

## 5 大数定律与中心极限定理

### 5.0 大数定律

**定义1:** 若随机变量序列  $X_1, X_2, \dots, X_n, \dots$  中任意  $n$  个随机变量都是相互独立的, 则称此随机变量序列  $\{X_n\}$  是相互独立的。

**定义2:** 随机变量序列依概率收敛于常数  $\mu$ : 设  $\{X_n\}$  为随机变量序列, 若对任意的正数  $\varepsilon$ , 恒有

$$\lim_{n \rightarrow \infty} P\{|X_n - \mu| < \varepsilon\} = 1,$$

则称此随机变量序列  $\{X_n\}$  依概率收敛于常数  $\mu$ 。

#### 5.0.0 切比雪夫不等式

**定理1:** 设随机变量  $X$  具有数学期望  $E(X) = \mu$  及方差  $D(X) = \sigma^2$ , 则对于任何正数  $\varepsilon$ , 有

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}, \text{ 或}$$
$$P(|X - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}.$$

#### 5.0.1 大数定律

**定理2(切比雪夫大数定律):** 设  $X_1, X_2, \dots$  是相互独立的随机变量序列, 它们的数学期望与方差都存在, 并且方差一致有上界, 即存在某一常数  $K$ , 使得

$$D(X_i) \leq K, i = 1, 2, \dots,$$

则对任意正数  $\varepsilon$  有

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)\right| < \varepsilon\right\} = 1.$$

切比雪夫大数定律表明, 独立随机变量序列  $\{X_n\}$ , 如果方差有共同的上界, 则  $\frac{1}{n} \sum_{i=1}^n X_i$  与其数学期望  $\frac{1}{n} \sum_{i=1}^n E(X_i)$  偏差很小的概率接近于 1, 即当  $n$  充分大时, 随机变量  $X_1, X_2, \dots, X_n$  的算术平均值  $\frac{1}{n} \sum_{i=1}^n X_i$  差不多不再是随机的了, 取值接近于其数学期望的概率接近于 1。

作为切比雪夫大数定律的特例, 有下面的定理:

**定理3(独立同分布下的大数定律):** 设  $X_1, X_2, \dots$  是独立同分布的随机变量序列, 它们的数学期望与方差都存在, 且  $E(X_i) = \mu$ ,  $D(X_i) = \sigma^2$ ,  $i = 1, 2, \dots$  则对任意正数  $\varepsilon$ , 有

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| < \varepsilon\right\} = 1,$$

即当  $n \rightarrow \infty$  时, 随机变量  $X_1, X_2, \dots, X_n$  的算术平均值依概率收敛于数学期望  $\mu$ 。

**定理4(贝努利大数定律):** 设  $n_A$  是  $n$  次独立重复试验中事件  $A$  发生的次数,  $p$  是每次试验中事件  $A$  发生的概率, 则对任意的正数  $\varepsilon$ , 有

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{n_A}{n} - p\right| < \varepsilon\right\} = 1.$$

贝努利大数定律表明, 当重复试验次数  $n$  充分大时, 事件  $A$  发生的频率  $\frac{n_A}{n}$  与事件  $A$  发生的概率  $p$  有较小偏差的概率很大, 或者说事件  $A$  发生的频率  $\frac{n_A}{n}$  与事件  $A$  发生的概率  $p$  有较大偏差的概率很小, 这为在实际应用中用事件发生的频率去估计事件发生的概率提供了理论依据。

## 5.1 中心极限定理

在概率论中, 习惯于把和的分布收敛于正态分布这一类定理都叫做中心极限定理。

下面给出独立同分布随机变量序列的中心极限定理, 也称 列维 - 林德伯格(Levy - Lindberg)定理。

**定理5(独立同分布的中心极限定理):** 设  $X_1, X_2, \dots$  是独立同分布的随机变量序列, 它们的数学期望与方差都存在, 且  $E(X_i) = \mu$ ,  $D(X_i) = \sigma^2$ ,  $i = 1, 2, \dots$  记随机变量  $Y_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma}$  的分布函数为  $F_n(x)$ , 则对任意实数  $x$ , 有

$$\lim_{n \rightarrow \infty} F_n(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x).$$

**说明:** 由于无穷个随机变量之和可能趋于  $\infty$ , 故不研究  $n$  个独立同分布的随机变量之和本身, 而考虑它的标准化的随机变量

$$Y_n = \frac{\sum_{i=1}^n X_i - E(\sum_{i=1}^n X_i)}{\sqrt{D(\sum_{i=1}^n X_i)}} = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma}$$

由定理知, 当  $n$  充分大时,  $Y_n$  近似服从标准正态分布  $N(0, 1)$ , 故  $\sum_{i=1}^n X_i$  近似服从正态分布  $N(n\mu, n\sigma^2)$ , 从而独立同分布随机变量的平均值  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  近似服从正态分布  $N(\mu, \frac{\sigma^2}{n})$ , 最后这个结论在数理统计中有重要作用。

下面的定理6是定理5的特殊情况, 它是最早的中心极限定理, 叫做 棣莫弗 - 拉普拉斯定理。

**定理6(棣莫弗 - 拉普拉斯定理):** 设  $X_1, X_2, \dots$  为相互独立并具有相同两点分布的随机变量序列, 且  $P\{X_i = 1\} = p$ ,  $P\{X_i = 0\} = 1 - p$ ,  $0 < p < 1$ ,  $i = 1, 2, \dots$  则对任意实数  $x$ , 有

$$\lim_{n \rightarrow \infty} P\left\{\frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}} \leq x\right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x).$$