概率论与数理统计

第五章　大数定律及中心极限定理

检查员逐个地检查某产品,每次花10秒钟检查一个,但也可能有的产品需要再花10秒种重新检查一次,假设每个产品需要复检的概率为0.5,求在8小时内检查员检查的产品个数多于1900个的概率是多少?

分析 在8小时内检查员检查的产品个数多于1900个的概率等于检查员检查1900个产品的时间小于8小时的概率,检查每个产品花费的时间可认为是相互独立的,由独立同分布的中心极限定理计算.

解 设X_i表示“检查第i个产品花费的时间”(秒),即

$$X_i=\begin{cases}10, & \text{不需复检},\\ 20, & \text{需复检},\end{cases}\qquad i=1,2,\cdots 1900.$$

则$X_1, X_2, \cdots, X_n$相互独立同分布, $X=\sum_{i=1}^{1900} X_i$为检查1900个产品所花费的时间,且

$$E(X_i)=10\times 0.5+20\times 0.5=15,\ D(X_i)=E(X_i^2)-[E(X_i)]^2\}=25.$$

于是

$$P\{X\le 8\times 3600\}=P\{\sum_{i=1}^{1900} X_i \le 28800\}$$

$$=P\{\frac{\sum_{i=1}^{1900} X_i-1900\times 15}{5\sqrt{1900}}\le \frac{28800-1900\times 15}{5\sqrt{1900}}\}$$

$$=P\{\frac{\sum_{i=1}^{1900} X_i-28500}{50\sqrt{19}}\le \frac{6}{\sqrt{19}}\}\approx \Phi(1.376)\approx 0.91559$$

故8小时内检查的个数多于1900个的概率是0.91559.

中心极限定理

求解步骤：

1.正确选择独立同分布随机变量$X_1, X_2, \cdots, X_n$；

2.将 $\sum_{k=1}^{n} X_k$ 标准化：

$$Y_n = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \sim N(0,1)$$

3.近似计算：

$$P\{a \le \sum_{k=1}^{n} X_k \le b\} = P\{\frac{a - n\mu}{\sqrt{n}\sigma} \le \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \le \frac{b - n\mu}{\sqrt{n}\sigma}\}$$

$$\approx \Phi(\frac{b - n\mu}{\sqrt{n}\sigma}) - \Phi(\frac{a - n\mu}{\sqrt{n}\sigma}).$$

§2　中心极限定理

- 中心极限定理
- 小结

定理5(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

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

$$= \Phi(x).$$

请注意：

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

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

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6(棣莫弗一拉普拉斯定理)

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　$\eta_n = \sum_{k=1}^{n} X_k$，其中 $X_1, X_2, \cdots, X_n$ 相互独立且服从同一 $(0-1)$ 分布.

$$P\{X_k = i\} = p^i (1-p)^{1-i},\ i = 0, 1.$$

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

根据定理4得

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

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

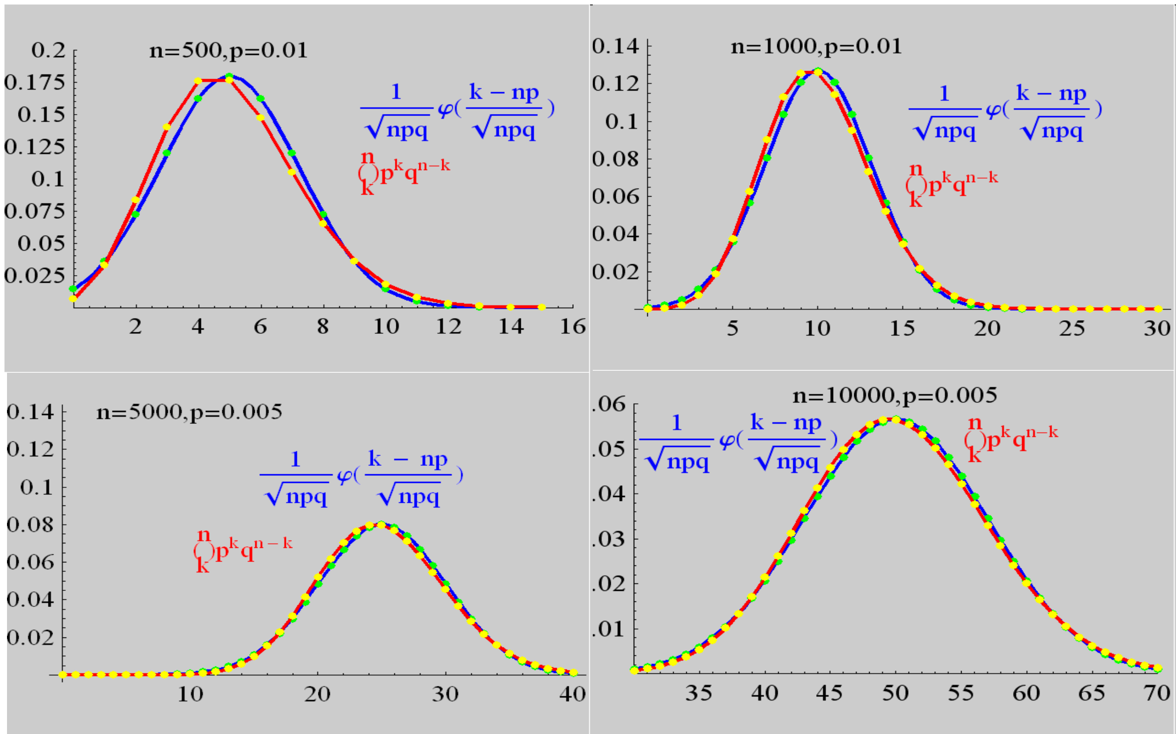
定理6表明:

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.

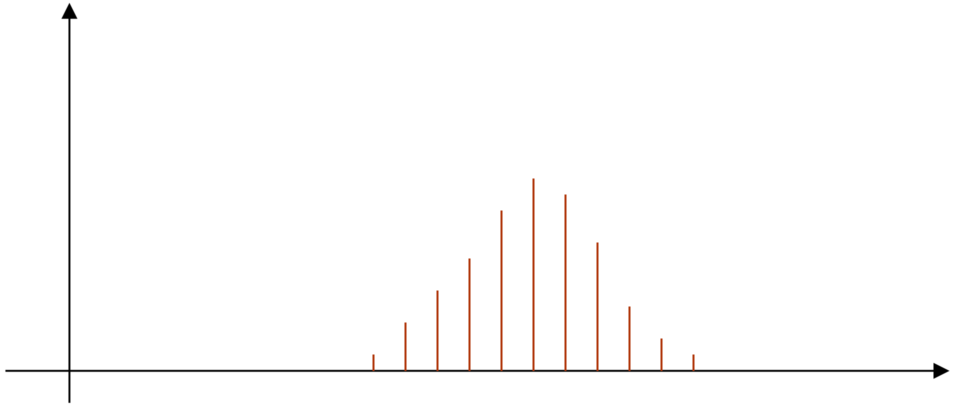
$$P\{a < \eta_n \leqslant b\} = P\left\{\frac{a-np}{\sqrt{np(1-p)}} < \frac{\eta_n - np}{\sqrt{np(1-p)}} \leqslant \frac{b-np}{\sqrt{np(1-p)}}\right\}$$

$$\approx \Phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-np}{\sqrt{np(1-p)}}\right)$$

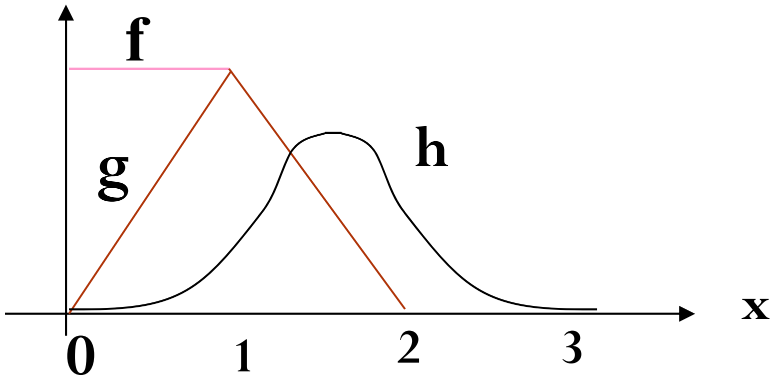
下面的图形表明:正态分布是二项分布的逼近.



下面演示不难看到中心极限定理的客观背景



例:20个0-1分布的和的分布



几个(0,1)上均匀分布的和的分布

$X_1 \sim f(x)$　　$X_1+X_2 \sim g(x)$

$X_1+X_2+X_3 \sim h(x)$

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 $P\{V > 105\}$ 的近似值．

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

由定理4，　随机变量

$$Z=\frac{\sum_{k=1}^{20}V_k-20\times5}{\sqrt{100/12}\sqrt{20}}=\frac{V-20\times5}{\sqrt{100/12}\sqrt{20}}$$

近似服从正态分布 $N(0,1)$，于是

$$P\{V>105\}$$

$$=P\{\frac{V-20\times5}{(10/\sqrt{12})\sqrt{20}}>\frac{105-20\times5}{(10/\sqrt{12})\sqrt{20}}\}$$

$$=P\{\frac{V-100}{(10/\sqrt{12})\sqrt{20}}>0.387\}$$

$$=1-P\left\{\frac{V-100}{(10/\sqrt{12})\sqrt{20}}\le 0.387\right\}$$

$$\approx 1-\int_{-\infty}^{0.387}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t$$

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

$$=0.348.$$

即有

$$P\{V>105\}\approx 0.348\,.$$

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次纵摇角度大于3^{o}的概率是多少？

解　将船舶每遭受一次波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}的次数记为X，

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

其分布律为

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

$$k=0,1,\cdots,90000.$$

所求概率为

$$P\{29\,500 \le X \le 30\,500\}$$

$$=\sum_{k=29\,500}^{30\,500}\binom{90\,000}{k}\left(\frac{1}{3}\right)^{k}\left(\frac{2}{3}\right)^{90\,000-k}.$$

直接计算很麻烦，利用棣莫弗一拉普拉斯定理

$$P\{29\,500 \le X \le 30\,500\}$$

$$= P\left\{\frac{29\,500 - np}{\sqrt{np(1-p)}} \le \frac{X - np}{\sqrt{np(1-p)}} \le \frac{30\,500 - np}{\sqrt{np(1-p)}}\right\}$$

$$\approx \int_{\frac{29\,500-np}{\sqrt{np(1-p)}}}^{\frac{30\,500-np}{\sqrt{np(1-p)}}} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2}{2}} \mathrm{d}t$$

$$= \Phi\left(\frac{30\,500 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{29\,500 - np}{\sqrt{np(1-p)}}\right)$$

其中 $n = 90\,000, \quad p = \frac{1}{3},$

即有

$$P\{29\,500 \le X \le 30\,500\}$$

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

$$= 0.9995.$$

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

而 $X=\sum\limits_{k=1}^{400}X_k,$ 由定理4, 随机变量

$$\frac{\sum\limits_{k=1}^{400}X_k-400\times1.1}{\sqrt{400}\sqrt{0.19}}=\frac{X-400\times1.1}{\sqrt{400}\sqrt{0.19}}$$

近似服从正态分布 $N(0,1)$, 于是

$$P\{X > 450\}$$

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

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

$$\approx 1 - \Phi(1.147)$$

$$= 0.1251.$$

(2) 以 Y 记有一名家长来参加会议的学生数,

则 $Y \sim b(400, 0.8)$, 由定理6得,

$$P\{Y \le 340\}$$

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

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

$$\approx \Phi(2.5)$$

$$= 0.9938 .$$

例4　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 $X_k = \begin{cases} 1 & \text{第}\boldsymbol{k}\text{次取到号码}0 \\ 0 & \text{否则} \end{cases}$，$k$=1,2, …

(1) 至少应取球多少次才能使“0”出现的频率在0.09-0.11之间的概率至少是0.95?

(2)用中心极限定理计算在100次抽取中,数码“0”出现次数在7和13之间的概率.

（1）解：设应取球n次，0出现频率为$\frac{1}{n}\sum_{k=1}^{n}X_k$

$$E(\frac{1}{n}\sum_{k=1}^{n}X_k)=0.1,\quad D(\frac{1}{n}\sum_{k=1}^{n}X_k)=\frac{0.09}{n}$$

由独立同分布中心极限定理：

$$\frac{\sum_{k=1}^{n}X_k-0.1n}{0.3\sqrt{n}}=\frac{\frac{1}{n}\sum_{k=1}^{n}X_k-0.1}{0.3/\sqrt{n}}\overset{\text{近似地}}{\sim}N(0,1)$$

$$P\{0.09\le\frac{1}{n}\sum_{k=1}^{n}X_k\le 0.11\}$$

$$= P\{|\frac{1}{n}\sum_{k=1}^{n} X_k - 0.1| \le 0.01\}$$

$$= P\{|\frac{\frac{1}{n}\sum_{k=1}^{n} X_k - 0.1}{0.3/\sqrt{n}}| \le \frac{\sqrt{n}}{30}\} \approx 2\Phi(\frac{\sqrt{n}}{30}) - 1$$

欲使 $2\Phi(\frac{\sqrt{n}}{30}) - 1 \ge 0.95$ 即 $\Phi(\frac{\sqrt{n}}{30}) \ge 0.975$

查表得 $\frac{\sqrt{n}}{30} \ge 1.96$

从中解得 $n \ge 3458$

即至少应取球3458次才能使“0”出现的频率在0. 09-0. 11之间的概率至少是0. 95.

（2）解：在100次抽取中，数码“0”出现次数为$\sum_{k=1}^{100} X_k$

由中心极限定理,

$$\frac{\sum_{k=1}^{100} X_k - \sum_{k=1}^{100} E(X_k)}{\sqrt{\sum_{k=1}^{100} D(X_k)}} \overset{\text{近似地}}{\sim} N(0,1)$$

其中$E(X_k)=0.1, D(X_k)=0.09$

即 $$\frac{\sum_{k=1}^{100} X_k - 10}{3} \overset{\text{近似地}}{\sim} N(0,1)$$

$$P(7 \le \sum_{k=1}^{100} X_k \le 13) = P(-1 \le \frac{\sum_{k=1}^{100} X_k - 10}{3} \le 1)$$

$$\approx \Phi(1) - \Phi(-1)$$

$$= 2\Phi(1) - 1 = 0.6826$$

即在100次抽取中，数码“0”出现次数在7和13之间的概率为0.6826.

小结

中心极限定理

- 独立同分布中心极限定理

$$\begin{cases} E(X_k)=\mu,\quad D(X_k)=\sigma^2 \\ \Rightarrow \sum_{k=1}^{n} X_k \overset{\text{近似地}}{\sim} N(n\mu, n\sigma^2) \end{cases}$$

- 棣莫弗－拉普拉斯中心极限定理

$$\begin{cases} \eta_n \sim B(n,p) \\ \Rightarrow \eta_n \overset{\text{近似地}}{\sim} N(np, np(1-p)) \end{cases}$$

- 李雅普诺夫中心极限定理

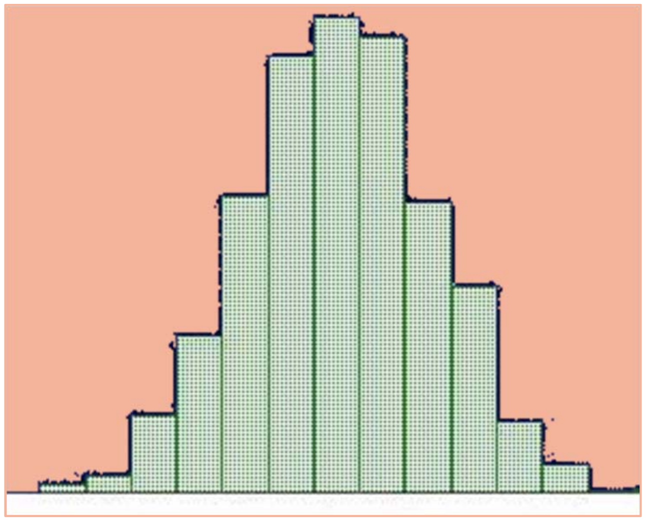
$$\begin{cases} E(X_k)=\mu_k, D(x_k)=\sigma_k{}^2 \\ \Rightarrow \sum_{k=1}^{n} X_k \overset{\text{近似地}}{\sim} N(\sum_{k=1}^{n}\mu_k, B_n{}^2) \end{cases}$$

注 随机变量$X_1, X_2, \cdots$是相互独立的

这一节我们介绍了中心极限定理

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.

在后面的课程中，我们还将经常用到中心极限定理.



作业：课后习题 5、11、13、14

练习：

1．设$X_1, X_2, \cdots X_n \cdots$是独立同分布的随机变量序列,且均值为$\mu$,方差为$\sigma^2$,那么当 n充分大时，近似有 $\overline{X} \sim$ $N(\mu, \frac{\sigma^2}{n})$ 或 $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim$ $N(0,1)$ 。特别的，当同为正态分布时，对于任意的n，都精确有

$\overline{X} \sim$ $N(\mu, \frac{\sigma^2}{n})$ 或$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim$ $N(0,1)$

2．设$X_1,X_2,\cdots X_n\cdots$是独立同分布的随机变量序列,且$E(X_i)=\mu,D(X_i)=\sigma^2$，那么$\frac{1}{n}\sum_{i=1}^{n}X_i^2$依概率收敛于________

解：$E(X_i^2)=D(X_i)+\left[E(X_i)\right]^2=\sigma^2+\mu^2(i=1,2,\cdots)$

且$X_1^2,X_2^2,\cdots,X_n^2,\cdots$相互独立

故 $$\frac{1}{n}\sum_{i=1}^{n}X_i^2\xrightarrow{P}\mu^2+\sigma^2$$

3 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，试用中心 极限定理估算夜晚同时开灯数在6800到7200之间的概率.

解： 设X表示夜晚同时开灯的盏数

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

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

故 $\boldsymbol{P}\{\ 6800 < \boldsymbol{X} < 7200\ \}$

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

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

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

$$= 2\Phi(4.3644) - 1$$

$$= 1$$

4 银行为支付某日即将到期的债券准备一笔现金, 已知这批债券共发放了500张, 每张需付本息1000元,设持券人(1人1券)到期到银行领取本息的概率为0.4.问银行于该日应准备多少现金才能以99.9%的把握满足客户的兑换?

解 设X为该日到银行领取本息的总人数,则

$X \sim B(500,0.4)$, 所需支付现金为$1000X$,设银行该日应准备现金x元,依题意有 $P\{1000X \le x\} \ge 0.999.$

由棣莫弗—拉普拉斯中心极限定理知

$$P\{1000X \le x\} = P\{X \le \frac{x}{1000}\}$$

$$= P\{\frac{X - 500\times 0.4}{\sqrt{500\times 0.4\times 0.6}} \le \frac{\frac{x}{1000} - 500\times 0.4}{\sqrt{500\times 0.4\times 0.6}}\}$$

$$= P\{\frac{X - 200}{\sqrt{120}} \le \frac{x - 200000}{2000\sqrt{30}}\}$$

$$\approx \Phi(\frac{x - 200000}{2000\sqrt{30}}) \ge 0.999.$$

即 $\dfrac{x - 200000}{2000\sqrt{30}} \ge 3.1,$

得 $x \ge 233958.799.$

因此银行于该日应准备现金234000元才能以99.9%的把握满足客户的兑换.