5.1. 大数定律 Law of large numbers

切比雪夫不等式(Chebyshev's Inequality)

若随机变量 X 的期望值 $\mu = E(X)$ 方差 $\sigma^2 = Var(X)$, 那么

$$P\left(|X-\mu| \geq k\sigma\right) \leq \frac{1}{k^2}, k > 0$$

或

$$P\left(|X-\mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2}, \varepsilon > 0$$

随机变量 X 落在其期望值两侧 k 倍标准差距离之外的概率至多是 $1/k^2$.

证明:

令 X 为连续型随机变量，概率密度为 $f(x)$

$$\sigma^2 = Var(X) = E([X-\mu]^2) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,dx$$

$$= \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)\,dx + \int_{\mu-k\sigma}^{\mu+k\sigma} (x-\mu)^2 f(x)\,dx + \int_{\mu+k\sigma}^{+\infty} (x-\mu)^2 f(x)\,dx$$

$$\geq \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)\,dx + \int_{\mu+k\sigma}^{+\infty} (x-\mu)^2 f(x)\,dx$$

$$\geq \int_{-\infty}^{\mu-k\sigma} (k\sigma)^2 f(x)\,dx + \int_{\mu+k\sigma}^{+\infty} (k\sigma)^2 f(x)\,dx$$

$$= (k\sigma)^2 P\,(|X-\mu| \geq k\sigma)$$

因此, $P\,(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}\,, \varepsilon > 0$

例 5.1 设 X 表示掷一颗骰子所出现的点数，若给定 $\varepsilon=1,2$，请实际计算 $P\{|X-E(X)|\geqslant\varepsilon\}$，并验证切比雪夫不等式成立.

$$P\left(|X-\mu|\geq\varepsilon\right)\leq\frac{\sigma^2}{\varepsilon^2},\varepsilon>0$$

解 因为 X 的分布律是 $P\{X=k\}=\frac{1}{6}, k=1,2,\cdots,6$，所以

$$E(X)=\frac{7}{2},\quad D(X)=\frac{35}{12},$$

$$P\left\{\left|X-\frac{7}{2}\right|\geqslant 1\right\}=P\{X=1\}+P\{X=2\}+P\{X=5\}+P\{X=6\}=\frac{2}{3};$$

$$P\left\{\left|X-\frac{7}{2}\right|\geqslant 2\right\}=P\{X=1\}+P\{X=6\}=\frac{1}{3}.$$

$\varepsilon=1$: $\frac{D(X)}{\varepsilon^2}=\frac{35}{12}>\frac{2}{3}$,　　$\varepsilon=2$: $\frac{D(X)}{\varepsilon^2}=\frac{1}{4}\times\frac{35}{12}=\frac{35}{48}>\frac{1}{3}$.

例 5.2 设电站供电网有 10 000 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 6 800 盏与 7 200 盏之间的概率.

解 设 X 表示在夜晚同时开着的灯的数目，它服从参数为 $n=10\,000$，$p=0.7$ 的二项分布.若要准确计算，应该用伯努利公式：

$$P\{6\,800<X<7\,200\}=\sum_{k=6\,801}^{7\,199}C_{10\,000}^{k}\times(0.7)^{k}\times(0.3)^{10\,000-k}.$$

如果用切比雪夫不等式估计：

$$E(X)=np=10\,000\times0.7=7\,000,$$

$$D(X)=npq=10\,000\times0.7\times0.3=2\,100,$$

$$P\{6\,800<X<7\,200\}=P\{|X-7\,000|<200\}\geqslant1-\frac{2\,100}{200^{2}}\approx0.95.$$

事实上，切比雪夫不等式的估计只说明概率大于 0.95，后面将[illegible]率约为 0.999 99.切比雪夫不等式在理论上具有重大意义，但估计的精确度不高.

定义 5.1　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 ε，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 $Y_n \xrightarrow{P} a$.

或 随机变量序列 Y_n 以概率1收敛到常数 a，如果对任意 $\varepsilon > 0$

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

几种等价说法:

以概率1收敛， convergence with probability 1

几乎必然收敛 convergence almost surely,

convergence almost always,

几乎处处收敛 convergence almost everywhere

切比雪夫大数定律

假设随机变量序列 $\{X_n\colon n \geq 1\}$ 相互独立，对所有的$i = 1,2,\ldots,$
$E(X_i)$ 及 $Var(X_i) < l$ 都存在，则对任给的 $\varepsilon > 0$，

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

特殊情况：随机变量序列 $\{X_n\colon n \geq 1\}$ 相互独立，且对$i = 1,2,\ldots,$
$E(X_i) = \mu$，$Var(X_i) = \sigma^2$，那么对任意的 $\varepsilon > 0$，有

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

切比雪夫大数定律说明: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.这意味着,经过算术平均以后得到的随机变量 $\frac{\sum_{i=1}^{n} X_i}{n}$ 将比较密地聚集在它的数学期望 $\frac{\sum_{i=1}^{n} E(X_i)}{n}$ 的附近,它与数学期望之差依概率收敛到 0.

推论（切比雪夫大数定律的特殊情况） 设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，且具有相同的数学期望和方差：

$$E(X_k) = \mu,\ D(X_k) = \sigma^2,\quad k = 1, 2, \cdots.$$

作前 n 个随机变量的算术平均 $Y_n = \frac{1}{n}\sum_{k=1}^{n} X_k$，则对于任意正数 ε，有

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

辛钦大数定律 (Khinchin′s law of large numbers)

设随机变量序列 $\{X_n: n \geq 1\}$ 相互独立，服从同一分布，且具有相同的数学期望 $E(X_i) = \mu, k = 1,2,3, \ldots$则对任意的正数 $\varepsilon > 0$,

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

应用：在不变的条件下多次重复测量某一物理量，取其算术平均值为要测量的物理量的值。

注 定理 5.2 及其推论中要求随机变量 $X_k (k = 1,2,\cdots,n)$ 的方差存在. 但在随机变量服从同一分布的情形下，并不需要这一要求

伯努利大数定理 Bernoulli′s law of large numbers

设 n_A 是 n 次独立重复试验中事件 A 发生的次数，p 为事件 A 在每次试验中发生的概率，则对任意的正数 $\varepsilon > 0$，

$$\lim_{n\to\infty} P\left(\left|\frac{n_A}{n} - p\right| < \varepsilon\right) = 1 \quad \text{or,} \quad \lim_{n\to\infty} P\left(\left|\frac{n_A}{n} - p\right| \geq \varepsilon\right) = 0$$

相对频率 $\frac{n_A}{n}$ 以概率1收敛到 p (converges to p with probability 1).

伯努利大数定律告诉我们，事件 A 发生的频率 $\frac{n_A}{n}$ 依概率收敛于事件 A 发生的概率 p，因此，本定律从理论上证明了大量重复独立试验中，事件 A 发生的频率具有稳定性

伯努利大数定理 是辛钦大数定律的特殊情况.

第二节 中心极限定理

定理 5.5（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：

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

则随机变量

$$Y_n=\frac{\sum\limits_{k=1}^{n}X_k-E\left(\sum\limits_{k=1}^{n}X_k\right)}{\sqrt{D\left(\sum\limits_{k=1}^{n}X_k\right)}}=\frac{\sum\limits_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}$$

的分布函数 $F_n(x)$ 对于任意 x，满足：

$$\lim_{n\to\infty}F_n(x)=\lim_{n\to\infty}P\left\{\frac{\sum\limits_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}\leqslant x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt.\quad(5-7)$$

可知，当 n 充分大时，近似地，有 $$Y_n = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n\sigma^2}} \sim N(0,1).$$

或者说，当 n 充分大时，近似地，有 $$\sum_{k=1}^{n} X_k \sim N(n\mu, n\sigma^2)$$

如果用 $X_1, X_2, \cdots, X_n$ 表示相互独立的各随机因素. 假定它们都服从相同的分布（不论服从什么分布），且都有有限的期望与方差（每个因素的影响有一定限度），则当 n 充分大时，

$\sum_{k=1}^{n} X_k$ 便近似地服从正态分布

例 5.3 一个螺丝钉的重量是一个随机变量,期望值是 100 克,标准差是 10 克.求一盒(100 个)同型号螺丝钉的重量超过 10.2 千克的概率.

解 设一盒重量为 X,盒中第 i 个螺丝钉的重量为 X_i,$i=1,2,\cdots,100$. $X_1,X_2,\cdots,X_{100}$ 相互独立,$E(X_i)=100$,$\sqrt{D(X_i)}=10$,则有 $X=\sum_{i=1}^{100}X_i$,且

$$E(X)=100\cdot E(X_i)=10\ 000(\text{克}),\quad \sqrt{D(X)}=100(\text{克}).$$

根据定理 5.5,有

$$P\{X>10\ 200\}=P\left\{\frac{X-10\ 000}{100}>\frac{10\ 200-10\ 000}{100}\right\}=1-P\left\{\frac{X-10\ 000}{100}\leqslant 2\right\}$$

$$\approx 1-\Phi(2)=1-0.977\ 2=0.022\ 8.$$

例 5.4 对敌人的防御地进行100次轰炸，每次轰炸命中目标的炸弹数目是一个随机变量，其期望值是2，方差是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 令第 i 次轰炸命中目标的炸弹数为 X_i，100次轰炸中命中目标的炸弹数 $X=\sum_{i=1}^{100} X_i$，应用定理5.5，X 渐近服从正态分布，期望值为200，方差为169，标准差为13. 所以

$$P\{180 \leqslant X \leqslant 220\} = P\{|X-200| \leqslant 20\} = P\left\{\left|\frac{X-200}{13}\right| \leqslant \frac{20}{13}\right\}$$

$$\approx 2\Phi(1.54)-1 = 0.8764.$$

Remarks:

1. The central limit theorem states that the probability of the average of $X_1, X_2, \ldots, X_n$, properly scaled (i.e., with subtraction of $n\mu$ and then division by $\sqrt{n}\sigma$), being less than "x," will converge to the cumulative distribution function of a standard Normal random variable, evaluated at x.

2. We do not need the random variables $X_1, X_2, \ldots,$ to be Normal..

3. This theorem basically says that sums of n independent random variables (of any type) are distributed similarly to a Normal random variable when n is large. (There is no minimum n necessary before the CLT applies, but the CLT is more effective, the larger n is.)

定理 5.6（李雅普诺夫（Lyapunov）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：

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

记 $B_n^2=\sum_{k=1}^{n}\sigma_k^2$，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

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

的分布函数 $F_n(x)$ 对于任意 x，满足：

$$\lim_{n\to\infty}F_n(x)=\lim_{n\to\infty}P\left\{\frac{\sum_{k=1}^{n}X_k-\sum_{k=1}^{n}\mu_k}{B_n}\leqslant x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t.$$

这个定理说明，随机变量 $$Z_n = \frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} \mu_k}{B_n}$$

当 n 很大时，近似地服从正态分布 $N(0,1)$. 因此，当 n 很大时，

$$\sum_{k=1}^{n} X_k = B_n Z_n + \sum_{k=1}^{n} \mu_k$$

近似地服从正态分布 $N\left(\sum_{k=1}^{n} \mu_k, B_n^2\right)$.

这表明无论随机变量 $X_k (k = 1,2,\cdots)$ 具有怎样的分布，

只要满足定理的条件，它们的和 $\sum_{k=1}^{n} X_k$

当 n 很大时，就近似地服从正态分布.

下面介绍另一个中心极限定理.

定理 5.7 设随机变量 X 服从参数为 $n,p\ (0<p<1)$ 的二项分布，则

1°（拉普拉斯(Laplace)定理） 局部极限定理：当 $n\to\infty$ 时，

$$P\{X=k\}\approx\frac{1}{\sqrt{2\pi npq}}\mathrm{e}^{-\frac{(k-np)^2}{2npq}}=\frac{1}{\sqrt{npq}}\varphi\left(\frac{k-np}{\sqrt{npq}}\right),\qquad(5-10)$$

其中 $p+q=1,k=0,1,2,\cdots,n,\varphi(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}$.

2°（棣莫弗-拉普拉斯(De Moivre -Laplace)定理） 积分极限定理：对于任意的 x，恒有

二项分布以正态分布为极限.

$$\lim_{n\to\infty}P\left\{\frac{X-np}{\sqrt{np(1-p)}}\leqslant x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t.\qquad(5-11)$$

例 5.5 10台机器独立工作，每台停机的概率为 0.2，求 3 台机器同时停机的概率.

解 10台机器中同时停机的数目 X 服从二项分布，$n=10, p=0.2$，则 $np=2$，$\sqrt{npq}\approx 1.265$.

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

(2) 若用局部极限定理近似计算：

$$P\{X=3\}\approx\frac{1}{\sqrt{npq}}\varphi\left(\frac{k-np}{\sqrt{npq}}\right)=\frac{1}{1.265}\varphi\left(\frac{3-2}{1.265}\right)\approx\frac{1}{1.265}\varphi(0.79)\approx 0.2308.$$

(2) 的计算结果与(1)相差较大，这是由于 n 不够大.

Laplacian Theory (拉普拉斯定理)

$$P(X=k)\approx\frac{1}{\sqrt{np(1-p)}}\varphi\left(\frac{k-np}{\sqrt{np(1-p)}}\right)$$

where $k=0,1,2,\dots,n$

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

例 5.2 设电站供电网有 10 000 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 6 800 盏与 7 200 盏之间的概率.

应用定理 5.7 De Moivre-Laplacian Theory (棣莫弗-拉普拉斯定理)

$\boldsymbol{X \sim b(n,p)}$.

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

解 $np = 7\,000$，$\sqrt{npq} \approx 45.83$，则

$$P\{6\,800 < X < 7\,200\} = P\{|X - 7\,000| < 200\}$$

$$\approx P\left\{\left|\frac{X-7\,000}{45.83}\right| < 4.36\right\}$$

$$\approx 2\Phi(4.36) - 1 = 0.999\,99.$$

例 5.7 产品为废品的概率为 $p=0.005$，求 10 000 件产品中废品数不大于 70 的概率.

解 10 000 件产品中的废品数 X 服从二项分布，$n=10\,000$，$p=0.005$，$np=50$，$\sqrt{npq}\approx 7.053$，则

$$P\{X\leqslant 70\}\approx\Phi\left(\frac{70-50}{7.053}\right)\approx\Phi(2.84)=0.997\,7.$$

正态分布和泊松分布虽然都是二项分布的极限分布，但后者以 $n\to\infty$，同时 $p\to 0$，$np\to\lambda$ 为条件，而前者则只要求 $n\to\infty$ 这一条件. 一般说来，对于 n 很大，p(或 q) 很小的二项分布($np\leqslant 5$) 用正态分布来近似计算不如用泊松分布计算精确. 见下例

例 5.8 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炮弹数目 X 服从二项分布,$n=500$,$p=0.01$,$np=5$,$\sqrt{npq}\approx 2.2$. 下面用 3 种方法计算并加以比较.

(1) 用二项分布公式计算:

$$P\{X=5\}=\mathrm{C}_{500}^{5}\times(0.01)^{5}\times(0.99)^{495}\approx 0.176\,35.$$

(2) 用泊松公式计算,直接查表可得:

$$np=\lambda=5,\quad k=5,\quad P_{5}(5)\approx 0.175\,467.$$

(3) 用拉普拉斯局部极限定理计算:

$$P\{X=5\}\approx\frac{1}{\sqrt{npq}}\varphi\left(\frac{5-np}{\sqrt{npq}}\right)\approx 0.179\,3.$$

可见,后者不如前者精确.