§5 大数定律及中心极限定理

§5.1 大数定律

一、问题的引入

本章要解决的问题

答复

1. 为何能以某事件发生的频率作为该事件的 概率的估计?
2. 为何能以样本均值作为总体期望的估计?

大数定律

3. 为何正态分布在概率论中占有极其重要的地位?
4. 大样本统计推断的理论基础是什么?

中心极限定理

复习

方差的概率意义：

刻画了随机变量X的取值的分散（或集中）的程度

切比雪夫不等式

定理　设随机变量 X 具有数学期望 $E(X)=\mu$，方差 $D(X)=\sigma^2$，则对于任意正数 ε，不等式

$$P\{|X-\mu|\geq\varepsilon\}\leq\frac{\sigma^2}{\varepsilon^2}\quad\Leftrightarrow\quad P\{|X-\mu|<\varepsilon\}\geq 1-\frac{\sigma^2}{\varepsilon^2}.$$

切比雪夫不等式

成立.

易知 $D(X)$ 越大 X的取值越分散.

二、基本定理


切比雪夫

定理一（契比雪夫大数定理）

设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，具有数学期望 $E(X_k)$ 和方差 $D(X_k)$ $(k = 1, 2, \cdots)$，若存在一常数 c，使得 $D(X_k) \le c$，则对于任意给定的 $\varepsilon > 0$，恒有

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

定理一推论（契比雪夫定理的特殊情况）

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：$E(X_k)=\mu$, $D(X_k)=\sigma^2\ (k=1,2,\cdots)$, 作前 n 个随机变量的算术平均 $\overline{X}=\frac{1}{n}\sum_{k=1}^{n} X_k$，则对于任意正数 ε 有

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

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

$$D\left[\frac{1}{n}\sum_{k=1}^{n}X_k\right]=\frac{1}{n^2}\sum_{k=1}^{n}D(X_k)=\frac{1}{n^2}\cdot n\sigma^2=\frac{\sigma^2}{n},$$

由**契比雪夫不等式**可得

$$P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu\right|<\varepsilon\right\}\geq 1-\frac{\sigma^2/n}{\varepsilon^2},$$

在上式中令 $n\to\infty$, 并注意到概率不能大于1, 则

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

关于定理一的说明:

1、定理中$\{|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu|<\varepsilon\}$是指一个随机事件，当$n\to\infty$时，这个事件的概率趋 于1.

2、定理以数学形式证明了 随机变量$X_1,\cdots X_n$的算术平均$\overline{\mathrm{X}}=\frac{1}{n}\sum_{i=1}^{n}X_i$接近数学期望E（$\mathrm{X}_k$）$=\mu$（$k=1,2\cdots,n$），这种接近说明其具 有的稳定性 .

这种稳定性的含义说明 算术平均值是依概率收敛的意义下逼近某一 常数.

依概率收敛序列的定义

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

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

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

$$Y_n \xrightarrow{P} a.$$

依概率收敛序列的性质:

设 $X_n \xrightarrow{P} a,\quad Y_n \xrightarrow{P} b,$

又设函数 $g(x,y)$ 在点 (a,b) 连续,

则 $g(X_n, Y_n) \xrightarrow{P} g(a,b).$

请注意 ：

$\{X_n\}$依概率收敛于a，意味着对任意给定的$\varepsilon > 0$，当n充分大时，事件$|X_n - X| < \varepsilon$的概率很大，接近于1；并不排除事件$|X_n - X| \geq \varepsilon$的发生，而只是说他发 生的可能性很小 .

定理一的另一种叙述:

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: $E(X_k) = \mu$, $D(X_k) = \sigma^2$ $(k = 1, 2, \cdots)$, 则序列 $\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$ 依概率收敛于 μ, 即 $\overline{X} \xrightarrow{P} \mu$.

（这个接近是概率意义下的接近）

即在定理条件下, n个随机变量的算术平均, 当n无限增加时, 几乎变成一个常数.

问题 ：

设$\mathbf{n}_A$是n重贝努里试验中事件A发生的次数，p是事件A发生的概率，

$\frac{n_A}{n}$ 是事件A发生的频率.

伯努利

事件发生的频率能否代 替事件的概率，频率是否具有稳定性呢？

定理二（伯努利大数定理）

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

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

证明　　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 $n_A = X_1 + X_2 + \cdots + X_n$,

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

且 X_k 服从以 p 为参数的 $(0-1)$ 分布,

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

根据定理一有

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

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

关于伯努利定理的说明:

伯努利定理表明事件发生的频率 $\frac{n_A}{n}$ 依概率收敛于事件的概率 p, 它以严格的数学形式表达了频率的稳定性.

故而当 n 很大时, 事件发生的频率与概率有较大偏差的可能性很小. 在实际应用中, 当试验次数很大时, 便可以用事件发生的频率来代替事件的概率.

定理三（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 $\lim\limits_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| < \varepsilon\right\} = 1.$

关于辛钦定理的说明:

(1) 与定理一相比, 不要求方差存在;

(2) 伯努利定理是辛钦定理的特殊情况.

注意：大数定理都是独立随机变量之和的平均值依概率收敛问题

三、典型例题

例1 设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立,

具有如下分布律:

X_n	$-na$	0	na
P	$\frac{1}{2n^2}$	$1-\frac{1}{n^2}$	$\frac{1}{2n^2}$

问是否满足契比雪夫定 理?

解 独立性依题意可知, 检验是否具有数学期望?

$$E(X_n) = -na^2 \cdot \frac{1}{2n^2} + 0 \cdot (1 - \frac{1}{n^2}) + na^2 \cdot \frac{1}{2n^2} = 0,$$

说明每一个随机变量都有数学期望,

检验是否具有有限方差?

$$\because \begin{array}{c|ccc} X_n^{\ 2} & (na)^2 & 0 & (na)^2 \\ \hline P & \dfrac{1}{2n^2} & 1-\dfrac{1}{n^2} & \dfrac{1}{2n^2} \end{array}$$

$$\therefore E(X_n^2) = 2(na)^2 \cdot \frac{1}{2n^2} = a^2,$$

$$\therefore D(X_n) = E(X_n^2) - [E(X_n)]^2 = a^2.$$

说明离散型随机变量有有限方差,

故满足契比雪夫定理的条件.

例2 设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 独立同分布，且 $E(X_k)=0,\ D(X_k)=\sigma^2, k=1,2,\cdots$，证明对任意正数 ε 有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k^{\ 2}-\sigma^2\right|<\varepsilon\right\}=1.$$

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

所以 $X_1^{\ 2}, X_2^{\ 2}, \cdots, X_n^{\ 2}, \cdots$ 也是相互独立的，

由 $E(X_k)=0$，得 $E(X_k^{\ 2})=D(X_k)+[E(X_k)]^2=\sigma^2$，

由辛钦定理知

对于任意正数 ε，有 $\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k^{\ 2}-\sigma^2\right|<\varepsilon\right\}=1.$

例3 在一个罐子中,装有10个编号为0-9的同样的球，从罐中有放回地抽取若干次，每次抽一个，并记下号码.

设 $X_k = \begin{cases} 1 & \text{第}k\text{次取到号码 }0 \\ 0 & \text{否则} \end{cases}$ ，$k=1,2,\dots$

问对序列$\{X_k\}$能否应用大数定律？

解:

X	1	0
p_k	0.1	0.9

$E(X_k)=0.1,\quad k=1,2,\dots$

诸X_k独立同分布，且期望存在，故能使用大数定律. 即对任意的$\varepsilon>0$,

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{k=1}^{n} X_k - 0.1| < \varepsilon\} = 1$$

四、小结

三个大数定理

- 契比雪夫定理
- 伯努利大数定理
- 辛钦定理

频率的稳定性是概率定义的客观基础，而伯努利大数定理以严密的数学形式论证了频率的稳定性.

§ 5.2　中心极限定理

一、问题的引入

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,其概率分布情况如何呢?

自从高斯指出测量误差服从正态分布之后，人们发现，正态分布在自然界中极为常见.

高斯

如果一个随机变量是由大量相互独立的随机因素的综合影响所造成，而每一个别因素对这种综合影响中所起的作用不大. 则这种随机变量一般都服从或近似服从正态分布.

现在我们就来研究独立随机变量之和所特有的规律性问题.

当n无限增大时，这个和的极限分布是什么呢？

由于无穷个随机变量之和可能趋于∞，故我们不研究n个随机变量之和本身而考虑它的标准化的随机变量.即考虑随机变量$X_k(k=1,\cdots n)$的和$\sum_{k=1}^{n}X_k$

$$Y_n=\frac{\sum_{k=1}^{n}X_k-E(\sum_{k=1}^{n}X_k)}{\sqrt{D(\sum_{k=1}^{n}X_k)}}$$

讨论Y_n的极限分布是否为标准 正态分布

在概率论中，习惯于把和的分布收敛于正态分布这一类定理都叫做**中心极限定理**.

二、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：$E(X_k)=\mu, D(X_k)=\sigma^2, (k=1,2,\cdots)$，则随机变量之和 $\sum_{k=1}^{n} X_k$ 的标准化变量

$$Y_n = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$$ 的分布函数 $F_n(x)$ 对于任意 x 满足

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

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

注　1、定理表明，独立同分布的随机变量之和 $\sum_{k=1}^{n} X_k$，

当n充分大时，随机变量之和与其标准化变量分别有

$$\sum_{k=1}^{n} X_k \overset{\text{近似地}}{\sim} N(n\mu, n\sigma^2) \quad ; \qquad \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \overset{\text{近似地}}{\sim} N(0,1).$$

2、独立同分布中心极限定理的另一种形式可写为

$$\overline{X} \overset{\text{近似地}}{\sim} N(\mu, \sigma^2/n) \text{ 或 } \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \overset{\text{近似地}}{\sim} N(0,1), \text{ 其中} \overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

3、虽然在一般情况下，我们很难求出 $\sum_{k=1}^{n} X_k$ 的分布的确切形式，但当n很大时，可以求出近似分布.

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

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

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

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

定理二表明:

1、定理中随机变量之和 $\sum_{k=1}^{n} X_k$ 及其标准化变量 Z_n 在 n 很大时，分别近似服从

$$\sum_{k=1}^{n} X_k \overset{\text{近似地}}{\sim} N\left(\sum_{k=1}^{n}\mu_k, B_n^2\right);\quad Z_n \overset{\text{近似地}}{\sim} N(0,1)$$

2、随机变量 X_k 无论服从什么分布，只 要满足定理条件，随机变量之 和 $\sum_{k=1}^{n} X_k$ 当 n 很大时，就近似服从正态分布，这就 是为什么正态分布在概 率论中所占的重要地位的 一个基本原因.

定理三(德莫佛－拉普拉斯定理)（定理一特殊情况）

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

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

证明: 根据第四章第二节例6知(书P103): $\eta_n = \sum_{k=1}^{n} X_k$,

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

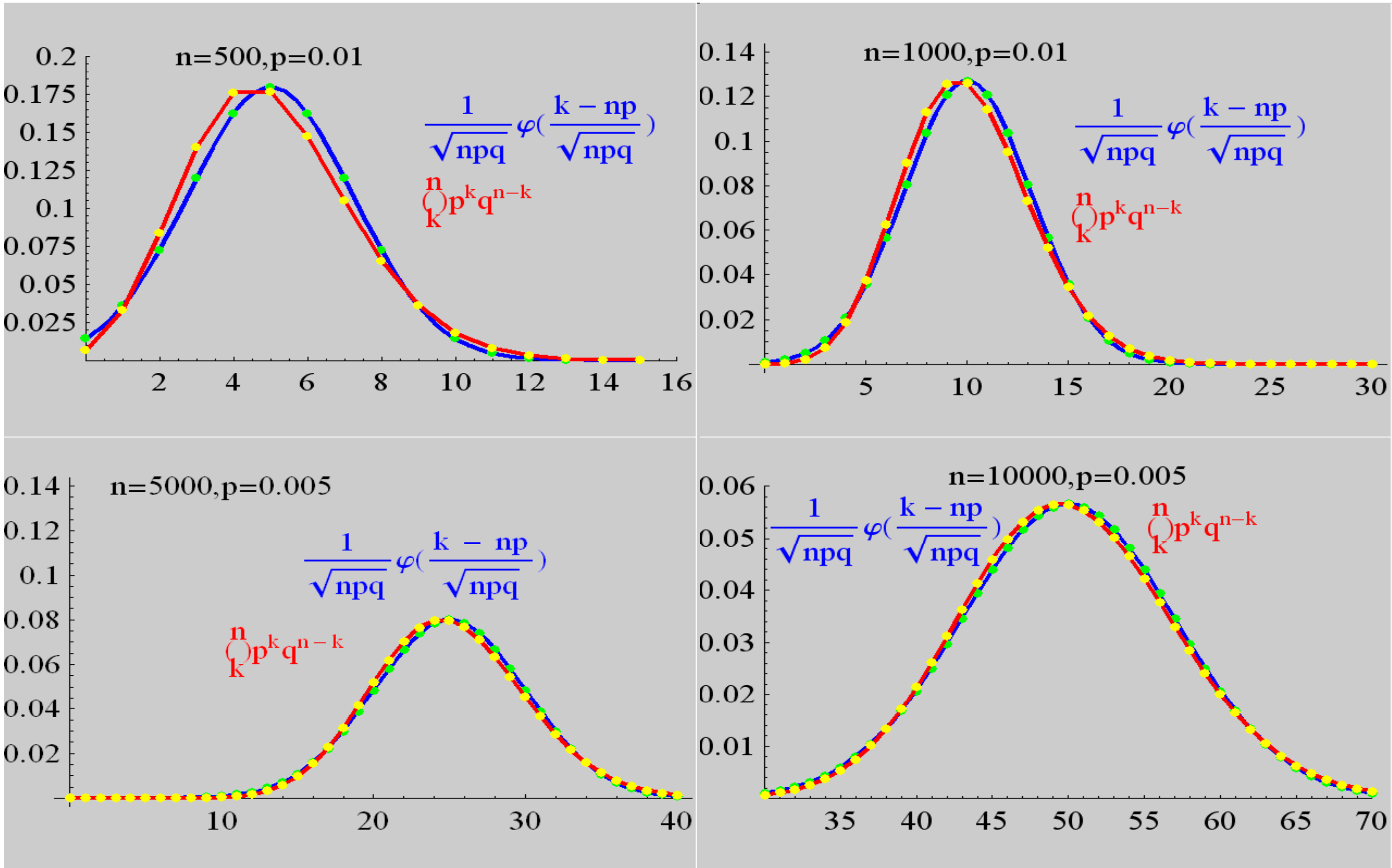
$$\lim_{n\to\infty} P\left\{\frac{\eta_n-np}{\sqrt{np(1-p)}}\le x\right\}=\lim_{n\to\infty} P\left\{\frac{\sum_{k=1}^{n}X_k-np}{\sqrt{np(1-p)}}\le x\right\}$$

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

定理三表明: 当n很大，$0<p<1$是一个定值时（或者说，$np(1-p)$也不太小时），二项变量η_n 的分布近似正态分布 $N(np,np(1-p))$.

即 $$\eta_n \overset{\text{近似地}}{\sim} N(np,np(1-p))$$

下面的图形表明:正态分布是二项分布的逼近.



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 $(0,1)$ 上服从均匀分布, 记 $V=\sum_{k=1}^{20} V_k$, 求 $P\{V>105\}$ 的近似值.

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

由定理一, 随机变量 Z 近似服从正态分布 $N(0,1)$,

其中 $Z=\dfrac{\sum_{k=1}^{20}V_k-20\times 5}{\sqrt{\frac{100}{12}}\sqrt{20}}=\dfrac{V-20\times 5}{\sqrt{\frac{100}{12}}\sqrt{20}}$

$$\therefore P\{V>105\}=P\{\frac{V-20\times 5}{\sqrt{\frac{100}{12}}\sqrt{20}}>\frac{105-20\times 5}{\sqrt{\frac{100}{12}}\sqrt{20}}\}$$

$$=P\{\frac{V-20\times 5}{\sqrt{\frac{100}{12}}\sqrt{20}}>0.387\}=1-P\{\frac{V-100}{\sqrt{\frac{100}{12}}\sqrt{20}}\le 0.387\}$$

$$\approx 1-\int_{-\infty}^{0.387}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t=1-\Phi(0.387)=0.348.$$

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^{o} 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^{o} 的概率是多少?

解 将船舶每遭受一次海浪的冲击看作一次试验,

并假设各次试验是独立的,

在90 000次波浪冲击中纵摇角大于 3^{o} 的次数为 X,

则 X 是一个随机变量, 且 $X \sim b(90000, \frac{1}{3})$.

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

$k=1,\cdots,90000.$

所求概率为

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

直接计算很麻烦，利用**德莫佛—拉普拉斯定理**

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

$$=P\left\{\frac{29500-np}{\sqrt{np(1-p)}}<\frac{X-np}{\sqrt{np(1-p)}}\le\frac{30500-np}{\sqrt{np(1-p)}}\right\}$$

$$\approx \int_{\frac{29500-np}{\sqrt{np(1-p)}}}^{\frac{30500-np}{\sqrt{np(1-p)}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t$$

$$= \Phi\left(\frac{30500-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{29500-np}{\sqrt{np(1-p)}}\right)$$

$$\because n = 90000, \quad p = \frac{1}{3},$$

$$\therefore P\{29500 < X \le 30500\} \approx \Phi\left(\frac{5\sqrt{2}}{2}\right) - \Phi\left(-\frac{5\sqrt{2}}{2}\right)$$

$$= 0.9995.$$

例3 对于一个学生而言, 来参加家长会的家长人数是一个随机变量. 设一个学生无家长、1名家长、 2名家长来参加会议的概率分别为0.05，0.8，0.15. 若学校共有400名学生, 设各学生参加会议的家长数相互独立, 且服从同一分布. (1) 求参加会议的家长数 X 超过450的概率; (2) 求有1名家长来参加会议的学生数不多于340的概率.

解 (1) 以 X_k $(k=1,2,\cdots,400)$ 记第 k 个学生来参加会议的家长数,

则 X_k 的分布律为

X_k	0	1	2
p_k	0.05	0.8	0.15

易知 $E(X_k)=1.1,\ D(X_k)=0.19,\quad (k=1,2,\cdots,400)$

而 $X=\sum_{k=1}^{400}X_k$，根据**独立同分布的中心极限定理，**

随机变量 $$\frac{\sum_{k=1}^{400}X_k-400\times 1.1}{\sqrt{400}\sqrt{0.19}}=\frac{X-400\times 1.1}{\sqrt{400}\sqrt{0.19}}$$

近似服从正态分布 $N(0,1)$,

于是 $P\{X>450\}$

$$=P\left\{\frac{X-400\times 1.1}{\sqrt{400}\sqrt{0.19}}>\frac{450-400\times 1.1}{\sqrt{400}\sqrt{0.19}}\right\}$$

$$=1-P\left\{\frac{X-400\times 1.1}{\sqrt{400}\sqrt{0.19}}\leq 1.147\right\}$$

$$\approx 1-\Phi(1.147)=0.1357;$$

(2) 以 Y 记有一名家长来参加会议的学生数，

则 $Y \sim b(400, 0.8)$，由**德莫佛－拉普拉斯定理**知，

$P\{X \le 350\}$

$$= P\left\{\frac{Y - 400 \times 0.8}{\sqrt{400 \times 0.8 \times 0.2}} \le \frac{340 - 400 \times 0.8}{\sqrt{400 \times 0.8 \times 0.2}}\right\}$$

$$= P\left\{\frac{Y - 400 \times 0.8}{\sqrt{400 \times 0.8 \times 0.2}} \le 2.5\right\} \approx \Phi(2.5) = 0.9938 .$$

例4：设供电网有10000盏电灯，夜晚每盏电灯开灯的概率均为0.7，并且彼此开闭与否相互独立，试用切比雪夫不等式和中心极限定理分别估算夜晚同时开灯数在6800到7200之间的概率

解：(1)设X表示夜晚同时开灯的盏 数

则$X \sim b(10000,0.7)$ $E(X)=7000$, $D(X)=2100$,

由切比雪夫不等式

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

故 $P\{6800<X<7200\}=P\{|X-7000|<200\}$

$$\geq 1-\frac{2100}{200^2}=0.9475$$

(2) 由中心极限定理求：

设X表示夜晚同时开灯的盏 数

则$X \sim b(10000,0.7) \quad E(X)=7000,\ D(X)=2100,$

由中心极限定理 $\dfrac{X-np}{\sqrt{np(1-p)}} \overset{\text{近似}}{\sim} N(0,1)$

故 $P\{\ 6800 < X < 7200\ \}$

$$= P\left\{\frac{6800-7000}{\sqrt{2100}} < \frac{X-7000}{\sqrt{2100}} < \frac{7200-7000}{\sqrt{2100}}\right\}$$

$$\approx \Phi\left(\frac{200}{\sqrt{2100}}\right) - \Phi\left(\frac{-200}{\sqrt{2100}}\right)$$

$$= 2\Phi\left(\frac{200}{\sqrt{2100}}\right) - 1 = 2\Phi(4.3644) - 1 = 1$$

比较几个近似计算的结果

二项分布(精确结果)
$$P\left(\left|\frac{X}{6000}-\frac{1}{6}\right|<0.01\right)\approx 0.9590$$

中心极限定理
$$P\left(\left|\frac{X}{6000}-\frac{1}{6}\right|<0.01\right)\approx 0.9624$$

Poisson 分布
$$P\left(\left|\frac{X}{6000}-\frac{1}{6}\right|<0.01\right)\approx 0.9379$$

Chebyshev 不等式
$$P\left(\left|\frac{X}{6000}-\frac{1}{6}\right|<0.01\right)\geq 0.7685$$

例5 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$\mu=\alpha_2$，$\sigma^2=\dfrac{\alpha_4-\alpha_2^2}{n}$的正态分布.

四、小结

三个中心极限定理 $\begin{cases} \text{独立同分布的中心极限定理} \\ \text{李雅普诺夫定理} \\ \text{德莫佛－拉普拉斯定理} \end{cases}$

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增加时, 其和的分布趋于正态分布.