

第五章 大数定律与中心极限定理

5.1 大数定律的概念

大数定律的概念

例1 掷一颗骰子, 出现1点的概率是 $1/6$, 在掷的次数比较少时, 出现1点的频率可能与 $1/6$ 相差很大, 但是在掷的次数很多时, 出现1点的频率接近 $1/6$ 是必然的.

例2 测量一个长度 a , 一次测量的结果不见得就等于 a , 量了若干次, 其算术平均值仍不见得等于 a , 但当测量次数很多时, 算术平均值接近于 a 几乎是必然的.

这两个例子说明：

在大量随机现象中, 不仅看到了随机事件的频率具有稳定性, 而且还看到大量测量值的平均结果也具有稳定性。这种稳定性就是本章所要讨论的大数定律的客观背景。即无论个别随机现象的结果如何, 或者它们在进行过程中的个别特征如何, 大量随机现象的平均结果实际上与每一个别随机现象的特征无关, 并且几乎不再是随机的了。

大数定律以确切的数学形式表达了这种规律性, 并论证了它成立的条件, 即从理论上阐述了这种大量的、在一定条件下的、重复的随机现象呈现的规律性即稳定性. 由于大数定律的作用, 大量随机因素的总体作用必然导致某种不依赖于个别随机事件的结果.

5.2 切贝谢夫不等式

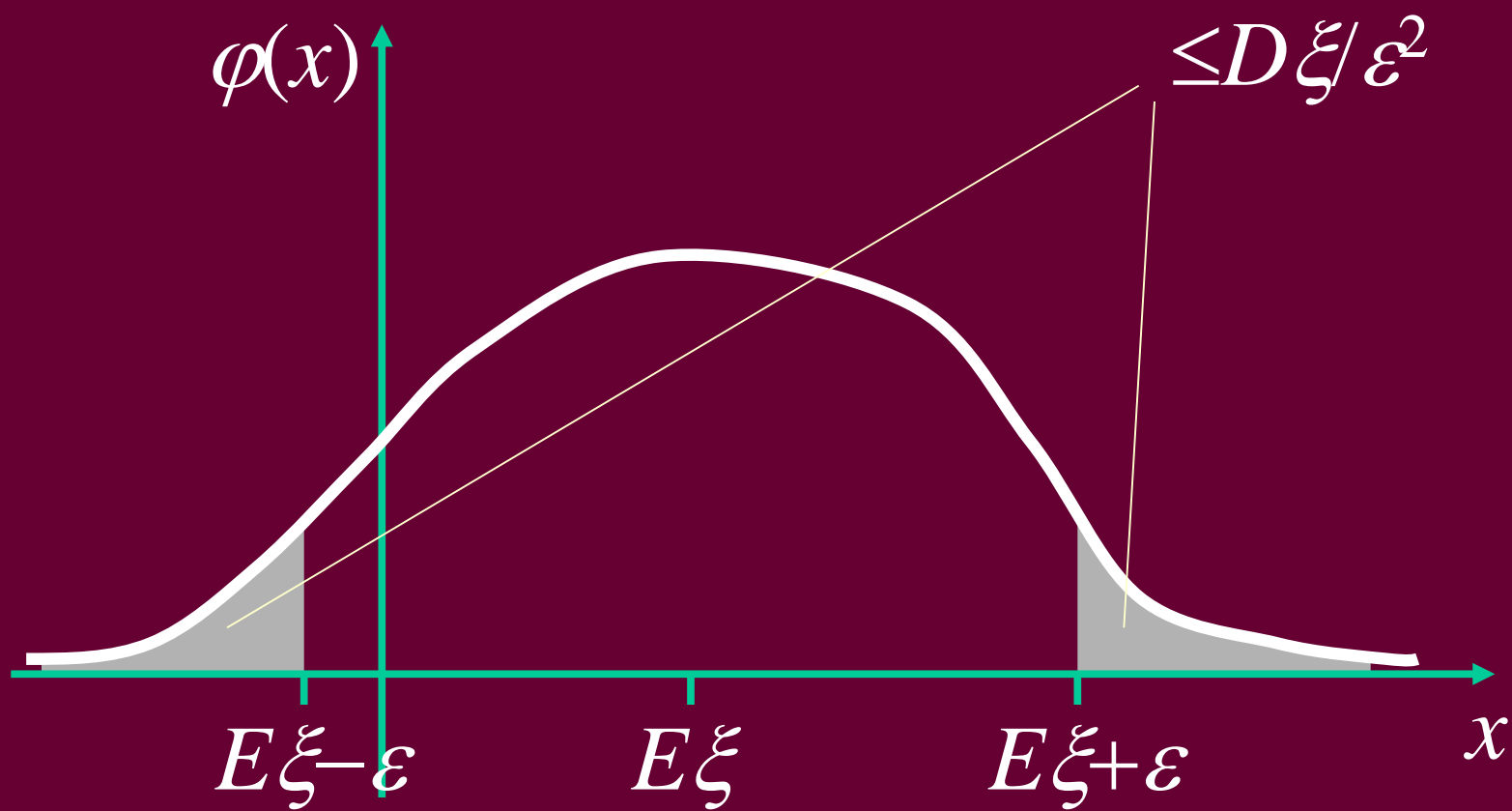
切贝谢夫不等式

设随机变量 ξ 有期望值 $E\xi$ 及方差 $D\xi$, 则任给 $\varepsilon > 0$, 有

$$P(|\xi - E\xi| \geq \varepsilon) \leq \frac{D\xi}{\varepsilon^2}$$

$$P(|\xi - E\xi| < \varepsilon) \geq 1 - \frac{D\xi}{\varepsilon^2}$$

示意图



例1 设 ξ 是掷一颗骰子所出现的点数, 若给定 $\varepsilon=1,2$, 实际计算 $P(|\xi-E\xi|\geq\varepsilon)$, 并验证切贝谢夫不等式成立.

解 因 $P(\xi=k)=1/6, (k=1,2,3,4,5,6)$

$$E\xi = \frac{1+2+3+4+5+6}{6} = \frac{7}{2}, E\xi^2 = \frac{1+4+9+16+25+36}{6} = \frac{91}{6}$$

$$D\xi = E\xi^2 - (E\xi)^2 = \frac{91}{6} - \frac{49}{4} = \frac{182-147}{12} = \frac{35}{12}$$

$$\varepsilon=1: \frac{D\xi}{\varepsilon^2} = \frac{35}{12} > \frac{2}{3} = P(|\xi - \frac{7}{2}| \geq 1)$$

$$\varepsilon=2: \frac{D\xi}{\varepsilon^2} = \frac{35}{4 \times 12} = \frac{35}{48} > \frac{1}{3} = P(|\xi - \frac{7}{2}| \geq 2)$$

例2 设电站供电所有10000盏电灯, 夜晚每一盏灯开灯的概率都是0.7, 而假定开关时间彼此独立, 估计夜晚同时开着的灯数在6800与7200之间的概率.

解 令 ξ 为同时开灯的数目, 则 $\xi \sim B(10000, 0.7)$

$$P(6800 < \xi < 7200) = \sum_{k=6801}^{7199} C_{10000}^k \times 0.7^k \times 0.3^{10000-k}$$

如果用切贝谢夫不等式估计:

$$E\xi = np = 10000 \times 0.7 = 7000$$

$$D\xi = npq = 10000 \times 0.7 \times 0.3 = 2100$$

$$P\{6800 < \xi < 7200\} = P(|\xi - 7000| < 200)$$

$$\geq 1 - \frac{2100}{200^2} \approx 0.95$$

可见只要有供应7200盏灯的电力就够用.

5.3 切贝谢夫定理

定义5.1 若存在常数 a ,使对于任何

$$\varepsilon > 0, \text{有} \lim_{n \rightarrow \infty} P(|\xi_n - a| < \varepsilon) = 1$$

则称随机变量序列 $\{\xi_n\}$ 依概率收敛于 a

定理5.1(切贝夫定理)设 $\xi_1, \xi_2 \cdots$ 是相互独立的随机变量序列,各有数学期望 $E\xi_1, E\xi_2 \cdots$ 及方差

$D\xi_1, D\xi_2 \cdots$ 并且对于所有 $k=1, 2, \cdots$ 都有 $D\xi_k < l$,

其中 l 是与 k 无关的常数,则任给 $\varepsilon > 0$,有

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n \xi_k - \frac{1}{n} \sum_{k=1}^n E\xi_k \right| < \varepsilon \right\} = 1 \quad (5.2)$$

随机变量的
算术平均值

随机变量期望
的算术平均值

切贝谢夫定理说明:在定理的条件下,当n充分大时,n个独立随机变量的平均数这个随机变量的离散程度是很小的.这意味,经过算术平均后得到的随机变量

$$\frac{1}{n} \sum_{k=1}^n \xi_k \xrightarrow{p} \frac{1}{n} \sum_{k=1}^n E \xi_k$$

将比较密地聚集在它的数学期望的附近. 它与数学期望之差, 当时 $n \rightarrow \infty$, 依概率收敛到0. 这就是大数定律. 切贝谢夫定理为这一定律作出了精确的数学公式. 它也称为切贝谢夫大数定律.

切贝谢夫定理的一个推论通常称为贝努里大数定律.

定理5.2(贝努里大数定律) 在独立试验序列中, 当试验次数 n 无限增加时, 事件 A 的频率 ξ/n (ξ 是 n 次试验中事件 A 发生的次数), 依概率收敛于它的概率 $P(A)$. 即对于任意给定的 $\varepsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{\xi}{n} - p \right| < \varepsilon \right\} = 1 \quad (5.3)$$

如果事件A的概率很小,则正如贝努里定理指出的,事件A的频率也是很小的,即事件A很少发生.例如 $P(A)=0.001$,则在1000次试验中只能希望事件A发生一次.

在实际中概率很小的随机事件在个别试验中几乎是不可能发生的.因此,人们常常忽略了那些概率很小的事件发生的可能性.这个原理叫作小概率事件的实际不可能性原理(简称小概率原理).

它在国家经济建设事业中有着广泛的应用.至于“小概率”小到什么程度才能看作实际上不可能发生,则要视具体问题的要求和性质而定.从小概率事件的实际不可能性原理容易得到下面的重要结论:如果随机事件的概率很接近1,则可以认为在个别试验中这事件几乎一定发生

定理5.3 (**辛钦大数定律**) 如果 ξ_1, ξ_2, \dots 是相互独立并且具有相同分布的随机变量. 有 $E \xi_k = a$ ($k=1, 2, \dots$), 有

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n \xi_k - a \right| < \varepsilon \right\} = 1 \quad (5.4)$$

$$Y_n = \frac{1}{n} \sum_{k=1}^n \xi_k \xrightarrow{P} a$$

这一定理使算术平均值的法则有了理论依据.假使要测量某一个物理量 μ ,在不变的条件下重复测量 n 次,得到的观测值 x_1, x_2, \dots, x_n

是不完全相同的.这些结果可以看作是服从同一分布并且期望值为 μ 的 n 个相互独立的随机变量 $\xi_1, \xi_2, \dots, \xi_n$ 的试验数值。由定理3可知,当 n 充分大时,取 $\frac{1}{n} \sum_{k=1}^n x_k$

作为的 μ 近似值,可以认为所发生的误差是很小的,即对于同一个随机变量 ξ 进行 n 次独立观察,则所有观察结果的算术平均数依概率收敛于随机变量的期望值.

§ 5.4 中心极限定理

在客观实际中有许多随机变量, 它们是由大量的相互独立的随机因素的综合影响所形成的。而其中每一个别因素在总的影响中所起的作用都是微小的。这种随机变量往往近似地服从正态分布, 这种现象就是中心极限定理的客观背景。

正态分布在随机变量的各种分布中, 占有特别重要的地位. 在某些条件下, 即使原来并不服从正态分布的一些独立的随机变量, 它们的和的分布, 当随机变量的个数无限增加时, 也是趋于正态分布的.

在概率论里, 把研究在什么条件下, 大量独立随机变量和的分布以正态分布为极限这一类定理称为中心极限定理.

一般说来, 如果某些偶然因素对总和的影响是均匀的, 微小的, 即没有一项起特别突出的作用, 那么就可以断定描述这些大量独立的随机因素的总和的随机变量是近似的服从正态分布.

这是数理统计中大样本的理论基础, 用数学形式来表达就是李雅普诺夫定理.

二.几个常用的中心极限定理

1. 定理5.4 (李雅普诺夫Liapunov定理)

设 ξ_1, ξ_2, \dots 是相互独立的随机变量, 有期望及方差

$$E\xi_k = a_k \quad D\xi_k = \sigma_k^2 \neq 0 \quad k=1, 2, \dots,$$

$$\lim_{n \rightarrow \infty} P \left\{ \frac{\sum_{k=1}^n \xi_k - \sum_{k=1}^n a_k}{\sqrt{\sum_{k=1}^n \sigma_k^2}} \leq x \right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi_0(x) \quad (5.5)$$

这个定理的实际意义是:如果一个随机现象由众多的随机因素所引起,每一因素在总的变化里起着不显著的作用,就可以推断,描述这个随机现象的随机变量近似的服从正态分布.由于这些情况很普遍,所以有相当多一类随机变量遵从正态分布,从而正态分布成为概率统计中最重要的分布.

例1 一个螺丝钉的重量是一个随机变量, 期望值是一两, 标准差是0.1两. 求一盒(100个)螺丝钉的重量超过10.2斤的概率.

解 设一盒重量为 ξ , 盒中第 i 个螺丝钉的重量为 ξ_i , ($i=1,2,\dots,100$). ξ_1,\dots,ξ_{100} 相互独立,

$$E\xi_i = 1, \sqrt{D\xi_i} = 0.1, \text{ 则 } \xi = \sum_{i=1}^{100} \xi_i, E\xi = 100(\text{两}),$$

$$\sqrt{D\xi} = \sqrt{0.1^2 \times 100} = 1(\text{两}), \text{ 则近似 } \xi \sim N(100, 1)$$

$$P(\xi > 102) = P\left(\frac{\xi - 100}{1} > 2\right) \approx 1 - \Phi_0(2) = 0.022750$$

例2 对敌人的防御地段进行100次轰炸, 每次轰炸命中目标的炸弹数目是一个随机变量, 其期望值为2, 方差为1.69. 求在100次轰炸中有180颗到220炸弹命中目标的概率.

解 令第*i*次轰炸命中目标的次数为 ξ_i , 100次轰炸命中目标次数 $\xi = \xi_1 + \xi_2 + \dots + \xi_{100}$. $E\xi = 200$, $D\xi = 169$, 近似有 $\xi \sim N(200, 13^2)$

$$\begin{aligned} P(180 \leq \xi \leq 220) &= P(|\xi - 200| \leq 20) \\ &= P\left(\left|\frac{\xi - 200}{13}\right| \leq \frac{20}{13}\right) \approx 2\Phi_0(1.54) - 1 = 0.87644 \end{aligned}$$

定理5.5 拉普拉斯定理, 设 $\xi \sim B(n, p)$

(1) 局部极限定理: 当 $n \rightarrow \infty$ 时

$$P(\xi = k) \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}} = \frac{1}{\sqrt{npq}} \varphi_0\left(\frac{k-np}{\sqrt{npq}}\right)$$

(2) 积分极限定理: 当 $n \rightarrow \infty$ 时

$$P(a < \xi < b) \approx \Phi(b) - \Phi(a)$$

$$= \Phi_0\left(\frac{b-np}{\sqrt{npq}}\right) - \Phi_0\left(\frac{a-np}{\sqrt{npq}}\right)$$

例3 10部机器独立工作, 每部停机的概率为0.2.
求3部机器同时停机的概率.

解 10部机器中同时停机的数目 $\xi \sim B(10, 0.2)$

$$n = 10, p = 0.2, np = 2, \sqrt{npq} \approx 1.265.$$

(1) 直接计算: $P(\xi = 3) = C_{10}^3 \times 0.2^3 \times 0.8^7 \approx 0.2013$

(2) 用局部极限定理近似计算:

$$\begin{aligned} P(\xi = 3) &= \frac{1}{\sqrt{npq}} \varphi_0 \left(\frac{k - np}{\sqrt{npq}} \right) = \frac{1}{1.265} \varphi_0 \left(\frac{3 - 2}{1.265} \right) \\ &= \frac{1}{1.265} \varphi_0(0.79) = 0.2308 \end{aligned}$$

例4 设电站供电所有10000盏电灯, 夜晚每一盏灯开灯的概率都是0.7, 而假定开关时间彼此独立, 估计夜晚同时开着的灯数在6800与7200之间的概率.

解 开着的灯数 $\xi \sim B(10000, 0.7)$

$$np = 7000, \sqrt{npq} \approx 45.83$$

$$P(6800 < \xi < 7200) = P(|\xi - 7000| < 200)$$

$$= P\left(\left|\frac{\xi - 7000}{45.83}\right| < 4.36\right)$$

$$= 2\Phi_0(4.36) - 1 = 0.99999$$

例5 产品为废品的概率为 $p=0.005$, 求10000件产品中废品数不大于70的概率.

解 10000件产品中的废品数 $\xi \sim B(10000, 0.005)$,

$$np = 50, \sqrt{npq} \approx 7.053$$

$$P(\xi \leq 70) = \Phi_0\left(\frac{70 - 50}{7.053}\right)$$

$$= \Phi_0(2.84) = 0.9977$$

例6 每颗炮弹命中飞机的概率为0.01, 求500发炮弹中命中5发的概率.

解 命中飞机的炮弹数目 $\xi \sim B(500, 0.01)$

$$np = 5, \sqrt{npq} \approx 2.225$$

(1) 用二项分布公式计算:

$$P(\xi = 5) = C_{500}^5 \times 0.01^5 \times 0.99^{495} = 0.17635$$

(2) 用普哇松公式计算, 查附表1可得

$$P_5(5) \approx 0.175467$$

(3) 用拉普拉斯局部极限定理计算

$$P(\xi = 5) \approx \frac{1}{\sqrt{npq}} \varphi_0 \left(\frac{5 - np}{\sqrt{npq}} \right) = 0.1793$$

例：假设各零件的重量都是随机变量，它们相互独立，且服从相同的分布，其数学期望为0.5kg，标准差为0.1kg，试用中心极限定理计算5000只零件的总重量超过2510kg的概率。(所求概率用标准正态分布函数 $\Phi_0(x)$ 的值表示)