第四章、概率极限定理

- 第九次

  §4.1 随机序列的收敛性

  §4.2 大数律和强大数律

- 第十次

  §4.1(续) 随机序列的收敛性

  §4.3 中心极限定理

§4.1 随机序列的收敛性

定义 (依概率收敛)

$\xi_n \xrightarrow{P} \eta$ 指$\forall \varepsilon > 0$, $P(|\xi_n - \eta| \geqslant \varepsilon) \to 0 (n \to \infty)$.

定义 (几乎必然收敛)

$\xi_n \xrightarrow{a.s.} \eta$ 指$P(\xi_n \to \eta) = 1$.

- $\xi_n \xrightarrow{\text{a.s.}} \eta \Leftrightarrow$存在$\Omega_0$ 使得$P(\Omega_0) = 0$, 且$\forall \omega \in \Omega \backslash \Omega_0$. $\xi_n(\omega) \to \eta(\omega)$,
- 定理1.1. 若$\xi_n \xrightarrow{\text{a.s.}} \eta$, 则$\xi_n \xrightarrow{P} \eta$.
  - 令$A_n = \{|\xi_n - \eta| \geqslant \varepsilon\}$.
    $P(A_n) \to 0$ vs $P(\cup_{m \geqslant n} A_m) \to 0$.
- 自习例1.1

§4.2 大数律和强大数律

假设$X_1, X_2, \cdots X_n, \cdots$ 是随机变量序列, 期望都存在. 令

$$S_n = X_1 + \cdots + X_n.$$

- Weak Law of Large Numbers (WLLN):
  若$\frac{1}{n}(S_n - ES_n) \xrightarrow{P} 0$, 则称$X_1, X_2, \cdots$ 满足WLLN.
- Strong Law of Large Numbers (SLLN):
  若$\frac{1}{n}(S_n - ES_n) \xrightarrow{\text{a.s.}} 0$, 则称$X_1, X_2, \cdots$ 满足SLLN.
- i.i.d.情形: $\frac{1}{n}S_n \to EX_1 = \mu$.
- 一般情形: $\frac{1}{b_n}(S_n - a_n) \to 0$, 理解为$S_n = a_n + o(b_n)$.

弱/强大数律的证明工具:

定理 (第二章、定理6.4, 定理7.1)

Markov's inequality:

假设$X \geqslant 0$, 期望存在. 则$P(X \geqslant C) \leqslant \frac{1}{C} EX$, $\forall C > 0$.

Chebyshev's inequality:

假设二阶矩存在. 则$P(|X - EX| \geqslant C) \leqslant \frac{1}{C^2}\mathrm{var}(X)$, $\forall C > 0$.

- $A = \{X \geqslant C\}$, $P(A) = E1_A$.
- $1_A \leqslant Y$, Y 满足: (a) $Y \geqslant 0$, (b) $Y|_A \geqslant 1$. 则$P(A) \leqslant EY$.
- 取$Y = \frac{X}{C}$, 或$\frac{(X-EX)^2}{C^2}$.
- 取$Y = \frac{(X-EX)^{2k}}{C^{2k}}$. $P(|X - EX| \geqslant C) \leqslant \frac{1}{C^{2k}} E(X - EX)^{2k}$.
- 取$Y = e^{a(X-C)}$, 其中$a > 0$. 则$P(X \geqslant C) \leqslant Ee^{a(X-C)}$.

定理 (Chebyshev's WLLN, 定理2.1)

假设$X_1, X_2, \cdots$ 相互独立，二阶矩都存在，且$\mathrm{var}(X_i) \leqslant C$, $\forall i$.
那么，$\frac{1}{n}(S_n - ES_n) \xrightarrow{P} 0$.

- 令$A_n = \{|\frac{1}{n}(S_n - ES_n)| \geqslant \varepsilon\}$. 需证明$P(A_n) \to 0$.
- 估计$P(A_n)$. 由切比雪夫不等式，
  $P(A_n) = P(|S_n - ES_n| \geqslant n\varepsilon) \leqslant \frac{1}{(n\varepsilon)^2}\mathrm{var}(S_n) \leqslant \frac{nC}{n^2\varepsilon^2}$,
  $P(A_n) = \frac{C}{\varepsilon^2} \cdot \frac{1}{n} \to 0$.
- “相互独立”可减弱为“两两不相关”.
- 可不妨设$EX_i = 0$, 否则用$Y_i := X_i - EX_i$ 代替X_i.

例1: 假设有n 种券, 集齐时间为S_n, 则$\frac{S_n}{n\log n} \xrightarrow{P} 1$.

- $S_n = X_1 + \cdots + X_n$.
- 假设X服从参数为p的几何分布，则$EX = \frac{1}{p}$, $\mathrm{var}(X) = \frac{1-p}{p^2} \leqslant \frac{1}{p^2}$.
- X_k服从参数为p_k的几何分布, $p_k = \frac{n-(k-1)}{n} = \frac{\ell}{n}$.
- $ES_n = n\sum_{\ell=1}^{n} \frac{1}{\ell} \approx n\log n$, $\mathrm{var}S_n \leqslant n^2 \sum_{\ell=1}^{n} \frac{1}{\ell^2} \leqslant Cn^2$.
- $P(|\frac{S_n - ES_n}{n\log n}| > \varepsilon) \leqslant \frac{\mathrm{var}(S_n)}{\varepsilon^2 n^2 (\log n)^2} \to 0$.
- 所以$\frac{S_n - ES_n}{n\log n} \xrightarrow{P} 0$, $\frac{S_n}{n\log n} \xrightarrow{P} 1$.
- 集齐一半的时间$T_n = X_1 + \cdots + X_{n/2}$.
  $ET_n \approx n\sum_{\ell=n/2}^{n} \frac{1}{\ell} \approx n\log 2$, $\mathrm{var}T_n \leqslant n^2 \sum_{\ell=n/2}^{n} \frac{1}{\ell^2} \leqslant \delta_n n^2$.
  $P(|\frac{T_n - ET_n}{n\log 2}| > \varepsilon) \leqslant \frac{\delta_n n^2}{\varepsilon n^2 (\log 2)^2} \to 0$, $\frac{T_n}{n\log 2} \xrightarrow{P} 1$.

定理 (Cantelli's SLLN, 定理2.2)

假设$X_1, X_2, \cdots$ 相互独立，四阶矩都存在，且$E(X_i - EX_i)^4 \leqslant C$, $\forall i$. 那么，$\frac{1}{n}(S_n - ES_n) \overset{a.s.}{\to} 0$.

- 不妨设$EX_i = 0$. $A_n = \{|\frac{1}{n}S_n| \geqslant \varepsilon\}$. 往证$P(\cup_{m\geqslant n}A_m) \to 0$.
- 估计$P(A_m)$, 利用$P(\cup_{m\geqslant n}A_m) \leqslant \sum_{m\geqslant n} P(A_m)$.
- $P(A_m) = P(S_m^4 \geqslant (m\varepsilon)^4) \leqslant \frac{1}{m^4\varepsilon^4}ES_m^4$.
- $ES_m^4 = \sum_{i,j,k,\ell=1}^m EX_iX_jX_kX_\ell$
  EX_r^4, $EX_r^3X_s$, $EX_r^2X_s^2$, $EX_r^2X_sX_t$, $EX_rX_sX_tX_u$.
- $ES_m^4 \leqslant mC + C_m^2C_4^2C \leqslant 4Cm^2$,
  $(EX_r^2X_s^2 \leqslant \sqrt{EX_r^4}\sqrt{EX_s^4}, \quad E(Y + xZ)^2 \geqslant 0)$.
- $P(A_m) \leqslant \frac{4C}{\varepsilon^4} \times \frac{1}{m^2}$, 从而$\sum_{m\geqslant n} P(A_m) \to 0$.

- Borel's SLLN(推论2.4). 单次小试验中事件A 发生的概率为p. 在独立重复小试验的大试验中, 前n 次小试验中A 发生的频率为$Y_n = \frac{1}{n}S_n$. 则$Y_n \xrightarrow{\text{a.s.}} p$.
- Kolmogorov's SLLN(定理2.4，证明不要求). 假设$X_1, X_2, \cdots$ 独立同分布.
  若$\mu = EX_1$ 存在(且有限), 则$\frac{1}{n}S_n \xrightarrow{\text{a.s.}} \mu$.
- 时间平均= 空间平均, (期望的含义).
- 反之亦然!
- $\mu = \pm\infty$ 亦然!
- 自习推论2.1, 2.2, 2.3, 2.4, 应用(1)~(3).

$\mu = +\infty$ 情形:

$EX^+ = \infty$, $EX^- < \infty$, 则$S_n/n \to \infty$, a.s..

- 不妨设$X \geqslant 0$ $(\because X = X^+ - X^-)$.
- $S_{n,M} = X_1 \wedge M + \cdots + X_n \wedge M$, $\forall M > 0$.
  由SLLN, $\frac{S_{n,M}}{n} \to E(X \wedge M)$.
- $S_n \geqslant S_{n,M}$, 所以$\liminf_n \frac{S_n}{n} \geqslant E(X \wedge M)$, $\forall M > 0$.
- $E(X \wedge M) = \int_0^\infty P(X \wedge M > x)dx = \int_0^M P(X > x)dx \to \infty$.

LLN的应用： 统计、随机模拟等的基础.

例2 (经验分布函数):

$F_n(x,\omega) = \frac{1}{n}\sum_{k=1}^{n} 1_{(-\infty,x]}(X_k(\omega)) \stackrel{\text{a.s.}}{\to} F(x)$.

例3 (Monte Carlo Method). 近似计算

$\int_D g(x_1,x_2,\cdots,x_d)dx_1\cdots dx_d$.

- 取一个矩形A使得$D\subset A$.
- ξ_j: 第j次从A中任取一点(均匀), 取到的点的位置.

$$1_D(\xi_j) = \begin{cases} 1, & \xi_j \in D, \\ 0, & \xi_j \notin D. \end{cases}$$

- $\frac{1}{n}\sum_{j=1}^{n} g(\xi_j)1_D(\xi_j) \stackrel{\text{a.s.}}{\to} Eg(\xi_1)1_D(\xi_1)$.
  $\because Eg(\xi_1)1_D(\xi_1) = \frac{1}{m(A)}\int_D g(x_1,x_2,\cdots,x_d)dx_1\cdots dx_d$.
  $\therefore \int_D g(x_1,x_2,\cdots,x_d)dx_1\cdots dx_d \approx \frac{m(A)}{n}\sum_{j=1}^{n} g(\xi_j)1_D(\xi_j)$.

例4. 更新定理 假设灯泡寿命$X_1, X_2, \cdots$ 独立同分布, 严格正, $EX_1 = \mu$. 则换灯泡的"频率"为$\frac{1}{\mu}$ (若t 时间内需要换N_t 个灯泡, 则$N_t/t \to \frac{1}{\mu}$).

- 第n 次换灯泡时间为$S_n = X_1 + \cdots + X_n$.
- 按N_t 的定义, $S_{N_t} \leqslant t < S_{N_t+1}$, 且$N_t \to \infty$.
- 存在$P(\Omega_0) = 0$ 使得$\forall \omega \in \Omega \backslash \Omega_0$, $\lim_{n\to\infty} \frac{1}{n} S_n(\omega) = \mu$, $\lim_t N_t(\omega) = \infty$, 从而

$$\frac{S_{N_t(\omega)}(\omega)}{N_t(\omega)}, \quad \frac{S_{N_t(\omega)+1}(\omega)}{N_t(\omega)} \xrightarrow{t\to\infty} \mu.$$

- $\frac{t}{N_t(\omega)} \to \mu$, 从而$\frac{N(t)}{t} \to \frac{1}{\mu}$.