概率极限理论

随机变量序列的收敛性

设概率空间 Ω, F, P 上的随机变量序列 X_n ，和随机变量 X 。

1. 如果 $P(\lim_{n\to+\infty} X_n = X) = 1$ ，则称 X_n 概率为 1 地收敛于 X 或几乎处处（几乎必然）收敛于 X 记作

$$\lim_{n\to+\infty} X_n \overset{a.e}{=} X \ \ or \ \ X_n \overset{a.e}{\to} X$$

   - 理解：首先 $\{X_n\}$ 需要收敛，即极限 $\lim_{n\to+\infty} X_n$ 必须存在，然后要求这个极限等于 X 。

2. 如果 $\forall \epsilon > 0,\ \ lim_{n_\to+\infty} P\{|X_n - X| < \epsilon\} = 1$ 则称 $\{X_n\}$ 依概率收敛于 X ，记作

$$lim_{n\to+\infty} X_n \overset{P}{=} X \ \ or \ \ X_n \overset{P}{\to} X$$

   - 理解：$\{X_n\}$ 并不需要收敛，即极限 $\lim_{n\to\infty} X_n$ 不一定存在。但是这种不收敛到 X 的 X_n 的出现概率几乎为 0（但不代表不存在，比如震荡的情形：$\mathbb{R}$ 上整数点对应的函数值是1，而其他点对应的函数值为0。）；

3. 设 X_n 和 X 的分布函数分别为 $F_n(x)$ 和 $F(x)$ ，如果对于连续点 x ，均有 $lim_{n\to+\infty} F_n(x) = F(x)$ ，则称 $\{X_n\}$ 依分布收敛于 X（亦称弱收敛），记作

$$lim_{n\to+\infty} X_n \overset{d}{=} X \ \ or \ \ X_n \overset{d}{\to} X$$

   - 理解：在这种收敛情形下，X_n 和 X 的取值并没有直接联系，只是两个值在统计意义下具有相同的概率分布，或者说没有办法根据 X 的值得到关于 X_n 的真实值的直接信息（反之亦然）。

4. 三种收敛方式的关系：
   - 概率为 1 的收敛 $\Rightarrow$ 依概率收敛，逆命题不成立，比如例子

$$Y_{ki}(\omega) = \begin{cases} 1, & \omega \in \left(\dfrac{i-1}{k}, \dfrac{i}{k}\right] \\ 0, & ohterwise \end{cases}$$
$$k = 1, 2, \cdots;\ \ i = 1, 2, \ldots, k$$
$$X_1 = Y_{11},\ X_2 = Y_{21},\ X_3 = Y_{22}, \cdots, X_n = Y_{ki}, \cdots$$

     显然 X_n 并不收敛到 0 ，因为 $X_n = 1$ 的点有无限个，故不是概率为 1 的收敛；但是同时这种点出现的概率随 n 趋于 0 ，因此是依概率的收敛。

   - 依概率的收敛 $\Rightarrow$ 依分布的收敛，但是逆命题不成立，比如

$$X_i = B(n,p)\ \ i = 1, 2, \ldots, n$$

     为独立同分布的随机变量序列，同时 $X \sim B(n,p)$ ，那么显然 X_n 和 X 的分布函数相同，但是

$$\begin{aligned} & P(|X_n - X| < \epsilon) \\ = & P(X_n = 0,\ \ X = 0) + P(X_n = 1,\ \ X = 1) \\ = & \frac{1}{2} \end{aligned}$$

     因此该序列不依概率收敛到 X。

大数定律

重要不等式

1. 设非负随机变量 X 的期望 $E(X)$ 存在，则对于任意实数 $\epsilon > 0$ 有

$$P(X \geq \epsilon) \leq \frac{E(X)}{\epsilon}$$

2. Markov 不等式：设随机变量 X 的 k 阶绝对原点矩 $E(|X|^k)$ 存在，则对于任意实数 $\epsilon > 0$

$$P(|X| \geq \epsilon) \leq \frac{E(|X|^k)}{\epsilon^k}$$

3. Chebyshev 不等式：设随机变量 X 的方差 $D(X)$ 存在，则对于任意实数 $\epsilon > 0$

$$P(|X - E(X)| \geq \epsilon) \leq \frac{D(X)}{\epsilon^2}$$

大数定律

设概率空间 (Ω, F, P) 上的随机变量序列 $\{X_n\}$，每个$E(X_k)$ 都存在，记

$$\bar{X_n} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

若 $n \to +\infty$ 时，

$$\bar{X_n} - E(\bar{X_n}) \xrightarrow{P} 0$$

即 $\forall \epsilon > 0$

$$\lim_{n \to +\infty} P\{|\bar{X_n} - E(\bar{X_n})| < \epsilon\} = 1$$

则称序列 $\{X_k\}$ 服从大数定律。

伯努利大数定律

- 表述：设 n_A 是 n 次独立重复试验中事件 A 发生的次数，p 是每次试验中 A 发生的概率，则 $\forall \epsilon > 0$

  $$\lim_{n \to +\infty} P(|\frac{n_A}{n} - p| \geq \epsilon) = 0$$

  即第 k 次试验（$k = 1, 2, \ldots$）构成的序列 X_k 服从大数定律。

- 证明：由 Chebyshev 不等式，

  $$P\{|\bar{X_n} - E(\bar{X_n})| < \epsilon\} = P(|\frac{n_A}{n} - p| \geq \epsilon) \leq \frac{D(\bar{X_n})}{\epsilon^2} = \frac{pq}{n\epsilon^2} \to 0$$

- 意义：解释了概率直观定义的合理性，即频率 $\frac{n_A}{n}$ 与 p 相差较大 $|\frac{n_A}{n} - p| \geq \epsilon$ 在 n 足够大时只是小概率事件，因此可以用频率来代替概率。这种频率稳定于概率由称为依概率稳定。

Chebyshav 大数定律

- 表述：设随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，记为 X_k ，且其具有相同的数学期望和方差

  $$E(X_k) = \mu,\ \ D(X_k) = \sigma^2,\ \ k = 1, 2, \ldots$$

  则 $\forall \epsilon > 0$ 有

  $$\lim_{n \to +\infty} P(|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu| \geq \epsilon) = 0$$

  即序列 X_k 服从大数定律。

- 证明：与伯努利大数定律类似，同样利用 Chebyshev 不等式

  $$P\{|\bar{X_n} - E(\bar{X_n})| < \epsilon\} = P(|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu| \geq \epsilon) \leq \frac{D(\bar{X_n})}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0$$

- 两条拓展结论：
  - 当 $X_1, X_2, \ldots, X_n, \ldots$ 相互独立，但不具有相同的数学期望与方差，可设

    $$E(X_k) = \mu_k,\ \ D(X_k) = \sigma_k^2 \leq \sigma^2,\ \ k = 1, 2, \cdots$$

    有

$$\lim_{n\to+\infty} P(|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n}\mu_k| \geq \epsilon) = 0$$

- $X_1, X_2, \cdots, X_n, \cdots$ 相互独立的条件可去掉，代之以

$$\frac{1}{n} D(\sum_{k=1}^{n} X_k) \overset{n\to+\infty}{\longrightarrow} 0$$

辛钦大数定律

- 表述：设随机变量序列 $X_1, X_2, \ldots, X_n, \ldots$ 独立同分布，设其数学期望是 μ ，则 $\forall \epsilon > 0$ 有

$$\lim_{n\to+\infty} P(|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu| \geq \epsilon) = 0$$

- 意义：辛钦定律表明，Chebyshev 大数定律在随机变量同分布的情形下，不要求方差存在。

强大数定律

设概率空间 (Ω, F, P) 上的随机变量序列 $\{X_n\}$ ，每个$E(X_k)$ 都存在， 记

$$\bar{X_n} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

若 $n \to +\infty$ 时，

$$\bar{X_n} - E(\bar{X_n}) \overset{a.e}{\to} 0$$

即 $\forall \epsilon > 0$

$$P\{\lim_{n\to+\infty}[\bar{X_n} - E(\bar{X_n})] = 0\} = 1$$

则称序列 $\{X_k\}$ 服从强大数定律。

柯尔莫哥洛夫强大数定律

1. Chebyshev 大数定律的强大数定律版本：设 X_n 相互独立，期望和方差存在，且 $\sum_{n=1}^{+\infty} \frac{D(X_n)}{n^2}$ 收敛，则 $\{X_n\}$ 服从强大数定律。

2. 辛钦大数定律的强大数定律版本：设 $X_n \quad n = 1, 2, \ldots$ ，独立同分布，且期望存在，则 $\{X_n\}$ 服从强大数定律。

3. 伯努利大数定律的强大数定律版本：设 n_A 是 n 次伯努利试验中事件 A 发生的次数，p 是事件 A 发生的概率，则

$$P(\lim_{n\to+\infty} \frac{n_A}{n} = p) = 1$$

中心极限定律

定义

设概率空间 (Ω, F, P) 上的随机变量序列 $\{X_k\}$ ，$E(X_k)$ 和 $D(X_k)$ 都存在，$k = 1, 2, \ldots, n$ 。随机变量 $X \sim N(0, 1)$ 。当 $n \to +\infty$ 时，随机变量 $\sum_{k=1}^{n} X_k$ 的标准化变量

$$\frac{\sum_{k=1}^{n} X_k - E(\sum_{k=1}^{n} X_k)}{\sqrt{D(\sum_{k=1}^{n} X_k)}} \overset{d}{\to} X$$

则称序列 $\{X_k\}$ 服从中心极限定理。

独立同分布中心极限定理

设随机变量序列 $X_1, X_2, \ldots, X_n, \ldots$ 独立同分布，且期望和方差都存在：

$$E(X_k) = \mu, \quad D(X_k) = \sigma^2, \ k = 1, 2, \cdots$$

则对于任意实数 x

$$\lim_{n\to+\infty} (\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

NOTE: 我们会发现无论是大数定律还是中心极限定理，一串独立同分布的随机变量序列满足定理所需要的条件都比不是同分布的序列更弱一些。