哈爾濱工業大學

第5章随机变量的数字特征与极限定理

第27讲 大数定律

依概率收敛

- 在第5讲中，把频率稳定值称为“统计概率”；在第24讲中，把算数平均值稳定值称为“数学期望”，这里“稳定”的含义是什么？

依概率收敛

- 例如，设A为某一试验的事件，将试验在相同的条件下重复进行n次，用m表示A出现的次数，m/n为A出现的频率，p为事件A发生的概率，当试验次数n充分大时，频率m/n稳定于概率p，可以写成 $\lim\limits_{n\to\infty}\frac{m}{n}=p.$?

即，

$\forall\varepsilon>0,\exists N,$当$n>N$时，有$\left|\frac{m}{n}-p\right|<\varepsilon.$

依概率收敛

这里频率是随机变量，频率m/n稳定于概率p, 应该从概率的角度来理解，即

$$\forall \varepsilon > 0, \text{当} n \text{充分大时，有} P\left\{\left|\frac{m}{n} - p\right| < \varepsilon\right\} = 1.$$

故 $$\lim_{n\to\infty} P\left\{\left|\frac{m}{n} - p\right| < \varepsilon\right\} = 1.$$

把它称为依概率收敛.

依概率收敛

- **定义1** 设$Z_1, Z_2, \cdots, Z_n, \cdots$是一个随机变量序列，$a$是一个常数，若对任意 $\varepsilon > 0$有

$$\lim_{n\to\infty} P(|Z_n - a| < \varepsilon) = 1,$$

或

$$\lim_{n\to\infty} P(|Z_n - a| \geq \varepsilon) = 0,$$

则称序列$Z_1, Z_2, \cdots, Z_n, \cdots$**依概率收敛**于$a$, 记为

$$\lim_{n\to\infty} Z_n \overset{P}{=} a \quad 或 \quad Z_n \xrightarrow{P} a \ (n \to \infty).$$

切比雪夫不等式

定理1 对任意随机变量X, 若$D(X)$存在，则对任意 $\varepsilon > 0$有

$$P[|X - E(X)| \geq \varepsilon] \leq \frac{D(X)}{\varepsilon^2}.$$

或

$$P[|X - E(X)| < \varepsilon] \geq 1 - \frac{D(X)}{\varepsilon^2}.$$

切比雪夫不等式

证 设X的概率密度为$f(x)$, 则

$$P[|X-E(X)|\geq\varepsilon]=\int_{|x-E(X)|\geq\varepsilon} f(x)dx$$

$$\leq\int_{|x-E(X)|\geq\varepsilon}\frac{[x-E(X)]^2}{\varepsilon^2}f(x)dx$$

$$\leq\frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty}[x-E(X)]^2 f(x)dx=\frac{D(X)}{\varepsilon^2}.$$

对离散型随机变量，只需把积分号换成求和号即可.

切比雪夫不等式

- 切比雪夫不等式应用范围广，但估计得比较粗糙.

例1 用切比雪夫不等式确定掷一匀称硬币时，需掷多少次，才能保证“正面”出现的频率在0. 4至0. 6之间的概率不小于0. 9.

解 设需掷n次, 正面出现次数为X, 则

$$X\sim B(n, 0.5),\ E(X)=0.5n,\ D(X)=0.25n.$$

切比雪夫不等式

求满足 $P(0.4 < \frac{X}{n} < 0.6) \geq 0.90$ 的n.

$X \sim B(n, 0.5)$, $E(X)=0.5n$, $D(X)=0.25n$.

$$P(0.4 < \frac{X}{n} < 0.6) = P(0.4n < X < 0.6n)$$

$$= P(|X - 0.5n| < 0.1n) \geq 1 - \frac{0.25n}{(0.1n)^2} = 1 - \frac{25}{n} \geq 0.9$$

所以 $n \geq 250$.

伯努利大数定律

定理2 设Y_n是n重伯努利试验中事件A发生的次数，$p(0<p<1)$是事件A发生的概率，则对任给的$\varepsilon > 0$，有

$$\lim_{n\to\infty} P\{|\frac{Y_n}{n} - p| \geq \varepsilon\} = 0.$$

或

$$\lim_{n\to\infty} P\{|\frac{Y_n}{n} - p| < \varepsilon\} = 1,$$

伯努利大数定律提供了用频率来确定概率的理论依据.

伯努利大数定律

证 由已知 $Y_n \sim B(n,p)$，故 $E(Y_n)=np, D(Y_n)=npq$

$(q=1-p)$,

从而 $E\left(\frac{Y_n}{n}\right)=p, D\left(\frac{Y_n}{n}\right)=\frac{1}{n^2}D(Y_n)=\frac{pq}{n}.$

由切比雪夫不等式得 $0 \le P\{|\frac{Y_n}{n}-p| \ge \varepsilon\} \le \frac{pq}{n\varepsilon^2}$,

故

$$\lim_{n\to\infty} P\{|\frac{Y_n}{n}-p| \ge \varepsilon\}=0.$$

独立同分布随机变量序列

- **定义2** 若随机变量序列$X_1, X_2, \cdots, X_n, \cdots$ 相互独立，对$n \geq 2$，且有相同的分布函数，则称 $X_1, X_2, \cdots, X_n, \cdots$ 是独立同分布的随机变量序列.

切比雪夫大数定律

定理3 设$X_1,X_2,\cdots,X_n,\cdots$是相互独立的随机变量序列. 它们都有有限的方差，并且方差有共同的上界，即 $D(X_i)\leq C(i=1,2,\cdots)$, 则对任意$\varepsilon>0$，有

$$\lim_{n\to\infty}P\{|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}(EX_i)|\geq\varepsilon\}=0$$

或

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

切比雪夫大数定律

推论 设$X_1,X_2,\ldots$是独立同分布的随机变量序列，且

$$E(X_i)=\mu, D(X_i)=\sigma^2 \quad (i=1,2,\cdots)$$

则对任给$\varepsilon>0$，有

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i-\mu|\geq\varepsilon\}=0$$

或

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i-\mu|<\varepsilon\}=1.$$

辛钦大数定律

定理4 设$X_1,X_2,\ldots$是独立同分布的随机变量序列，且

$E(X_i)=\mu,(i=1,2,\cdots)$ ，则对任给$\varepsilon>0$，

有

$$\lim_{n\to\infty}P\{|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu|\geq\varepsilon\}=0$$

或

$$\lim_{n\to\infty}P\{|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu|<\varepsilon\}=1.$$

这是随机变量序列的算术平均值稳定性的较确切的解释.

例2 设$X_1,X_2,\ldots$是独立同分布的随机变量序列，都服从参数为2的指数分布。则当 $n\to\infty$ 时，$Y_n=\frac{1}{n}\sum_{i=1}^{n}X_i^2$ 依概率收敛于什么？

解 由已知$X_1^2,X_2^2,\cdots$也独立同分布，

由$X_i\sim E(2)\Rightarrow E(X_i)=1/2, D(X_i)=1/4,$

$E(X_i^2)=D(X_i)+[E(X_i)]^2=1/4+(1/2)^2=1/2,$

由辛钦大数定律，$Y_n=\frac{1}{n}\sum_{i=1}^{n}X_i^2\xrightarrow{P}E(X_i^2)=\frac{1}{2}\qquad(n\to\infty).$

谢　　谢！