《概率论与数理统计》第十周要点

第五章大数定律与中心极限定理——复习要点

1．大数定律

(1) 频率稳定性与依概率收敛 大量独立重复试验中，事件 A 的频率

$$\nu_n(A) = \frac{n(A)}{n}$$

会稳定在某一常数附近。形式化地，若随机变量序列 Y_n 满足

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} P(|Y_n - c| < \varepsilon) = 1,$$

则称 Y_n 依概率收敛于 c，记为 $Y_n \xrightarrow{P} c$。

(2) 弱大数定律（辛钦大数定律） 设 $X_1, X_2, \ldots$ 独立同分布，$E(X_k) = \mu$，$D(X_k) = \sigma^2 < \infty$，定义样本均值

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k,$$

则

$$\overline{X}_n \xrightarrow{P} \mu.$$

(3) 伯努利大数定律 在独立重复试验中，事件 A 的指示变量为 X_k，其均值为 $p = P(A)$。事件 A 的频率为

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k.$$

则

$$\overline{X}_n \xrightarrow{P} p.$$

2．中心极限定理

(1) 独立同分布的中心极限定理 设 $X_1, X_2, \ldots$ 独立同分布，$E(X_k) = \mu$，$D(X_k) = \sigma^2 > 0$。令

$$S_n = \sum_{k=1}^{n} X_k, \qquad Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}},$$

则对任意实数 x 有

$$\lim_{n\to\infty} P(Z_n \le x) = \Phi(x),$$

即

$$Z_n \xrightarrow{d} N(0,1).$$

(2) 算术平均的中心极限定理 样本平均满足

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1).$$

(3) 李雅普诺夫中心极限定理 设 $X_1, \ldots, X_n$ 独立但不要求同分布，$E(X_k) = \mu_k$，$D(X_k) = \sigma_k^2$，令

$$B_n^2 = \sum_{k=1}^{n} \sigma_k^2.$$

若存在 $\delta > 0$ 使

$$\lim_{n\to\infty} \frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} E(|X_k - \mu_k|^{2+\delta}) = 0,$$

则

$$\frac{S_n - \sum \mu_k}{B_n} \xrightarrow{d} N(0,1).$$

(4) 棣莫弗-拉普拉斯定理（二项分布的正态近似） 若 $X \sim B(n,p)$，则

$$P\left(X \le np + x\sqrt{np(1-p)}\right) \approx \Phi(x).$$

3. 中心极限定理的典型应用

例 1：均匀分布求和的近似 若 $X_i \sim U(0,10)$，$n = 20$，则可用中心极限定理近似求

$$P(S_{20} > 105) \approx 0.348.$$

例 2：大规模二项分布近似正态 若 $X \sim B(90000, 1/3)$，求

$$P(29500 < X < 30500),$$

由 CLT 得近似 0.9995。

例 3：学校家长参会人数模型 400 名学生的家长到会人数，通过计算 $E(X_i)$、$D(X_i)$ 后应用中心极限定理近似求事件概率。

4. 本章小结

- 大数定律说明大量重复试验的“频率稳定性”；
- 中心极限定理说明“正态分布的普遍性”；
- 不论单个变量的分布如何，只要符合条件，其和或均值都可近似正态；
- 正态逼近是统计推断的理论基础。