证明有关子列依概率收敛问题 (**命题 1.4**).

$\vdots$

$\vdots$

例 1.5. 设 $X_1, \cdots X_n$ 独立同分布, 证明:

$$\frac{X_n}{n} \xrightarrow{\text{a.s.}} 0 \Longleftrightarrow \mathbb{E}|X_1| < \infty.$$

证明. 我们有

$$\frac{X_n}{n} \xrightarrow{\text{a.s.}} 0 \Longleftrightarrow \forall \varepsilon > 0, \mathbb{P}\left(\frac{|X_n|}{n} > \varepsilon, \text{i.o.}\right) = 0 \Longleftrightarrow \forall \varepsilon > 0, \sum_{n=1}^{+\infty} \mathbb{P}\left(\frac{|X_n|}{\varepsilon} > n\right) < \infty.$$

利用 (**见第 5 次习题课讲义引理 2.1**)

$$\sum_{m=1}^{\infty} \mathbb{P}(|X| \geq m) \leq \mathbb{E}[|X|] \leq 1 + \sum_{m=1}^{\infty} \mathbb{P}(|X| \geq m) = \sum_{m=0}^{\infty} \mathbb{P}(|X| \geq m),$$

有

$$\frac{\mathbb{E}\left[|X_n|\right]}{\varepsilon} - 1 \leqslant \sum_{n=1}^{+\infty} \mathbb{P}\left(\frac{|X_n|}{\varepsilon} > n\right) \leqslant \frac{\mathbb{E}\left[|X_n|\right]}{\varepsilon}.$$

因此

$$\frac{X_n}{n} \xrightarrow{\text{a.s.}} 0 \Longleftrightarrow \forall \varepsilon > 0, \sum_{n=1}^{+\infty} \mathbb{P}\left(\frac{|X_n|}{\varepsilon} > n\right) < \infty \Longleftrightarrow \mathbb{E}|X_1| < \infty.$$

□

注意到

$$X_n \xrightarrow{\text{a.s.}} X \Longleftrightarrow \forall \varepsilon > 0, \mathbb{P}(|X_n - X| > \varepsilon, \text{i.o.}) = 0 \Longleftrightarrow \mathbb{P}\left(\lim_{n\to\infty} \bigcup_{k=n}^{+\infty} \{|X_k - X| > \varepsilon\}\right) = 0$$

$$\Longleftrightarrow \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k=n}^{+\infty} \{|X_k - X| > \varepsilon\}\right) = 0 \Longleftrightarrow \lim_{n\to\infty} \mathbb{P}\left(\sup_{k\geqslant n} |X_k - X| > \varepsilon\right) = 0.$$

由此推得如下命题:

命题 1.3. 我们有

(1) $X_n \xrightarrow{\text{a.s.}} X \iff \sup\limits_{k\geqslant n}|X_k - X| \xrightarrow{\mathrm{P}} 0$. 特别地, 若 $|X_n|\downarrow 0$, 则

$$X_n \xrightarrow{\text{a.s.}} 0 \iff \sup_{k\geqslant n}|X_k| \xrightarrow{\mathrm{P}} 0 \iff X_n \xrightarrow{\mathrm{P}} 0.$$

(2) 利用 Cauchy 列等价转换,

$$\begin{aligned} X_n \xrightarrow{\text{a.s.}} X &\iff \sup_{k\geqslant n}|X_k - X| \xrightarrow{\mathrm{P}} 0 \\ &\iff \sup_{k\geqslant n}|X_k - X_n| \xrightarrow{\text{a.s.}} 0 \quad \iff \sup_{k\geqslant n}|X_k - X_n| \xrightarrow{\mathrm{P}} 0 \\ &\iff \sup_{k,m\geqslant n}|X_k - X_m| \xrightarrow{\text{a.s.}} 0 \iff \sup_{k,m\geqslant n}|X_k - X_m| \xrightarrow{\mathrm{P}} 0 \end{aligned}$$

问题: 对于一般情形, 能否把几乎处处收敛转化成依概率收敛问题?

引理 1.1. 我们有

$$X_n \xrightarrow{\mathrm{P}} X \iff \forall \varepsilon > 0, \sup_{k\geqslant n}\mathbb{P}(|X_k - X_n| > \varepsilon) \xrightarrow{n\to+\infty} 0,$$

注意: 区分 $\sup\limits_{k\geqslant n}\mathbb{P}(|X_k - X_n| > \varepsilon)\&\mathbb{P}(\sup\limits_{k\geqslant n}|X_k - X_n| > \varepsilon)$, 两者不一样!

证明. $\Longrightarrow$: 对 $\forall\varepsilon > 0$, 由 $X_n \xrightarrow{\mathrm{P}} X$ 知, 对 $\forall\delta > 0, \exists n_0 \in \mathbb{N}^*$, 当 $n \geqslant n_0$ 时, 有 $\mathbb{P}(|X_n - X| > \frac{\varepsilon}{2}) < \frac{\delta}{2}$. 故当 $k \geqslant n \geqslant n_0$ 时,

$$\mathbb{P}(|X_k - X_n| > \varepsilon) \leqslant \mathbb{P}(|X_k - X - (X_n - X)| > \varepsilon) \leqslant \mathbb{P}\left(|X_k - X| > \frac{\varepsilon}{2}\right) + \mathbb{P}\left(|X_n - X| > \frac{\varepsilon}{2}\right) < \delta.$$

$\Longleftarrow$: $\forall k \in \mathbb{N}^*$, $\exists n_k \uparrow +\infty$, 使得

$$\mathbb{P}(|X_{n_{k+1}} - X_{n_k}| > 2^{-k}) < 2^{-k} \quad \forall k \in \mathbb{N}^*$$

所以

$$\sum_{k=1}^{+\infty}\mathbb{P}(|X_{n_{k+1}} - X_{n_k}| > 2^{-k}) < 1 \implies \mathbb{P}(|X_{n_{k+1}} - X_{n_k}| > 2^{-k}, \text{i.o.}) = 0$$

记 $\Omega_0 = \left\{|X_{n_{k+1}} - X_{n_k}| > 2^{-k}, \text{i.o.}\right\}$, 则当 $\omega \notin \Omega_0$ 时,

$$\sum_{k=1}^{+\infty}|X_{n_{k+1}}(\omega) - X_{n_k}(\omega)| < \infty$$

因此 $X_{n_k}(\omega)$ 为 Cauchy 列. 故存在极限 $X(\omega)$, 满足 $X_{n_k}(\omega) \to X(\omega)$. 所以

$$X_{n_k} \xrightarrow{\text{a.s.}} X.$$

因此当 $n > n_k$ 时, 利用

$$\mathbb{P}(|X_n - X| > \varepsilon) \leqslant \mathbb{P}\left(|X_{n_k} - X| > \frac{\varepsilon}{2}\right) + \mathbb{P}\left(|X_n - X_{n_k}| > \frac{\varepsilon}{2}\right)$$

可知

$$X_n \xrightarrow{\mathrm{P}} X.$$

$\square$

命题 1.4. 我们有

(1) $X_n \xrightarrow{\mathrm{P}} X \implies$ 存在子列 $\{n_k\}, X_{n_k} \xrightarrow{\text{a.s.}} X$.

(2) $X_n \xrightarrow{\mathrm{P}} X \iff$ 对任意子列 $\{n_k\}, \exists \{n_k'\} \subset \{n_k\}, X_{n_k'} \xrightarrow{\text{a.s.}} X$.

证明. (1) $\implies$: 参考**引理 1.1** 的证明;
(2) $\implies$: 利用 (1) 及

$$a_n \to a \iff \forall \{n_k\}, \exists \{n_k'\} \subset \{n_k\}, a_{n_k'} \to a.$$

$\impliedby$: 反证. 若 $X_n \xrightarrow{\mathrm{P}} X$ 不成立, 则 $\exists \varepsilon > 0, \delta > 0, \limsup\limits_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) > \delta$, 因此

$$\exists \{n_k\}, \mathbb{P}(|X_{n_k} - X| > \varepsilon) > \delta \implies \forall \{n_k'\} \subset \{n_k\}, \mathbb{P}(|X_{n_k'} - X| > \varepsilon) > \delta.$$

矛盾. □

注. 因此, 随机变量列的几乎处处收敛给出了另一个关于子列依概率收敛的刻画条件.

证明 r 阶收敛:

(1) 利用**命题 1.1** 中的矩不等式, 如 **Hölder 不等式**、**Minkowski 不等式**、C_r **不等式**等. 看一看笔记和作业题中的几个例子.

$$X_n \xrightarrow{\mathrm{r}} X, Y_n \xrightarrow{\mathrm{r}} Y \implies X_n + Y_n \xrightarrow{\mathrm{r}} X + Y, X_n Y_n \xrightarrow{\mathrm{r}} XY.$$

(2) 借助矩不等式和截尾等技术, 运用单调收敛定理、控制收敛定理、Fatou 引理等证明.

$\vdots$

$\vdots$

已学过的定理: 若 $\exists k > 0$, 使得 $\mathbb{P}(|X_n| \leqslant k) = 1, \forall n \in \mathbb{N}^*$, 则

$$X_n \xrightarrow{\mathrm{P}} X \implies X_n \xrightarrow{\mathrm{r}} X.$$

更一般地情形留给读者自行验证: 若 $|X_n| \leqslant Y$, Y 绝对可积且 $X_n \xrightarrow{\mathrm{P}} X$, 则 $X_n \xrightarrow{1} X$.

注. r 阶收敛不是本门课重点, 加之多数人没有系统学习实变的内容, 以及没有引入一致可积等概念, 因此对这门课而言一般来说不会专门在这方面出过难的题目, 后续高等概率论课程还会深入讲解这部分内容.

例 1.6. 设随机变量 $X_1, \cdots, X_n$ 独立同分布, 且服从 $[0, a]$ 上的均匀分布, 其中 $a > 0$. 记

$$M_n = \max\{X_1, \cdots, X_n\},$$

分别在 a.s., r 阶收敛的意义下证明 $M_n \to a$.

解. 显然 $M_n < a$, 对 $\forall 0 < \varepsilon < a$ 有

$$\mathbb{P}(M_n \leqslant a - \varepsilon) = \prod_{k=1}^{n} \mathbb{P}(X_k \leqslant a - \varepsilon) = \left(\frac{a - \varepsilon}{a}\right)^n.$$

因此

$$\sum_{n=1}^{+\infty}\mathbb{P}(M_n<a-\varepsilon)=\sum_{n=1}^{+\infty}\left(\frac{a-\varepsilon}{a}\right)^n<+\infty.$$

由 Borel-Cantelli 引理知 $\mathbb{P}(M_n\leqslant a-\varepsilon,\text{i.o.})=0$. 因此 $M_n\xrightarrow{\text{a.s.}}a$.
同时有

$$\begin{aligned}\mathbb{E}[|M_n-a|^r]&=\mathbb{E}[|M_n-a|^r I_{\{|M_n-a|\leqslant\varepsilon\}}]+\mathbb{E}[|M_n-a|^r I_{\{|M_n-a|>\varepsilon\}}]\\&\leqslant\varepsilon^r+a^r\mathbb{P}(|M_n-a|>\varepsilon)\\&=\varepsilon^r+a^r\left(\frac{a-\varepsilon}{a}\right)^n.\end{aligned}$$

令 $n\to+\infty$, 再令 $\varepsilon\to 0$, 则有

$$M_n\xrightarrow{r}a.$$

□

证明/利用依分布收敛:

(1) 定义: $F_{X_n}\xrightarrow{w}F_X$. 例子见作业题 **Grimmett5.12.32，5.12.39**

(2) 利用连续性定理, 用特征函数证明或者 CLT、Lindeberg-Feller CLT 等, 内容详见**第 7 次习题课讲义 2.3 节**. 例子见作业题 **Grimmett5.12.41**

(3) 弱大数律 (**注意条件是 i.i.d 且一阶矩存在**).

(4) 利用 $X_n\xrightarrow{\text{P}}X\Longrightarrow X_n\xrightarrow{\text{D}}X$.

(5) 利用依分布收敛的等价条件, 各种性质转化, 具体见后.

⋮

⋮

例 1.7. 随机变量 $X_1,X_2,\cdots X_n$ 独立同分布且 $\mathbb{E}[X_1]=0,\mathbb{E}[X_1^2]=1$, $S_n=\sum_{k=1}^{n}X_k$, 则

$$\frac{S_n}{\sqrt{n}}\xrightarrow{\text{D}}N(0,1),$$

但不存在随机变量 Z, 使得

$$\frac{S_n}{\sqrt{n}}\xrightarrow{\text{P}}Z.$$

证明. 利用 i.i.d CLT 即可推出 $\frac{S_n}{\sqrt{n}}\xrightarrow{\text{D}}N(0,1)$, 现反证法证明后一个命题: 若存在随机变量 Z, 使得 $\frac{S_n}{\sqrt{n}}\xrightarrow{\text{P}}Z$ 成立, 则由极限分布的唯一性及前一个命题知 $Z\sim N(0,1)$. 一方面,

$$\frac{S_{2n}-S_n}{\sqrt{2n}}=\frac{S_{2n}}{\sqrt{2n}}-\frac{1}{\sqrt{2}}\frac{S_n}{\sqrt{n}}\xrightarrow{\text{P}}\left(1-\frac{1}{\sqrt{2}}\right)Z.$$

另一方面,

$$\frac{S_{2n}-S_n}{\sqrt{2n}}\xrightarrow{\text{D}}\frac{S_n}{\sqrt{2n}}\xrightarrow{\text{P}}\frac{1}{\sqrt{2}}Z,$$

矛盾. 因此不存在 Z 满足 $\frac{S_n}{\sqrt{n}}\xrightarrow{\text{P}}Z$. □

另外对于连续性随机变量, 也可以通过证明密度函数列的逐点收敛性推出依分布收敛:

命题 1.5. 设 $\{X_n\}, X$ 均为连续型随机变量, 记 X_n, X 的密度函数是 $f_n(x), f(x)$, 若对 $\forall x \in \mathbb{R}$, 均有 $f_n(x) \to f(x)$, 则有 $X_n \xrightarrow{D} X$.

证明. 对 $\forall x \in \mathbb{R}$, 利用 $|X| = X^+ + X^- = 2X^+ - X$, 有

$$
\begin{aligned}
|\mathbb{P}(X_n \leqslant x) - \mathbb{P}(X \leqslant x)| = \left|\int_{-\infty}^{x} (f_n(y) - f(y))\mathrm{d}y\right| &\leqslant \int_{-\infty}^{+\infty} |f_n(y) - f(y)|\mathrm{d}y = 2\int_{-\infty}^{+\infty} (f(y) - f_n(y))^+\mathrm{d}y \\
&\leqslant 2\int_{-\infty}^{+\infty} f(y)\mathrm{d}y < +\infty.
\end{aligned}
$$

利用控制收敛定理, 有

$$
\lim_{n\to+\infty}\int_{-\infty}^{+\infty} |f_n(y) - f(y)|\mathrm{d}y = 2\lim_{n\to+\infty}\int_{-\infty}^{+\infty} (f(y) - f_n(y))^+\mathrm{d}y = 2\int_{-\infty}^{+\infty}\lim_{n\to+\infty}(f(y) - f_n(y))^+\mathrm{d}y = 0.
$$

进而

$$
\lim_{n\to+\infty} |\mathbb{P}(X_n \leqslant x) - \mathbb{P}(X \leqslant x)| = 0 \implies X_n \xrightarrow{D} X.
$$

□

回顾已讲过的关于依分布收敛的内容:

引理 1.2. $X_n \xrightarrow{D} X \iff$ 对任意子列 $\{n_k\}$, 存在子子列 $\{n_k'\} \subset \{n_k\}$, 使得

$$
X_{n_k'} \xrightarrow{D} X.
$$

定理 1.2 (Skorokhod 表示定理). 设 $X_n \xrightarrow{D} X$, 则存在 $(\Omega, \mathcal{F}, \mathbb{P})$ 及其上的 Y_n, Y 满足:

(1) Y_n 与 Y 同分布, Y 与 X 同分布,

(2) $Y_n \xrightarrow{\text{a.s.}} Y$.

定理 1.3. $X_n \xrightarrow{D} X \iff \forall g \in \mathrm{C_b}(\mathbb{R})$(有界连续函数), 有 $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)], n \to \infty$.

下面看一个和定理 1.3 有关的例子:

例 1.8. 设 $X_n \xrightarrow{\mathrm{D}} X$, 且存在 $r, C > 0$, 使得 $\mathbb{E}[|X_n|^r] \leqslant C$, 证明: 对 $\forall 0 < s < r$, 均有

$$
\lim_{n\to+\infty} \mathbb{E}[|X_n|^s] = \mathbb{E}[X^s].
$$

证明. 利用定理 1.3, 对 $\forall g \in \mathrm{C_b}(\mathbb{R})$(有界连续函数), 有 $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)], n \to \infty$. 设 $M > 0$, 定义

$$
g_M(x) = \begin{cases} |x|^s, & |x| < M, \\ M^s, & |x| \geqslant M. \end{cases}
$$

则 $g_M(x)$ 是有界连续函数, 以及当 $M \to +\infty$ 时 $0 \leqslant g_M(x) \uparrow |x|^s$, 因此有

$$
\lim_{n\to+\infty} \mathbb{E}[g_M(X_n)] = \mathbb{E}[g_M(X)].
$$

同时我们有

$$0 \leqslant \mathbb{E}[|X_n|^s] - \mathbb{E}[g_M(X_n)] \leqslant \mathbb{E}[|X_n|^s I_{\{|X_n| \geqslant M\}}] \leqslant \frac{1}{M^{r-s}}\mathbb{E}[|X_n|^r I_{\{|X_n| \geqslant M\}}] \leqslant \frac{C}{M^{r-s}},$$

因此

$$\mathbb{E}[g_M(X_n)] \leqslant \mathbb{E}[|X_n|^s] \leqslant \mathbb{E}[g_M(X_n)] + \frac{C}{M^{r-s}}.$$

先令 $n \to +\infty$, 再令 $M \uparrow +\infty$, 结合 $\lim\limits_{M \to +\infty} \mathbb{E}[g_M(X)] = \mathbb{E}[|X|^s]$(利用单调收敛定理), 得

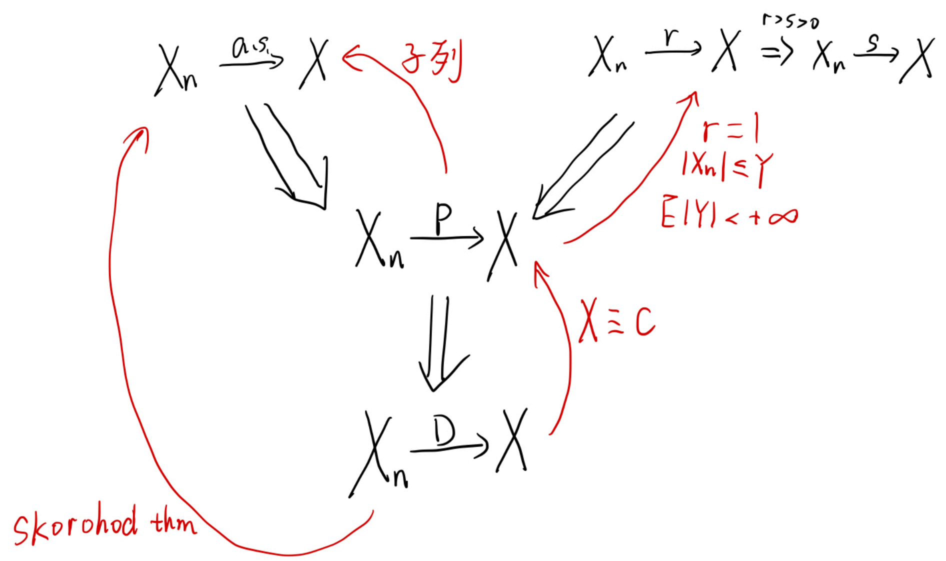
$$\lim_{n \to +\infty} \mathbb{E}[|X_n|^s] = \mathbb{E}[X^s].$$

□

注. 这里的截尾处理非常精妙, 自行体会.

1.3 结论拾零及应用

本节主要以例子来呈现. 先回顾一下各收敛间的关系:



例 1.9. 设 $f : \mathbb{R} \to \mathbb{R}$ 是连续函数, 证明:

(1) $X_n \xrightarrow{\text{a.s.}} X \implies f(X_n) \xrightarrow{\text{a.s.}} f(X)$.

(2) $X_n \xrightarrow{\text{P}} X \implies f(X_n) \xrightarrow{\text{P}} f(X)$.

(3) $X_n \xrightarrow{\text{D}} X \implies f(X_n) \xrightarrow{\text{D}} f(X)$.

证明. (1) $\exists \Omega_0 \subset \Omega$, 使得 $\mathbb{P}(\Omega_0) = 1$ 且对 $\forall \omega \in \Omega_0$, 均有 $X_n(\omega) \to X(\omega)$, 由 f 的连续性知 $f(X_n(\omega)) \to f(X(\omega))$, 因此 $f(X_n) \xrightarrow{\text{a.s.}} f(X)$.
(2) 反证: 若命题不成立, 则存在子列 $\{n_k\}$, $\varepsilon, \delta > 0$, 满足

$$\mathbb{P}(|f(X_{n_k}) - f(X)| > \varepsilon) > \delta$$

因为 $X_{n_k} \xrightarrow{\mathrm{P}} X$, 利用**命题 1.4**, 存在子列 $\{n_k'\} \subset \{n_k\}$, 使得 $X_{n_k'} \xrightarrow{\text{a.s.}} X$. 由 (1) 知 $f(X_{n_k'}) \xrightarrow{\text{a.s.}} f(X)$, 从而 $f(X_{n_k'}) \xrightarrow{\mathrm{P}} f(X)$, 矛盾.
(3) 利用 Skorokhod 表示定理, 在一个概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 上存在 Y_n, Y, 使得 $Y_n \overset{D}{=\!=} X_n, Y \overset{D}{=\!=} X$ 且 $Y_n \xrightarrow{\text{a.s.}} Y$. 由 (1) 知 $f(Y_n) \xrightarrow{\text{a.s.}} f(Y)$, 因此 $f(Y_n) \xrightarrow{\mathrm{D}} f(Y) \implies f(X_n) \xrightarrow{\mathrm{D}} f(X)$.

□

例 1.10. 设 $X_1 \cdots X_n$ 是随机变量列, $N_1, N_2, \cdots$ 是取值为正整数值的随机变量, 证明:

(1) 若 $X_n \xrightarrow{\text{a.s.}} X$ 且 $N_k \xrightarrow{a.s.} \infty$, 则 $X_{N_k} \xrightarrow{\text{a.s.}} X$.

(2) 若 $X_n \xrightarrow{\text{a.s.}} X$ 且 $N_k \xrightarrow{P} \infty$, 则 $X_{N_k} \xrightarrow{\mathrm{P}} X$.

(3) 若 $X_n \xrightarrow{\mathrm{P}} X$, $N_k \xrightarrow{P} \infty$ 且 $\{X_n\}$ 与 $\{N_k\}$ 独立, 则 $X_{N_k} \xrightarrow{\mathrm{P}} X$.

(4) 若 $X_n \xrightarrow{\mathrm{D}} X$, $N_k \xrightarrow{P} \infty$ 且 $\{X_n\}$ 与 $\{N_k\}$ 独立, 则 $X_{N_k} \xrightarrow{\mathrm{D}} X$.

证明. (1) $\exists \Omega_0 \subset \Omega$, 使得 $\mathbb{P}(\Omega_0) = 1$ 且对 $\forall \omega \in \Omega_0$, 均有 $X_n(\omega) \to X(\omega)$ 及 $N_k(\omega) \to +\infty$. 故对 $\forall \varepsilon > 0, \exists N \in \mathbb{N}^*$, 当 $n > N$ 时, 有 $|X_n(\omega) - X(\omega)| < \varepsilon$ 且 $\exists K \in \mathbb{N}^*$, 当 $k > K$ 时, 有 $N_k(\omega) > n$, 从而 $|X_{N_k}(\omega) - X(\omega)| < \varepsilon$. 故 $X_{N_k}(\omega) \to X(\omega)$, 因此 $X_{N_k} \xrightarrow{\text{a.s.}} X$.
(2) 对 $\forall \varepsilon > 0$, 有

$$\mathbb{P}(|X_{N_k} - X| > \varepsilon) \leqslant \mathbb{P}(N_k \leqslant n) + \mathbb{P}(|X_{N_k} - X| > \varepsilon, N_k > n) \leqslant \mathbb{P}(N_k \leqslant n) + \mathbb{P}(\sup_{m \geqslant n} |X_m - X| > \varepsilon)$$

先令 $k \to +\infty$, 再令 $n \to +\infty$, 结合**命题 1.3** 即得.
(3) 对 $\forall \varepsilon > 0, \delta > 0, \exists N \in \mathbb{N}^*$, 当 $n > N$ 时, 有 $\mathbb{P}(|X_n - X| > \varepsilon) < \delta$, $\exists K \in \mathbb{N}^*$, 当 $k > K$ 时, 有 $\mathbb{P}(N_k \leqslant N) < \delta$. 因此

$$\begin{aligned}
\mathbb{P}(|X_{N_k} - X| > \varepsilon) &\leqslant \mathbb{P}(N_k \leqslant N) + \sum_{n=N+1}^{+\infty} \mathbb{P}(|X_{N_k} - X| > \varepsilon, N_k = n) \\
&= \mathbb{P}(N_k \leqslant N) + \sum_{n=N+1}^{+\infty} \mathbb{P}(|X_n - X| > \varepsilon)\mathbb{P}(N_k = n) \\
&\leqslant 2\delta.
\end{aligned}$$

令 $\delta \to 0$ 即得.
(4) 设 $\phi_n(t) = \mathbb{E}[\mathrm{e}^{\mathrm{i}X_n t}], \phi(t) = \mathbb{E}[\mathrm{e}^{\mathrm{i}Xt}], \psi_k(t) = \mathbb{E}[\mathrm{e}^{\mathrm{i}X_{N_k} t}] = \mathbb{E}[\mathbb{E}[\mathrm{e}^{\mathrm{i}X_{N_k} t} | N_k]] = \sum_{n=1}^{+\infty} \mathbb{P}(N_k = n)\phi_n(t)$, 则

$$|\psi_k(t) - \phi(t)| \leqslant \sum_{n=1}^{+\infty} \mathbb{P}(N_k = n)|\phi_n(t) - \phi(t)|$$

对 $\forall \varepsilon > 0, \exists N \in \mathbb{N}^*$, 则对 $n > N$, 有 $|\phi_n(t) - \phi(t)| < \varepsilon$, $\exists K \in \mathbb{N}^*$. 当 $k > K$ 时, 有 $\mathbb{P}(N_k \leqslant N) < \dfrac{\varepsilon}{M}$ (其中 $M = \max\limits_{1 \leqslant i \leqslant n} |\phi_i(t) - \phi(t)|$), 因此

$$|\psi_k(t) - \phi(t)| \leqslant \frac{\varepsilon}{M} \max_{1 \leqslant i \leqslant n} |\phi_i(t) - \phi(t)| + \mathbb{P}(N_k > N)\varepsilon \leqslant 2\varepsilon.$$

令 $\varepsilon \to 0$ 即得. □

定理 1.4 (Slutsky 引理). $X_n \xrightarrow{\mathrm{D}} X, Y_n \xrightarrow{\mathrm{P}} b, Z_n \xrightarrow{\mathrm{P}} c$. 则

$$X_nY_n + Z_n \xrightarrow{\mathrm{D}} bX + c.$$

注 1. 若 $b \neq 0$, 则有 $Y_n^{-1} \xrightarrow{\mathrm{P}} b^{-1}$. 另外可以配合**例 1.9** 食用.

注 2. Slutsky 引理在研究参数估计、统计量的渐进性质中会经常用到, 作业题 **Grimmett7.2.5(a)** 给出了其简单形式, 我们来看几个相关的例子:

例 1.11 (作业题 Grimmett 7.3.13). 设均值为 μ 的随机变量列 $\{X_n\}$ 独立同分布且二阶矩有限, 令 $\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$, 证明:

$$\sum_{k=1}^{n}(X_k - \mu) \Big/ \sqrt{\sum_{k=1}^{n}(X_k - \overline{X})^2} \xrightarrow{\mathrm{D}} N(0,1).$$

证明. 设其方差为 σ^2, 即证

$$\sum_{k=1}^{n}\frac{X_k - \mu}{\sqrt{n}\sigma} \Big/ \sqrt{\sum_{k=1}^{n}\frac{(X_k - \overline{X})^2}{n\sigma^2}} \xrightarrow{\mathrm{D}} N(0,1).$$

注意到

$$\sum_{k=1}^{n}\frac{(X_k - \overline{X})^2}{n\sigma^2} = \frac{1}{n\sigma^2}\sum_{k=1}^{n}(X_k - \mu)^2 - \frac{2}{n\sigma^2}(\overline{X} - \mu)\sum_{k=1}^{n}(X_k - \mu) + \frac{1}{\sigma^2}(\overline{X} - \mu)^2.$$

利用 CLT 知

$$\sum_{k=1}^{n}\frac{X_k - \mu}{\sqrt{n}\sigma} \xrightarrow{\mathrm{D}} N(0,1).$$

利用弱大数律知

$$\frac{1}{n\sigma^2}\sum_{k=1}^{n}(X_k - \mu)^2 \xrightarrow{\mathrm{P}} 1, \quad \frac{1}{n}\sum_{k=1}^{n}(X_k - \mu) \xrightarrow{\mathrm{P}} 0, \quad \overline{X} - \mu \xrightarrow{\mathrm{P}} 0.$$

结合**定理 1.4(Slutsky 引理)** 知

$$\sum_{k=1}^{n}(X_k - \mu) \Big/ \sqrt{\sum_{k=1}^{n}(X_k - \overline{X})^2} \xrightarrow{\mathrm{D}} N(0,1).$$

□

例 1.12. 随机变量列 $\{X_n\}$ 相互独立且满足

$$\mathbb{P}(X_n = \pm n) = \frac{1}{2n^3}, \quad \mathbb{P}(X_n = \pm 1) = \frac{1}{2} - \frac{1}{2n^3}, \quad n \in \mathbb{N}^*.$$

证明:

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_k \xrightarrow{\mathrm{D}} N(0,1).$$

证明. 引入截尾处理后的随机变量: $Y_i = \mathrm{sgn}X_i$, 注意到

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_k = \frac{1}{\sqrt{n}}\sum_{k=1}^{n} Y_k + \frac{1}{\sqrt{n}}\sum_{k=1}^{n}(X_k - Y_k)$$

一方面, 利用 CLT 知,

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} Y_k \xrightarrow{\mathrm{D}} N(0,1).$$

另一方面, 对 $\forall\varepsilon > 0$, 有

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(X_k - Y_k)\right| > \varepsilon\right) &\leqslant \frac{1}{\sqrt{n}\varepsilon}\mathbb{E}[|\sum_{k=1}^{n}(X_k - Y_k)|] \leqslant \frac{1}{\sqrt{n}\varepsilon}\sum_{k=1}^{n}\mathbb{E}[|X_k - Y_k|] \\ &\leqslant \frac{1}{\sqrt{n}\varepsilon}\sum_{k=1}^{+\infty}\frac{1}{k^2} \xrightarrow{n\to+\infty} 0.\end{aligned}$$

因此

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(X_k - Y_k) \xrightarrow{\mathrm{P}} 0.$$

结合**定理 1.4(Slutsky 引理)** 知,

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_k \xrightarrow{\mathrm{D}} N(0,1).$$

□

注. 截尾法的妙用在这里又体现了出来. 这里也可以类似**例 1.3** 一样转化.

例 1.13 (Δ 方法)**.** 随机变量列 $\{X_n\}$, 且存在常数 a 和 $\sigma^2 > 0$, 使得

$$\sqrt{n}(X_n - a) \xrightarrow{\mathrm{D}} N(0,\sigma^2).$$

若函数 g 在 a 处可导, 且 $g'(a) \neq 0$, 则

$$\sqrt{n}(g(X_n) - g(a)) \xrightarrow{\mathrm{D}} N(0,\sigma^2(g'(a))^2).$$

证明. 参考 **苏淳, 冯群强：概率论（第三版）, 例 6.3.4**. 这也是 Slutsky 引理的一个重要应用. □

例 1.14 (**Kolmogorov 不等式**)**.** 设 $\{X_k, 1 \leqslant k \leqslant n\}$ 独立并满足 $\mathbb{E}[X_k] = 0, \mathbb{E}[X_k^2] < \infty$, 记 $S_n = \sum_{k=1}^{n} X_k$. 对 $\forall\varepsilon > 0$,

(1) 证明: $\mathbb{P}(\max_{1\leqslant k\leqslant n}|S_k| \geqslant \varepsilon) \leqslant \frac{1}{\varepsilon^2}\sum_{k=1}^{n}\mathbb{E}[X_k^2]$.

(2) 若存在 $c > 0$, 使得 $|X_k| \leqslant c, \forall 1 \leqslant k \leqslant n$, 证明:

$$\mathbb{P}(\max_{1\leqslant k\leqslant n}|S_k| \geqslant \varepsilon) \geqslant 1 - \frac{(\varepsilon + c)^2}{\sum_{k=1}^{n}\mathbb{E}[X_k^2]}.$$

证明. (1) 记 $T=\inf\{k\in\mathbb{N}^*:|S_k|\geqslant\varepsilon\}$, 则

$$\mathbb{P}(\max_{1\leqslant k\leqslant n}|S_k|\geqslant\varepsilon)=\mathbb{P}(T\leqslant n)=\sum_{k=1}^n\mathbb{P}(T=k)=\sum_{k=1}^n\mathbb{E}[I_{\{T=k\}}]\leqslant\mathbb{E}\left[\left(\frac{|S_k|}{\varepsilon}\right)^2 I_{\{T=k\}}\right]=\frac{1}{\varepsilon^2}\sum_{k=1}^n\mathbb{E}[S_k^2 I_{\{T=k\}}]$$

而

$$\begin{aligned}\mathbb{E}[S_n^2 I_{\{T=k\}}]&=\mathbb{E}[(S_k+S_n-S_k)^2 I_{\{T=k\}}]=\mathbb{E}[S_k^2 I_{\{T=k\}}]+\mathbb{E}[(S_n-S_k)^2 I_{\{T=k\}}]+2\mathbb{E}[S_k(S_n-S_k)I_{\{T=k\}}]\\&\geqslant\mathbb{E}[S_k^2 I_{\{T=k\}}].\end{aligned}$$

其中 $\mathbb{E}[S_k(S_n-S_k)I_{\{T=k\}}]=\mathbb{E}[S_k I_{\{T=k\}}]\mathbb{E}[S_n-S_k]=0$, 这是因为 $\{T=k\}$ 只决定于 $\{X_1,\cdots X_k\}$ $(\sigma(X_1,\cdots X_k))$. 因此

$$\mathbb{P}(\max_{1\leqslant k\leqslant n}|S_k|\geqslant\varepsilon)=\frac{1}{\varepsilon^2}\sum_{k=1}^n\mathbb{E}[S_k^2 I_{\{T=k\}}]\leqslant\frac{1}{\varepsilon^2}\sum_{k=1}^n\mathbb{E}[S_n^2 I_{\{T=k\}}]=\frac{1}{\varepsilon^2}\mathbb{E}[S_n^2 I_{\{T\leqslant n\}}]\leqslant\frac{\mathbb{E}[S_n^2]}{\varepsilon^2}=\frac{1}{\varepsilon^2}\sum_{k=1}^n\mathbb{E}[X_k^2].$$

(2) 我们有

$$\begin{aligned}\mathbb{E}[S_n^2 I_{\{T=k\}}]&=\mathbb{E}[S_k^2 I_{\{T=k\}}]+\mathbb{E}[(S_n-S_k)^2 I_{\{T=k\}}]\leqslant(\varepsilon+c)^2\mathbb{P}(T=k)+\mathbb{E}[(S_n-S_k)^2]\mathbb{P}(T=k)\\&\leqslant\mathbb{P}(T=k)(\mathbb{E}[S_n^2]+(\varepsilon+c)^2),\end{aligned}$$

其中 $\mathbb{E}[S_k^2 I_{\{T=k\}}]\leqslant(\varepsilon+c)^2\mathbb{P}(T=k)$ 是因为在 $\{T=k\}$ 下有 $|S_k|\leqslant|S_{k-1}+X_k|\leqslant\varepsilon+c$. 对 k 从 1 至 n 累加得

$$\mathbb{E}[S_n^2 I_{\{T\leqslant n\}}]\leqslant\mathbb{P}(T\leqslant n)(\mathbb{E}[S_n^2]+(\varepsilon+c)^2).$$

另一方面,

$$\mathbb{E}[S_n^2 I_{\{T\leqslant n\}}]=\mathbb{E}[S_n^2]-\mathbb{E}[S_n^2 I_{\{T>n\}}]\geqslant\mathbb{E}[S_n^2]-\varepsilon^2\mathbb{P}(T>n)=\mathbb{E}[S_n^2]-\varepsilon^2+\varepsilon^2\mathbb{P}(T\leqslant n).$$

因此

$$\mathbb{P}(T\leqslant n)\geqslant\frac{\mathbb{E}[S_n^2]-\varepsilon^2}{\mathbb{E}[S_n^2]+(\varepsilon+c)^2-\varepsilon^2}=1-\frac{(\varepsilon+c)^2}{\mathbb{E}[S_n^2]+(\varepsilon+c)^2-\varepsilon^2}\geqslant 1-\frac{(\varepsilon+c)^2}{\mathbb{E}[S_n^2]}.$$

□

注. 这里引入了新的概念停时: $T=\inf\{1\leqslant k\leqslant n:|S_k|\geqslant\varepsilon\}$, 这是一个概率尤其是随机过程中非常基本且重要的工具 (直观理解: 调节 "信息" 变化).

利用 Kolmogorov 不等式可以得到一级数定理:

定理 1.5 (一级数定理)**.** 设 $\{X_n\}$ 独立, $\mathbb{E}[X_n]=0$, $\sum_{n=1}^{\infty}\mathbb{E}X_n^2<\infty$, 则 $\sum_{n=1}^{\infty}X_n$ a.s. 收敛.

证明. 记 $S_n=\sum_{k=1}^n X_k$, 利用 Kolmogorov 不等式 (例 1.14), 有

$$\mathbb{P}(\max_{M\leqslant m\leqslant N}|S_m-S_M|\geqslant\varepsilon)\leqslant\frac{\mathbb{E}[|S_N-S_M|^2]}{\varepsilon^2}=\frac{1}{\varepsilon^2}\sum_{m=M+1}^{N}\mathbb{E}[X_m^2].$$

令 $N\uparrow+\infty$, 则有

$$\mathbb{P}(\max_{m\geqslant M}|S_m-S_M|\geqslant\varepsilon)\leqslant\frac{1}{\varepsilon^2}\sum_{m=M+1}^{+\infty}\mathbb{E}[X_m^2]\xrightarrow{M\to+\infty}0.$$

结合**命题 1.3**, 得

$$\max_{m\geqslant M}|S_m-S_M|\xrightarrow{\mathrm{P}}0\implies S_m\text{ 为 a.s. Cauchy 列}\implies S_m\text{ a.s. 收敛}\implies\sum_{n=1}^{\infty}X_n\text{ a.s. 收敛.}$$

□

1.4 用截尾术证明弱大数律

课堂上我们利用特征函数及连续性定理方法证明了弱大数律 (先证明依分布收敛于 μ, 再依概率收敛等价转换). 这里我们再展示另一种证明方法——截尾术:

定理 1.6 (弱大数律)**.** 设 $X_1,\cdots,X_n$ 独立同分布, $\mu=\mathbb{E}[X_1]$ 且 $\mathbb{E}[|X_1|]<+\infty$, 记 $S_n=\sum_{k=1}^{n}X_k$, 则

$$\frac{S_n}{n}\xrightarrow{P}\mu.$$

证明. 引入经截断处理的随机变量:

$$X_n^{(1)}=X_nI_{\{|X_n|\leqslant M\}},\quad X_n^{(2)}=X_nI_{\{|X_n|>M\}}.$$

其中 $X_n=X_n^{(1)}+X_n^{(2)}$. 并设 $S_n^{(1)}=\sum_{i=1}^{n}X_i^{(1)},S_n^{(2)}=\sum_{i=1}^{n}X_i^{(2)}$. 则对 $\forall\varepsilon>0$, 有

$$\begin{aligned}\mathbb{P}\left(\left|\frac{S_n}{n}-\mu\right|>\varepsilon\right)&=\mathbb{P}\left(\left|\frac{(S_n^{(1)}-\mathbb{E}[S_n^{(1)}])+(S_n^{(2)}-\mathbb{E}[S_n^{(2)}])}{n}\right|>\varepsilon\right)\\&\leqslant\mathbb{P}\left(\left|\frac{S_n^{(1)}-\mathbb{E}[S_n^{(1)}]}{n}\right|>\frac{\varepsilon}{2}\right)+\mathbb{P}\left(\left|\frac{S_n^{(2)}-\mathbb{E}[S_n^{(2)}]}{n}\right|>\frac{\varepsilon}{2}\right)\end{aligned}$$

现在我们分别处理这两部分:

$$\mathbb{P}\left(\left|\frac{S_n^{(1)}-\mathbb{E}[S_n^{(1)}]}{n}\right|>\frac{\varepsilon}{2}\right)\leqslant\frac{\mathrm{Var}(S_n^{(1)})}{(\frac{1}{2}\varepsilon n)^2}=\frac{4\mathrm{Var}(X_1^{(1)})}{\varepsilon^2 n}\leqslant\frac{4\mathbb{E}[(X_1^{(1)})^2]}{\varepsilon^2 n}\leqslant\frac{4M^2}{\varepsilon^2 n}\xrightarrow{n\to+\infty}0.$$

同时有

$$\mathbb{P}\left(\left|\frac{S_n^{(2)}-\mathbb{E}[S_n^{(2)}]}{n}\right|>\frac{\varepsilon}{2}\right)\leqslant\frac{\mathbb{E}|S_n^{(2)}-\mathbb{E}[S_n^{(2)}]|}{\frac{\varepsilon}{2}n}\leqslant\frac{4\mathbb{E}[|X_1|I_{\{|X_1|>M\}}]}{\varepsilon}\xrightarrow{M\to+\infty}0,$$

其中

$$\mathbb{E}|S_n^{(2)}-\mathbb{E}[S_n^{(2)}]|\leqslant\mathbb{E}[|S_n^{(2)}|]+|\mathbb{E}[S_n^{(2)}]|\leqslant 2\mathbb{E}[|S_n^{(2)}|]\leqslant 2\sum_{k=1}^{n}\mathbb{E}[|X_k^{(2)}|]=2n\mathbb{E}[|X_1^{(2)}|].$$

因此令 $n\to+\infty$, 再令 $M\to+\infty$, 有

$$\mathbb{P}\left(\left|\frac{S_n}{n}-\mu\right|>\varepsilon\right)\to 0.$$

故 $\dfrac{S_n}{n}\xrightarrow{P}\mu$. □

注 1. 分别处理并运用 Markov 不等式等矩不等式. 因为 $X_n^{(1)}$ 有界, 因此其常用于高阶矩方法估计, 而尾巴部分 $X_n^{(2)}$ 通过 $\mathbb{E}[|X|]<\infty$ 用一阶矩来控制.

注 2. 也可以这样截尾: $Y_n = X_n I_{\{|X_n|\leqslant n\}}$. 由**例 1.3** 知 $\frac{1}{n}\sum_{k=1}^{n}(X_k - Y_k) \xrightarrow{\text{a.s.}} 0$. 因此

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\text{P}} \mathbb{E}[X] \Longleftrightarrow \frac{1}{n}\sum_{k=1}^{n} Y_k \xrightarrow{\text{P}} \mathbb{E}[X].$$

而

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[Y_k] = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[X_1 I_{\{|X_1|\leqslant k\}}] \to \mathbb{E}[X]\ (a_n \to a \Longrightarrow \frac{1}{n}\sum_{k=1}^{n} a_k \to a)$$

只需证 $\frac{1}{n}\sum_{k=1}^{n}(Y_k - \mathbb{E}[Y_k]) \xrightarrow{\text{P}} 0$, 由 CHebyshev 不等式, 只需证 $\frac{1}{n^2}\sum_{k=1}^{n}\mathbb{E}[Y_k^2] \xrightarrow{n\to+\infty} 0$, 即证

$$\sum_{i=1}^{\infty}\frac{1}{i^2}\mathbb{E}[Y_i^2] < \infty.$$

记 $B_{ij} = \{j-1 \le X_i < j\}$, 则

$$\begin{aligned}
\sum_{i}\frac{1}{i^2}\mathbb{E}[Y_i^2] &= \sum_{i=1}^{\infty}\frac{1}{i^2}\sum_{j=1}^{i}\mathbb{E}[Y_i^2 I_{B_{ij}}] \le \sum_{i=1}^{\infty}\frac{1}{i^2}\sum_{j=1}^{i} j^2\mathbb{P}(B_{ij}) \\
&= \sum_{i=1}^{\infty}\frac{1}{i^2}\sum_{j=1}^{i} j^2\mathbb{P}(B_{1j}) \xlongequal{\text{Fubini}} \sum_{j=1}^{\infty}\left(\sum_{i=j}^{\infty}\frac{1}{i^2}\right) j^2\mathbb{P}(B_{1j}) \\
&\le \sum_{j=1}^{\infty}\frac{2}{j}\cdot j^2\mathbb{P}(B_{1j}) = 2\sum_{j=1}^{\infty} j\mathbb{P}(B_{1j}) \\
&\le 2(1+\mathbb{E}[X_1]) < \infty
\end{aligned}$$

利用**注 2** 的证明思路, 我们还可以得到弱大数律最精确的结果, 因此弱大数律并不是真正“弱”爆了.

定理 1.7. 设 $X_1, \cdots, X_n$ 独立同分布, 且满足

$$x\mathbb{P}(|X_1| > x) \xrightarrow{x\to+\infty} 0.$$

记 $S_n = \sum_{k=1}^{n} X_k$, $\mu_n = \mathbb{E}[X_1 I_{\{|X_1|\leqslant n\}}]$ 则

$$\frac{S_n}{n} - \mu_n \xrightarrow{P} 0.$$

证明. 具体可以参考 **DurrettTheorem 2.2.12**. □

1.5 再谈 a.s. 收敛和强大数律

这节我们着重讲一讲证明 a.s. 收敛的处理技巧. 首先 Borel-Cantelli 引理是一个最基本也是必要的工具, 关键是怎么去使用? 这里时常要辅助**子序列**, **截尾**, 独立复制与对称化等等方法来处理. 另外, 如何估计 $\mathbb{P}(|X_n - X| > \varepsilon)$ 也是一门学问, 一般来说要借助各种矩不等式或其他方法及放缩技巧解决. Markov 不等式 (包括其推论和 Chebyshev 不等式) 是最常见的一种, 对于随机变量来说, 矩的阶数越高带来的信息越多 (高阶矩存在直接推出低阶矩存在), 所以对矩选取合适的阶数也是至关重要的. 本节主要从选 “矩” 的角度运用 Markov 不等式和子序列方法这两个层面进行讲解, 最后再重新回到上课已讲过的强大数律中.

子序列方法

1. 先找子列 $\{n_k\}$, 结合 Borel-Cantelli 引理, 满足

$$\sum_k \mathbb{P}(|X_{n_k}-X|>\varepsilon)<\infty, \forall\varepsilon>0 \implies X_{n_k}\xrightarrow{\text{a.s.}} X.$$

2. 对于一般项 n, 寻找子列中离其前后最近两的项消除其影响. 设 $n_k\leqslant n<n_{k+1}$, 下面两种是比较常见的处理方法:

   - $\sup\limits_{n_k\leqslant n<n_{k+1}}|X_n-X_{n_k}|\xrightarrow{\text{a.s.}}0.$
   - 如果这里 X_n 是 $\dfrac{S_n}{n}$ 型的, 则可以利用单调性控制 (前提 $S_{n_k}\leqslant S_n\leqslant S_{n_k+1}$):

   $$\frac{S_{n_k}}{n_{k+1}}\leqslant\frac{S_n}{n}\leqslant\frac{S_{n_k+1}}{n_k}.$$

   同时为了利用夹逼原理, 还要满足两侧均满足 a.s 收敛且收敛值相同.

   满足以上两条件之一, 再结合 $X_{n_k}\xrightarrow{\text{a.s.}}X$ 就能推出 $X_n\xrightarrow{\text{a.s.}}X$.

我们通过几个例子来展现子序列方法:

例 1.15. 随机变量 $X_1,\cdots,X_n$ 独立同分布, $\mathbb{P}(X_1=1)=\mathbb{P}(X_1=-1)=\dfrac{1}{2}$, 记 $S_n=\sum\limits_{k=1}^n X_k$, 证明:

$$\frac{S_n}{n}\xrightarrow{\text{a.s.}}0.$$

证明. 先证：$\dfrac{S_{n^2}}{n^2}\xrightarrow{\text{a.s.}}0$. 对 $\forall\varepsilon>0$, 利用 Markov 不等式, 有

$$\mathbb{P}\left(\left|\frac{S_{n^2}}{n^2}\right|>\varepsilon\right)\leqslant\frac{1}{\varepsilon^2n^4}\mathbb{E}[S_{n^2}^2]=\frac{1}{\varepsilon^2n^4}\sum_{k=1}^{n^2}\mathbb{E}[X_k^2]=\frac{1}{\varepsilon^2n^2}\implies\sum_{n=1}^{\infty}\mathbb{P}\left(\left|\frac{S_{n^2}}{n^2}\right|>\varepsilon\right)<+\infty.$$

由 Borel-Cantelli 引理知, $\dfrac{S_{n^2}}{n^2}\xrightarrow{\text{a.s.}}0$.
回到一般: 对 $\forall k\in\mathbb{N}^*$, $\exists n\in\mathbb{N}^*$, 使得 $n^2\leqslant k<(n+1)^2$. 而

$$\sup_{n^2\leqslant k<(n+1)^2}\frac{|S_k-S_{n^2}|}{n^2}=\sup_{n^2\leqslant k<(n+1)^2}\frac{\left|\sum\limits_{i=n^2+1}^{k}X_i\right|}{n^2}\leqslant\frac{(n+1)^2-n^2}{n^2}=\frac{2n+1}{n^2}\xrightarrow{\text{a.s.}}0(n\to+\infty)$$

因此

$$\sup_{n^2\leqslant k<(n+1)^2}\left|\frac{S_k}{k}\right|\leqslant\sup_{n^2\leqslant k<(n+1)^2}\frac{|S_k-S_{n^2}|}{n^2}+\frac{|S_{n^2}|}{n^2}\xrightarrow{\text{a.s.}}0(n\to+\infty)\implies\frac{S_k}{k}\xrightarrow{\text{a.s.}}0.$$

□

注. 这里只展现子序列的方法, 也可以通过更高阶矩的估计来完成.

例 1.16. 设零均值随机变量 $X_1,\cdots,X_n$ 两两不相关, $\sup\limits_n\mathbb{E}[X_n^2]\leqslant M<\infty$, 记 $S_n=\sum\limits_{k=1}^n X_k$, 定义 $D_n=\sup\limits_{n^2\leqslant k<(n+1)^2}|S_k-S_{n^2}|$, 证明:

(1) $\dfrac{D_n}{n^2}\xrightarrow{\text{a.s.}}0$;

(2) $\dfrac{S_n}{n}\xrightarrow{\text{a.s.}}0$.

证明. (1) 注意到

$$\mathbb{E}[D_n^2]=\mathbb{E}(\sup_{n^2\leqslant k<(n+1)^2}|S_k-S_{n^2}|)^2\leqslant\sum_{k=n^2+1}^{(n+1)^2}\mathbb{E}[|S_k-S_{n^2}|^2]\leqslant(2n+1)\mathbb{E}[|S_{(n+1)^2-1}-S_{n^2}|^2].$$

故对 $\forall\varepsilon>0$, 利用 Markov 不等式, 有

$$\mathbb{P}\left(\left|\frac{D_n}{n^2}\right|>\varepsilon\right)\leqslant\frac{\mathbb{E}[D_n^2]}{\varepsilon^2n^4}\leqslant\frac{2n+1}{\varepsilon^2n^4}\mathbb{E}[|S_{(n+1)^2-1}-S_{n^2}|^2]\leqslant\frac{(2n+1)^2M}{n^4\varepsilon^2}\leqslant\frac{C}{n^2}\xrightarrow{n\to+\infty}0.$$

其中 C 为某个正常数. 因此 $\sum_{n=1}^{+\infty}\mathbb{P}\left(\left|\frac{D_n}{n^2}\right|>\varepsilon\right)<+\infty$. 由 Borel-Cantelli 引理, $\frac{D_n}{n^2}\xrightarrow{\text{a.s.}}0$.

(2) 先证: $\frac{S_{n^2}}{n^2}\xrightarrow{\text{a.s.}}0$. 对 $\forall\varepsilon>0$, 利用 Markov 不等式, 有

$$\mathbb{P}\left(\left|\frac{S_{n^2}}{n^2}\right|>\varepsilon\right)\leqslant\frac{1}{\varepsilon^2n^4}\mathbb{E}[S_{n^2}^2]=\frac{1}{\varepsilon^2n^4}\sum_{k=1}^{n^2}\mathbb{E}[X_k^2]\leqslant\frac{M}{\varepsilon^2n^2}\implies\sum_{n=1}^{\infty}\mathbb{P}\left(\left|\frac{S_{n^2}}{n^2}\right|>\varepsilon\right)<+\infty.$$

由 Borel-Cantelli 引理, $\frac{S_n}{n^2}\xrightarrow{\text{a.s.}}0$.
回到一般: 对 $\forall k\in\mathbb{N}^*$, $\exists n\in\mathbb{N}^*$, 使得 $n^2\leqslant k<(n+1)^2$. 因此

$$\sup_{n^2\leqslant k<(n+1)^2}\left|\frac{S_k}{k}\right|\leqslant\frac{D_n}{n^2}+\frac{|S_{n^2}|}{n^2}\xrightarrow{\text{a.s.}}0(n\to+\infty)\implies\frac{S_k}{k}\xrightarrow{\text{a.s.}}0.$$

□

注. 一个致命的细节错误: $\mathbb{E}(\sup_{n^2\leqslant k<(n+1)^2}|S_k-S_{n^2}|)^2\leqslant\sup_{n^2\leqslant k<(n+1)^2}\mathbb{E}[|S_k-S_{n^2}|^2]$. 反例大家自行思考.

例 1.17. 设事件 $A_1,A_2\cdots$ 两两独立, 且 $\sum_{n=1}^{\infty}\mathbb{P}(A_n)=\infty$. 证明:

$$\frac{\sum_{k=1}^{n}I_{A_k}}{\sum_{k=1}^{n}\mathbb{P}(A_k)}\xrightarrow{\text{a.s.}}1.$$

证明. 记 $S_n=\sum_{k=1}^{n}I_{A_k},\mathbb{E}[S_n]=\sum_{k=1}^{n}\mathbb{P}(A_k),\mathbb{E}[S_n]\uparrow+\infty$. 则对 $\forall\varepsilon>0$, 利用 Chebyshev 不等式, 有

$$\mathbb{P}\left(\left|\frac{S_n}{\mathbb{E}[S_n]}-1\right|>\varepsilon\right)\leqslant\frac{\mathrm{Var}(S_n)}{\varepsilon^2(\mathbb{E}[S_n])^2}=\frac{\sum_{k=1}^{n}\mathrm{Var}(I_{A_k})}{\varepsilon^2(\mathbb{E}[S_n])^2}\leqslant\frac{\sum_{k=1}^{n}\mathbb{E}[I_{A_k}]}{\varepsilon^2(\mathbb{E}[S_n])^2}=\frac{1}{\varepsilon^2\mathbb{E}[S_n]}.$$

定义子列 $\{n_k\}$, 满足 $n_k=\inf\{n:\mathbb{E}[S_n]\geqslant k^2\}$, 利用 $k^2\leqslant\mathbb{E}[S_{n_k}]\leqslant\mathbb{E}[S_{n_k-1}+X_{n_k}]<k^2+1$, 则有

$$\sum_{k=1}^{+\infty}\mathbb{P}\left(\left|\frac{S_{n_k}}{\mathbb{E}[S_{n_k}]}-1\right|>\varepsilon\right)\leqslant\sum_{k=1}^{+\infty}\frac{1}{\varepsilon^2\mathbb{E}[S_{n_k}]}\leqslant\sum_{k=1}^{+\infty}\frac{1}{\varepsilon^2k^2}<+\infty.$$

由 Borel-Cantelli 引理知 $\dfrac{S_{n_k}}{\mathbb{E}[S_{n_k}]} \xrightarrow{\text{a.s.}} 1$.

回到一般: 对 $\forall n \in \mathbb{N}^*$, 存在 $k \in \mathbb{N}^*$, 使得 $n_k \leqslant n < n_{k+1}$. 注意到

$$\frac{S_{n_k}}{\mathbb{E}[S_{n_{k+1}}]} \leqslant \frac{S_n}{\mathbb{E}[S_n]} \leqslant \frac{S_{n_{k+1}}}{\mathbb{E}[S_{n_k}]}.$$

利用 $\dfrac{\mathbb{E}[S_{n_{k+1}}]}{\mathbb{E}[S_{n_k}]} \in \left[1, \dfrac{(k+1)^2+1}{k^2}\right]$ 知

$$\frac{S_{n_{k+1}}}{\mathbb{E}[S_{n_k}]} = \frac{S_{n_{k+1}}}{\mathbb{E}[S_{n_{k+1}}]} \cdot \frac{\mathbb{E}[S_{n_{k+1}}]}{\mathbb{E}[S_{n_k}]} \xrightarrow{\text{a.s.}} 1 (k \to +\infty), \qquad \frac{S_{n_k}}{\mathbb{E}[S_{n_{k+1}}]} \xrightarrow{\text{a.s.}} 1.$$

夹逼得 $\dfrac{S_n}{\mathbb{E}[S_n]} \xrightarrow{\text{a.s.}} 1$. □

选 "矩": 没有固定的套路, 但要学会大胆尝试!

考虑随机变量 $X_1, \cdots X_n$, 记 $S_n = \sum\limits_{k=1}^{n} X_k$, 一般可以先中心化处理故不妨设它们零均值. 对 $\delta > 0$, 往往题目待证 $\dfrac{S_n}{n^\delta} \xrightarrow{\text{P}} 0$ 或 $\dfrac{S_n}{n^\delta} \xrightarrow{\text{a.s.}} 0$, 进而会涉及到: 对 $\forall \varepsilon > 0$, $\mathbb{P}\left(\left|\dfrac{S_n}{n^\delta}\right| > \varepsilon\right)$ 的估计问题. 常见的一种处理方法就是利用题干条件已知的矩信息, 通过 Markov 不等式解决. 具体为:

$$\mathbb{P}\left(\left|\frac{S_n}{n^\delta}\right| > \varepsilon\right) \leqslant \frac{\mathbb{E}[|S_n|^r]}{\varepsilon^r n^{\delta r}}, r > 0\text{待定}$$

我们只需要找到一个满足要求的正数 r(一般选整数), 若证依概率收敛, 说明 $\dfrac{\mathbb{E}[|S_n|^r]}{\varepsilon^r n^{\delta r}} \xrightarrow{n \to +\infty} 0$ 即可; 若证几乎处处收敛, 利用 Borel-Cantelli 引理知, 只需说明 $\sum\limits_{n=1}^{+\infty} \dfrac{\mathbb{E}[|S_n|^r]}{\varepsilon^r n^{\delta r}} < +\infty$ 即可.

难点就在于如何利用题干已知的矩信息选取 r, 我们给一个例子让大家自行体会:

例 1.18. 随机变量 $X_1, \cdots, X_n$ 独立同分布, $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \dfrac{1}{2}$, 证明:

(1) $\dfrac{1}{n^2} \sum\limits_{k=1}^{n} k X_k \xrightarrow{\text{a.s.}} 0$;

(2) 对 $\forall \delta > 0$, 均有 $\dfrac{1}{n^{\frac{1}{2}+\delta}} \sum\limits_{k=1}^{n} X_k \xrightarrow{\text{a.s.}} 0$.

思路: (1) 记 $T_n = \sum\limits_{k=1}^{n} k X_k$, 待定 $r \in \mathbb{N}^*$, 考察: $\dfrac{\mathbb{E}[|T_n|^r]}{n^{2r}}$, 看 $\sum\limits_{n=1}^{+\infty} \dfrac{\mathbb{E}[|T_n|^r]}{n^{2r}}$ 是否收敛.

- $r = 1$, $\dfrac{\mathbb{E}[|T_n|]}{n^2} \leqslant \dfrac{1}{n^2} \sum\limits_{k=1}^{n} k \mathbb{E}[|X_k|] = \dfrac{n+1}{2n}$, 排除;
- $r = 2$, $\dfrac{\mathbb{E}[T_n^2]}{n^4} = \dfrac{1}{n^4}\left(\sum\limits_{k=1}^{n} k^2 \mathbb{E}[X_k^2] + \sum\limits_{1 \leqslant i \neq j \leqslant n} ij \mathbb{E}[X_i X_j]\right) = \dfrac{1}{n^4} \sum\limits_{k=1}^{n} k^2 = O(n^{-1})$, 排除;
- $r = 3$, $\dfrac{\mathbb{E}[|T_n|^3]}{n^6} \leqslant \dfrac{1}{n^6}\left(\sum\limits_{k=1}^{n} k^3 \mathbb{E}[|X_k|^3] + C_1 \sum\limits_{1 \leqslant i \neq j \leqslant n} i^2 j \mathbb{E}[|X_i|^2 |X_j|] + \sum\limits_{1 \leqslant i \neq j \neq k \leqslant n} ijk \mathbb{E}[|X_i||X_j||X_k|]\right) = \dfrac{1}{n^6}\left(\sum\limits_{k=1}^{n} k^3 + O(n^5) + O(n^6)\right) = O(1)$, 排除;

- $r=4,\ \dfrac{\mathbb{E}[T_n^4]}{n^8}=\dfrac{1}{n^8}\left(\sum_{k=1}^{n}k^4\mathbb{E}[X_k^4]+C_1\sum_{1\leqslant i\neq j\leqslant n}i^3j\mathbb{E}[X_i^3X_j]+C_2\sum_{1\leqslant i\neq j\leqslant n}i^2j^2\mathbb{E}[X_i^2X_j^2]\right)$
$$+\frac{1}{n^8}\left(C_3\sum_{1\leqslant i\neq j\neq k\leqslant n}i^2jk\mathbb{E}[X_i^2X_jX_k]+\sum_{1\leqslant i\neq j\neq k\neq l\leqslant n}ijkl\mathbb{E}[X_iX_jX_kX_l]\right)$$
$$=\frac{1}{n^8}\left(\sum_{k=1}^{n}k^4+C_2\sum_{1\leqslant i\neq j\leqslant n}i^2j^2\right)\leqslant\frac{1}{n^8}\left(\sum_{k=1}^{n}k^4+C_2\sum_{1\leqslant i\leqslant n}i^2\sum_{1\leqslant j\leqslant n}j^2\right)=O(n^{-2}),\text{ 成立.}$$

证明. (1) 记 $T_n=\sum_{k=1}^{n}kX_k$, 结合思路, 对 $\forall\varepsilon>0$, 利用 Markov 不等式, 我们有

$$\sum_{n=1}^{+\infty}\mathbb{P}\left(\left|\frac{T_n}{n^2}\right|>\varepsilon\right)\leqslant\sum_{n=1}^{+\infty}\frac{\mathbb{E}[T_n^4]}{\varepsilon^4n^8}\leqslant\sum_{n=1}^{+\infty}\frac{C}{\varepsilon^4n^2}<+\infty,$$

其中 C 为大于 0 的某个常数. 利用 Borel-Cantelli 引理即得 $\dfrac{T_n}{n^2}\xrightarrow{\text{a.s.}}0$.

(2) 记 $S_n=\sum_{k=1}^{n}X_k$, 利用**第 6 次习题课讲义定理 2.7**(矩方法, 用到组合计数的 idea), 有

$$\mathbb{E}\left[\left(\frac{S_n}{\sqrt{n}}\right)^k\right]\xrightarrow{n\to+\infty}\gamma_k,\qquad\gamma_{2m-1}=0,\qquad\gamma_{2m}=(2m-1)!!.$$

取正偶数 k, 使得 $\delta k>1$, 对 $\forall\varepsilon>0$, 当 n 充分大时, 有

$$\mathbb{P}\left(\left|\frac{S_n}{n^{\frac12+\delta}}\right|>\varepsilon\right)\leqslant\frac{1}{\varepsilon^kn^{\delta k}}\mathbb{E}\left[\left(\frac{S_n}{\sqrt{n}}\right)^k\right]\leqslant\frac{2\gamma_k}{\varepsilon^kn^{\delta k}}\implies\sum_{n=1}^{+\infty}\mathbb{P}\left(\left|\frac{S_n}{n^{\frac12+\delta}}\right|>\varepsilon\right)\leqslant\sum_{n=1}^{+\infty}\frac{2\gamma_k}{\varepsilon^kn^{\delta k}}<+\infty.$$

利用 Borel-Cantelli 引理即得 $\dfrac{S_n}{n^{\frac12+\delta}}\xrightarrow{\text{a.s.}}0$. □

思考: (1) 中阶数 n^{-2} 能否放得再更紧一些 (找到更精确的收敛速度)?

注. 高阶矩展开往往会带有 “组合计数” 的想法在里面, 另外注意偶阶矩和奇阶矩处理时微妙的区别.

最后回到强大数律:

定理 1.8 (强大数律). 设 $X_1,\cdots,X_n$ 独立同分布, $\mu=\mathbb{E}[X_1]$ 且 $\mathbb{E}[|X_1|]<+\infty$, 记 $S_n=\sum_{k=1}^{n}X_k$, 则

$$\frac{S_n}{n}\xrightarrow{\text{a.s.}}\mu.$$

回顾证明思路: 先设 $X_i\geqslant0$:

1. 截尾术: 引入 $Y_n=X_nI_{\{X_n<n\}}=\begin{cases}X_n & X_n<n\\ 0 & X_n\geq n\end{cases}$, 则

$$\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{\text{a.s.}}\mu\iff\frac{1}{n}\sum_{i=1}^{n}Y_i\xrightarrow{\text{a.s.}}\mu.$$

2. 子序列方法：找一个 a.s. 收敛的子列: 对 $\alpha > 1$, 令 $\beta_k = [\alpha^k]$, 则

$$\alpha^k \leq \beta_k < \alpha^k + 1, \quad \frac{\beta_{k+1}}{\beta_k} \to \alpha, k \to \infty$$

且

$$\frac{1}{\beta_n} \sum_{k=1}^{\beta_n} Y_k \xrightarrow{\text{a.s.}} \mu.$$

3. a.s. 收敛对一般 n 都成立 设 $S'_n = \sum_{k=1}^{n} Y_k$, 则 S'_n 关于 n 单调增, 取 $\beta_m \leq n < \beta_{m+1}$, 由

$$\frac{S'_{\beta_m}}{\beta_{m+1}} \leq \frac{S'_n}{n} \leq \frac{S'_{\beta_{m+1}}}{\beta_m}$$

从而

$$\frac{\beta_m}{\beta_{m+1}} \frac{S'_{\beta_m}}{\beta_m} \leq \frac{S'_n}{n} \leq \frac{S'_{\beta_{m+1}}}{\beta_{m+1}} \frac{\beta_{m+1}}{\beta_m}$$

所以

$$\frac{1}{\alpha}\mu \leq \liminf_{n\to\infty} \frac{S'_n}{n} \leq \limsup_{n\to\infty} \frac{S'_n}{n} \leq \alpha\mu, \text{a.s.}$$

令 $\alpha \downarrow 1$ 知,

$$\lim_{n\to\infty} \frac{S'_n}{n} = \mu, \text{a.s.}$$

最后, $X_i = X_i^+ - X_i^-$ 即可.

注. 实际上, 强大数律的证明就是这节讲过的各种技巧的组合: 截尾、子序列方法等.

2 作业

7.3.6 (Weierstrass 逼近定理) 设函数 $f : [0, 1] \to \mathbb{R}$, 随机变量 S_n 服从二项分布 $B(n, x)$, 证明:

$$\lim_{n\to+\infty} \sup_{0\leqslant x\leqslant 1} \left| f(x) - \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} \right| = 0.$$

证明. 引入随机变量 $Z_{n,x} = f(x) - f\left(\frac{S_n}{n}\right)$. 即证:

$$\sup_{0\leqslant x\leqslant 1} |\mathbb{E}[Z_{n,x}]| \xrightarrow{n\to+\infty} 0.$$

由题意知 f 在 $[0, 1]$ 上一致连续, 且存在 $M > 0$, 满足 $|f| \leqslant M$. 对 $\forall \varepsilon > 0, \exists \delta > 0$, 使得当 $|y - z| < \delta$ 且 $y, z \in [0, 1]$ 时, 有 $|f(y) - f(z)| < \varepsilon$. 引入事件 $A = \left\{ \left| \frac{S_n}{n} - x \right| > \delta \right\}$. 对 $\forall x \in [0, 1]$, 一方面,

$$|\mathbb{E}[Z_{n,x} I_{A^c}]| \leqslant \mathbb{E}[|Z_{n,x}| I_{A^c}] \leqslant \varepsilon.$$

另一方面, 利用 Chebyshev 不等式, 有

$$|\mathbb{E}[Z_{n,x} I_A]| \leqslant \mathbb{E}[|Z_{n,x}| I_A] \leqslant 2M\mathbb{P}(A) \leqslant 2M \cdot \frac{Var(S_n)}{n^2\delta^2} = \frac{2Mx(1-x)}{n\delta^2}.$$

因此

$$\sup_{0\leqslant x\leqslant 1} |\mathbb{E}[Z_{n,x}]| \leqslant \sup_{0\leqslant x\leqslant 1} (|\mathbb{E}[Z_{n,x} I_A]| + |\mathbb{E}[Z_{n,x} I_{A^c}]|) \leqslant \sup_{0\leqslant x\leqslant 1} \frac{2Mx(1-x)}{n\delta^2} + \varepsilon \leqslant \frac{M}{2n\delta^2} + \varepsilon \xrightarrow{n\to+\infty} \varepsilon.$$

最后令 $\varepsilon \to 0$ 即可. □

7.3.13 见本讲义**例 1.11**

3 补充习题

1 **Grimmett 7.3.8, 7.4.2, 7.11.18, 7.11.19, 7.11.22**

2 在函数空间 $L^1(\Omega,\mathcal{F},\mathbb{P})$ 上定义

$$d_1(X,Y)=\mathbb{E}[|X-Y|],\quad d_2(X,Y)=\mathbb{E}[|X-Y|\wedge 1],\quad d_3(X,Y)=\mathbb{E}\left[\frac{|X-Y|}{1+|X-Y|}\right]$$

(1) d_1,d_2,d_3 是 Ω 上的度量.

(2) $X_n \xrightarrow{1} X \Longleftrightarrow d_1(X_n,X)\to 0$.

(3) $X_n \xrightarrow{\mathrm{P}} X \Longleftrightarrow d_2(X_n,X)\to 0 \Longleftrightarrow d_3(X_n,X)\to 0$.

证明. (3) 第一个 $\Longleftrightarrow$ 引入示性函数拆分成两部分, 分别估计即可. 第二个 $\Longleftrightarrow$ 见 **Grimmett 7.11.10**. □

3 设对 $\forall i=1,2,\cdots,m$, 有 $X_{n,i}\xrightarrow{\mathrm{P}} X_i$, 函数 $g:\mathbb{R}^m\to\mathbb{R}$ 连续, 则

$$g(X_{n_1},\cdots,X_{n,m})\xrightarrow{\mathrm{P}} g(X_1,\cdots,X_m).$$

提示. 利用**命题 1.4**.

4 设 $\{X_n\}$ 为非负随机变量列, $\mathbb{E}[X_1]=\mu, S_n=\sum_{k=1}^{n}X_k$, 令 $N(t)=\max\{n:S_n\leqslant t\}$, 证明:

(1) $\mathbb{P}(\lim_{n\to+\infty}N(t)=\infty)=1$.

(2) $\dfrac{N(t)}{t}\xrightarrow{\text{a.s.}}\dfrac{1}{\mu}$.

证明. 见 **Grimmett7.5.2** . □

5 设 $\{X_n\}$ 独立同服从参数为 p 的 Bernoulli 分布, 令 $Y_k=X_kX_{k+1}, S_n=\sum_{k=1}^{n}Y_k$. 证明:

$$\frac{S_n}{n}\xrightarrow{\text{a.s.}} p^2.$$

提示. 先中心化:$Z_n=Y_n-p^2$, 然后类似**例 1.18**, 算到四阶矩后用 Markov 不等式再利用 Borel-Cantelli 引理即可.

4 利用 Kolmogorov 三级数定理证明强大数律 *

命题 4.1. 设 $\{X_n\}$ 独立, $\mathbb{E}[X_n]=0$, 且存在 $C>0$, 使得 $|X_n|\leqslant C$, 则

$$\sum_{n=1}^{\infty}\mathrm{Var}(X_n)<+\infty \Longleftrightarrow \sum_{n=1}^{\infty}X_n \text{ a.s. 收敛.}$$

证明. $\Longrightarrow$: 见**定理 1.5(一级数定理)**;

$\Longleftarrow$: 反证: 设 $\sum\limits_{n=1}^{\infty}\mathrm{Var}(X_n)=+\infty$, 利用**例 1.14** 有

$$\mathbb{P}(\max_{1\leqslant k\leqslant m}|X_{n+1}+\cdots+X_{n+k}|\geqslant\varepsilon)\geqslant 1-\frac{(\varepsilon+c)^2}{\sum\limits_{k=n+1}^{n+m}\mathbb{E}[X_k^2]}\xrightarrow{m\to+\infty}1.$$

因此

$$\mathbb{P}(\sup_{k\geqslant 1}|X_{n+1}+\cdots+X_{n+k}|\geqslant\varepsilon)=1 \implies \sum_{n=1}^{\infty}X_n \text{ a.s. 发散.}$$

矛盾. □

更一般地, 有

命题 4.2. 设 $\{X_n\}$ 独立, 且存在 $C>0$, 使得 $|X_n|\leqslant C$, 则

$$\begin{cases}\sum\limits_{n=1}^{\infty}\mathrm{Var}(X_n)<+\infty\\ \sum\limits_{n=1}^{\infty}\mathbb{E}[X_n]<\infty.\end{cases} \Longleftrightarrow \sum_{n=1}^{\infty}X_n \text{ a.s. 收敛.}$$

证明. $\Longrightarrow$: 利用**定理 1.5(一级数定理)** 即可;

$\Longleftarrow$: 设 $\left\{X_n'\right\}$ 是 $\{X_n\}$ 的独立复制, 则 $\widetilde{X_n}:=X_n-X_n'$ 是对称的随机变量, 且

$$\sum_{n=1}^{\infty}X_n \text{ a.s. 收敛} \Longleftrightarrow \sum_{n=1}^{\infty}X_n' \text{ a.s. 收敛.}$$

因此 $\sum\limits_{n=1}^{\infty}\widetilde{X_n}$ a.s. 收敛. 利用

$$\mathbb{E}[|X_n|]<\infty \implies \mathbb{E}[\widetilde{X_n}]=0,$$

由**命题 4.1** 得

$$\sum_{n=1}^{\infty}\mathrm{Var}(\widetilde{X_n})<+\infty.$$

同时,

$$\mathbb{E}[X_n^2]<\infty \implies \mathbb{E}[\widetilde{X_n}^2]=\mathrm{Var}(\widetilde{X_n})=2\mathrm{Var}(X_n) \implies \sum_{n=1}^{\infty}\mathrm{Var}(X_n)<+\infty.$$

利用**定理 1.5(一级数定理)**, 结合 $\sum\limits_{n=1}^{\infty}X_n$ a.s. 收敛, 得

$$\sum_{n=1}^{\infty}(X_n-\mathbb{E}[X_n]) \text{ a.s. 收敛} \implies \sum_{n=1}^{\infty}\mathbb{E}[X_n]<\infty.$$

□

由此我们得到三级数定理:

定理 4.1 (Kolmogorov 三级数定理)**.** 设 $\{X_n\}$ 为相互独立的随机变量, 则 $\sum\limits_{n=1}^{\infty} X_n$ a.s. 收敛的充分条件是: 存在常数 $c>0$, 使得

$$\begin{cases} 1^{\mathrm{o}}.\ \sum\limits_{n=1}^{\infty} \mathbb{P}(|X_n|>c)<\infty, \\ 2^{\mathrm{o}}.\ \sum\limits_{n=1}^{\infty} \mathbb{E}[X_n I_{\{|X_n|\leqslant c\}}]<\infty, \\ 3^{\mathrm{o}}.\ \sum\limits_{n=1}^{\infty} \mathrm{Var}(X_n I_{\{|X_n|\leqslant c\}})<\infty. \end{cases}$$

而必要条件是: 对 $\forall c>0$, 上述级数收敛.

证明. 充分性: 令 $Y_n = X_n I_{\{|X_n|\leqslant c\}}$, 则由 1^{o} 知

$$\sum_{n=1}^{+\infty} \mathbb{P}(X_n \neq Y_n) = \sum_{n=1}^{+\infty} \mathbb{P}(|X_n|>c)<\infty.$$

因此 $\sum\limits_{n=1}^{\infty} X_n$ 和 $\sum\limits_{n=1}^{\infty} Y_n$ 有相同的收敛性. 结合**命题 4.2**, 由 2^{o} 和 3^{o} 知 $\sum\limits_{n=1}^{\infty} Y_n$ a.s. 收敛. 所以 $\sum\limits_{n=1}^{\infty} X_n$ a.s. 收敛.

必要性: $\sum\limits_{n=1}^{\infty} X_n$ a.s. 收敛 $\implies X_n \xrightarrow{\text{a.s.}} 0 \iff \sum\limits_{n=1}^{+\infty} \mathbb{P}(|X_n|>c)<\infty$, 1^{o} 成立. 因此 $\sum\limits_{n=1}^{\infty} X_n$ 和 $\sum\limits_{n=1}^{\infty} Y_n$ 有相同的收敛性, 故 $\sum\limits_{n=1}^{\infty} Y_n$ a.s. 收敛. 结合**命题 4.2** 知 2^{o} 和 3^{o} 成立. □

三级数定理本身非常有用, 在此不作列举. 现在我们要通过三级数定理来证明强大数律. 在此之前先引入非常重要的一个引理:

引理 4.1 (Kronecker 引理)**.** 设 $\{a_n\}$ 为正实数序列且 $a_n \uparrow \infty$, $\{x_n\}$ 为实数列. 若 $\sum\limits_{n=1}^{\infty} \dfrac{x_n}{a_n}$ 收敛, 则

$$\lim_{n\to+\infty} \frac{1}{a_n}\sum_{k=1}^{n} x_n = 0.$$

证明. 令 $b_n = \sum\limits_{k=1}^{n} \dfrac{x_k}{a_k}$, 则 $b_n \xrightarrow{n\to+\infty} \sum\limits_{k=1}^{\infty} \dfrac{x_k}{a_k} := b$. 故

$$\frac{1}{a_n}\sum_{k=1}^{n} x_k = \frac{1}{a_n}\sum_{k=1}^{n} a_k(b_k - b_{k-1} = \frac{1}{a_n}\left(a_n b_n - \sum_{k=0}^{n-1} b_k(a_k - a_{k-1})\right) \xrightarrow{n\to+\infty} b - b = 0.$$

□

我们来看强大数律的证明:

定理 4.2 (强大数律)**.** 设 $X_1,\cdots,X_n,X$ 独立同分布, 且 $\mathbb{E}[|X|]<+\infty$, 则

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\text{a.s.}} \mathbb{E}[X].$$

证明. 不妨设 $\mathbb{E}[X]=0$, 要证 $\dfrac{\sum\limits_{k=1}^{n} X_k}{n} \xrightarrow{\text{a.s.}} 0$. 由**引理 4.1(Kronecker 引理)** 知只需证: $\sum\limits_{n=1}^{+\infty} \dfrac{X_n}{n}$ a.s. 收敛, 再由**定理 4.1(Kolmogorov 三级数定理)** 可知等价于证明

$$\begin{cases} 1^{\mathrm{o}}. \sum\limits_{n=1}^{\infty} \mathbb{P}(|X_n|>n)<\infty, \\ 2^{\mathrm{o}}. \sum\limits_{n=1}^{\infty} \mathbb{E}[\dfrac{X_n}{n} I_{\{|X_n|\leqslant n\}}]<\infty, \\ 3^{\mathrm{o}}. \sum\limits_{n=1}^{\infty} \operatorname{Var}\left(\dfrac{X_n}{n} I_{\{|X_n|\leqslant n\}}\right)<\infty. \end{cases}$$

1^{o} : $\sum\limits_{n=1}^{\infty} \mathbb{E}[I_{\{|X|>n\}}]=\mathbb{E}[\sum\limits_{n=1}^{\infty} I_{\{|X|>n\}}]=\mathbb{E}[\sum\limits_{n<|x|} 1] \leqslant \mathbb{E}[|X|]<\infty$.

3^{o} : $\sum\limits_{n=1}^{\infty} \operatorname{Var}\left(\dfrac{X_n}{n} I_{\{|X_n|\leqslant n\}}\right) \leqslant \sum\limits_{n=1}^{\infty} \mathbb{E}\left(\dfrac{X_n}{n} I_{\{|X_n|\leqslant n\}}\right)^2=\sum\limits_{n=1}^{\infty} \mathbb{E}[\dfrac{X^2}{n^2} I_{\{|X|\leqslant n\}}]=\mathbb{E}[\sum\limits_{n=1}^{\infty} \dfrac{X^2}{n^2} I_{\{|X|\leqslant n\}}]=\mathbb{E}[X^2 \sum\limits_{n\geqslant|X|} \dfrac{1}{n^2}] \leqslant \mathbb{E}[X^2 \cdot \dfrac{C}{|X|}] \leqslant C\mathbb{E}[|X|]<\infty$.

2^{o}: 令 $Y_n=X_n I_{\{|X_n|\leqslant n\}}$, 则

$$\sum_{n=1}^{\infty} \operatorname{Var}\left(\frac{Y_n}{n}\right)<\infty \Longrightarrow \sum_{n=1}^{\infty} \frac{Y_n-\mathbb{E}[Y_n]}{n} \text{ a.s. 收敛.}$$

再由**引理 4.1(Kronecker 引理)** 知

$$\frac{\sum\limits_{k=1}^{n}(Y_k-\mathbb{E}[Y_k])}{n} \xrightarrow{\text{a.s.}} 0$$

结合

$$\frac{\sum\limits_{k=1}^{n} \mathbb{E}[Y_k]}{n}=\frac{\sum\limits_{k=1}^{n} \mathbb{E}[X I_{\{|X|\leqslant k\}}]}{n} \xrightarrow{n\to+\infty} \mathbb{E}[X].$$

可知

$$\frac{1}{n} \sum_{k=1}^{n} Y_k \xrightarrow{\text{a.s.}} \mathbb{E}[X].$$

再由 1^{o} 知

$$\frac{1}{n} \sum_{k=1}^{n} X_k \xrightarrow{\text{a.s.}} \mathbb{E}[X].$$

□

5 一些可以忽略的定理证明 *

定理 5.1 (Skorokhod 表示定理)**.** 设 $X_n \xrightarrow{D} X$, 则存在 $(\Omega, \mathcal{F}, \mathbb{P})$, 其上 Y_n, Y 满足:

(1) Y_n 与 Y 同分布, Y 与 X 同分布

(2) $Y_n \xrightarrow{\text{a.s.}} Y$

证明. Fact: F 为分布函数, 定义“反函数”

$$F^{-1}(y) := \sup\{x : F(x) < y\}, y \in (0,1)$$

则 $F^{-1}(y) \leq x \Leftrightarrow y \leq F(x)$, 且当 U 为 $(0,1)$ 上均匀分布时, $F^{-1}(U)$ 的分布函数为 F.

下面 CHECK $F^{-1}(y) > x \Leftrightarrow y > F(x)$

$\Leftarrow$, 由 $y > F(x)$ 及右连续性可知, $\exists \delta > 0$,s.t. $y > F(x+\delta)$, 再由 $F^{-1}(y)$ 定义知: $x + \delta \leq F^{-1}(y)$, 从而 $x < F^{-1}(y)$.

$\Rightarrow$, 由定义知, $\exists x^* \in \{x : F(x) < y\}$, s.t. $x < x^*$, 从而 $F(x) \leq F(x^*) < y$.

令 $Y_n(u) = F_n^{-1}(u), Y(u) = F^{-1}(u), u \in (0,1)$, 则

$$u \leq F_n(x) \Leftrightarrow Y_n(u) \leq x \tag{1}$$

$$u \leq F(x) \Leftrightarrow Y(u) \leq x \tag{2}$$

进而 X_n 与 Y_n, X 与 Y 同分布.

Step1. $\forall \epsilon > 0, u \in (0,1), x \in \mathrm{C}_F$,s.t.

$$Y(u) - \epsilon < x < Y(u)$$

(2)表明 $F(x) < u$, 结合 $F_n(x) \to F(x)$, 可知对充分大 n, $F_n(x) < u$, 再由 (1)知, $Y_n(u) > x$, 从而

$$Y(u) - \epsilon < x < Y_n(u)$$

于是 $\liminf\limits_{n\to\infty} Y_n(u) \geq Y(u)$

Step2. $u < u' < 1, x \in \mathrm{C}_F$,s.t. Y 在 u 连续

$$Y(u') < x < Y(u') + \epsilon$$

由(2)知 $u < u' \leq F(x)$, 结合 $F_n(x) \to F(x)$, 可知对充分大 n,$F_n(x) > u$, 再由 (1)知, $Y_n(u) \leq x$, 进而

$$Y_n(u) \leq x < Y(u') + \epsilon$$

取上极限 $\limsup\limits_{n\to\infty} Y_n(u) \leq Y(u')$, 因 Y 单调增, 可取 $u' \in \mathrm{C}_F$ 且 $u' \downarrow u$, 于是 $\limsup\limits_{n\to\infty} Y_n(u) \leq Y(u)$, 从而 $\lim\limits_{n\to\infty} Y_n(u) = Y(u)$ 在所有的连续点处成立, 因为 Y 是单调增, 不连续点可数, 所以不连续点的概率测度为 0, 证毕. □

定理 5.2 (**Slutsky 引理**). $X_n \xrightarrow{\mathrm{D}} X, Y_n \xrightarrow{\mathrm{P}} b, Z_n \xrightarrow{\mathrm{P}} c$. 则

$$X_nY_n + Z_n \xrightarrow{\mathrm{D}} bX + c.$$

证明. $b = 0$ 留给读者思考, 以下设 $b > 0, Y_n \geqslant 0$. 对 $\forall 0 < \varepsilon < b$, 均有

$$\begin{aligned}\mathbb{P}(X_nY_n + Z_n \leqslant x) &\leqslant \mathbb{P}(|Y_n - b| > \varepsilon) + \mathbb{P}(|Z_n - c| > \varepsilon) + \mathbb{P}(X_nY_n + Z_n \leqslant x, |Y_n - b| \leqslant \varepsilon, |Z_n - c| \leqslant \varepsilon) \\ &\leqslant \mathbb{P}(|Y_n - b| > \varepsilon) + \mathbb{P}(|Z_n - c| > \varepsilon) + \mathbb{P}\left(X_n \leqslant \frac{x - c + \varepsilon}{b - \varepsilon}\right)\end{aligned}$$

$$\xrightarrow{n\to+\infty} \mathbb{P}\left(X \leqslant \frac{x-c+\varepsilon}{b-\varepsilon}\right).$$

取 $0<\varepsilon<b$ 使得 $\dfrac{x-c+\varepsilon}{b-\varepsilon} \in C_{F_x}$, 并令 $\varepsilon \downarrow 0$, 有

$$\limsup_{n\to\infty} \mathbb{P}(X_nY_n+Z_n \leqslant x) \leqslant \mathbb{P}\left(X \leqslant \frac{x-c}{b}\right) = \mathbb{P}(bX+c \leqslant x)$$

另一方面,

$$\begin{aligned}
\mathbb{P}(X_nY_n+Z_n \leqslant x) &\geqslant \mathbb{P}(X_nY_n+Z_n \leqslant x, |Y_n-b| \leqslant \varepsilon, |Z_n-c| \leqslant \varepsilon) \\
&\geqslant \mathbb{P}\left(X_n \leqslant \frac{x-c-\varepsilon}{b+\varepsilon}, |Y_n-b| \leqslant \varepsilon, |Z_n-c| \leqslant \varepsilon\right) \\
&= \mathbb{P}\left(X_n \leqslant \frac{x-c-\varepsilon}{b+\varepsilon}\right) - \mathbb{P}\left(|Y_n-b| > \varepsilon\right) - \mathbb{P}\left(|Z_n-c| > \varepsilon\right) \\
&\xrightarrow{n\to+\infty} \mathbb{P}\left(X \leqslant \frac{x-c-\varepsilon}{b+\varepsilon}\right).
\end{aligned}$$

取 $0<\varepsilon<b$ 使得 $\dfrac{x-c-\varepsilon}{b+\varepsilon} \in C_{F_x}$, 并令 $\varepsilon \downarrow 0$, 有

$$\liminf_{n\to\infty} \mathbb{P}(X_nY_n+Z_n \leqslant x) \geqslant \mathbb{P}\left(X < \frac{x-c}{b}\right) = \mathbb{P}(bX+c<x) \overset{x\in C_F}{=\!=\!=} \mathbb{P}(bX+c \leqslant x).$$

$\square$