定理3.5.4　切比雪夫不等式

定理　设随机变量 X 具有数学期望 $E(X)$,方差 $D(X)$, 则对于任意正数 ε, 不等式

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

成立.

切比雪夫不等式

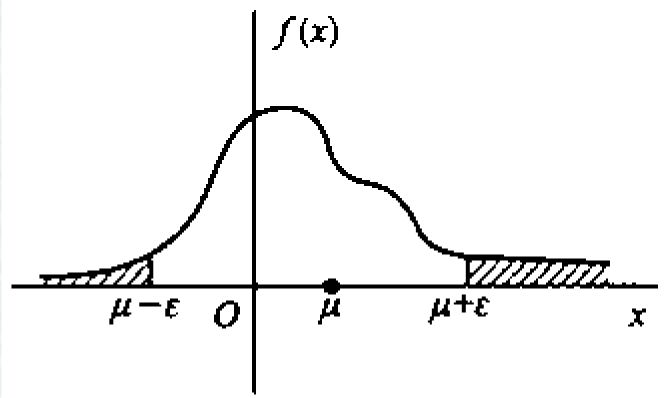
切比雪夫不等式也可以写成

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

证明 取连续型随机变量的情况来证明.

设 X 的概率密度为 $f(x)$,记$E(X)=\mu$,则有

$$P\{|X-\mu|\geq\varepsilon\}=\int_{|x-\mu|\geq\varepsilon} f(x)\,\mathrm{d}x$$



$$\leq\int_{|x-\mu|\geq\varepsilon}\frac{|x-\mu|^2}{\varepsilon^2}f(x)\,\mathrm{d}x$$

$$\leq\frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty}(x-\mu)^2 f(x)\,\mathrm{d}x=\frac{1}{\varepsilon^2}D(X).$$

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

例3.5.5 一加法器同时收到 20 个噪声电压 V_k $(k=1,2,\cdots 20)$，设它们是相互独立的随机变量，且都在区间 $(0,10)$ 上服从均匀分布，记 $V=\sum_{k=1}^{20} V_k$，试用切比雪夫不等式估计 $P\{80<V<120\}$.

解 $\because E(V_k)=5,\quad D(V_k)=\dfrac{100}{12}\ (k=1,2,\cdots,20).$

$$E(V)=E\left(\sum_{k=1}^{20} V_k\right)=20E(V_k)=20\times 5=100$$

$$D(V)=D\left(\sum_{k=1}^{20} V_k\right)=20D(V_k)=20\times\frac{100}{12}=\frac{500}{3}$$

$$E(V)=E\left(\sum_{k=1}^{20}V_k\right)=20E(V_k)=20\times5=100$$

$$D(V)=D\left(\sum_{k=1}^{20}V_k\right)=20D(V_k)=20\times\frac{100}{12}=\frac{500}{3}$$

故由切比雪夫不等式知

$$P\{80<V<120\}=P\{|V-100|<20\}$$

$$\geq 1-\frac{500/3}{20^2}=\frac{7}{12}$$

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

注 3.2.3 切比雪夫不等式的意义在于,只需根据随机变量的两个数字特征:期望和方差,即可对随机变量的取值偏离均值超过ε做一个估计,却不涉及到X的分布.但需指出的是,正是因为不涉及X的分布,用切比雪夫估计出的值是比较粗糙的,比如,设$X\sim N(\mu,\sigma^2)$,根据$3\sigma-$准则,

$$P(|X-\mu|\geq 3\sigma)\approx 0.003,$$

但由切比雪夫不等式

$$P(|X-\mu|\geq 3\sigma)\leq \frac{\sigma^2}{(3\sigma)^2}=\frac{1}{9},$$

因此切比雪夫不等式只是在一些对精度要求不太高的情况下使用起来比较方便.

第四章　大数定律与中心极限定理

第一节　大数定律

第二节　中心极限定理

第一节　大数定律

一、问题的引入

二、基本定理

- 伯努利大数定律
- 切比雪夫大数定律
- 辛钦大数定律

$x_1, x_2, \cdots, x_n$

一、问题的引入

实例1 测量一个物体的长度，真实值为a，记录下n次测量结果：$x_1, x_2, \cdots, x_n$，当n很大时，算术平均值 $\frac{1}{n}\sum_{i=1}^{n} x_i \approx a$ 。

实例2 频率的稳定性

随着试验次数的增加, 事件发生的频率逐渐稳定于某个常数.

启示:从实践中人们发现大量测量值的算术平均值有稳定性.

定义4.1.1

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$\{Y_n\}$依概率收敛于a, 记为$Y_n \xrightarrow{P} a.$

伯努利大数定律

设 n_A 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

$$\lim_{n\to\infty} P\left\{\left|\frac{n_A}{n} - p\right| < \varepsilon\right\} = 1 \text{ 或 } \lim_{n\to\infty} P\left\{\left|\frac{n_A}{n} - p\right| \geq \varepsilon\right\} = 0.$$

亦称 $\dfrac{n_A}{n} \xrightarrow{P} p.$

分析 引入随机变量

$$X_k = \begin{cases} 0, & \text{若在第} k \text{次试验中 } A \text{ 不发生}, \\ 1, & \text{若在第} k \text{次试验中 } A \text{ 发生}, k = 1, 2, \cdots. \end{cases}$$

显然 $n_A = X_1 + X_2 + \cdots + X_n \sim b(n, p)$

$$E\left(\frac{n_A}{n}\right) = p, \quad D\left(\frac{n_A}{n}\right) = \frac{p(1-p)}{n}$$

由切比雪夫不等式可得

$$P\left\{\left|\frac{n_A}{n} - p\right| \geq \varepsilon\right\} \leq \frac{p(1-p)}{n\varepsilon^2}$$

在上式中令 $n \to \infty$, 并注意到概率大于等于0, 则

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

显然 $n_A = X_1 + X_2 + \cdots + X_n \sim b(n, p)$

这里$X_1, X_2, \cdots, X_n, \cdots$是相互独立的，

且X_k服从以 p 为参数的 $(0-1)$ 分布，

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

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

即 $$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}(X_1 + X_2 + \cdots + X_n) - p\right| < \varepsilon\right\} = 1,$$

亦即 $$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - E\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right)\right| < \varepsilon\right\} = 1,$$

切比雪夫大数定律

可换成“相互独立”

设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 两两不相关，

期望 $E(X_k)$和方差$D(X_k)$存在，若存在常数C,

使得$D(X_k) \leq C$，$(k = 1,2,\cdots)$,

则对于任意正数ε,有

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

即 $\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k) \xrightarrow{P} 0$

辛钦大数定律

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

即 $\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \mu$

$$E\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n} E(X_k) = \frac{1}{n}\cdot n\mu = \mu$$

大数定律描述了在某种条件下，n个随机变量的算术平均值与其数学期望之差 $\frac{1}{n}\sum_{k=1}^{n} X_k - E\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right)$ 依概率收敛到0。

n个r.v.的算术平均值，它仍是一个r.v.；但当试验次数n无限增大时，此r.v.将趋向于某个常数，即它的数学期望（随机性消失了）。

第二节　中心极限定理

一、基本定理

二、典型例题

二、基本定理

定理4.2.1（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

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

的分布函数 $F_n(x)$ 对于任意 x 满足

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

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

定理4. 2. 1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4.2.1也表明：当n充分大时

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \overset{\text{近似}}{\sim} N(0,1)$$

由此可见,当n充分大时

$$\sum_{i=1}^{n} X_i \overset{\text{近似}}{\sim} N\left(n\mu, n\sigma^2\right)$$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \overset{\text{近似}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

定理4.2.2（德莫佛—拉普拉斯定理）

设随机变量 η_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

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

证明 因为 $\eta_n \sim b(n, p)$，所以 $\eta_n = \sum_{k=1}^{n} X_k$,

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

$$P(X_k = 1) = p, \quad P(X_k = 0) = 1 - p$$

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

根据定理一得

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

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

定理4. 2. 2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.

补：李雅普诺夫中心极限定理

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

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

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

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

中心极限定理表明, 在相当一般的条件下, 当独立随机变量的个数n充分大时， 不论随机变量服从什么分布，其和 $\sum_{i=1}^{n} X_i$ 近似服从正态分布 $N\left(E\left(\sum_{i=1}^{n} X_i\right), D\left(\sum_{i=1}^{n} X_i\right)\right)$.

例4.2.1 计算机进行加法计算时，把每个加数取最接近它的整数来计算．设所有的取整误差是相互独立的随机变量，且都服从 $[-0.5,0.5]$ 上的均匀分布．若将1500个数相加，求误差总和的绝对值超过15的概率.

解 设X_i表示第i个数的取整误差, $i=1,2,\ldots,1500$

则$X_i \sim U[-0.5,0.5]$, $E(X_i)=0$, $D(X_i)=\dfrac{1}{12}$.

又设X表示误差总和，则$X=\sum\limits_{i=1}^{1500} X_i$

由中心极限定理知：X近似服从正态分布

$$X=\sum_{i=1}^{1500}X_i \overset{\text{近似}}{\sim} N\left(0,\ \frac{1500}{12}\right)$$

$$\therefore P\{|X|>15\}=1-P\{-15\le X\le 15\}$$

$$\approx 1-\left[\Phi\left(\frac{15}{\sqrt{1500/12}}\right)-\Phi\left(\frac{-15}{\sqrt{1500/12}}\right)\right]$$

$$\approx 1-\left[2\Phi(1.34)-1\right]$$

$$=1-(2\times 0.9099-1)=0.1802$$

例4.2.2 设某保险公司为某企业的一万名员工设计了一款公共交通意外保险，每人每年支付120元保费. 已知在一年内这些人死亡的概率为0.006，投保人死亡后，保险公司需向家属支付10000元，试求：

（1）保险公司一年的利润不少于60万元的概率;

（2）保险公司亏本的概率.

解 设参加保险的一万人中一年内死亡的人数为 X,

则 $X \sim b(10\,000, 0.006)$.

（1）所求概率为

$$P\{120\times 10000-10000X\geq 600000\}=P\{0\leq X\leq 60\}$$

直接计算很麻烦

由**德莫佛－拉普拉斯中心极限定理**知：

$$X\overset{\text{近似}}{\sim}N(10000\times 0.006,\quad 10000\times 0.006\times 0.994)$$

$$\therefore P\{0\leq X\leq 60\}\approx\Phi\left(\frac{60-60}{\sqrt{59.64}}\right)-\Phi\left(\frac{0-60}{\sqrt{59.64}}\right)$$

$$\approx\Phi(0)-\Phi(-7.77)\approx 0.5-0=0.5$$

（2）所求概率为

$$P\{120\times 10000-10000X<0\}=P\{X>120\}$$

由**德莫佛一拉普拉斯中心极限定理**知：

$$X\overset{\text{近似}}{\sim}N\left(10000\times 0.006,\quad 10000\times 0.006\times 0.994\right)$$

$$\therefore P\{X>120\}\approx 1-\Phi\left(\frac{120-60}{\sqrt{59.64}}\right)\approx 1-\Phi(7.77)$$

$$\approx 1-1=0.$$

例4.2.3 某单位内部有260架电话分机, 每个分机有4%的时间要用外线通话， 可以认为各个电话分机是否用外线是相互独立的, 问总机至少要备有多少条外线才能以95%以上的把握保证各个分机在外线时不必等候?

解 设X表示260架电话分机中同时使用外线的分机数,

则$X \sim b(260,\ 0.04)$

要求至少需备x条外线,使得$P(X \le x) \ge 0.95$

$X \sim b(260,\ 0.04)$

由**德莫佛－拉普拉斯中心极限定理**知：

$$X \overset{近似}{\sim} N\left(260\times 0.04,\quad 260\times 0.04\times 0.96\right)$$

$$P\{X \le x\} \approx \Phi\left(\frac{x-260\times 0.04}{\sqrt{260\times 0.04\times 0.96}}\right) \ge 0.95$$

查表 $\Phi(1.65) = 0.9505$,

故 $\dfrac{x-260\times 0.04}{\sqrt{260\times 0.04\times 0.96}} \ge 1.65$,　　得 $x \ge 15.61$.

第四章 大数定律及中心极限定理
习 题 课

一、主要内容

二、重点与难点

三、典型例题

一、主要内容

大数定律

- 辛钦大数定律
- 伯努利大数定律
- 切比雪夫大数定律

中心极限定理

- 独立同分布的中心极限定理
- 德莫佛－拉普拉斯中心极限定理

二、重点与难点

理解大数定律，记住其结论

中心极限定理及其运用.

三、典型例题

例1 一加法器同时收到 20 个噪声电压 V_k $(k=1,2,\cdots 20)$, 设它们是相互独立的随机变量, 且都在区间 $(0,10)$ 上服从均匀分布, 记 $V=\sum_{k=1}^{20} V_k$, 求 $P\{V>105\}$ 及 $P\{80<V<120\}$的近似值.

解 $\because E(V_k)=5,\quad D(V_k)=\dfrac{100}{12}\ (k=1,2,\cdots,20).$

由中心极限定理知：V近似服从正态分布

$$E(V)=E\left(\sum_{k=1}^{20}V_k\right)=20E(V_k)=20\times5=100$$

$$D(V)=D\left(\sum_{k=1}^{20}V_k\right)=20D(V_k)=20\times\frac{100}{12}=\frac{500}{3}$$

$$V\overset{\text{近似}}{\sim}N\left(100,\quad\frac{500}{3}\right)$$

$$\therefore P\{V>105\}\approx1-\Phi\left(\frac{105-100}{\sqrt{500/3}}\right)$$

$$\approx1-\Phi(0.39)=0.3483.$$

$$V \overset{\text{近似}}{\sim} N\left(100, \quad \frac{500}{3}\right)$$

$$P\{80 < V < 120\} \approx \Phi\left(\frac{120-100}{\sqrt{500/3}}\right) - \Phi\left(\frac{80-100}{\sqrt{500/3}}\right)$$

$$\approx \Phi(1.55) - \Phi(-1.55) \quad = 2\Phi(1.55) - 1$$

$$= 2 \times 0.9394 - 1 = 0.8788$$

例2 一船舶在某海区航行, 已知每遭受一次海浪的冲击, 纵摇角大于 3° 的概率为1/3, 若船舶遭受了90 000次波浪冲击, 问其中有29 500～30 500次纵摇角大于 3° 的概率是多少?

解 在90 000次波浪冲击中纵摇角大于 3° 的次数为 X,

则 $X \sim b(90\,000, \frac{1}{3})$.

分布律为 $P\{X=k\} = C_{90000}^{k}\left(\frac{1}{3}\right)^{k}\left(\frac{2}{3}\right)^{90000-k}$,

$k = 0, 1, \cdots, 90\,000.$

所求概率为

$$P\{29500 < X \leq 30500\} = \sum_{k=29501}^{30500} \binom{90000}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{90000-k}.$$

直接计算很麻烦，利用**德莫佛－拉普拉斯定理**

$$X \overset{\text{近似}}{\sim} N\left(90000 \times \frac{1}{3},\quad 90000 \times \frac{1}{3} \times \frac{2}{3}\right)$$

$$\therefore P\{29500 < X \leq 30500\}$$

$$\approx \Phi\left(\frac{30500-30000}{\sqrt{20000}}\right) - \Phi\left(\frac{29500-30000}{\sqrt{20000}}\right)$$

$$\therefore P\{29500 < X \le 30500\}$$

$$\approx \Phi\left(\frac{30500-30000}{\sqrt{20000}}\right)-\Phi\left(\frac{29500-30000}{\sqrt{20000}}\right)$$

$$=\Phi\left(\frac{5\sqrt{2}}{2}\right)-\Phi\left(-\frac{5\sqrt{2}}{2}\right)$$

$$\approx 2\Phi(3.54)-1 = 2\times 0.9998-1=0.9996.$$

例3 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 在6000粒种子中良种的数目为 X,

则 $X \sim b(6000, \frac{1}{6})$. 根据题意, 所求概率为

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

$$X \sim b(6000, \frac{1}{6}).$$

由中心极限定理有: $X \overset{\text{近似}}{\sim} N\left(1000,\ 1000\times\frac{5}{6}\right)$,

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

$$\approx 2\Phi\left(\frac{6\sqrt{6}}{\sqrt{5000}}\right)-1 \approx 2\Phi(0.21)-1$$

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