AI 中的数学

第十五讲

方聪，概率统计部分参考章复熹和张原老师课件

2024 年秋季

1 随机序列的收敛性

2 大数定律与中心极限定理

§4.1 随机序列的收敛性

- 定义 1.1. 若

$$\lim_{n\to\infty} P\left(|\xi_n-\eta|\geqslant\varepsilon\right)=0,\quad \forall\varepsilon>0,$$

则称 ξ_n 依概率收敛于 η, 记为 $\xi_n\xrightarrow{P}\eta$.

- 定义 1.2. 若

$$P\left(\lim_{n\to\infty}\xi_n=\eta\right)=1$$

则称 ξ_n 几乎必然收敛于 η, 记为 $\xi_n\xrightarrow{\text{a.s.}}\eta$.

- 定义 1.3. 若

$$\lim_{n\to\infty} F_{\xi_n}(x)=F_\eta(x),\quad \forall x\in C\left(F_\eta\right),$$

则称 ξ_n 依分布收敛于 η, 记为 $\xi_n\xrightarrow{d}\eta$.

- 定理 1.1. 若 $\xi_n \xrightarrow{\text{a.s.}} \eta$, 则 $\xi_n \xrightarrow{P} \eta$.
- 例 1.1 表明 $\xi_n \xrightarrow{P} \eta$ 不能推出 $\xi_n \xrightarrow{\text{a.s.}} \eta$.
- 定理 1.2. 若 $\xi_n \xrightarrow{P} \eta$, 则 $\xi_n \xrightarrow{d} \eta$.
- 例 1.2 表明 $\xi_n \xrightarrow{d} \eta$ 不能推出 $\xi_n \xrightarrow{P} \eta$.

- 几乎必然收敛，指不收敛的点 ω 是空集 (概率为 0)
- 依概率收敛要求随机变量不同的概率越来越小
- 弱收敛只关注分布与 ω 无关

例（不做要求）：对于 $(0,1)$ 上均匀分布，考虑下列随机变量序列：对任何正整数 k 及 $j=1,\cdots,2^k$，令

$$X_{k1}=\begin{cases}1, & 0<\omega<\frac{1}{2^k},\\ 0, & \text{其他};\end{cases}\quad,\quad X_{kj}=\begin{cases}1, & \frac{j-1}{2^k}<\omega<\frac{j}{2^k},\\ 0, & \text{其他};\end{cases}\quad (j>1).$$

这些 $X_{kj}: k\geqslant 1, j=1,\cdots,2^k\}$ 可排成一个序列：
$X_{11},X_{12},X_{21},X_{22},X_{23},X_{24},\cdots$（按照字典排列法，将第一个足标从小到大排，若相同则按第二个足标从小到大排），将该序列记为 $\xi_1,\xi_2,\cdots$，其中 $\xi_n=X_{k_n j_n}$，则对任何 $\varepsilon\in(0,1)$ 有：

$$P(|\xi_n|\geqslant\varepsilon)=P(\xi_n=1)=\frac{1}{2^k_n}$$

在 $n\to\infty$ 时 $k_n\to\infty$，故有 $\lim_{n\to\infty}P(|\xi_n|\geqslant\varepsilon)=0$，这表明 $\xi_n\xrightarrow{P}0$.
而对于任何 $\omega\in(0,1)$，$\lim_{n\to\infty}\xi_n(\omega)$ 不存在。$\xi_n\xrightarrow{a.s.}\eta$ 不成立。

例：设随机变量 $X \sim N(0,1)$，令

$$\xi_{2n-1} = X, \quad \xi_{2n} = X \quad (n = 1, 2, \cdots).$$

易知所有的 ξ_n 有相同的分布函数 $\phi(x)$，为标准正态分布函数，显然 $\xi_n \xrightarrow{\omega} \eta$. 但是对 $\varepsilon > 0$，有

$$P(|\xi_n - X| \geqslant \varepsilon) = \begin{cases} 0, & n\text{是奇数}, \\ P(|X| \geqslant \frac{\varepsilon}{2}), & n\text{是偶数}. \end{cases}$$

可见 $\xi_1, \xi_2, \cdots$ 并不依概率收敛于 X。

- 假设 $X_1, X_2, \cdots$ 是随机变量序列，令

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

- 定义 1.4. 若 $EX_n, n = 1, 2, \cdots$ 都存在，且

$$\frac{1}{n}(S_n - ES_n) \xrightarrow{P} 0$$

则称 $X_1, X_2, \cdots$ 服从 (弱) 大数律 (Weak Law of Large Numbers, WLLN)

- 定义 1.5. 若 $EX_n, n = 1, 2, \cdots$ 都存在，且

$$\frac{1}{n}(S_n - ES_n) \xrightarrow{\text{a.s.}} 0,$$

则称 $X_1, X_2, \cdots$ 服从强大数律 (SLLN).

定义：若对任意 $n \geqslant 2$ 都有 $X_1, \cdots, X_n$ 相互独立，则称 $X_1, X_2, \cdots$ 是相互独立的随机变量序列。

若 $X_1, X_2, \cdots$ 相互独立，且 $X_n \overset{d}{=} X_1, \forall n \geqslant 2$，则称 $X_1, X_2, \cdots$ 独立同分布，记为 i.i.d. (independent and identically distributed).

例：设随机变量序列 $\{X_n\}$ 独立同分布，其密度函数为

$$p(x)=\begin{cases}\frac{1}{\beta}, & 0<x<\beta,\\ 0, & \text{其他}.\end{cases}$$

其中常数 $\beta>0$，令 $Y_n=\max\{X_1,X_2,\cdots,X_n\}$，证明 $Y_n \xrightarrow{P} \beta$.

例：设随机变量序列 $\{X_n\}$ 独立同分布，其密度函数为

$$p(x)=\begin{cases}\frac{1}{\beta}, & 0<x<\beta,\\ 0, & \text{其他}.\end{cases}$$

其中常数 $\beta>0$，令 $Y_n=\max\{X_1,X_2,\cdots,X_n\}$，证明 $Y_n \xrightarrow{P} \beta$.

证明：因为当 $x<0$ 时，有 $P(Y_n\leqslant x)=0$，当 $x\geqslant\beta$ 时，有 $P(Y_n\leqslant x)=1$，当 $0\leqslant x<\beta$ 时，有

$$P(Y_n\leqslant x)=\prod_{i=1}^{n}P(X_i\leqslant x)=\prod_{i=1}^{n}\int_0^x\frac{1}{\beta}dx=\left(\frac{x}{\beta}\right)^n,$$

所以对任意的 $\varepsilon>0(\varepsilon<\beta)$，当 $n\to\infty$ 时，有

$$P(|Y_n-\beta|\geqslant\varepsilon)=P(Y_n\leqslant\beta-\varepsilon)=\left(\frac{\beta-\varepsilon}{\beta}\right)^n\to 0,$$

所以有 $Y_n\xrightarrow{P}\beta$.

1 随机序列的收敛性

2 大数定律与中心极限定理

定理

Chebyshev's WLLN, 定理 2.1 假设 $X_1, X_2, \cdots$ 相互独立，且 $\operatorname{var}(X_i) \leqslant M, \forall i$. 那么，

$$\frac{1}{n}(S_n - ES_n) \xrightarrow{P} 0.$$

- 令 $A_n = \left\{\left|\frac{1}{n}(S_n - ES_n)\right| \geqslant \varepsilon\right\}$. 需验证 $P(A_n) \to 0$.
- 由切比雪夫不等式，

$$P(A_n) = P(|S_n - ES_n| \geqslant n\varepsilon) \leqslant \frac{1}{(n\varepsilon)^2}\operatorname{var}(S_n)$$
$$\leqslant \frac{nM}{n^2\varepsilon^2} = \frac{M}{\varepsilon^2} \cdot \frac{1}{n} \to 0.$$

- “相互独立”可减弱为“两两不相关”.

推论：设 $X_1, X_2, \cdots$ 独立同分布, $\mathrm{var}\left(X_1\right) < \infty$, 则当 $n \to \infty$ 时，

$$\frac{S_n}{n} \xrightarrow{P} E(X_1).$$

推论：（伯努利大数律）单次试验中 A 发生的概率为 p, 设在 n 次试验中事件 A 发生了 ν_n 次，则当 $n \to \infty$ 时，

$$\frac{\nu_n}{n} \xrightarrow{P} p.$$

推论：设 $X_1, X_2, \cdots$ 独立同分布, $\mathrm{var}(X_1) < \infty$, 则当 $n \to \infty$ 时，

$$\frac{S_n}{n} \xrightarrow{P} E(X_1).$$

推论：（伯努利大数律）单次试验中 A 发生的概率为 p, 设在 n 次试验中事件 A 发生了 ν_n 次，则当 $n \to \infty$ 时，

$$\frac{\nu_n}{n} \xrightarrow{P} p.$$

证明：令

$$X_i = \begin{cases} 1, & \text{第}i\text{次试验中}A\text{发生}, \\ 0, & \text{第}i\text{次试验中}A\text{不发生}, \end{cases} \quad (i = 1, 2, \cdots),$$

则 $\frac{\nu_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 由于 $X_1, X_2, \cdots$ 是独立同分布的随机变量序列，$E(X_i) = p$，$\mathrm{var}(X_i) = p(1-p)(i = 1, 2, \cdots)$，故由上一推论知本推论成立。

如果不假定 $E(X_1)$ 存在，上述推论是否成立？

例：设 $X_1, X_2, \cdots$ 独立同分布，密度为 $p(x) = \frac{1}{\pi(x^2+1)}$. 可以证明，$\frac{S_n}{n}$ 与 X_1 有相同的密度. 于是，对任何 a 和 $\varepsilon > 0$，有

$$P\left(\left|\frac{S_n}{n} - a\right| > \varepsilon\right) = P\left(|X_1 - a| > \varepsilon\right) \text{ 不趋于}0.$$

故 $\frac{S_n}{n}$ 不能以概率收敛于 a。

定理

Cantelli's SLLN, 定理 *2.2*, 引理 *2.1* 假设 $X_1, X_2, \cdots$ 相互独立, EX_i 存在, 且 $E\left(X_i - EX_i\right)^4 \leqslant M, \forall i$. 那么,

$$\frac{1}{n}\left(S_n - ES_n\right) \xrightarrow{a.s.} 0.$$

- 推论 2.3. 设 $X_1, X_2, \cdots$ 独立同分布, EX_1^4 存在, 则 $\frac{S_n}{n} \xrightarrow{\text{a.s.}} EX_1$.
- 推论 2.4. 单次小试验中事件 A 发生的概率为 p. 在独立重复试验中, 前 n 次试验中 A 发生的频率 $\xrightarrow{\text{a.s.}} p$.
- 定理 2.4. (Kolmogorov's SLLN). 假设 $X_1, X_2, \cdots$ 独立同分布, 期望存在, 则 $\frac{1}{n}S_n \xrightarrow{\text{a.s.}} EX_1$.
- 时间平均 $=$ 空间平均, (期望的含义).

例：设 $X_1, X_2, \cdots$ 是相互独立的随机变量序列，$X_1 \equiv 0$，对一切 $n \geqslant 2$，X_n 只取三个可能值 $n, -n, 0$，且

$$P(X_n) = n = P(X_n = -n) = \frac{1}{2n \ln n}, \quad P(X_n = 0) = 1 - \frac{1}{n \ln n}.$$

证明 $X_1, X_2, \cdots$ 服从切比雪夫大数定律。

例：设 $X_1, X_2, \cdots$ 是相互独立的随机变量序列，$X_1 \equiv 0$，对一切 $n \geqslant 2$，X_n 只取三个可能值 $n, -n, 0$，且

$$P(X_n) = n = P(X_n = -n) = \frac{1}{2n\ln n}, \quad P(X_n = 0) = 1 - \frac{1}{n\ln n}.$$

证明 $X_1, X_2, \cdots$ 服从切比雪夫大数定律。

证明：

$$E(X_n) = 0, \quad \operatorname{var}(X_1) = 0, \quad \operatorname{var}(X_n) = \frac{n}{\ln n}(n = 2, 3, \cdots).$$

令 $S_n = \sum_{i=1}^n X_i$，则 $\operatorname{var}(S_n) = \sum_{i=1}^n \operatorname{var}(X_i) = \sum_{i=2}^n \frac{i}{\ln i}$. 由于 $x \geqslant 3$ 时 $\frac{x}{\ln x}$ 是 x 的增函数，故 $\operatorname{var}(S_n) \leqslant \frac{2}{\ln 2} + \frac{n^2}{\ln n}$，利用切比雪夫不等式，有

$$P\left(\left|\frac{S_n - E(S_n)}{n}\right| \geqslant \varepsilon\right) \leqslant \frac{1}{n^2\varepsilon}\operatorname{var}(S_n) \to 0 \quad (\varepsilon > 0, n \to \infty).$$

这表明 $X_1, X_2, \cdots$ 服从切比雪夫大数定律。

- 参考习题书：茆诗松，程依明，濮晓龙. 概率论与数理统计教程. 高等教育出版社.
- 作业都会讲
- 考试难度与作业相当，最后依情况给出考点范围与复习题
- 应用：统计是概率的应用，机器学习是统计的应用