

大数定律与中心极限定理

Didnelpsun

目录

1	依概率收敛	1
2	大数定律	1
2.1	切比雪夫大数定律	1
2.2	伯努利大数定律	2
2.3	辛钦大数定律	2
3	中心极限定理	2
3.1	列维-林德伯格定理	3
3.2	棣莫弗-拉普拉斯定理	3

这些定理与定律是针对极大量数据的概率分析，是概率论向数理统计的过渡。

1 依概率收敛

定义：设随机变量 X 与随机变量序列 $\{X_n\}$ ($n = 1, 2, 3 \dots$)，如果对任意的 $\epsilon > 0$ ，有 $\lim_{n \rightarrow \infty} P\{|X_n - X| \geq \epsilon\} = 0$ 或 $\lim_{n \rightarrow \infty} P\{|X_n - X| < \epsilon\} = 1$ ，则称随机变量序列 $\{X_n\}$ 依概率收敛于随机变量 X ，记为 $\lim_{n \rightarrow \infty} X_n = X(P)$ 或 $X_n \xrightarrow{P} X (n \rightarrow \infty)$ 。

2 大数定律

在满足一定的条件下，大数定律均为 $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} E\left(\frac{1}{n} \sum_{i=1}^n X_i\right)$ 。
所以大数定律一般是考定律成立条件与结论正确性。

2.1 切比雪夫大数定律

定义：假设随机变量序列 $\{X_n\}$ ($n = 1, 2, 3 \dots$) 是相互独立的，若方差 DX_i ($i \geq 1$) 存在且一致有上界，即存在常数 C ，使得 $DX_i \leq C$ 对一切 $i \geq 1$ 均成立，则 $\{X_n\}$ 服从大数定律： $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \frac{1}{n} \sum_{i=1}^n EX_i$ 。即 $\bar{X} \xrightarrow{P} E\bar{X}$ 。

例题：设 $X_1, X_2, \dots, X_n, \dots$ 为相互独立的随机变量序列， X_n 服从参数为 n 的指数分布 ($n \leq 1$)，则下列随机变量序列中不服从切比雪夫大数定律的是 ()。

- A. $X_1, \frac{1}{2}X_2, \dots, \frac{1}{n}X_n, \dots$ B. $X_1, X_2, \dots, X_n, \dots$
C. $X_1, 2X_2, \dots, nX_n, \dots$ D. $X_1, 2^2X_2, \dots, n^2X_n, \dots$

解：切比雪夫大数定律要求有两点，一个是随机变量序列有解，一个是方差存在上界，即 $DX_i \leq C$ 。因为题目说明相互独立，所以只用考虑方差上界。

$$\because X_n \sim E(n), \therefore EX_n = \frac{1}{n}, DX_n = \frac{1}{n^2}.$$

$$\text{对于 A, } D\left(\frac{1}{n}X_n\right) = \frac{1}{n^2}DX_n = \frac{1}{n^4} \leq 1. \text{ 对于 B, } DX_n = \frac{1}{n^2} \leq 1.$$

$$\text{对于 C, } D(nX_n) = n^2 \frac{1}{n^2} = 1, \text{ 对于 D, } D(n^2X_n) = n^4 \frac{1}{n^2} = n^2 \xrightarrow{n \rightarrow \infty} \infty.$$

所以选择 D。

2.2 伯努利大数定律

定义：假设 μ_n 是 n 重伯努利试验中事件 A 发生的次数，在每次试验中事件 A 发生的概率为 p ($0 < p < 1$)，则 $\frac{\mu_n}{n} \xrightarrow{P} p$ ，即对任意的 $\epsilon > 0$ ，有 $\lim_{n \rightarrow \infty} P \left\{ \left| \frac{\mu_n}{n} - p \right| < \epsilon \right\} = 1$ 。

可以看作通过 n 重伯努利试验，一个事件的概率会逼近一个固定的值。

2.3 辛钦大数定律

定义：假设随机变量序列 $\{X_n\}$ ($n = 1, 2, 3 \dots$) 是相互独立的同分布的，如果 $EX_i = \mu$ ($i = 1, 2, \dots$) 存在，则 $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \mu$ ，即对任意的 $\epsilon > 0$ 有 $\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| < \epsilon \right\} = 1$ 。也可以转换为即 $\bar{X} \xrightarrow{P} E\bar{X}$ 。

例题：假设随机变量序列 $X_1, X_2, \dots, X_n, \dots$ 相互独立，根据辛钦大数定律，当 $n \rightarrow \infty$ 时， $\frac{1}{n} \sum_{i=1}^n X_i$ 依概率收敛于数学期望，只要 $\{X_n\}()$ 。

A. 有相同的数学期望 B. 服从同一离散型分布

C. 服从同一泊松分布 D. 服从同一连续型分布

解：辛钦大数定律要求三点：随机变量序列独立、拥有同样分布、期望存在。

已知题目表示变量相互独立，所以只用证明有同样分布、有期望就可以。

对于 BD 而言满足是有分布的，但是此时不一定有期望，所以 BD 不行。

对于 A 有相同期望，只要求有期望就可以了，相同期望不一定同一分布。

对于 D 服从同一分布，且泊松分布期望存在。

例题：将一枚骰子重复投掷 n 次，当 $n \rightarrow \infty$ 时， n 次掷出的点数的算术平均值 \bar{X} 依概率收敛于何值？

解：根据题目，投掷是独立事件，发生概率是离散的同分布，且期望存在 $= \frac{1}{6} \sum_{i=1}^6 i = 3.5$ ，所以使用辛钦大数定律。

所以根据辛钦大数定律 $\bar{X}_n \xrightarrow{P} E\bar{X}_n = EX_i = 3.5$ 。

3 中心极限定理

中心极限定理总结来看均为：若 X_i 独立同分布于某一分布 F ，则 $\sum_{i=1}^n X_i \xrightarrow{n \rightarrow \infty} N(n\mu, n\sigma^2)$ 。

3.1 列维-林德伯格定理

定义：假设 $\{X_n\}$ 是独立分布的随机变量序列，若 $EX_i = \mu$, $DX_i = \sigma^2 > 0$ ($i = 1, 2, \dots$) 存在，则对任意的实数 x ，有 $\lim_{n \rightarrow \infty} P \left\{ \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \leq x \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt = \Phi(x)$ 。(正态分布标准化)

定理要求：独立、同分布、期望方差存在。

3.2 棣莫弗-拉普拉斯定理

定义：假设随机变量 $Y_n \sim B(n, p)$ ($0 < p < 1$, $n \geq 1$)，则对任意实数 x ，有 $\lim_{n \rightarrow \infty} P \left\{ \frac{Y_n - np}{\sqrt{np(1-p)}} \leq x \right\} = \frac{1}{\sqrt{pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt = \Phi(x)$ 。

例题：生产线生产的产品成箱包装，每箱质量是随机的。假设每箱平均重 50 千克，标准差为 5，若用载重为 5 吨的汽车承运，试用中心极限定理说明每辆汽车最多可以装多少箱才能保证不超载的概率大于 0.977。($\Phi(2) = 0.977$)

解：设 X_i 为第 i 箱质量，所以 $EX_i = 50$, $DX_i = 25$ 。

记 $T_n = \sum_{i=1}^n X_i$, $ET_n = 50n$, $DT_n = 25n$ 。

根据中心极限定理得到： $P\{T_n \leq 5000\} = P\left\{\frac{T_n - 50n}{5\sqrt{n}} \leq \frac{5000 - 50n}{5\sqrt{n}}\right\} \approx \Phi\left(\frac{5000 - 50n}{5\sqrt{n}}\right) > 0.977 = \Phi(2)$ 。
即 $\frac{5000 - 50n}{5\sqrt{n}} \geq 2$ ，即 $n \leq 98$ ，即选 98。