第五章 大数定律及中心极限定理

一、重点与难点

二、主要内容

三、典型例题

一、重点与难点

1.重点

中心极限定理及其运用.

2.难点

证明随机变量服从大数定律.

二、主要内容

大数定律

- 定理一
  - 定理一的另一种表示
- 定理二
- 定理三

中心极限定理

- 定理一
- 定理二
- 定理三

大数定律

迄今为止,人们已发现很多大数定律(laws of large numbers)所谓大数定律，简单地说，就是大量数目的随机变量所呈现出的规律，这种规律一般用随机变量序列的某种收敛性来刻画。

依概率收敛的意义

依概率收敛即依概率1收敛。随机变量序列$\{X_n\}$依概率收敛于a，说明对于任给的$\varepsilon>0$，当n很大时，事件“$|X_n-a|<\varepsilon$”的概率接近于1，但正因为是概率，所以不排除小概率事件“$|X_n-a|\ge\varepsilon$”发生。所以说依概率收敛是不确定现象中关于收敛的一种说法。

依概率收敛

设$Y_1,Y_2,\cdots,Y_n,\cdots$是一个随机变量序列，$a$是一个常数，若对于任意正数$\varepsilon$，有

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

则称序列$Y_1,Y_2,\cdots,Y_n,\cdots$依概率收敛于$a$，记为

$$Y_n\xrightarrow{p}a$$

依概率收敛的性质

设$X_n\xrightarrow{p}a$, 设$Y_n\xrightarrow{p}b$，又设函数$g(x,y)$在点(a,b)连续，

则 $$g(X_n,Y_n)\xrightarrow{p}g(a,b)$$

特别地，若 $g(X_n, Y_n) = cX_n + dY_n$，$c$，$d$ 为常数，

则 $cX_n + dY_n \xrightarrow{p} ca + db$；

若 $g(X_n, Y_n) = X_nY_n$，则 $X_nY_n \xrightarrow{p} ab$

若 $g(X_n, Y_n) = \dfrac{X_n}{Y_n}$，$Y_n \neq 0$，则 $\dfrac{X_n}{Y_n} \xrightarrow{p} \dfrac{a}{b}$ $(b \neq 0)$

切比雪夫大数定律

定理　设 $X_1, X_2, \cdots, X_n, \cdots$ 是相互独立的随机变量序列，数学期望 $E(X_i)$ 和方差 $D(X_i)$ 都存在 $(i = 1, 2, \cdots)$，且 $D(X_i) < C (i = 1, 2, \cdots)$，则对任意给定的 $\varepsilon > 0$，有

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

切比雪夫大数定律的特殊情况

定理　设 $X_1, X_2, \cdots, X_n, \cdots$ 是相互独立的随机变量序列，有相同的数学期望和方差，$E(X_i) = \mu, D(X_i) = \sigma^2$ $(i = 1, 2, \cdots)$。则对任意给定的 $\varepsilon > 0$，有

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

即　$\overline{X} \xrightarrow{p} \mu$

伯努利大数定律

定理　设 Y_n 是 n 次独立重复试验中事件 A 发生的次数，p 是事件 A 在每次试验中发生的概率，则对于任意正数 $\varepsilon > 0$，有

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

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

伯努利大数定理给出了当 n 很大时，A 发生的频率 $\dfrac{Y_n}{n}$ 依概率收敛于 A 的概率这一结论，证明了频率的稳定性。

辛钦定理

设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，服从同一分布，且具有数学期望 $E(X_k) = \mu$ $(k = 1, 2, \cdots)$，则对于任意正数 ε，有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| < \varepsilon\right\} = 1.$$

大数定律在概率论中的意义

大数定律给出了在试验次数很大时频率和平均值的稳定性，从理论上肯定了用算术平均值代替均值，用频率代替概率的合理性。它既验证了概率论中一些假设的合理性，又为数理统计中用样本推断总体提供了理论根据，所以说，大数定律是概率论中最重要的基本定律。

中心极限定理

实例: 考察射击命中点与靶心距离的偏差.

这种偏差是大量微小的偶然因素造成的微小误差的总和, 这些因素包括: 瞄准误差、测量误差、子弹制造过程方面 (如外形、重量等) 的误差以及射击时武器的振动、气象因素(如风速、风向、能见度、温度等) 的作用, 所有这些不同因素所引起的微小误差是相互独立的, 并且它们中每一个对总和产生的影响不大.

问题: 某个随机变量是由大量相互独立且均匀小的随机变量相加而成的, 研究其概率分布情况.

基本定理

定理一: 独立同分布的中心极限定理

设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立, 服从同一分布, 且具有数学期望和方差: $E(X_k)=\mu$, $D(X_k)=\sigma^2>0\ (k=1,2,\cdots)$, 则随机变量之和的

标准化变量 $Y_n=\dfrac{\sum_{k=1}^{n}X_k-E\left(\sum_{k=1}^{n}X_k\right)}{\sqrt{D\left(\sum_{k=1}^{n}X_k\right)}}=\dfrac{\sum_{k=1}^{n}X_k-n\mu}{\sqrt{n}\,\sigma}$

的分布函数 $F_n(x)$ 对于任意 x 满足

$$\lim_{n\to\infty}F_n(x)=\lim_{n\to\infty}P\left\{\frac{\sum_{k=1}^{n}X_k-n\mu}{\sqrt{n}\,\sigma}\le x\right\}$$

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

定理表明:

当 $n\to\infty$, 随机变量序列 Y_n 的分布函数收敛于标准正态分布的分布函 数.

定理二 (李雅普诺夫定理)

设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立, 它们具有数学期望和方差:

$$E(X_k)=\mu_k,\quad D(X_k)=\sigma_k^2\neq 0\ (k=1,2,\cdots),$$

记 $\quad B_n^2=\sum_{k=1}^{n}\sigma_k^2,$

若存在正数 δ, 使得当 $n\to\infty$ 时,

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

则随机变量之和的标准化变量

$$Z_n=\frac{\sum_{k=1}^{n}X_k-E\left(\sum_{k=1}^{n}X_k\right)}{\sqrt{D\left(\sum_{k=1}^{n}X_k\right)}}=\frac{\sum_{k=1}^{n}X_k-\sum_{k=1}^{n}\mu_k}{B_n}$$

的分布函数 $F_n(x)$ 对于任意 x 满足

$$\lim_{n\to\infty}F_n(x)=\lim_{n\to\infty}P\left\{\frac{\sum_{k=1}^{n}X_k-\sum_{k=1}^{n}\mu_k}{B_n}\le x\right\}$$

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

定理二表明:

无论各个随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 服从什么分布, 只要满足定理的条件, 那么它们的和 $\sum_{k=1}^{n}X_k$ 当 n 很大时, 近似地服从正态分布.

(如实例中射击偏差服从正态分布)

下面介绍的定理三是定理一的特殊情况.

定理三(德莫佛－拉普拉斯定理)

设随机变量 $\eta_n\ (n=1,2,\cdots)$ 服从参数为 n,p $(0<p<1)$的二项分布, 则 对于任意 x, 恒有

$$\lim_{n\to\infty} P\left\{\frac{\eta_n - np}{\sqrt{np(1-p)}} \le x\right\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x).$$

证明 $\quad \eta_n = \sum_{k=1}^{n} X_k,$

其中 $X_1, X_2, \cdots, X_n$ 是相互独立的、服从同一(0—1) 分布的随机变量 , 分布律为

$$P\{X_k = i\} = p^i (1-p)^{1-i}, \qquad i = 0, 1.$$

$\because E(X_k) = p,\quad D(X_k) = p(1-p) \quad (k=1,2,\cdots,n),$

根据定理一得

$$\lim_{n\to\infty} P\left\{\frac{\eta_n - np}{\sqrt{np(1-p)}} \le x\right\} = \lim_{n\to\infty} P\left\{\frac{\sum_{k=1}^{n} X_k - np}{\sqrt{np(1-p)}} \le x\right\}$$

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

定理三表明:

正态分布是二项分布的极限分布, 当n充分大时, 可以利用该定理来计算二项分布的概率.

三、典型例题

例 假设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的简单随机样本,已知 $E(X^k) = \alpha_k (k=1,2,3,4)$. 证明:当 n 充分大时,随机变量 $Z_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ 近似服从正态分布,并指出其分布参数.

解 因为 $X_1, X_2, \cdots, X_n$ 独立同分布,

所以 $X_1^2, X_2^2, \cdots, X_n^2$ 也独立同分布,

且 $E(X_i^2) = \alpha_2,$

$$D(X_i^2) = E(X_i^4) - (EX_i^2)^2 = \alpha_4 - \alpha_2^2,$$

根据独立同分布的中心极限定理知

$$V_n = \frac{\sum_{i=1}^{n} X_i^2 - n\alpha_2}{\sqrt{n(\alpha_4 - \alpha_2^2)}} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \alpha_2}{\sqrt{\frac{1}{n}(\alpha_4 - \alpha_2^2)}} = \frac{Z_n - \alpha_2}{\sqrt{\frac{1}{n}(\alpha_4 - \alpha_2^2)}}$$

的极限分布是标准正态分布.

故当 n 充分大时,

V_n近似服从标准正态分布,

从而当 n 充分大时,

$Z_n = \sqrt{\frac{1}{n}(\alpha_4 - \alpha_2^2)}\, V_n + \alpha_2$ 近似服从

参数为 $\mu = \alpha_2,\ \sigma^2 = \frac{\alpha_4 - \alpha_2^2}{n}$ 的正态分布.

例 某保险公司的老年人寿保险有1万人参加,每人每年交200元. 若老人在该年内死亡,公司付给家属1万元. 设老年人死亡率为0.017,试求保险公司在一年内的这项保险中亏本的概率.

解 设 X 为一年中投保老人的死亡数,

则 $X \sim B(n, p),$

其中 $n = 10000,\quad p = 0.017,$

由德莫佛－拉普拉斯定理知,

保险公司亏本的概率

$$P\{10000X > 10000\times 200\} = P\{X > 200\}$$

$$= P\left\{\frac{X-np}{\sqrt{np(1-p)}} > \frac{200-np}{\sqrt{np(1-p)}}\right\}$$

$$= P\left\{\frac{X-np}{\sqrt{np(1-p)}} > 2.321\right\}$$

$$\approx 1-\Phi(2.321)\approx 0.01 .$$

概率论与数理统计

例 现有一批种子,其中良种占$\frac{1}{6}$,今在其中任选 6000 粒,试问在这些种子中良种所占的比例与$\frac{1}{6}$之差的绝对值小于$\frac{1}{1000}$的概率是多少?

解 令 $X_i = \begin{cases} 1, & 第\ i\ 粒是良种, \\ 0, & 第\ i\ 粒不是良种, \end{cases} \quad i = 1,2,\cdots,n.$

则 $P(X_i = 1) = \frac{1}{6}$, 记 $Y_n = \sum_{i=1}^{n} X_i$,

则 $Y_n \sim B\left(n, \frac{1}{6}\right)$, $n = 6000.$

概率论与数理统计

根据题意,所求概率为

$$P\left(\left|\frac{Y_n}{6000}-\frac{1}{6}\right| \le \frac{1}{1000}\right) = P(|Y_n - 1000| \le 6),$$

因为 $Y_n \sim B\left(6000, \frac{1}{6}\right)$,

由中心极限定理有:

Y_n近似服从 $N\left(1000,\ 1000\times\frac{5}{6}\right)$,

概率论与数理统计

所以 $P\left(\left|\frac{Y_n}{6000}-\frac{1}{6}\right| \le \frac{1}{1000}\right)$

$$= P\left(\left|\frac{Y_n - 1000}{\sqrt{1000\times 5/6}}\right| \le \frac{6}{\sqrt{1000\times 5/6}}\right)$$

$$= 2\Phi\left(\frac{6\sqrt{6}}{\sqrt{5000}}\right) - 1 = 2\Phi(0.208) - 1$$

$$= 2\times 0.5832 - 1 = 0.1664 .$$