5

大数定律与中心极限定理

概率论与数理统计是研究随机现象的统计规律的科学，在随机现象的统计规律中，频率的稳定性最引人注目.由此可以揭示随机现象本身许多固有的规律性.极限定理中的大数定律就是刻画频率稳定性的理论.而在概率论与数理统计的基础理论与应用中，中心极限定理占有极其重要的地位.

5.1 大数定律

5.1.1 问题的提出

重复试验中事件的频率的稳定性，是大量随机现象的统计规律性的典型表现.人们在实践中认识到频率具有稳定性，进而由频率的稳定性预见概率的存在；由频率的性质推断概率的性质，并在实际应用中用频率的值来估计概率的值.

其实，在大量随机现象中，不但事件的频率具有稳定性，而且大量随机现象的平均结果一般也具有这种稳定性；单个随机现象的行为对大量随机现象共同产生的总平均效果几乎不发生影响.这就是说，尽管单个随机现象的具体实现不可避免地引起随机偏差，然而在大量随机现象共同作用时，由于这些随机偏差互相抵消，补偿和拉平，致使总的平均结果趋于稳定. 例如，一个精密钳工在测量一个工件时，由于具有随机误差，他总是反复测量多次，然后用各次的平均值来作为测量的结果.而且经验表明：只要测量的次数足够多，总可以达到要求的精度.

概率论中，一切关于大量随机现象之平均结果稳定性的定理，统称为**大数定理**.

5.1.2 切比雪夫不等式与大数定律

为了讨论大数定律,首先我们给出一个重要的不等式:切比雪夫(Chebychev)不等式,为此,我们先证明一个称为马尔科夫不等式的结果.

定理 5.1.1(Markov)设 X 为只取非负值的随机变量,则对任一实数 $a>0$,

$$P\{X \geq a\} \leq \frac{E[X]}{a}.$$

证明　我们对 X 为连续型随机变量,有密度函数 $f(x)$ 的情形给出证明.

$$E[X]=\int_0^{\infty} xf(x)dx=\int_0^{a} xf(x)dx+\int_a^{\infty} xf(x)dx$$

$$\geq \int_a^{\infty} xf(x)dx \geq \int_a^{\infty} af(x)dx$$

$$=a\int_a^{\infty} f(x)dx=aP\{X \geq a\}.$$

例 5.1.1　设随机变量 X 且 $EX^4 \leq 100$.试给出 $P(X \geq 5)$ 的一个上界.

解 令 $Y = X^4$，于是 Y 是一个非负随机变量且均值不超过 100.
注意到如果 $X \geq 5$，则有 $Y \geq 625$．由上述关系式，得

$$P(X \geq 5) \leq P(Y \geq 625) \leq \frac{EY}{625} \leq \frac{100}{625}.$$

定理 5.1.2 (Chebychev) 设 X 为随机变量, 有有限的均值 μ 及方差 σ^2, 则对任一实数 $\varepsilon > 0$

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

证明 既然 $(X-\mu)^2$ 是非负随机变量, 故可用马尔科夫不等式得

$$P\{|X-\mu|^2 \geq \varepsilon^2\} \leq \frac{E[(X-\mu)^2]}{\varepsilon^2}.$$

但是, 由于 $|X-\mu|^2 \geq \varepsilon^2$ 的充要条件为 $|X-\mu| \geq \varepsilon$， 因此

$$P\{|X-\mu| \geq \varepsilon\} \leq \frac{E[(X-\mu)^2]}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}.$$

马尔科夫不等式和切比雪夫不等式的重要性在于,当我们仅仅知道概率分布的均值, 或者同时知道其均值及方差时,它们能使我们得到概率值的界. 如果对随机变量作些限制，我们可以给出更精细的概率不等式.例如，设 X 为非负随机变量，$EX = \mu, DX = \sigma^2$, 对 $b > 0$, 则

$$P\{X \geq \mu + b\sigma\} \leq \frac{1}{1+b^2}.$$

证明 令

$$I(X) = \begin{cases} 1, & X \geq \mu + b\sigma \\ 0, & \text{其它} \end{cases}$$

和

$$g(X) = \frac{[(X-\mu)b + \sigma]^2}{\sigma^2(1+b^2)^2}.$$

易知 $E[I(X)] = P\{X \geq \mu + b\sigma\}$, 注意到 $E[(X-\mu)b+\sigma]^2 = \sigma^2(b^2+1)$. 则

$$E[g(X)] = \frac{E[(X-\mu)b+\sigma]^2}{\sigma^2(1+b^2)^2} = \frac{\sigma^2(1+b^2)}{\sigma^2(1+b^2)^2} = \frac{1}{(1+b^2)}.$$

又 $$h(X) = g(X) - I(X) \geq 0, \text{ 如果 } X \geq 0,$$

所以 $P\{h(X)\geq 0\}=1$. 由期望的性质知

$$0\leq E[h(X)]=E[g(X)]-E[I(X)]$$

$$=\frac{1}{1+b^2}-P\{X\geq \mu+b\sigma\},$$

定理得证.

定义 5.1.1 设 $X_1,X_2,\cdots$ 是一列随机变量，令 $\overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i$ ，$n=1,2,\cdots$.若存在这样的常数列 $a_1,a_2,\cdots$ 对于任意的 $\varepsilon>0$，有

$$\lim_n P\{\left|\overline{X}_n-a_n\right|<\varepsilon\}=1,$$

则称序列 $X_1,X_2,\cdots,X_n$ 服从**大数定律**.记作

$$P-\lim_n[\overline{X}_n-a_n]=0,$$

上式的直观意义是：当 $n\to\infty$ 时，事件 $\{\left|\overline{X}_n-a_n\right|<\varepsilon\}$ 的概率趋于 1.

定义 5.1.2 设 $X_1,X_2,\cdots$ 是一列随机变量，令 $\overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i$ ，$n=1,2,\cdots$.若存在这样的常数 a 对于任意的 $\varepsilon>0$，有

$$\lim_n P\{\left|\overline{X}_n-a\right|<\varepsilon\}=1,$$

则称序列 $\overline{X}_1,\overline{X}_2,\cdots,\overline{X}_n,\cdots$ **依概率收敛**于 a .记作

$$P-\lim_n[\overline{X}_n-a]=0,\text{或}\ \overline{X}_n\xrightarrow{P}a$$

从直观上看，定义 5.1.1 中的常数列 $a_1,a_2,\cdots$ 取为 $E(\overline{X}_1),E(\overline{X}_2),\cdots$ 是合适的.

定理 5.1.3（马尔科夫）假设随机变量 $X_1,X_2,\cdots$ 满足条件：对任意的 $n\geq 1$, $E\left|X_n\right|<\infty$, $D(\sum_{i=1}^{n}X_i)<\infty$，且 $\lim_n n^{-2}D(\sum_{i=1}^{n}X_i)=0$. 那么 $X_1,X_2,\cdots$ 服从大数定律.

证明 令 $\overline{X}_n=n^{-1}\sum_{i=1}^{n}X_i$ ，对任意的 $\varepsilon>0$，由 Chebychev 不等式，有

$$P\{\left|\overline{X}_n-E(\overline{X}_n)\right|\geq\varepsilon\}\leq\frac{D(\overline{X}_n)}{\varepsilon^2}=\frac{D(\sum_{i=1}^{n}X_i)}{n^2\varepsilon^2}\to 0,\ (n\to\infty).$$

从而有

$$\lim_n P\{\left|\overline{X}_n-E\overline{X}_n\right|<\varepsilon\}=1.$$

所以 $X_1,X_2,\cdots$ 服从大数定律

定理 5.1.4 (Bernoulli 大数定律)在事件 A 发生的概率为 p 的 n 次重复试验 Bernoulli 概型中,令 μ_n 表示 n 次重复试验 A 发生的次数,则对任意给定的 $\varepsilon>0$,有

$$\lim_n P\{\left|\frac{\mu_n}{n}-p\right|<\varepsilon\}=1.$$

证明 令 X_i 表示第 i 次试验中 A 发生的次数,则 $X_1, X_2, \cdots$ 相互独立,且 $\mu_n=\sum_{i=1}^{n} X_i$, 由于

$$EX_i=p, \qquad DX_i=p(1-p).$$

故由 Markov 大数定律知,结论成立.

注: ⑴ Bernoulli 大数定律的结论虽然简单,但其意义却是相当深刻的. 它告诉我们当试验次数趋于无穷时,事件A发生的频率依概率收敛于A发生的概率,这样,频率接近于概率这一直观的经验就有了严格的数学意义.

⑵上述结果可以用几种等价的可选择的方法来叙述. 显然

$$P\{\left|\frac{\mu_n}{n}-p\right|<\varepsilon\}\geq 1-\frac{p(1-p)}{n\varepsilon^2}.$$

或

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

⑶重要的是要注意上面的收敛与通常微积分学里所说的那种收敛是不同的. 当我们说当 $n\to\infty$ 时, a^{-n} 收敛于零,我们是指当 n 充分大时, a^{-n} 变得始终任意地接近于零. 当我们说 μ_n/n 收敛于 p,是指 n 充分大时,可使事件

$$\{\omega:\left|\frac{\mu_n}{n}-p\right|<\varepsilon\}$$

的概率任意地逼近于 1.

⑷当我们提出下列问题时,还可以得到大数定律的另一形式: 为了至少有 0.95 的概率使频率与 p 之差小于 0.01 ,我们应把试验重复多少次?亦即,对于 $\varepsilon=0.01$,我们希望选择 n 使得 $1-p(1-p)/[n(0.01)^2]=0.95$. 由此求解 n 得 $n=p\ (1-p)/(0.01)^2(0.05)$.

用 δ 和 ε 分别代替 0.05 及 0.01 这两个特殊的值,我们有

$$P\{\left|\frac{\mu_n}{n}-p\right|<\varepsilon\}\geq 1-\delta.\ \text{当}\ n\geq\frac{p(1-p)}{\varepsilon^2\delta}.$$

还应强调,取 $n\geq p(1-p)/\varepsilon^2\delta$,并不意味对 $\left|\frac{\mu_n}{n}-p\right|$ 作出任何保证. 这不过是使 $\left|\frac{\mu_n}{n}-p\right|$ 有可能成为很小.

例 5.1.2 掷一枚均匀的骰子,为了至少有 95%的把握使六点朝上的频率与理论概率 1/6 之差落在 0.01 的范围之内,问需要掷多少次?

解 这里 $p=1/6, \varepsilon=0.01, \delta=0.05$,利用上述关系式求得

$$n\geq\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)/(0.01)^2(0.05)=27778.$$

注 (1)$\frac{\mu_n}{n}$是一个随机变量而不仅是一个观察值. 如果我们实际上把骰子掷了27778次, 然后计算六点朝上的频率, 这个数也不一定与1/6的差在0.01的范围内. 上述例子的要点是: 如果我们有100个人都掷骰子27778次, 则大约有95个人观察到的频率与1/6的差在0.01的范围内.

(2)在许多问题中我们并不知道p的值. 因此不能用上述n的界值. 在那种情形, 我们可以利用$p=1/2$时$1-p(1-p)$取得极大值1/4. 因此, 我们可以肯定地说, 当$n\geq\frac{1}{4\varepsilon^2\delta}$时有:

$$P\{\left|\frac{\mu_n}{n}-p\right|<\varepsilon\}\geq 1-\delta.$$

定理 5.1.5（Khintchine 大数定律）设$X_1,X_2,\cdots,X_n,\cdots$是随机变量序列，若期望$EX_n=\mu(n=1,2,\cdots)$有限，则对任意的$\varepsilon>0$有：

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

定理 5.1.6（Kolmogorov 加强大数定理）设$X_1,X_2,\cdots,X_n,\cdots$是独立同分布随机变量序列，$E|X_n|<\infty$的充分必要条件是

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

5.2 中心极限定理

5.2.1 中心极限定理的提法

n个相互独立同分布的随机变量之和的分布近似于正态分布，n愈大，此种近似程度愈好，这一重要现象可以从下面两个例子看出.

例 5.2.1 一棵均匀的骰子连掷n次，其中点数之和Y_n是n个相互独立同分布随机变量之和，即

$$Y_n=X_1+X_2+\cdots X_n$$

其中诸X_i的共同的概率分布为

X_i	1	2	3	4	5	6
P	1/6	1/6	1/6	1/6	1/6	1/6

这也是概率分布，其概率直方图是平顶的.

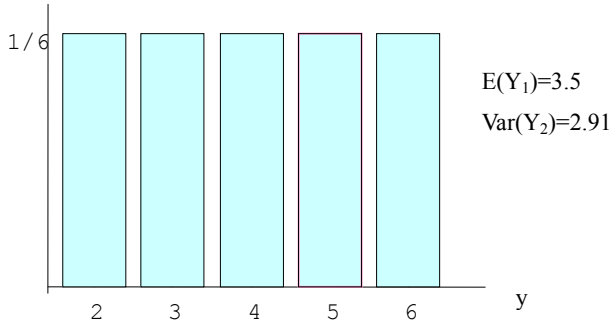

图 5-1 $Y_1=X_1$ 的概率分布

当 $n=2$ 时， $Y_2=X_1+X_2$ 的概率分布可用离散形式的卷积公式求得

$Y_2=X_1+X_2$	2	3	4	5	6	7	8	9	10	11	12
P	1/36	2/36	3/36	8/36	5/36	6/36	5/36	8/36	3/36	2/36	1/36

它的概率直方图呈单峰对称的阶梯形，且阶梯的每阶高度相等.

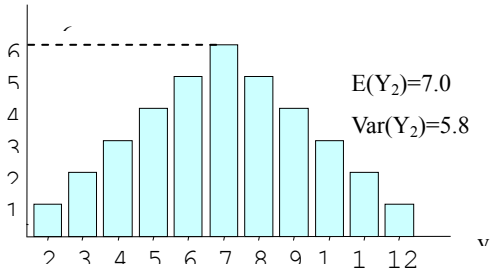

图 5-2$Y_2=X_1+X_2$ 的概率分布

当 $n=3$ 时， $Y_3=X_1+(X_2+X_3)$ 的概率分布和 $Y_4=(X_1+X_2)+(X_3+X_4)$ 的概率分布都可用卷积公式求得.它的概率直方图仍呈单峰对称的阶梯形，但台阶增多，每个台阶高度不等，中间台阶高度要比两侧台阶高度略高一点，从图 5-3 和图 5-4 上已呈现出正态分布的轮廓.

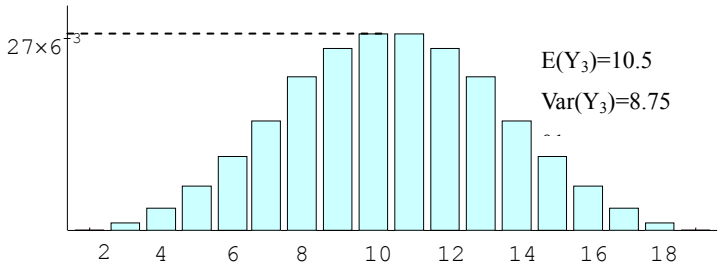

图 5-3 $Y_3=X_1+X_2+X_3$ 的概率分布

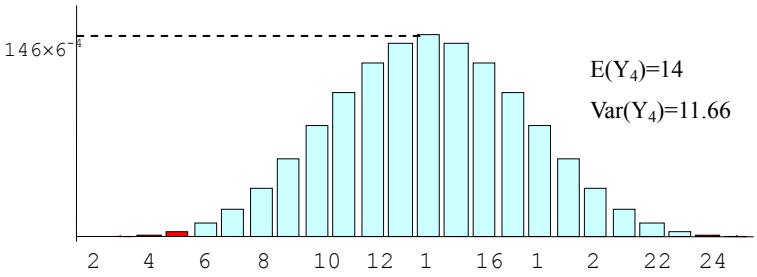

图 5-4 $Y_4=X_1+X_2+X_3+X_4$ 的概率分布

当 n 再增大时，可以想象 $Y=X_1+X_2+\cdots+X_n$ 的概率直方图的轮廓线与正态密度曲线更为接近，只是分布中心 $E(Y_n)$ 将随着 n 的增加不断地向右移动，而标准差 $\sigma(Y_n)$ 不断增大.假如对 Y_n 施行标准化变换后，所得

$$Y_n^* = \frac{Y_n - EY_n}{\sigma(Y_n)} = \frac{X_1 + X_2 + \cdots + X_n - E(X_1 + X_2 \cdots + X_n)}{\sqrt{Var(X_1 + X_2 + \cdots + X_n)}}$$

$$= \frac{X_1 + \cdots + X_n - nE(X_1)}{\sqrt{n}\sigma(Y_1)}$$

的分布有望接近标准正态分布 $N(0,1)$. 这种期望已被证明是正确的.在标准正态分布 $N(0,1)$ 的帮助下近似计算概率 $P(Y_n < a)$ 已不是很困难的事了.

譬如，当 $n = 100$ 时, $E(Y_{100}) = 100 \times 3.5 = 350$, $\sigma(Y_{100}) = \sqrt{100} \times 1.71 = 17.1$, 于是利用标准正态分布可得

$$P(Y_{100} \le 400) = P\left(\frac{Y_{100} - 350}{17.1} \le \frac{400 - 350}{17.1}\right)$$

$$= P(Y_{100}^* \le 2.9240) \approx \Phi(2.9240) = 0.9982$$

假如不利用正态近似，完成此种计算是很困难的，最后结果表明：连续 100 次掷骰子，其点数之和不超过 400 是几乎必然发生的事件.

例 5.2.2 设 $X_1, X_2, \cdots, X_n$ 是 n 个独立同分布的随机变量，其共同分布为区间（0，1）上的均匀分布，即诸 $X_i \sim U(0,1)$. 若取 $n = 100$，要求概率 $P(X_1 + X_2 + \cdots X_n \le 60) = ?$

要精确地求出上述概率，就要寻求 n 个独立同分布的随机变量，则在较小的 n 场合尚能用卷积公式写出 $p_n(y)$, 譬如

$$p_1(y) = \begin{cases} 1, & 0 < y < 1 \\ 0, & \text{其它场合} \end{cases};$$

$$p_2(y) = \begin{cases} y, & 0 < y < 1 \\ 2 - y, & 1 \le y < 2 \\ 0, & \text{其它场合} \end{cases}$$

对 $p_2(y)$ 和 $P_1(y)$，使用卷积公式又可得 $Y_3 = X_1 + X_2 + X_3$ 的密度函数

$$p_3(y) = \begin{cases} y^2/2, & 0 < y < 1 \\ -(y - 3/2)^2 + 3/4, & 1 \le y < 2 \\ (3 - y)^2/2, & 2 \le y < 3 \\ 0, & \text{其它场合} \end{cases}$$

这是一个连续函数，它的非零部分是由三段二次曲线相连，并连续处较为光滑.照此下去，可以看出，Y_n 的密度函数 $p_n(y)$ 是一个连续函数，它的非零部分是由 n 段 $n-1$ 次曲线相连.但是要具体写出 $p_n(y)$ 的表达式绝非易事.即使写出表达式，使用起来也很不方便.这样一来，要精确计算 $P(X_1 + X_2 + \cdots X_n \le 60)$ 就发生困难.图 5-4 给人们提供了一条解决这个问题的思路，随着 n 增大， $p_n(y)$ 的图形愈来愈接近正态曲线.

如例 5.2.2 一样，当 n 增大时 Y_n 的密度函数 $p_n(y)$ 中的 $E(Y_n)$ 右移，标准差 $\sigma(Y_n)$ 增

大，为了克服这些障碍，使用标准化技术就可使极限分布稳定于标准正态分布 $N(0,1)$. 用此极限分布计算上述概率已不是很难的事了.

由于均匀分布 $U(0,1)$ 的期望与标准差分别为

$$E(X_1)=0.5, \qquad \sigma(X_1)=\sqrt{1/12}=0.2887$$

当 $n=100$ 时， $E(Y_{100})=100\times 0.5=50$, $\sigma(Y_{100})=\sqrt{100}\times 0.2887=2.887$. 于是

$$\begin{aligned} P(Y_{100}\le 60) &= P\left(\frac{Y_{100}-50}{2.887}\le\frac{60-50}{2.887}\right)\\ &\approx \Phi(3.464)=0.9997 \end{aligned}$$

这个概率很接近于 1，说明事件“ $X_1+X_2+\cdots X_n\le 60$ ”几乎必然要发生的.

5.2.2 中心极限定理

定理 5.2.1 设 $X_1, X_2, \cdots$ 为独立同分布的随机变量序列,均值为 μ ，方差为 σ^2 ,则

$$Z_n=\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}.$$

具有渐近分布 $N(0,1)$.也就是说,当 $n\to\infty$ 时

$$P\left\{\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}\le x\right\}\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt.$$

推论 5.2.1 （De Moivre-Laplace ）设 n 重 Bernoulli 试验中,事件 A 在每次试验中出现的概率为 p，$0<p<1$，μ_A 为 n 次试验中事件 A 出现的次数(即 $\mu_A\sim B(n,p)$)，对所有 $x\in\mathrm{R}$, 有

$$\lim_n P\left(\frac{\mu_A-np}{\sqrt{np(1-p)}}\le x\right)=\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt.$$

证明 令

$$X_i=\begin{cases}1, & \text{在}i\text{次试验中事件}A\text{出现}\\ 0, & \text{在}i\text{次试验中事件}A\text{不出现}\end{cases}$$

则 $X_1, X_2, \cdots$ 独立同分布,且 $\mu_A=\sum_{i=1}^{n}X_i$,注意到

$$EX_i=p, \quad DX_i=p(1-P)$$

故由定理 5.2.1 得结论成立.

注 (1) 定理 5.2.1 的结论告诉我们,只有当 n 充分大时, Z_n 才近似服从标准正态分布 $N(0,1)$.而 n 较小时.此种近似不能用.

(2) 在中心极限定理中,所谈及的一般条件可以非正式地概括为:在总和中的每个单独的项为总和的变化提供了一个不可忽视的量,而每一个单独的项都不可能给总和作出很大的贡献.

⑶ 中心极限定理说明: 为使总和能用正态分布近似, 被加项不必是正态分布. 事实证明: 在许多问题中所考虑的随机变量可以表示为 n 个独立随机变量之和, 从而它的分布可以用正态分布去近似.

例如, 在任一给定时间, 一个城市的耗电量是由大量单独的耗电者需用电量的总和; 在一个贮水池中的储水量可以看成是极大数量的单独供水池的供水量的总和; 在一个物理实验中的测量误差是由许多不可能观测到的、而可看作是可加的小误差所组成; 一个悬浮于一种液体中的小质点受到分子的碰撞, 而使它在随机的方向作随机大小的位移, 而该质点的位置（在一定长的时间以后）可以看作为各个位移的总和.

例 5.2.3 某保险公司有 10000 个同龄又同阶层的人参加人寿保险. 已知该类人在一年内死亡的概率为 0.006.每个参加保险的人在年初付 12 元保险费,而在死亡时家属可从公司得到 1000 元的保险费.问在此项业务活动中,

①保险公司亏本的概率多大?

②保险公司一年的利润不少于 40000 元，60000 元，8000 元的概率各为多大?

解 在参加人寿保险中把第 i 个人在一年内死亡记为 "$X_i=1$", 活着记为 "$X_i=0$". 则 X_i 是一个服从二项分布 $b(1,0.006)$ 的随机变量, 其和 $X_1+\cdots+X_n$ 表示一年内死亡总人数.

另一方面, 保险公司在该项业务中每年共收入 $10000\times12=120000$ 元, 故仅当每年死亡人数超过120人时公司才会亏本; 当每年死亡的人数不超过80人时公司可获利不少于 40000 元. 由此可知, 所求的概率分别为

$$P(X_1+\cdots+X_{10000}>120)$$

$$P(X_1+\cdots+X_{10000}\le 80)$$

由于 X_i 是独立同分布随机变量, $X_i\sim b(1,0.006)$. 由推论 5.2.1 知

$$P(X_1+\cdots+X_{10000}>120)$$

$$=P\left(\frac{X_1+\cdots+X_{10000}-10000\times0.0060}{\sqrt{10000\times0.006\times(1-0.006)}}>\frac{120-60}{\sqrt{59.64}}\right)$$

$$\approx 1-\Phi(7.8341)=0$$

$$P(X_1+\cdots+X_{10000}\le 80)$$

$$\approx\Phi\left(\frac{80-60}{\sqrt{59.64}}\right)=\Phi(2.6545)=0.9960.$$

同样可得保险公司一年的利润不少于 60000 元及 8000 元的概率各为 0. 5, 0. 005261

例 5.2.4 为了精确起见, 度量一个物理量许多次, 每次度量都产生一个随机误差, 合理的假设是, 在选择适当的单位下, 它是在-1 和+1 之间均匀分布着的. 现在如果我们取 n 次度量的算术平均值, 求它与真值的差小于一个微小的正数 δ 的概率是多少.

解 设 m 表示真值, $X_i, 1\le i\le n$ 表示实际的度量, 于是

$$X_i=m+\varepsilon_i$$

其中 ε_i 是在[-1, +1]中均匀分布的随机变量. 因此

$$E(\varepsilon_i)=\int_{-1}^{1}x\,dx=0,\quad \sigma^2=E(\varepsilon_i^2)=\int_{-1}^{1}\frac{1}{2}x^2dx=\frac{1}{3}.$$

$$EX_i=m,\qquad \sigma^2(X_i)=\frac{1}{3}.$$

我们要计算 $P\{|S_n-nm|<\delta\}$ 的渐进值, 为此将此式变为

$$P\left\{\left|\frac{S_n-nm}{\sqrt{n/3}}\right|<\delta\sqrt{3n}\right\}\approx 2\Phi(\delta\sqrt{3n})-1$$

5.2.3* 若干应用

1．正态随机数的产生有各种方法. 下面介绍一种常用的方法：设 $\{X_k\}$ 独立同分布，都服从[0,1]上的均匀分布，则 $EX_k=0.5$, $\sigma=1/\sqrt{12}$，由中心极限定理，n 很大时，$\eta=\dfrac{\sum_{k=1}^{n}X_k-n/2}{\sqrt{n}/\sqrt{12}}$ 近似服从标准正态分布，事实上取 n=12 就够了. 于是取区间 [0, 1] 上 12 个均匀随机数，则 $\eta=\sum_{k=1}^{12}X_k-6$ 即近似为标准正态随机数.

2．对数理统计学的许多分支，如参数（区间）估计、假设检验、抽样调查等，中心极限定理都有着重要的作用. 事实上，它也是保险精算等学科的理论基础之一. 假定某保险公司为某险种推出保险业务，现有 n 个顾客投保，第 i 份保单遭受风险后损失索赔量记为 X_i. 对该保险公司而言，随机理赔量应该是所有保单索赔量之和，记为 S，即

$$S=\sum_{i=1}^{n}X_i$$

弄清 S 的概率分布对保险公司进行保费定价至关重要. 在实际问题中，通常假定所有保单索赔相互独立. 这样，当保单总数 n 充分大时，我们并不需要计算 S 的精确分布（一般情况下这是困难甚至不可能的）. 此时，可应用中心极限定理，对 S 进行正态逼近：

$$\frac{S-ES}{\sqrt{DS}}$$

渐近具有正态分布 $N(0,1)$，并以此来估计一些保险参数.

例 5.2.5 某保险公司发行一年期的保险索赔金分别为 1 万元与 2 万元的两种人身意外险. 索赔概率 q_{1r} 及投保人数 n_{1r} 如下表所示（金额单位：万元）.

类别 k	索赔概率 q_{1r}	索赔额 b_{1r}	投保数 n_{1r}

1	0.02	1	500
2	0.02	2	500
3	0.10	1	300
4	0.10	2	500

保险公司希望只有 0.05 的可能使索赔金额超过所收取的保费总额. 设该保险公司按期望值原理进行保费定价，即保单 i 的保费 $\pi(X_i)=(1+\theta)EX_i$. 要求估计 θ.

解 计算 $S=\sum_{i=1}^{1800}X_i$ 的均值与方差

$$ES=\sum_{i=1}^{1800}EX_i=\sum_{k=1}^{4}n_kb_kq_k$$

$$=500\cdot1\cdot0.02+500\cdot2\cdot0.02+300\cdot1\cdot0.10+500\cdot2\cdot0.10=160,$$

$$DS=\sum_{i=1}^{1800}DX_i=\sum_{k=1}^{4}n_kb_k^2q_k(1-q_k)$$

$$=500\cdot1^2\cdot0.02\cdot0.98+500\cdot2^2\cdot0.02\cdot0.98$$

$$+300\cdot1^2\cdot0.10\cdot0.90+500\cdot2^2\cdot0.10\cdot0.90=256,$$

由此得保费总额

$$\pi(S)=(1+\theta)ES=160(1+\theta)$$

依题意， 我们有 $P(S\le(1+\theta)ES)=0.95$，也即

$$P\left(\frac{S-ES}{\sqrt{DS}}\le\frac{\theta ES}{\sqrt{DS}}\right)=P\left(\frac{S-ES}{\sqrt{DS}}\le10\theta\right)=0.95$$

将 $\frac{S-ES}{\sqrt{DS}}$ 近似看作标准正态随机变量，查表可得 $10\theta=1.645$，故 $\theta=0.1645$.

补充与注记

1. 在 18 和 19 世纪，极限定理一直是概率论研究的中心课题. 贝努里大数定律是第一个从数学上被严格证明的概率论定律，它由贝努里在其 1713 年出版的名著《推测术》中详细给出. 大数定律这个名称则是泊松（Poisson 1781–1840）于 1837 年提出的. 中心极限定理这个名词 1920 年由波利亚（*poo'lya*）给出，用于统称随机变量序列部分和的分布渐近于正态分布的一类定理. 它是概率论中最为重要的一类定理，并有着广泛的实际背景. 最初的中心极限定理是关于 n 重贝努里试验的，1716 年，德莫佛对 $p=1/2$ 的情形作了讨论，随后拉普拉斯将其推广到 $0<p<1$ 的情形. 从 19 世纪中叶到 20 世纪初期，一批著名的俄国数学家对概率论的发展做出了重要贡献. 他们运用严格的、强有力的数学分析工具，如富里埃变换等，将贝努里大数律、德莫佛-拉普拉斯中心极限定

理推广到一般随机变量和的情形.

2. 在 18 世纪以前，证明贝努里大数律是一件相当困难的事情，它涉及到下列和式的计算:

$$\sum_{k:|(k/n)-p|\geq\varepsilon}\binom{n}{k}p^k q^{n-k}.$$

直到德莫佛-拉普拉斯的重要发现以后，贝努里大数律才有了新的、较为简单的证明. 事实上，德莫佛-拉普拉斯证明了如下的局部和整体中心极限定理：对足够大的 n 和 $k/n\sim p$

$$\binom{n}{k}p^k q^{n-k}\sim\frac{1}{\sqrt{2\pi npq}}e^{\frac{(k-np)^2}{2npq}},$$

$$\sum_{k:|k/n-p|\geq\tau}\binom{n}{k}p^k q^{n-k}\sim\frac{1}{\sqrt{2\pi}}\int_{-\tau\sqrt{n/pq}}^{\tau\sqrt{n/pq}}e^{-x^2/2}dx$$

从上述渐近结果，我们不难得到贝努里大数定律.

3. 伯努利与大数定律

现在我们来介绍伯努利《推测术》中最重要的部分——包含了如今被称之为“伯努利大数定律”的第 4 部分. 回到前面的缶中抽球模型: 缶中有大小、质地一样的球 $a+b$ 个，其中白球 a 个，黑球 b 个，“抽出之球为白球”的概率为 p，则有 $p=a/(a+b)$. 假设有放回地从缶中抽球 n 次，记 X_n 为抽到白球的次数，以 X_n/n 估计 p. 这种估计法现今仍是数理统计学中最基本的方法之一. 此处的条件是，每次抽取时都要保证缶中 $a+b$ 个球的每一个有同等机会被抽出，但这一点在实践中并不见得容易保证. 例如，产生中奖号码时可能要用复杂的装置. 在实际工作中，统计学家有时用一种叫做“随机数表”的工具. 这是一本很厚的书，各页按行、列排列着数字 0,1,2,…,9，它们是用“充分随机”的方法产生的. 在使用时，“随机地”翻到一页并随机地点到一个位置，以此处的数字确定抽出的对象.

伯努利企图证明的是: 用 X_n/n 估计 p 可以达到事实上的确定性——他称为道德确定性. 其确切含义是：任意给定两个数 $\varepsilon>0$ 和 $\delta>0$，总可以取足够大的抽样次数 n，使事件 $\{|X_n/n-p|>\varepsilon\}$ 的概率不超过 δ. 这意思就很显然：$|X_n/n-p|>\varepsilon$ 表明估计误差未达到指定的接近程度 ε，但这种情况发生的可能性可以“随心所欲地小”（代价是加大 n）. 为忠实于伯努利的表达形式，应指出两点：一是伯努利把 ε 限定于 $(a+b)^{-1}$，虽然其证明对一般 ε 也有效. 但他做这一模型限定与所用缶子模型的特殊性有关：必要时把缶中的白、黑球分别改为 ra 和 rb 个，则 p 不变，$(a+b)^{-1}$ 改为 $(ra+rb)^{-1}$，只须取 r 足够大，便可使 $(ra+rb)^{-1}$ 任意小. 其次，伯努利欲证明的是：对任给的 $c>0$，只要抽取次数足够大，就可使

$$P\left\{\left|\frac{X_n}{n}-p\right|\leq\varepsilon\right\}>cP\left\{\left|\frac{X_n}{n}-p\right|>\varepsilon\right\}\qquad(5.2.1)$$

这与前面所说是一回事. 因为由上式得

$$P\left\{\left|\frac{X_n}{n}-p\right|>\varepsilon\right\}<\frac{1}{1+c}. \tag{5.2.2}$$

取 c 充分大，可使(5.2.2)式右边小于 δ.

另外要指出的是: 伯努利使用的这个缸子模型使被估计的 p 值只能取有理数，因而有损于结果的普遍性. 但其证明对任意的 p 成立，故这一细节并不重要.

伯努利上述对事实上确定性数学的理解，即(5.2.1)式，有一个很值得赞赏的地方，即他在概率论的发展刚刚起步的阶段，就给出了问题的一个适当的提法. 因为，既然我们欲证明的是当 n 充分大时，X_n/n 和 p 可以任意接近，则一个看来更直接了当的提法是

$$\lim_n \frac{X_n}{n}=p \tag{5.2.3}$$

而这不可能实现. 因为原则上不能排除“每次抽到白球”的可能性，这时 X_n/n 总为 1，不能收敛到 $p<1$. 或者退一步: 要求(5.2.3)式成立的概率为 1，这一结论是对的，但直到 1909 年才由波莱尔给予证明，证明的难度比伯努利的提法大得多. 设想一下，如果当时伯努利就采用该提法，他也许在有生之年不能完成这一工作. 由于波莱尔的结论比伯努利的结论强，现今人们又把他们的结论分别称之为强大数定律和弱大数定律.

习题 A

1．利用 Chebychev 不等式证明：能以大于 0.97 的概率断言，掷 1000 次均匀硬币，正面出现的次数在 400 到 600 次之间.

2．在每次试验中，事件 A 发生的概率为 0.75，利用切比雪夫不等式求：1º 在 1000 次独立试验中，事件 A 发生的次数在 700－800 之间的概率；2º n 为多大时才能保证在 n 次重复独立试验中事件 A 出现的频率在 0.74－0.76 之间的概率至少为 0.90.

3．设随机变量 X 服从幂—指分布，其概率密度为

$$f(x)=\begin{cases}\dfrac{x^m e^{-x}}{m!}, x>0;\\ 0, \qquad x\le 0.\end{cases}$$

利用切比雪夫不等式证明：

$$P[0<X<2(m+1)]\ge\frac{m}{m+1}.$$

4．设随机变量 X 的概率密度为

$$f(x)=\begin{cases}xe^{-x}, x>0\\ 0, \quad x\le 0\end{cases}$$

用 Chebychev 不等式证明 $P\{0<X<4\}\ge 1/2$

5. 从正态总体 $N(3.4, 36)$中抽取容量为 n 的样本，如果要求其样本均值位于区间(1.4, 5.4)内的概率不小于 0.95，问样本容量 n 至少应取多大？

6. 某电视机厂每月生产 10000 台电视机，但它的显象管车间的正品率为 0.8，为了以 0.997 的概率保证出厂的电视机都装上正品的显象管，该车间每月应生产多少只显象管？

7. 保险公司为 50 个集体投保人提供医疗保险，假设他们的医疗费相互独立，且医疗花费（单位：百元）服从相同的分布律 $\begin{bmatrix}0 & 0.5 & 1.5 & 3\\ 0.1 & 0.3 & 0.4 & 0.2\end{bmatrix}$. 当花费超过 1 百元时，保险公司支付超过 1 百元的部分；当花费不超过 1 百元时，由投保人自己负担，如果以总支付费 X；的期望值 EX 作为预期的总支付费，那么保险公司应收取总保险费为 $(1+\theta)EX$，其中 θ 为相对附加保险费率，为使保险公司获利的概率不小于 95%，问 θ 至少应取多大？

8. 某保险公司对 20 岁男青年卖保险，每年交 300 元，约定：若在今后 5 年内投保人死亡，则其家属可得 1000000 元保险金. 关于死亡的分布，据统计有以下记录：

死亡年龄	20	21	22	23	24
概率	0.00180	0.00182	0.00185	0.00189	0.00195
公司损失	-99700	-99400	-99100	-98800	-98500

历史资料表明一个人若能活到 25 岁并一直投保，则平均保险公司可获利 1500 元. 试问：
（1）20 岁男青年能活过 25 岁以上的概率有多大？
（2）收 300 元保险费，而一旦死亡要赔 10 万元，两者差距似乎很大，而公司还能获利，

为什么？设有十万人投保能获利多少？

（3）试求对每个 20 岁投保人，大致可获利多少？

（4）为了准备获利 1000000 元，应征集多少 20 岁男青年投保？

9.某灯泡厂生产的灯泡的平均寿命原为 2000 小时，标准差为 250 小时，经过技术改革采用新工艺使平均寿命提高到 2250 小时，标准差不变.为了确认这一改革的成果，上级技术部门派人前来检查，办法如下：任意挑选若干只灯泡，如这些灯泡的平均寿命超过 2200 小时，就正式承认改革有效，批准采用新工艺.如欲使检查通过的概率超过 0.997，至少应检查多少只灯泡？

10．某工厂生产两种硬币：均匀的和偏重的.每个偏重硬币正面出现的机会占 55%.我们现有这个厂生产的一枚硬币，但不知道它是均匀的还是偏重的.为确定这枚硬币的类型，我们作如下统计试验：、将此硬币抛 1000 次，如果正面出现了 525 次或更多，则认为它是偏重的. 如果正面出现的次数少于 525，就认为它是均匀的.假定这枚硬币事实上是均匀的，我们将得到错误的概率是多少？若硬币本来是偏重的？

11.某药厂断言，该工厂生产的某种药品对于治疗一种疑难的疾病的治愈率为 0.8.某医院试用了这种药品，任意抽查了 100 个服用次药品的病人，如果其中多于 75 人治愈，医院就接受药厂的这一断言，否则就拒绝之.问：

（1）若实际上次药品对这种疾病的治愈率为 0.8，那么，医院接受这一断言的概率是多少？

（2）若实际上次药品对这种疾病的治愈率为 0.7，那么，医院接受这一断言的概率是多少？

12.在制作统计报表时要计算 10000 个数之和.先把每个数保留 m 位小数，假设由此产生个数的误差相互独立，且均匀分布在 $(-\frac{1}{2}\times 10^{-m}, \frac{1}{2}\times 10^{-m})$ 上，求一个区间，使总计的误差落在这个区间内的概率大于 0.997.

13．某保险公司多年的统计资料表明，在索赔中被盗索赔户占 20%，以 X 表示在随机抽查的 100 个索赔户中，因被盗向保险公司索赔的户数.

（1）写出 X 的概率分布；

（2）利用 D-L 定理，求被盗索赔户不少于 14 户且不多于 30 户的概率近似值.

14．设投掷一枚均匀的骰子 100 次.令 X_i 表示第 i 次掷出的点书、数.求 $P\left\{\prod_{i=1}^{100} X_i \le a^{100}\right\}, (1<a<6)$ 的近似值.

15．某商店负责供应某地区 1000 人所需商品,其中一商品在一段时间内每人需用一件的概率为 0.6,假定在这一段时间内个人购买与否彼此无关,问商店应预备多少件这样的商品,才能以 99.7%的概率保证不会脱销(假定该商品在某一段时间内每人最多可以买一件).

16．银行为支付某日即将到期的债卷须准备一笔现金，已知这批债卷共发放了 500 张，每张须付本息 1000 元，设持卷人（1 人 1 张）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.

17*．设$U_1,U_2,\cdots,U_n$独立同分布，且$U_1\sim U(0,1)$.试利用中心极限定理证明：

$$P(\exp(-n-c\sqrt{n})\le U_1\cdots U_n\le P(\exp(-n+c\sqrt{n})\approx 2\Phi(c)-1,\qquad n>>1.$$

18*．设$X\sim E(\lambda)$,试利用 Markov 不等式证明

（1）$P(X\ge c\beta)\le\dfrac{1}{e^{c\beta}(1-\beta t)}$, $c\in R$, $0\le t\le\dfrac{1}{\beta}$;

（2）$P(X\ge c\beta)\le ce^{1-c}$.

习题 B

选择题

1．设X为非负随机变量且$EX^2=1.1, DX=0.1$,则一定有

（A）$P\{-1<X<1\}\ge 0.9$;　　(B) $P\{0<X<2\}\ge 0.9$;

(C) $P\{X+1\ge 1\}\le 0.9$;　　(D) $P\{|X|\ge 1\}\le 0.1$.

2．设$X_1,X_2,\cdots,X_{500}$为独立同分布的随机变量序列,且$X_1\sim B(1,p)$,则下列不正确的为

(A)$\dfrac{1}{500}\sum\limits_{i=1}^{500}X_i\approx p\ (P)$;(B) $\sum\limits_{i=1}^{500}X_i\sim B(500,p)$;(C) $P\left\{a<\sum\limits_{i=1}^{500}X_i<b\right\}\approx\Phi(b)-\Phi(a)$;

(D) $P\left\{a<\sum\limits_{i=1}^{500}X_i<b\right\}\approx\Phi\left(\dfrac{b-500p}{\sqrt{500p(1-p)}}\right)-\Phi\left(\dfrac{a-500p}{\sqrt{500p(1-p)}}\right)$.

(其中 P 表示在概率意义下的近似)

3．设$\{X_1,X_2,\cdots,X_n,\cdots\}$是相互独立的随机变量序列，$X_n$服从参数为 n 的指数分布（n=1,2, ……），则随机变量序列$\{X_1,2^2X_2,\cdots,n^2X_n,\cdots\}$：

(A) 服从切比雪夫大数定律;
(B) 服从辛钦大数定律;
(C) 同时服从切比雪夫大数定律和辛钦大数定律;
(D) 既不服从切比雪夫大数定律，也不服从辛钦大数定律.

4．设随机变量序列$X_1,\cdots,X_n,\cdots$独立同分布，其分布函数为$F(x)=a+\dfrac{1}{\pi}\arctan\dfrac{x}{b}$,$b\ne 0$,则辛钦大数定理对此序列

(A) 当常数a,b取适当的数值时适用；(B)不适用；(C)适用；(D)无法判别.

填空题

1．设一次试验成功的概率为 p，进行 100 次独立重复试验，当时 p=_____，成

功次数的标准差的值最大，其最大值为_____

2．设随机变量 X 的方差为 2，则根据切比雪夫不等式有估计 $P\{|X-EX|\geq 2\}\leq$ _____.

3．设 $X_1, X_2, \cdots$ 相互独立同分布，且 $EX_n=0$，则 $\lim\limits_{n\to\infty} P\{\sum\limits_{i=1}^{n} X_i < n\}=$_____

4．设 $X_1, X_2, \cdots$ 为独立同分布的随机变量序列，均值为 μ 方差 $\sigma^2>0$，则

则 $P\left\{\dfrac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}>0\right\}=$_____

5．设 $X_1, X_2, \cdots$ 为独立同分布的随机变量序列，且服从参数为 λ 的指数分布，则

$P\left\{\dfrac{\lambda(X_1+\cdots+X_n)-n}{\sqrt{n}}\leq x\right\}=$_____.

6．已知 X 的期望为 5，而均方差为 2，估计 $P\{2<X<8\}\geq$_____.另设 $E(X)=-2, E(Y)=2, D(X)=1, D(Y)=4, \rho_{XY}=-0.5$，试估计 $P(|X+Y|\geq 6)\leq$ _____.

7．设 $X\sim b(n,p)$，则由大数定理（或频率的稳定性）知 $\forall\varepsilon>0$，$\lim\limits_{n\to\infty} P\{|X-np|>n\varepsilon\}=$_____．现有 N 位学生相互独立地做实验，各自的实验误差均服从 $[0,1]$ 的均匀分布，结果发现其中恰好有 100 位学生的实验误差小于 $\dfrac{1}{3}$，用上面的大数定理近似计算 $N=$_____.

8．某班上有 100 位学生各有一部手机，上课时都开机．假设每部手机上课时间内收到电话的次数都服从平均次数为 1 的泊松分布（各人间相互独立），用中心极限定理近似计算上课时不会有电话干扰的概率为_____，该近似计算的（绝对）误差为_____.

第五章　综合例题

例 1．设 X 为连续型随机变量，c 为常数，$\varepsilon>0$，求证

$$P\{|X-c|\geq\varepsilon\}\leq\frac{E|X-c|}{\varepsilon}$$

分析　此类概率不等式的证明，一般考虑用切比雪夫不等式或直接从定义用类似切比雪夫不等式的方法来证.

证　设 X 的密度函数为 $f(x)$，则

$$\begin{aligned}P\{|X-c|\geq\varepsilon\}&=\int_{|x-c|\geq\varepsilon}f(x)dx\\&\leq\int_{|x-c|\geq\varepsilon}\frac{|x-c|}{\varepsilon}f(x)dx\leq\int_{-\infty}^{\infty}\frac{|x-c|}{\varepsilon}f(x)dx\\&=\frac{1}{\varepsilon}\int_{-\infty}^{\infty}|x-c|f(x)dx=\frac{1}{\varepsilon}E|X-c|\end{aligned}$$

例 2．设随机变量 X 和 Y 的数学期望都是 2, 方差分别为 1 和 4, 相关系数为 0.5, 则根据切比雪夫不等式有 $P\{|X-Y|\geq 6\}\leq$ _____.

解　$\frac{1}{12}$.

由于　　$E(X-Y)=0,\ D(X-Y)=DX+DY-2\rho_{XY}\sqrt{DXDY}=3,$

故

$$P\{|X-Y|\geq 6\}\leq 3/36=1/12\,.$$

例 3．设在独立重复试验中，每次试验中事件 A 发生的概率为 1/4.问是否用 0.925 的概率确信在 1000 次试验中 A 发生的次数在 200 到 300 之间?

分析 在 1000 次试验中事件 A 发生的次数 $X\sim B(1000,1/4)$, 且

$$E(X)=1000\times 1/4=250\quad D(X)=1000\times 1/4\times(1-1/4)=375/2$$

而　　$P\{200\leq X\leq 300\}=P\{|X-250|\leq 50\}$

利用 Chebychev 不等式得

$$P\{200\leq X\leq 300\}=P\{|X-250|\leq 50\}\geq 1-\frac{D(X)}{50^2}=0.925$$

所以可用 0.925 的概率确信在 1000 次试验中 A 发生的次数在 200 到 300 之间.

解　如分析所述，由 Chebychev 不等式即可得

例 4．分布用切比雪夫不等式与隶美弗—拉普拉斯中心极限定理确定：当掷一枚硬币时，需要掷多少次，才能保证出现正面的频率在 0.4~0.6 之间的概率不小于 90%。

解　设 X 为 n 次掷硬币正面出现的次数，则 $X\sim B(n,p)$，其中 $p=\frac{1}{2}$

（1）由切比雪夫不等式知

$$P\left\{0.4 \le \frac{X}{n} \le 0.6\right\} = P\left\{\left|\frac{X}{n} - 0.5\right| \le 0.1\right\} = P\{|X - 0.5n| \le 0.1n\}$$

$$\ge 1 - \frac{D(X)}{(0.1n)^2} = 1 - \frac{n \times \frac{1}{4}}{0.01 \cdot n^2} = 1 - \frac{25}{n}$$

令 $1 - \frac{25}{n} \ge 90\%$. 则得 $n \ge 250$ 。

（2） 由隶美弗-拉普拉斯的中心极限定理，得：

$$P\{0.4 \le \frac{X}{n} \le 0.6\}$$

$$= P\{0.4n \le X \le 0.6n\}$$

$$= P\{\frac{0.4n - 0.5n}{\sqrt{0.25n}} \le \frac{X - 0.5n}{\sqrt{0.25n}} \le \frac{0.6n - 0.5n}{\sqrt{0.25n}}$$

$$\approx 2\Phi(\frac{0.1n}{0.5\sqrt{n}}) - 1$$

$$= 2\Phi(\frac{\sqrt{n}}{5}) - 1 \ge 90\% \Rightarrow \Phi(\frac{\sqrt{n}}{5}) \ge 0.95.$$

查表知： $\frac{\sqrt{n}}{5} \ge 1.6$ 。

$$n \ge 67.64$$

$$\Rightarrow n \ge 68$$

例 5. (1)一个复杂系统由 100 个相互独立的元件组成, 在系统运行期间每个元件损坏的概率为 0.10, 又知为使系统正常运行, 至少必须有 85 个元件工作, 求系统的可靠度; (2)上述系统假如由 n 个相互独立的元件组成, 而且又要求至少有 80%的元件工作才能使整个系统正常运行, 问 n 至少为多大时才能保证系统的可靠度不小于 0.95.

解 (1)设 $X_i = \begin{cases} 1, \text{第}i\text{个元件没有损坏} \\ 0, \quad \text{第}i\text{个元件损坏} \end{cases}$, S 为系统正常运行时完好的元件个数, 于是

$$S = \sum_{i=1}^{100} X_i \text{ 服从 } b(100,0.9),$$

因而

$$ES = 100 \times 0.9 = 90, \quad DS = npq = 100 \times 0.9 \times 0.1 = 9.$$

故所求的概率为

$$P(S > 85) = 1 - P(S \le 85) = 1 = P\left\{\frac{S - 90}{\sqrt{9}} \le \frac{85 - 90}{\sqrt{9}}\right\} = 1 - \Phi\left(-\frac{5}{3}\right) = 0.952.$$

(2)此时 $S \sim b(n,0.9)$, 要求 $P(S \ge 0.8n) \ge 0.95$, 而

$$P(S \geq 0.8n) = 1 - P\left\{\frac{S-0.9n}{0.3\sqrt{n}} \leq \frac{0.8n-0.9n}{0.3\sqrt{n}}\right\} = 1-\Phi\left(-\frac{\sqrt{n}}{3}\right) = \Phi\left(\frac{\sqrt{n}}{3}\right).$$

故 $\Phi\left(\frac{\sqrt{n}}{3}\right) \geq 0.95$，查表得 $\frac{\sqrt{n}}{3} \geq 1.65, \Rightarrow n \geq 24.5,$ 取 n =25

例 6. 一加法器同时收到 20 个噪声电压 $V_i, (i = 1,2,\cdots,20)$，设它们是相互独立且都服从区间（0，10）上的均匀分布，求总和噪声电压超过计划 105（伏）的概率.

解 记 $V = \sum_{i=1}^{20} V_i$，因 $V_1, V_2, \cdots, V_{20}$ 是相互独立且都服从（0，10）上的均匀分布，且

$$\mu = E(V_i) = 5, \sigma_i = D(V_i) = \frac{100}{12}, i = 1,2,\cdots,20$$

由独立同分布中心极限定理知

$$V = \sum_{i=1}^{20} V_i \xrightarrow{n\to\infty} N(20\times 5, 20\times\frac{100}{12}) = N(100, \frac{500}{3}),$$

故

$$P(V > 105) \approx 1 - P(V \leq 105) = 1 - \Phi(\frac{105-100}{\sqrt{500/3}})$$
$$= 1 - \Phi(0.39) = 0.3483.$$

例 7. 假设 $X_1, X_2 \cdots, X_n$ 是来自总体 X 的简单随机样本；已知 $EX^k = \alpha_k (k = 1,2,3,4),$ 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.

分析 此题主要考查对中心极限定理的理解与运用.

解 依题意知 $X_1, X_2, \cdots, X_n$ 独立同分布，从而其函数 $X_1^2, X_2^2, \cdots, X_n^2$ 也是独立同分布，且

$$EX_i^2 = EX^2 = \alpha_2, DX_i^2 = EX_i^4 - (EX_i^2)^2 = \alpha_4 - \alpha_2^2,$$
$$EZ_n = \frac{1}{n}\sum_{i=1}^{n} EX_i^2 = \alpha_2,$$
$$DZ_n = D(\frac{1}{n}\sum_{i=1}^{n} X_i^2) = \frac{1}{n^2}\sum_{i=1}^{n} DX_i^2 = \frac{1}{n}(\alpha_4 - \alpha_2^2)$$

由中心极限定理

$$U_n = \frac{Z_n - \alpha_2}{\sqrt{(\alpha_4 - \alpha_2^2)/n}}$$

的极限分布为标准正态分布，即当 n 充分大时，Z_n 近似地服从参数为 $(\alpha_2, \frac{\alpha_4 - \alpha_2^2}{n})$ 的正态分布.

例 8. 设随机变量 $X_i, 1 \leq i \leq n,$ 独立同分布，且分布密度为 f(x)，记 $p = P\{\sum_{i=1}^{n} X_i \leq x\}$，当 n 充分大时，则有

A. p 可以根据 $f(x)$ 计算；　　B . p 不可以根据 f(x)计算；

C. p 一定可以用中心极限定理近似计算；　D. p 一定不可以用中心极限定理近似计算

解　由于 $X_i, 1 \le i \le n,$ 独立同分布，它们的联合概率密度等于各边缘密度的乘积.因此 p 可以如下计算：

$$p = \int_{x_1+\cdots+x_n \le x} \cdots \int f_1(x_1) \cdots f_n(x_n) dx_1 \cdots dx_n$$

由于不知道 $X_i, 1 \le i \le n.$ 的期望和方差是否存在，故无法判断能否用中心极限定理.

综上所述，选 A.