大数定律及中心极限定理

大数定律

依概率收敛：设 $X_1, X_2 \ldots, X_n, \ldots$ 为一系列随机变量，若存在常数 α 使得对任意 $\varepsilon > 0$ 有

$$\lim_{n\to\infty} P(|X_n - X| \geqslant \varepsilon) = 0$$

则称随机变量序列 $\{X_n\}$ 依概率收敛到 α ，记为 $X_n \xrightarrow{P} \alpha$

考虑随机变量序列 $\{X_n\}$ ，设

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

若对任意的 $\varepsilon > 0$ 有

$$\lim_{n\to\infty} P\left(\left|\overline{X} - E\left[\overline{X}\right]\right| \geqslant \varepsilon\right) = 0$$

则称 $\{X_n\}$ **服从大数定律**，即其样本均值依概率收敛到其期望，$\overline{X} - E[\overline{X}] \xrightarrow{P} 0$

切比雪夫大数定律：设 $\{X_n\}$ 为**两两互不相关**的随机变量序列，且存在常数 C 使得对每个随机变量 X_i 有 $D[X_i] \leqslant C$ （**方差一致上有界**），则 $\{X_n\}$ 服从大数定律

特别的，当 $\{X_n\}$ 为一系列**独立同分布**的随机变量，且有 $E[X_n] = \mu, D[X_n] = \sigma^2 < \infty$ ，则 $\{X_n\}$ 满足大数定律，即 $\overline{X} \xrightarrow{P} \mu$ ，该结论称为**辛钦大数定律**，事实上只需要期望存在即可得出该结论

Bernoulli 大数定律：设 μ_n 为 n 重 Bernoulli 试验中事件 A 发生的次数，p 为事件 A 发生的概率，则对任意的 $\varepsilon > 0$ 有

$$\lim_{n\to\infty} P\left(\left|\frac{\mu_n}{n} - p\right| \geqslant \varepsilon\right) = 0$$

即事件 A 发生的频率依概率收敛到 A 发生的概率，所以在试验次数很大时用事件 A 的频率作为其概率的近似是合理的。

中心极限定理

列维-林德伯格中心极限定理：设 $\{X_n\}$ 是独立同分布的随机变量序列，且 $E[X_n]=\mu, D[X_n]=\sigma^2$ ，则对任意 $x\in(-\infty,+\infty)$ 有

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

即 $\sum_{k=1}^n X_k \sim N(n\mu, n\sigma^2)$

实际问题中，只要试验数 n 足够大即可将独立同分布的随机变量的和作为正态分布来处理

也可写作

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$
$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$$

同理，将二项分布看作 n 个独立的 (0, 1) 分布之和，对 $X\sim B(n,p)$ 有

$$X \sim N(np, np(1-p))$$