概率论讨论班(第⑩次)
大数定律

桶寄001戴沄汐

2022 年 5 月 24 日

目录

1 简单的极限定理

1.1 服从弱大数定律的随机变量序列

首先要说明的是，所谓“大数定律”并不是某一条定律，而是一类问题，用于刻画一列随机变量的收敛情况。

本节介绍服从弱大数定律的随机变量序列。

定义 1.1. 如果随机变量序列$\{X_n, n \in \mathbb{N}\}$的部分和

$$S_n = \sum_{j=1}^{n} X_j$$

满足

$$\frac{S_n - \mathbb{E}(S_n)}{n} \xrightarrow{\mathbb{P}} 0 \tag{1}$$

我们就称弱大数定律对序列$\{X_n, n \in \mathbb{N}\}$成立

注 1.1. (1)式一个自然的推广如下：

$$\frac{S_n - a_n}{b_n} \xrightarrow{\mathbb{P}} 0$$

其中$\{a_n\}$是一个实数列，$\{b_n\}$是趋于无穷的一个正数列.

接下来我们将给出定理说明(1)式将会在何种情况下成立.在这之前，需要补充一个定义.

定义 1.2. *如果随机变量$X, Y \in L^2$(即具有有限的二阶矩),且*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) \tag{2}$$

*则称它们***不相关***.*

注 1.2. (2)式等价于

$$\mathbb{E}\{(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\} = 0,$$

即

$$\mathbf{Cov}(X, Y) = 0$$

这与不相关的另一种定义相符.

条件中的$X, Y \in L^2$是必要的，它保证了$\mathbb{E}(X)$与$\mathbb{E}(Y)$有限，同时还保证了$\mathbb{E}(XY)$有限(由Cauchy-schwartz不等式)，这使得(2)式是有意义的.

定理 1.1. *如果序列$\{X_n, n \in \mathbb{N}\}$不相关,且它们的二阶矩有共同的界，则(1)式在$L^2$中成立，因而也按概率成立.*

证明. 首先考虑到序列$\{X_n, n \in \mathbb{N}\}$与序列$\{X_n - \mathbb{E}(X_n), n \in \mathbb{N}\}$是一一对应的，故不失一般性，我们可以设对每一个$n$，$\mathbb{E}(X_n) = 0$.(否则考察$Y_n = X_n - \mathbb{E}(X_n)$就好了) 因此我们简化为证明

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} 0$$

考虑到在上一章中，我们学习了如果序列$\{X_n, n \in \mathbb{N}\}$是$L^r$收敛的，那么它必依概率收敛. 故考虑序列$Z_n = S_n/n$,那么$\mathbb{E}(Z_n^2) \to 0$即可得到$S_n/n \xrightarrow{\mathbb{P}} 0$.下面我们证明$\mathbb{E}(S_n^2) = o(n^2)$.

首先我们有

$$\begin{aligned}\mathbb{E}(S_n^2) &= \mathbb{E}((\sum_{j=1}^{n} X_j)^2) \\ &= \mathbb{E}(\sum_{j=1}^{n} X_j^2 + 2\sum_{1\leq j<k\leq n} X_j X_k) \\ &= \sum_{j=1}^{n}\mathbb{E}(X_j^2) + 2\sum_{1\leq j<k\leq n}\mathbb{E}(X_j X_k).\end{aligned} \tag{3}$$

注意到上式中有n^2项，故即使有$X_n \in L^2, n \in \mathbb{N}$,即其中的所有项都以一个固定的常数为上界，也仅能得出$\mathbb{E}(S_n^2) = O(n^2)$. 如果(3)中的"混合项"可以消去那就好了. 这是可以办到的，因为序列$\{X_n, n \in \mathbb{N}\}$不相关.

由不相关的定义，我们有$\mathbb{E}(X_j X_k) = \mathbb{E}(X_j)\mathbb{E}(X_k)$, 而由证明开头的假设，$\mathbb{E}(X_n) = 0, n \in \mathbb{N}$，即知$\mathbb{E}(X_j X_k) = 0$. □

1.2 服从强大数定律的随机变量序列

类似定义1.1，我们有

定义 1.3. *当(1)式是几乎处处收敛时，即*

$$\frac{S_n - \mathbb{E}(S_n)}{n} \xrightarrow{a.e.} 0 \tag{4}$$

我们就称强大数定律对序列$\{X_n, n \in \mathbb{N}\}$成立.

更进一步，可以通过证明几乎处处收敛加强定理1.1的结论. 下面这个定理将引入一个神奇的方法——取子序列证明的方法.

定理 1.2. *在定理1.1的同样假设下，(4)式也成立.*

证明. 不失一般性，我们仍然可以假设每一个n，$\mathbb{E}(X_n) = 0$. 设X_n的二阶矩的共同的界为M，由(3)式我们有

$$\mathbb{E}(S_n^2) \leq Mn,$$

由Chebyshev不等式，对任意的$\varepsilon > 0$，我们有

$$\mathbb{P}(|S_n| > n\varepsilon) \leq \frac{Mn}{n^2\varepsilon^2} = \frac{M}{n\varepsilon^2}.$$

如果我们对上式直接求和，则上式右边的级数是发散的. 不过如果我们考虑子列$\{n^2\}$，则

$$\sum_n \mathbb{P}(|S_{n^2}|> n^2\varepsilon) = \sum_n \frac{M}{n^2\varepsilon^2} < \infty.$$

从而由Borel-Cantelli引理，我们有

$$\mathbb{P}(|S_{n^2}|> n^2\varepsilon \quad i.o.) = 0. \tag{5}$$

又考虑到第四章的一个定理，上式等价于

$$\frac{S_{n^2}}{n^2} \to 0 \quad a.e. \tag{6}$$

于是我们就对一个子列证明了我们想要的结果;

现在我们需要证明S_k与最近的S_{n^2}之间的差别小得可以忽略.

对每个n，令

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

则有

$$\begin{aligned}
\mathbb{E}(D_n^2) &\le \sum_{k=n^2+1}^{(n+1)^2} \mathbb{E}(|S_k - S_{n^2}|^2) \\
&\le 2n\mathbb{E}(|S_{(n+1)^2} - S_{n^2}|^2) \\
&= 2n \sum_{i=n^2+1}^{(n+1)^2} \mathbb{E}(X_i^2) \\
&\le 4n^2 M
\end{aligned}$$

从而由Chebyshev不等式，

$$\mathbb{P}(|D_n|> n^2\varepsilon) \le \frac{4n^2M}{n^4\varepsilon^2} = \frac{4M}{n^2\varepsilon^2}.$$

完全类似上面的步骤，对上式求和后用Borel-Cantelli引理即得

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

又因为对$n^2 \le k < (n+1)^2$有

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

即对任意的n证明了结论. □

这两个定理说明了不相关且二阶矩有限的随机变量序列既服从弱大数定律也服从强大数定律. 实际上，对任意的随机变量序列$\{X_n, n \in \mathbb{N}\}$，如果$\mathbb{E}(X_n^2) \to 0$，则序列服从弱大数定律但未必满足强大数定律.(留作练习)

强大数定律应用的一个重要例子是通过所谓"正规数"来表述的.

定义 1.4. 将$[0,1]$中的每个实数展开为通常的十进制小数

$$\omega = 0.x_1x_2 \dots x_n \dots \tag{7}$$

规定有限小数的展开使用"有限"的表示形式. 对固定的$k : 0 \le k \le 9$，令v_k^n表示ω的前n位数字中等于k的数字个数. 考虑极限

$$\lim_{n\to\infty} \frac{v_k^n(\omega)}{n} = \varphi_k(\omega), \tag{8}$$

若该极限对每个k都存在且等于$1/10$，则ω称为**简单正规数**.

下面这个由Borel给出的定理说明了一个神奇的论断:几乎每一个数都是正规的.

定理 1.3. 除一个测度为零的*Borel*集外，$[0,1]$中的每一个数都是简单正规数.

证明. 考虑(7)式，显然序列$\{x_n, n \in \mathbb{N}\}$是独立随机变量的一个序列，且有

$$\mathbb{P}(x_n = k) = \frac{1}{10}, \quad k = 0, 1, \dots, 9$$

现对固定的k我们定义随机变量X_n为集合$\{\omega | x_n(\omega) = k\}$的示性函数，则$\mathbb{E}(X^2) = \mathbb{E}(X) = \frac{1}{10}$，且

$$S_n = \sum_{i=1}^{n} X_i(\omega)$$

即为定义1.4中的v_k^n. 显然$\{X_n\}$不相关且具有有限的二阶矩，故可使用定理1.2得到

$$\frac{S_n}{n} \xrightarrow{a.e.} \frac{1}{10}$$

沿用(8)中的记号，对于每个k我们得到了$\mathbb{P}(\varphi_k = \frac{1}{10}) = 1$.因而有

$$\mathbb{P}(\bigcap_{k=0}^{9}(\varphi_k = \frac{1}{10})) = 1$$

这意味着简单正规数集的测度为1. □

2 弱大数定律

2.1 Khinchin弱大数定律

上一节的形如(1)(4)的大数定律仅包含一阶矩，但我们的几个定理都对随机变量序列的二阶矩有所要求. 下面介绍一种不对二阶矩作任何要求的弱大数定律. 为此需要引进新的工具——由Khinchin引进的等价序列或由Lyapunov引进的特征函数.

定义 2.1. *设随机变量X的分布函数为$F(x)$，将*

$$f(t) = \int_{-\infty}^{\infty} e^{itx} \mathrm{d}F(x)$$

称为X的特征函数.

定义 2.2. *如果*

$$\sum_n \mathbb{P}(X_n \neq Y_n) < \infty$$

*就称随机变量序列$\{X_n\}$和$\{Y_n\}$**等价**.*

等价的随机变量具有一些优秀的性质，比如

命题 2.1. *如果$\{X_n\}$和$\{Y_n\}$等价，$a_n \uparrow \infty$，那么*

1. $\frac{1}{a_n} \sum\limits_{i=1}^{n} (X_i - Y_i) \to 0$;
2. *若*$\frac{1}{a_n} \sum\limits_{i=1}^{n} X_i$*依概率收敛于*$X$*，那么*$\frac{1}{a_n} \sum\limits_{i=1}^{n} Y_i$*也依概率收敛于*$X$.

证明. 仅提供思路. 1由Borel-Cantelli引理证明，2由1证明. □

有了工具，我们就可以着手了解本节开头提到的弱大数定律了.

定理 2.1. *设$\{X_n\}$是两两独立且同分布并具有有限均值m的随机变量序列，则我们有*

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} m \tag{9}$$

证明. 由$\mathbb{E}$的有界性和同分布,有

$$\sum_n \mathbb{P}(|X_n| > n) < \infty \tag{10}$$

我们引进一列随机变量$\{Y_n\}$满足

$$Y_n(\omega)=\begin{cases}X_n(\omega), & |X_n(\omega)|\leq n;\\ 0, & |X_n(\omega)|> n.\end{cases}$$

因为$\mathbb{P}(|X_n|> n)=\mathbb{P}(X_n\neq Y_n)$，故由(10)$\{X_n\}$和$\{Y_n\}$等价.这样"截断"的好处是，对每一个$n$，$Y_n$都是有限的，故$Y_n$有有限的二阶矩.

考虑随机变量$Z_n=Y_n-\mathbb{E}(Y_n)$，那么对每一个n，Z_n有有限的二阶矩且它们不相关。从而可以对Z_n使用定理5.1.1得到

$$\frac{\sum\limits_{i=1}^{n}(Y_i-\mathbb{E}(Y_i))}{n}\to 0(\text{按概率})$$

当$n\to\infty$时有$\mathbb{E}(Y_n)\to\mathbb{E}(X)=m$，因此有

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Y_i)\to m$$

结合命题2.1即得结论.

也可以用特征函数轻松地证明.

记$f_n(t)=\mathbb{E}exp\{it\frac{S_n-nm}{n}\}$，则有

$$f_n(t)=\prod_{k=1}^{n}\mathbb{E}exp\{it\frac{X_k-m}{n}\}=(\mathbb{E}exp\{it\frac{X_k-m}{n}\})^n$$

由于$\mathbb{E}(X-m)=0$，所以由特征函数在$t=0$处的taylor展开知对任何$t\in\mathbb{R}$，都有

$$\mathbb{E}exp\{it\frac{X_k-m}{n}\}=1+o(\frac{t}{n})$$

从而立得

$$f_n(t)=(1+o(\frac{t}{n}))^n\to 1,\quad n\to\infty$$

这等价于$\frac{S_n-nm}{n}\xrightarrow{\mathbb{P}}0$.(这一步的等价性实际上要借助依概率收敛于常数和依分布收敛的等价性和连续性定理得到.) □

2.2 完全独立随机变量的弱大数定律

对于完全独立的随机变量，其最一般形式的弱大数定律的充分必要条件是已知的.

定理 2.2. 设$\{X_n\}$是独立随机变量的一个序列，其分布函数为$\{F_n\}$，且$S_n = \sum_{i=1}^n X_i$.设$\{b_n\}$是单调递增趋于$+\infty$的一列给定的实数.下面*3*个式子

1.
$$\lim_{n\to\infty}\sum_{i=1}^n \mathbb{P}(|X_i|\geq b_n)=0;$$

2.
$$\lim_{n\to\infty}\frac{1}{b_n^2}\sum_{i=1}^n \boldsymbol{Var}X_iI_{\{|X_i|<b_n\}};$$

3.
$$a_n=\sum_{i=1}^n \mathbb{E}X_iI_{\{|X_i|<b_n\}}.$$

等价于
$$\frac{S_n-a_n}{b_n}\xrightarrow{\mathbb{P}}0. \tag{11}$$

证明. 只证充分性.对任给的$\varepsilon>0$，由Chebyshev不等式和2式知
$$\mathbb{P}(\frac{1}{b_n}|\sum_{i=1}^n X_iI_{\{|X_i|<b_n\}}-\mathbb{E}X_iI_{\{|X_i|<b_n\}}|\geq\varepsilon)\leq\frac{1}{\varepsilon^2b_n^2}\sum_{i=1}^n \mathbf{Var}X_iI_{\{|X_i|<b_n\}}\to 0.$$

此式和定理中第3式共同蕴含
$$\frac{1}{b_n}(\sum_{i=1}^n X_iI_{\{|X_i|<b_n\}}-a_n)\xrightarrow{\mathbb{P}}0. \tag{12}$$

与此同时，由定理中第1式，对任给的$\varepsilon \in (0,1)$，有

$$\begin{aligned}
&\mathbb{P}(\frac{1}{b_n}|S_n - \sum_{i=1}^{n} X_i I_{\{|X_i|<b_n\}}|\geq \varepsilon) \\
&= \mathbb{P}(\frac{1}{b_n}|\sum_{i=1}^{n} X_i I_{\{|X_i|\geq b_n\}}|\geq \varepsilon) \\
&\leq \mathbb{P}(|\sum_{i=1}^{n} I_{\{|X_i|\geq b_n\}}|\geq \varepsilon) \\
&\leq \mathbb{P}(\bigcup_{i=1}^{n}(|X_i|\geq b_n)) \\
&\leq \sum_{i=1}^{n} \mathbb{P}(|X_i|\geq b_n) \to 0.
\end{aligned}$$

这又说明了

$$\frac{1}{b_n}(S_n - \sum_{i=1}^{n} X_i I_{\{|X_i|<b_n\}}) \xrightarrow{\mathbb{P}} 0. \tag{13}$$

(12)(13)共同说明了充分性. 必要性的证明及其艰巨，在此略过. □

在本节结束之前，我们给出一个弱大数定律成立而强大数定律不成立的例子来作为定理2.2的应用.

例 1. 设$\{X_n\}$是具有共同分布函数F的随机变量，满足

$$\mathbb{P}(X_1 = n) = \mathbb{P}(X_1 = -n) = \frac{c}{n^2 \log n}, \quad n = 3, 4, \ldots$$

其中c为常数

$$\frac{1}{2}(\sum_{n=3}^{\infty} \frac{1}{n^2 \log n})^{-1}$$

那么$\{X_n\}$满足弱大数定律但不满足强大数定律.

证明. 由对称性，$\mathbb{E}X = 0$，故$a_n = 0$.

$$\sum_{i=1}^{n} \mathbb{P}(|X_i|\geq n) = n \sum_{k>n} \frac{c}{k^2 \log k} \sim \frac{c}{\log n}$$

$$\frac{1}{n^2} \sum_{i=1}^{n} \mathbf{Var} X_i I_{\{|X_i|<n\}} = \frac{1}{n^2} \cdot n \cdot \sum_{k=3}^{n} \frac{ck}{k \log k} \sim \frac{c}{\log n}$$

从而$\{X_n\}$满足定理2.2的前两式，所以有$S_n/n \xrightarrow{\mathbb{P}} 0$.

但另一方面，我们有

$$\mathbb{P}(|X_1|> n) \sim \frac{c}{n \log n},$$

又由同分布知

$$\sum_n \mathbb{P}(|X_n|> n) = \infty.$$

故由Borel-Cantelli引理知

$$\mathbb{P}(|X_n|> n \quad i.o.) = 1$$

而$|S_n - S_{n-1}|= X_n > n$蕴含$|S_n|> n/2$或$|S_{n-1}|> n/2$，故有

$$\mathbb{P}(|S_n|> \frac{n}{2} \quad i.o.) = 1$$

这说明S_n/n不可能几乎处处收敛于0. □

3 一个重要不等式

在开始学习强大数定律之前，我们先了解一个在强大数定律的研究中起着重要作用的不等式——Kolmogorov不等式.

定理 3.1 (Kolmogorov)**.** *设$\{X_n\}$是一列相互独立的随机变量，满足*

$$\mathbb{E}X_k = 0, \quad \mathbb{E}X_k^2 < \infty, \quad 1 \le k \le n.$$

则对任意的$\varepsilon > 0$，都有

$$\mathbb{P}(\max_{1\le k\le n} |S_k|> \varepsilon) \le \frac{\mathbb{E}(S_k^2)}{\varepsilon^2}.$$

证明. 记

$$\Lambda = (\max_{1\le k\le n} |S_k|> \varepsilon);$$
$$\Lambda_1 = (|S_1|> \varepsilon), \quad \Lambda_k = (\max_{1\le j<k} |S_j|< \varepsilon, |S_k|> \varepsilon)$$

显然Λ_k两两不交，且$\Lambda = \bigcup_{k=1}^n \Lambda_k$.因此我们还有$I(\Lambda) = \sum_{k=1}^n I(\Lambda_k)$和$\mathbb{P}(\Lambda) = \sum_{k=1}^n \mathbb{P}(\Lambda_k)$.所以

$$\mathbb{E}(S_n^2) \ge \mathbb{E}(S_n^2 I(\Lambda)) = \sum_{k=1}^n \mathbb{E}(S_n^2 I(\Lambda_k))$$

又由于$\{X_n\}$是相互独立的，所以$S_k I(\Lambda_k)$与$S_n - S_k$独立.从而

$$\begin{aligned}\mathbb{E}(S_n^2 I(\Lambda_k)) &= \mathbb{E}((S_n - S_k + S_k)^2 I(\Lambda_k)) \\ &= \mathbb{E}(S_k^2 I(\Lambda_k)) + \mathbb{E}((S_n - S_k)^2 I(\Lambda_k)) \\ &\geq \mathbb{E}(S_k^2 I(\Lambda_k)) \\ &\geq \varepsilon^2 \mathbb{P}(\Lambda_k).\end{aligned}$$

综合上式即得

$$\mathbb{E}(S_n^2) \geq \sum_{k=1}^{n} \mathbb{E}(S_n^2 I(\Lambda_k)) \geq \sum_{k=1}^{n} \varepsilon^2 \mathbb{P}(\Lambda_k) = \varepsilon^2 \mathbb{P}(\Lambda).$$

$\square$