第五章　　大数定律及中心极限定理

概率论与数理统计是研究随机现象统计规律性的学科.　随机现象的规律性只有在相同的条件下进行大量重复试验时才会呈现出来. 也就是说，要从随机现象中去寻求必然的法则，应该研究大量随机现象.

研究大量的随机现象，常常采用极限形式，由此导致对极限定理进行研究. 极限定理的内容很广泛，其中最重要的有两种:

大数定律 与 **中心极限定理**

下面我们先介绍大数定律

§1　大数定律

大数定律的客观背景

大量的随机现象中频率的稳定性

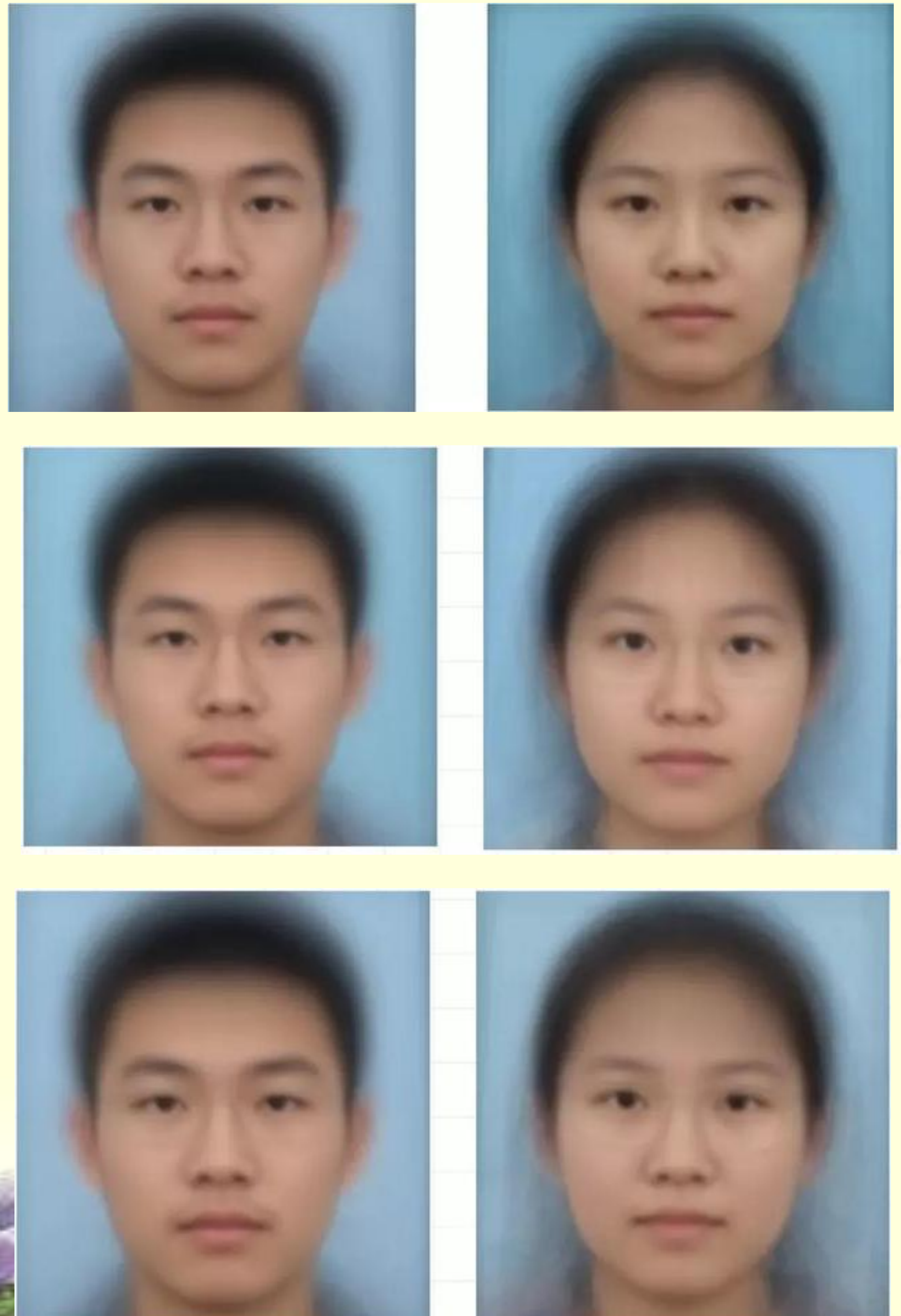
大量抛掷硬币
正面出现频率

生产过程中的
废品率

字母使用频率

……

平均脸



辛钦大数定律

设$X_1, X_2, \dots$是独立同分布的随机变量序列，且$E(X_i)=\mu$，$i=1, 2, \dots$，则对任给 $\varepsilon>0$,

辛钦

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

或

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$$

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

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

证明：

只在随机变量的方差$D(X_k)=\sigma^2, k=1,2,\ldots$存在条件下证明。

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

$$D(\frac{1}{n}\sum_{i=1}^{n} X_k) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_k) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}$$

由切比雪夫不等式

$$P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| < \varepsilon\} \geq 1 - \frac{\sigma^2/n}{\varepsilon^2}$$

令$n\to\infty$

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

意义

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

$X_1,\cdots X_n$ 的平均值 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

μ的偏差小于ε

表明当***n*非常大**时，事件 $\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| < \varepsilon\}$

发生可能性 很大

意义

1辛钦大数定律表明，当重复试验次数n 充分大时，样本均值与总体均值有较大偏差的概率很小.

2 辛钦大数定律给出了平均值稳定性的科学描述

3 辛钦大数定律为估计随机变量的期望值提供了一条实际可行的途径.

定义 设 $Y_1,Y_2,\cdots,Y_n,\cdots$ 是一系列随机变量，

Y是一随机变量，若$\forall\varepsilon>0$ 有

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

(或 $\lim\limits_{n\to\infty} P\left(\left|Y_n - Y\right| < \varepsilon\right) = 1$)

则称随机变量序列 $Y_1,Y_2,\cdots,Y_n,\cdots$ 依概率收敛

于常数 Y，记作

$$Y_n \xrightarrow{\quad P \quad} Y$$

辛钦大数定律可写为

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} p$$

定理 $\{X_n, n\geq 1\}$和$\{Y_n, n\geq 1\}$是两个随机变量序列，a，b是两个常数，如果

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

且$h(x)$在a点连续，$g(x, y)$在点(a, b)连续，则

$$h(X_n) \xrightarrow{P} h(a) \qquad g(X_n, Y_n) \xrightarrow{P} g(a, b)$$

推论1 $\{X_n, n\geq 1\}$和$\{Y_n, n\geq 1\}$是两个随机变量序列， a，b是两个常数，

如果 $X_n \xrightarrow{P} a$， $Y_n \xrightarrow{P} b$ 则有

(1) $X_n \pm Y_n \xrightarrow{P} a \pm b$

(2) $X_n \cdot Y_n \xrightarrow{P} a \cdot b$

(3) $X_n / Y_n \xrightarrow{P} a/b \quad (b \neq 0)$

贝努里大数定律

贝努里

雅各布第一·伯努利

设S_n是n重贝努里试验中事件A发生的次数，p是一次试验中事件A发生的概率，则对任给的$\varepsilon > 0$，

$$\lim_{n\to\infty} P\{|\frac{S_n}{n} - p| < \varepsilon\} = 1$$

或

$$\frac{S_n}{n} \xrightarrow{P} \mu$$

意义

1 贝努里大数定律表明，当重复试验次数n充分大时，事件A发生的**频率S_n/n**与事件A的**概率p**有较大偏差的概率很小.

2 贝努里大数定律给出了**频率稳定性**的科学描述

3 贝努里大数定律提供了**通过试验来确定事件概率**的方法.

例1.设X_1，X_2 ,… 独立同分布，且X_i的k阶矩$m_k=E(X_i^k)$存在，证明

$$\frac{1}{n}\sum_{i=1}^{n}X_i^k \xrightarrow{P} m_k$$

证明：令 $Y_i = X_i^k, i = 1,2,\ldots$ 则有Y_1，Y_2 ,… 独立同分布

且 $EY_i = EX_i^k = m_k, i = 1,2,\ldots$

所以由辛钦大数定律 $\frac{1}{n}\sum_{i=1}^{n}Y_i \xrightarrow{P} E(Y_1) = m_k$

也即 $\frac{1}{n}\sum_{i=1}^{n}X_i^k \xrightarrow{P} m_k$

例2. 设随机变量序列$X_1, X_2, \dots, X_n, \dots$独立同分布，都服均匀分布$U(0, 1)$。问当$n\to\infty$时，$\sqrt[n]{X_1X_2\cdots X_n}$依概率收敛吗？若收敛，请给出收敛的极限值，否则请说明理由。

解：令 $Y_n = \sqrt[n]{X_1X_2\cdots X_n}$

则有 $\ln Y_n = \frac{1}{n}\sum_{i=1}^{n}\ln X_i$　　令 $Z_n = \ln Y_n$

由于$X_1, X_2, \dots, X_n$独立同分布，所以$\ln X_1, \ln X_2, \dots, \ln X_n$独立同分布。

又因为$X_1 \sim U(0, 1)$，易知其密度函数为

$$f(x)=\begin{cases}1, & 0<x<1\\ 0, & \text{其它}\end{cases}$$

所以

$$E(\ln X_1)=\int_{-\infty}^{+\infty}(\ln x)f(x)dx=\int_0^1 \ln x dx=-1$$

故由辛钦大数定律得 $$\frac{1}{n}\sum_{i=1}^{n}\ln X_i \xrightarrow{P} -1$$

即 $$Z_n \xrightarrow{P} -1$$

取 $h(x)=e^x$ 由定理 可得 $$Y_n = e^{Z_n} \xrightarrow{P} e^{-1}$$

即 $$\sqrt[n]{X_1 X_2 \cdots X_n} \xrightarrow{P} e^{-1}$$

例3. 已知随机变量序列$X_1, X_2, \ldots, X_n, \ldots$独立，都服从正态分布$N(\mu,\sigma^2)$，令

$$\bar{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

证明:

$$\bar{X}_n^2\xrightarrow{P}\mu^2,\quad S_n^2\xrightarrow{P}\sigma^2$$

证明：由辛钦大数定律可得

$$\bar{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{P}EX_i=\mu$$

由定理 可得

$$\bar{X}_n^2\xrightarrow{P}\mu^2$$

又因为

$$\sum_{i=1}^{n}(X_i-\bar{X}_n)^2=\sum_{i=1}^{n}(X_i^2-2X_i\bar{X}_n+\bar{X}_n^2)=\sum_{i=1}^{n}X_i^2-2\sum_{i=1}^{n}X_i\bar{X}_n+\sum_{i=1}^{n}\bar{X}_n^2$$

$$=\sum_{i=1}^{n}X_i^2-2\bar{X}_n\sum_{i=1}^{n}X_i+\sum_{i=1}^{n}\bar{X}_n^2=\sum_{i=1}^{n}X_i^2-2n\bar{X}_n^2+n\bar{X}_n^2=\sum_{i=1}^{n}X_i^2-n\bar{X}_n^2$$

所以 $$S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2=\frac{1}{n-1}(\sum_{i=1}^{n}X_i^2-n\bar{X}_n^2)=\frac{n}{n-1}(\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}_n^2)$$

由例2知 $$\frac{1}{n}\sum_{i=1}^{n}X_i^2\xrightarrow{P}EX_i^2=DX_i+(EX_i)^2=\sigma^2+\mu^2$$

又因为 $\lim\limits_{n\to\infty}\dfrac{n}{n-1}=1$ $\qquad \bar{X}_n^2 \xrightarrow{P} \mu^2$

所以由定理及其推论可得

$$S_n^2=\frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}_n^2\right)\xrightarrow{P}1\times(\sigma^2+\mu^2-\mu^2)=\sigma^2$$

§2 中心极限定理

大数律讨论了随机序列部分和的依概率收敛性.

中心极限定理讨论对充分大的n，随机变量序列部分和 $X_1+X_2+\ldots+X_n$ 的概率分布问题.

中心极限定理的客观背景

在实际问题中，常常需要考虑许多随机因素所产生总影响.

例如：炮弹射击的落点与目标的偏差，就受着许多随机因素的影响.

如瞄准时的误差，

空气阻力所产生的误差，

炮弹或炮身结构所引起的误差等等.

对我们来说重要的是这些随机因素的总影响.

观察表明，如果一个量是由大量相互独立的随机因素的影响所造成，而每一个别因素在总影响中所起的作用不大．则这种量一般都服从或近似服从正态分布.

该结论得益于高斯对测量误差分布的研究.

现在我们就来研究，当n无限增大时，独立随机变量之和 $Y_n=\sum_{k=1}^{n} X_k$ 的分布的极限是什么？

例4. 二项分布

独立地重复某一试验，设

$$X_j=\begin{cases}1, & \text{当第j次试验成功,}\\ 0, & \text{当第j次试验不成功。}\end{cases}$$

则$\{X_j\}$ $iid \sim B(1,p)$（两点分布）。

令 $S_n = X_1 + X_2 + \ldots + X_n$

则S_n为n次独立试验中成功的次数,$S_n \sim B(n,p)$。

$n \to \infty$ **, S_n的分布形状很象正态分布。**

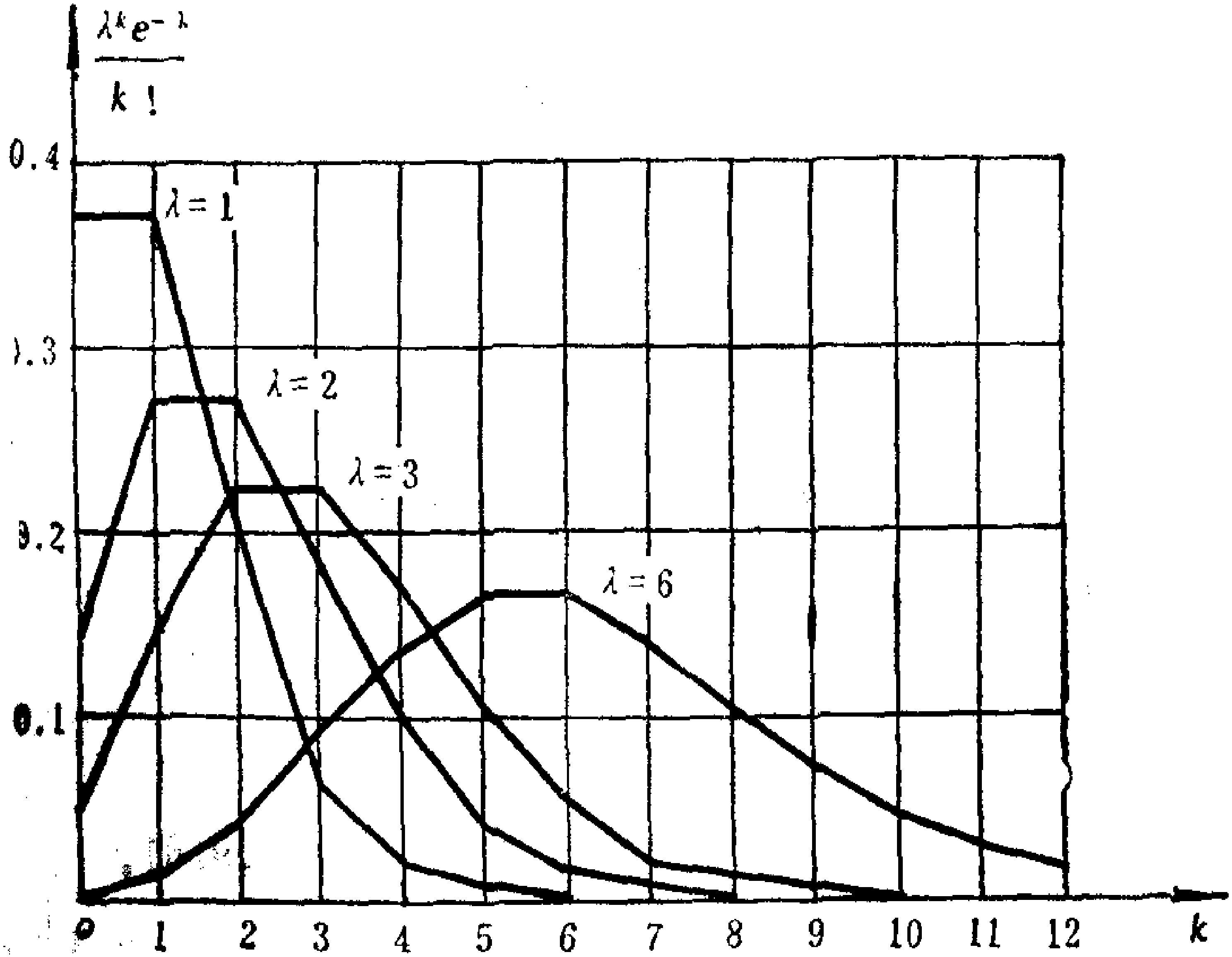
二项分布演示

例5 ***Poisson*（泊松）分布**

若 $\{X_j\}$ *iid* $P(\lambda)$，部分和

$$S_n = \sum_{i=1}^{n} X_i \sim P(n\lambda).$$

$n \to \infty$ **时, S_n的分布形状很象正态分布。**



$\frac{\lambda^k e^{-\lambda}}{k!}$
0.4
0.3
0.2
0.1
$\lambda = 1$
$\lambda = 2$
$\lambda = 3$
$\lambda = 6$
0
1
2
3
4
5
6
7
8
9
10
11
12
k

定理1（独立同分布的中心极限定理）

设随机序列 $\{X_j\}$ 独立同分布，有共同的数学期望 μ 和方差 σ^2. 部分和 $S_n = X_1 + X_2 + \ldots + X_n$，则 S_n 的标准化

$$\xi_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

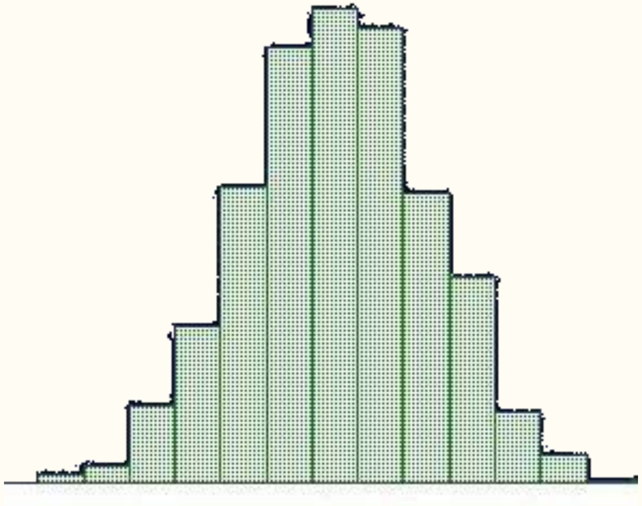
依分布收敛到标准正态分布. 即对任何 x,

$$\lim_{n\to\infty} P\{\xi_n \le x\} = \Phi(x). \qquad (2.1)$$

这里 $\Phi(x)$ 是标准正态分布的分布函数.

我们把结论(2. 1)记成 $\xi_n \xrightarrow{\mathrm{d}} N(0,1)$ ，其中的 d 表示依分布收敛.

中心极限定理是概率论中最著名的结果之一，它不仅提供了计算独立随机变量之和的近似概率的简单方法，而且有助于解释为什么很多自然群体的经验频率呈现出钟形曲线这一值得注意的事实.



中心极限定理的应用

可以用 $N(0,1)$ 近似计算关于 ξ_n 的概率，

用$N(n\mu, n\sigma^2)$ 近似计算关于 S_n 的概率。

例6. 近似计算

当辐射的强度超过每小时0.5毫伦琴(mr)时,辐射会对人的健康造成伤害. 设一台彩电工作时的平均辐射强度是0.036(mr/h), 方差是0.0081. 则家庭中一台彩电的辐射一般不会对人造成健康伤害. 但是彩电销售店同时有多台彩电同时工作时,辐射可能对人造成健康伤害. 现在有16台彩电同时工作,问这 16 台彩电的辐射量可以对人造成健康伤害的概率.

例6. （续）

解： 用X_i表示第i台彩电的辐射量(mr/h)，
则X_i的数学期望 μ =0.036, 方差σ^2 =0.0081.

$S_n=X_1+X_2+\ldots+X_{16}$ 是n=16台彩电的辐射量.

$$ES_n=16\cdot 0.036 \qquad DS_n=16\cdot 0.0081$$

由于$\{X_i\}$独立同分布，根据中心极限定理，

$$\xi_n=\frac{S_n-16\cdot 0.036}{\sqrt{16\cdot 0.0081}}$$

近似服从$N(0,1)$分布，于是

例6. （续）

$$P(S_n > 0.5) = P\left(\frac{S_n - 16\times 0.036}{\sqrt{16\times 0.0081}} > \frac{0.5 - 16\times 0.036}{\sqrt{16\times 0.0081}} \right)$$

$$= P\left(\xi_n > \frac{0.5 - 16\times 0.036}{\sqrt{16\times 0.0081}} \right)$$

$$= P(\xi_n > -0.211)$$

$$=1\text{-}P(\xi_n \le -0.211)$$

$$\approx 1\text{-}\Phi(\text{-}0.211)$$

$$=\Phi(0.211) = 0.58.$$

这16台彩电以大约58%的概率会对人造成健康伤害.

例7一加法器同时收到20个噪声电压 $V_i(i=1,2,\cdots,20)$，设它们是互相独立的随机变量，且都在区间(0,10)上服从均匀分布，记

$$S_n=\sum_{i=1}^{20}V_i$$

求 $P\{S_n>105\}$ 近似值。

解：$EV_i=5$，$DV_i=\dfrac{10^2}{12}$，$(i=1,2,\cdots,20)$，由中心极限定理知：

$$Y_n=\frac{S_n-20\cdot 5}{\sqrt{20\cdot 10^2/12}}$$

近似服从 $N(0,1)$ 分布

例7一加法器同时收到20个噪声电压 $V_i(i=1,2,\cdots,20)$，设它们是互相独立的随机变量，且都在区间(0,10)上服从均匀分布，记

$$S_n=\sum_{i=1}^{20}V_i$$

求$P\{S_n>105\}$近似值。

$$\mathrm{P}\{S_n>105\}=\mathrm{P}\left\{\frac{S_n\text{-}20\times 5}{\sqrt{10^2/12}\times\sqrt{20}}>\frac{105\text{-}20\times 5}{\sqrt{10^2/12}\times\sqrt{20}}\right\}$$

$$=\mathrm{P}\{Y_n>0.387\}=1-\mathrm{P}\{Y_n\le 0.387\}$$

$$\approx 1-\Phi(0.387)=0.348$$

二项分布的正态近似

定理2. 设$S_n \sim B(n,p)$, $p=1-q \in (0,1)$, 则

$$\frac{S_n - np}{\sqrt{npq}} \xrightarrow{d} N(0,1). \qquad (2.2)$$

证明： 令 $S_n = \sum_{i=1}^{n} X_i$,

其中 $X_1,\cdots,X_n$ 相互独立且都服从于(0-1)分布。

$EX_i = p$, $DX_i = pq$ 。

由定理1结论成立

例8 设一个系统由*100*个相互独立起作用的部件组成，每个部件的损坏率为*0.1*。为了使整个系统正常工作，至少必须有*85*个部件正常工作，求整个系统正常工作的概率。

解： 设 S_n是损坏的部件数，则 $S_n \sim B(100,0.1)$。

则整个系统能正常工作当且仅当 $S_n \le 15$.

由定理2得

$$Y_n = \frac{S_n - 100 \cdot 0.1}{\sqrt{100 \cdot 0.1 \cdot 0.9}}$$

近似服从$N(0,1)$分布

例8 设一个系统由*100*个相互独立起作用的部件组成，每个部件的损坏率为*0.1*。为了使整个系统正常工作，至少必须有*85*个部件正常工作，求整个系统正常工作的概率。

$$P\{S_n \le 15\} = P\left\{\frac{S_n - 100\times 0.1}{\sqrt{100\times 0.1\times 0.9}} \le \frac{15-100\times 0.1}{\sqrt{100\times 0.1\times 0.9}}\right\}$$

$$\approx \Phi\left(\frac{15-100\times 0.1}{\sqrt{100\times 0.1\times 0.9}}\right) = \Phi\left(\frac{5}{3}\right) = 0.952.$$

例9.某保险公司有10000个同龄又同阶层的人参加某种医疗保险。已知该类人在一年内生该种病的概率为0.006。每个参加保险的人在年初付12元保险费，而在生病时可从公司领得1000元。问在此项业务活动中，（1）保险公司赔钱的概率是多少？（2）保险公司获得利润（暂不计管理费）不少于40000的概率是多少?

解：令X表示10000个参加该种医疗保险的人中，一年内生该病的人数，易知

$$X \sim B(10000, 0.006)$$

$$EX = 10000 \times 0.006 = 60 \qquad DX = 10000 \times 0.006 \times 0.994 = 59.64$$

由棣莫佛-拉普拉斯定理，有 $\dfrac{X-60}{\sqrt{59.64}} \overset{\text{近似}}{\sim} N(0,1)$

$$\frac{X-60}{\sqrt{59.64}} \overset{\text{近似}}{\sim} N(0,1)$$

（1）$P\{10000\times 12-X\times 1000\le 0\} = P\{X\ge 120\}$

$$\approx 1-\Phi\left(\frac{120-60}{\sqrt{59.64}}\right)=1-\Phi(7.769)\approx 0$$

（2）$P\{10000\times 12-X\times 1000\ge 40000\}=P\{X\le 80\}$

$$\approx \Phi\left(\frac{80-60}{\sqrt{59.64}}\right)=\Phi(2.59)=0.9952$$

例10. (供电问题)某车间有200台车床, 在生产期间由于需要检修、调换刀具、变换位置及调换工件等常需停车. 设开工率为0.6, 并设每台车床的工作是独立的，且在开工时需电力1千瓦.

问应供应多少千瓦电力就能以99.9%的概率保证该车间不会因供电不足而影响生产?

解: 设需供应x千瓦电力,

用X表示在某时刻工作着的车床数,

依题意，$X \sim B(200,0.6)$, 现在的问题是:

求满足 $P(X \le x) \ge 0.999$ 的最小的x.

由定理2

$$\frac{X-np}{\sqrt{np(1-p)}}$$ 近似 $N(0,1)$,

这里 np=120, $np(1-p)$=48

于是 $P(X\le x)=$

$$=P\left(\frac{X-120}{\sqrt{48}}\le\frac{x-120}{\sqrt{48}}\right)$$

$$\approx\Phi\left(\frac{x-120}{\sqrt{48}}\right)$$

由 $\Phi(\frac{x-120}{\sqrt{48}})\geq 0.999$，

查正态分布函数表得

$\Phi(3.1)=0.999$

故 $\frac{x-120}{\sqrt{48}}\geq 3.1$,

从中解得$x\geq 141.5$

例11某单位有200台电话分机，每台分机有5%的时间要使用外线通话。假定每台分机是否使用外线是相互独立的，问该单位总机要安装多少条外线，才能以90%以上的概率保证分机用外线时不等待？

解： 设有S_n部分机同时使用外线，则有$S_n \sim B(n, p)$,

其中 $n = 200, p = 0.05, np = 10, \sqrt{np(1-p)} = 3.08.$

设有N条外线。

由题意有 $$P\{S_n \le N\} \ge 0.9$$

由定理2得

$$P\{S_n \le N\} = P\left\{\frac{S_n - np}{\sqrt{np(1-p)}} \le \frac{N - np}{\sqrt{np(1-p)}}\right\}$$

例11（续）

$$P\{S_n \le N\} = P\left\{\frac{S_n - np}{\sqrt{np(1-p)}} \le \frac{N - np}{\sqrt{np(1-p)}}\right\}$$

$$\approx \Phi\left(\frac{N - np}{\sqrt{np(1-p)}}\right) = \Phi\left(\frac{N-10}{3.08}\right).$$

查表得$\Phi(1.28) = 0.90.$

故 N 应满足条件 $\dfrac{N-10}{3.08} \ge 1.28,$

即 $N \ge 13.94.$ 取 $N = 14,$ 即至少要安装 14 条外线。

例12　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人数相互独立,且服从同一分布.

(1)参加会议的家长人数X超过450的概率;

(2)求有1名家长来参加会议的学生人数不多于340的概率.

解:

以X_k(k=1, 2,…, 400）表示第k个学生来参加会议的家长人数，则X_k的分布律为

X_k	0	1	2
P_k	0.05	0.8	0.15

$$E(X_k) = 1.1, \quad D(X_k) = 0.19, K = 1, 2, \ldots, 400.$$

$$X = \sum_{k=1}^{400} X_k.$$ 由中心极限定理

$$\frac{X - 400 \times 1.1}{\sqrt{400}\sqrt{0.19}}$$ 近似服从正态分布**N(0, 1)**,

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

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

$$\approx 1 - \Phi(1.147) = 0.1251.$$

Y:一名家长参加会议的学生人数,

$$Y \sim B(400, 0.8),$$

$$P(Y \le 340) = P\left(\frac{Y - 400\times 0.8}{\sqrt{400\times 0.8\times 0.2}} \le \frac{300 - 400\times 0.8}{\sqrt{400\times 0.8\times 0.2}}\right)$$

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

$$\approx \Phi(2.5) = 0.9938.$$

第五章

- 作业： 2,4,6,9,10

休息片刻继续下一讲