第五章 大数定律与中心极限定理

§1 切贝谢夫不等式

研究随机变量的离差与方差的关系。

设随机变量ξ有期望值Eξ与方差Dξ。

对任给ε>0, 有

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

$$P(|\xi-E\xi|<\varepsilon)\geq1-\frac{D\xi}{\varepsilon^2}$$

称为切贝谢夫不等式

证：若ξ是离散型随机变量。

$$P(\xi=x_k)=p_k$$

$$P(|\xi - E\xi| \geq \varepsilon) = \sum_{|x_k - E\xi| \geq \varepsilon} P(\xi = x_k)$$

$$\leq \sum_{|x_k - E\xi| \geq \varepsilon} \frac{(x_k - E\xi)^2}{\varepsilon^2} p_k$$

$$\leq \sum_{k} \frac{(x_k - E\xi)^2}{\varepsilon^2} p_k$$

$$= \frac{D\xi}{\varepsilon^2}$$

若ξ是连续型随机变量。

ξ的概率密度为$\varphi(x)$

$$P(|\xi - E\xi| \geq \varepsilon) = P(\xi \leq E\xi - \varepsilon) + P(\xi \geq E\xi + \varepsilon)$$

$$= \int_{-\infty}^{E\xi-\varepsilon} \varphi(x)dx + \int_{E\xi+\varepsilon}^{+\infty} \varphi(x)dx$$

$$\leq \int_{-\infty}^{E\xi-\varepsilon} \frac{(x-E\xi)^2}{\varepsilon^2} \varphi(x)dx + \int_{E\xi+\varepsilon}^{+\infty} \frac{(x-E\xi)^2}{\varepsilon^2} \varphi(x)dx$$

$$\leq \int_{-\infty}^{+\infty} \frac{(x-E\xi)^2}{\varepsilon^2} \varphi(x)dx$$

$$= \frac{D\xi}{\varepsilon^2}$$

- 例1　设ξ是掷一颗骰子所出现的点数，若给定ε＝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：$P(\xi=k)=\frac{1}{6}, k=1,2,\ldots,6$

$$E\xi=\frac{7}{2} \qquad D\xi=\frac{35}{12}$$

$$P\left(\left|\xi-\frac{7}{2}\right|\geq 1\right)=\frac{2}{3} \qquad P\left(\left|\xi-\frac{7}{2}\right|\geq 2\right)=\frac{1}{3}$$

$$\varepsilon=1\text{时}, \frac{D\xi}{\varepsilon^2}=\frac{35}{12} > \frac{2}{3}$$

$$\varepsilon=2\text{时}, \frac{D\xi}{\varepsilon^2}=\frac{35}{48} > \frac{1}{3}$$

例2　设电站供电网有10000盏电灯，夜晚每一盏灯开灯的概率都是0. 7，而假定开、关事件彼此独立，估计夜晚同时开着的灯数在6800与7200之间的概率。

解：令ξ表示夜晚同时开着的灯的数目。

$$\xi \sim B(10000, 0.7)$$

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

用切贝谢夫不等式估计：

$$E\xi = np = 7000 \qquad D\xi = npq = 2100$$

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

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

§2 大数定律

- 例1 掷一枚硬币，出现正面的概率为$\frac{1}{2}$

  掷的次数很多时，出现正面的频率接近$\frac{1}{2}$

  这种现象为频率的稳定性。

- 例2 测量一个长度a，一次测量，结果未必等于a

  测量多次，结果的计算平均值未必等于a

  测量次数很大时，算术平均值接近于a

  这种现象为平均结果的稳定性

  大量随机现象中的平均结果与每一个别随机现象无关，几乎不再随机。

- 定义1 若存在常数a，使对于任何ε>0，有

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

称随机变量序列$\{\xi_n\}$依概率收敛于a

- 例3 设ξ_n为两点分布

$$P\left(\xi = \frac{1}{n}\right) = 1 - \frac{1}{n} \qquad P(\xi = n+1) = \frac{1}{n}$$

对任给ε>0, n充分大时，必有n+1>ε且$\frac{1}{n} < \varepsilon$

$$\lim_{n\to\infty} P(|\xi_n - 0| < \varepsilon) = \lim_{n\to\infty} P\left(\xi = \frac{1}{n}\right)$$

$$= \lim_{n\to\infty}\left(1 - \frac{1}{n}\right) = 1$$

即$\{\xi_n\}$依概率收敛于0

定理1 (切贝谢夫定理)设$\xi_1,\xi_2,\ldots,$是相互独立的随机变量序列，各有数学期望$E\xi_1,E\xi_2,\ldots$及方差$D\xi_1,D\xi_2,\ldots$并且对于所有i=1, 2, . . . $D\xi_i<M$, M与i无关，则任给$\varepsilon>0$

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

此定理表明n个独立随机变量的平均值$\frac{1}{n}\sum_{i=1}^{n}\xi_i$

依概率收敛于其数学期望$\frac{1}{n}\sum_{i=1}^{n}E\xi_i$

也称为切贝谢夫大数定律。

它有如下重要的推论。

定理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$\frac{\xi}{n}$依概率收敛于A发生的概率P(A)=p

即 对任给ε>0 $\lim\limits_{n\to\infty} P\left(\left|\frac{\xi}{n}-p\right|<\varepsilon\right)=1$

证明见P107

大量重复试验中，事件发生的频率接近于概率。

若P(A)很小，则A发生的频率也很小

如P(A)=0.001,约在1000次试验中，A发生一次

在一次试验中认为A几乎不可能发生。

这称为小概率事件的实际不可能性原理。

- 定理3 (辛钦大数定律)如果$\xi_1,\xi_2,\ldots$是相互独立有相同分布的随机变量, 有$E\xi_i=a(i=1,2,\ldots)$则对任意给定的$\varepsilon>0$，有

$$\lim_{n\to\infty}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-a\right|<\varepsilon\right)=1$$

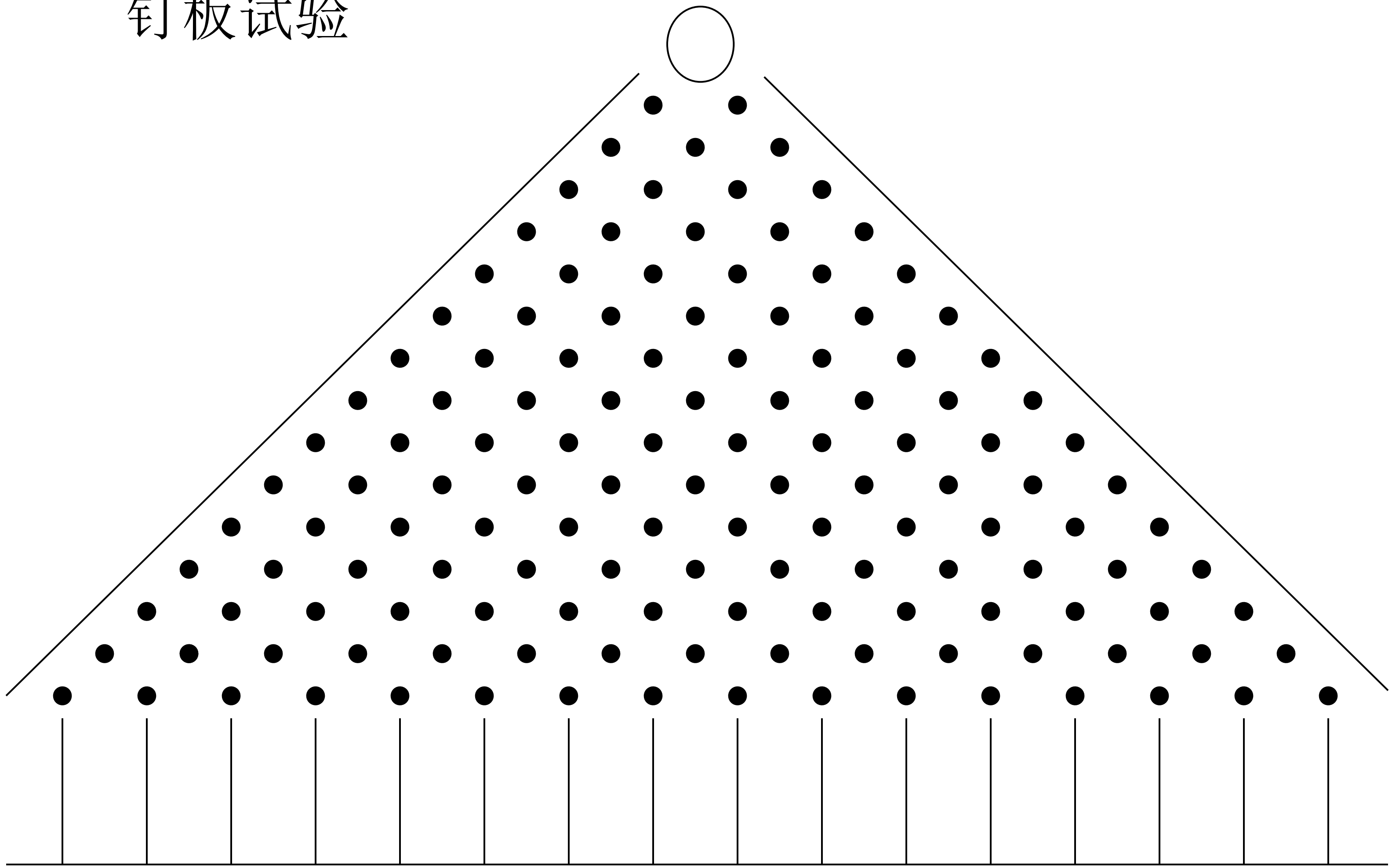
即$\xi_1,\xi_2,\ldots,\xi_n$的算术平均值依概率收敛于a

实际应用中，对某一量a，在不变条件下重复测量n次，得到观察值$x_1,\ldots,x_n$

当n充分大时，可用$\frac{1}{n}\sum_{i=1}^{n}x_i$作为a的近似值。

§3 中心极限定理

钉板试验



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。

一般地，若某项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，则这些大量独立偶然因素总和的随机变量近似服从正态分布。

设$\xi_1,\xi_2,\ldots$相互独立，$E\xi_i = a_i, D\xi_i = \sigma_i^2$

若每个ξ_i对总和$\xi = \sum_{i=1}^{n} \xi_i$影响不太大，则当n很大时，$\xi$近似服从正态分布。

由于$E\xi = \sum_{i=1}^{n} E\xi_i = \sum_{i=1}^{n} a_i, D\xi = \sum_{i=1}^{n} D\xi_i = \sum_{i=1}^{n} \sigma_i^2$

$$\xi \sim N\left(\sum_{i=1}^{n} a_i, \sum_{i=1}^{n} \sigma_i^{\ 2}\right) \qquad \frac{\xi - \sum_{i=1}^{n} a_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^{\ 2}}} \sim N(0,1)$$

这就是如下的李雅普诺夫定理：

- 定理1　设ξ_1，ξ_2,...相互独立，$E\xi_i = a_i, D\xi_i = \sigma_i^{\ 2}$

若某个ξ_i对总和$\sum_{i=1}^{n} \xi_i$影响不大，令$S_n = \sqrt{\sum_{i=1}^{n} \sigma_i^{\ 2}}$,则

$$\lim_{n \to \infty} P\left(\frac{1}{S_n} \sum_{i=1}^{n} (\xi_i - a_i) \le x\right) = \Phi_0(x)$$

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

解：设第i个螺丝钉重量为ξ_i，一盒重量为$\xi=\sum_{i=1}^{100}\xi_i$

$\xi_1, \ldots, \xi_{100}$相互独立，$E\xi_i = 0.1, D\xi_i = 0.1^2$

$$E\xi = \sum_{i=1}^{100} E\xi_i = 100(\text{两}) \qquad D\xi = \sum_{i=1}^{100} D\xi_i = 1$$

ξ近似服从正态分布$\xi \sim N(100,1)$

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

$$\approx 1 - \Phi_0(2) = 0.02275$$

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颗炸弹命中目标的概率。

解：第i次轰炸命中目标的次数为ξ_i

$$100\text{次轰炸命中目标的次数}\ \xi=\sum_{i=1}^{100}\xi_i$$

$$E\xi=\sum_{i=1}^{100}E\xi_i=200 \quad D\xi=\sum_{i=1}^{100}D\xi_i=169 \quad \sqrt{D\xi}=13$$

$$\xi\sim N(200,13^2)$$

$$P(180\le\xi\le 220)=P\left(\frac{|\xi-200|}{13}\le\frac{20}{13}\right)$$

$$=2\Phi_0(1.54)-1=0.87644$$

- 例3 某大型商场每天接待顾客10000人，设某位顾客的消费额(元)服从[100,1000]上的均匀分布，且顾客的消费额是独立的，试求该商场的销售额在平均销售额上、下浮动不超过20000元的概率。

解：第i位顾客消费额位ξ_i，商场销售额为ξ

$$\xi = \sum_{i=1}^{10000} \xi_i$$

$$E\xi_i = 550 \qquad D\xi_i = \frac{1}{12}(1000-100)^2 = \frac{900^2}{12}$$

$$E\xi = 5500000, D\xi = 10000 \times \frac{900^2}{12}$$

$$P(5500000-20000<\xi<5500000+20000)$$

$$=P\left(\left|\frac{\xi-5500000}{100\times 900 \big/ \sqrt{12}}\right|\leq\frac{20000}{100\times 900 \big/ \sqrt{12}}\right)$$

$$=2\Phi_0(0.77)-1$$

$$\approx 0.56$$

二项分布可以看成多个0-1分布之和

当n增加时，它以正态分布为极限。

- 定理2 (拉普拉斯定理)

(1)局部极限定理：当$n \to \infty$时

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

(2)积分极限定理：当$n \to \infty$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4 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：设同时停机的数目为ξ，它服从二项分布

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

(1)直接计算

$$P(\xi=3)=C_{10}^{3}0.2^{3}0.8^{7}\approx 0.2013$$

(2)用局部极限定理

$$P(\xi=3)=\frac{1}{\sqrt{npq}}\varphi_0\left(\frac{k-np}{\sqrt{npq}}\right)=\frac{1}{1.265}\varphi_0\left(\frac{3-2}{1.265}\right)$$

$$=\frac{1}{1.265}\varphi_0(0.79)=0.2308$$

相差较大，这是因为n较小。 一般要求$n\geq 30$

- 例5 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:500发炮弹中命中飞机的数目ξ服从二项分布

n=500 p=0.01

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

(1)直接计算

$$P(\xi = 5) = C_{500}^{5} 0.01^{5} 0.09^{495} = 0.17635$$

(2)用局部极限定理

$$P(\xi = 5) = \frac{1}{\sqrt{npq}} \varphi_0\left(\frac{k - np}{\sqrt{npq}}\right) = \frac{1}{2.225} \varphi_0\left(\frac{5 - 5}{2.225}\right)$$

$$= \frac{1}{2.225} \varphi_0(0) = 0.1793$$

(3)由于n很大，p很小，也可用Poisson分布计算

$\lambda = np = 5$ 查表得

$P_5(5)=0.175467$

比用正态分布更精确

正态分布与Poisson分布都是二项分布的极限分布。

用正态分布要求：$n \to \infty$

用Poisson分布要求：$n \to \infty, p \to 0, np \to \lambda$

对n很大，p(或q)很小的二项分布（$np \le 5$）

用Poisson分布近似计算比用正态分布精确

实际应用更多的是积分极限定理

- 例6 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： 废品数ξ服从二项分布

$$n=10000 \qquad p=0.005$$

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

ξ近似服从正态分布$N(50, 7.053^2)$

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

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

- 例7 某单位有200台电话分机，每台大约有5%时间使用外线。若各分机是否使用外线是相互独立的，问总机至少要装多少条外线才能使打外线的接通率达到90%？

解：用ξ表示需使用外线的分机数，它服从二项分布。

$$n=200, p=0.05, np=10, npq=9.5$$

设要装k条外线。

$$P(\xi\le k)=P\left(\frac{\xi-10}{\sqrt{9.5}}\le\frac{k-10}{\sqrt{9.5}}\right)=\Phi_0\left(\frac{k-10}{\sqrt{9.5}}\right)\ge 0.9$$

故 $\dfrac{k-10}{\sqrt{9.5}}\ge 1.30$

$\therefore k\ge 14$　　至少要装14条外线