大 数 定 律

12.1 大数定律

抛掷一枚均匀硬币，记 n 次抛掷中出现正面的次数为 μ_n，μ_n 是不确定的，但直观经验告诉我们，当 n 越来越大时，出现正面的频率 $\frac{\mu_n}{n}$ 将逐渐接近于 $\frac{1}{2}$。

利用计算机模拟抛硬币的过程：随机生成一个 0-1 之间的数，如果这个数大于 1/2 就认为是抛到了正面，否则即为抛到了背面。关于如何用计算机生成随机数，我们在 12.3 中会进一步介绍。模拟 n 次抛硬币的过程算一次试验，记录得到的正面的次数 μ_n。

表中给出了 n 分别等于 10, 50, 100 和 1000 时的一些模拟结果，对每个 n 重复 10 次试验，每次试验模拟抛硬币 10 次。当 $n=10$ 时，得到的频率 0.4 到 0.7 不等，与 1/2 有较明显的偏离；而当 $n=50$ 和 100 时，10 次试验的频率都在 0.4 到 0.6 之间；当 $n=1000$ 时，与 1/2 偏离最大的频率是 0.518。可以明显地看出随着 n 的增大，出现正面的频率越来越接近于 1/2。

$n=10$		$n=50$		$n=100$		$n=1000$	
μ_n	$\frac{\mu_n}{n}$	μ_n	$\frac{\mu_n}{n}$	μ_n	$\frac{\mu_n}{n}$	μ_n	$\frac{\mu_n}{n}$
7	0.7	22	0.44	56	0.56	518	0.518
6	0.6	22	0.44	46	0.46	517	0.517
7	0.7	20	0.40	42	0.42	504	0.504
4	0.4	29	0.58	51	0.51	503	0.503
5	0.5	23	0.46	54	0.54	498	0.498
4	0.4	23	0.46	53	0.53	495	0.495
6	0.6	26	0.52	56	0.56	504	0.504
4	0.4	24	0.48	53	0.53	490	0.490
7	0.7	27	0.54	56	0.56	514	0.514
4	0.4	23	0.46	40	0.40	504	0.504

****

“频率收敛于概率”，抽样次数越多频率越接近于概率，平均值越接近于期望。

考虑a个白球，b个黑球的盒子，摸到白球的概率为$\frac{a}{a+b}$。

概率为$\frac{a}{a+b}$的含义为：重复次数n，μ_n次抽到白球，则n越大，$\frac{\mu_n}{n}$越接近于$\frac{a}{a+b}$。

这个事实虽然感觉很显然，但是μ_n是不确定的，这种越来越接近的确切含义到底是什么，它与确定性的序列的极限是不同的。直到18世纪，数学家伯努利才给出了一个严格的数学描述和理论证明。

****

伯努利大数定律

定理1：设μ_n为n重伯努利试验中事件A发生的次数，p为每次试验中A出现的概率。则对任意的$\varepsilon>0$，有$\lim_{n\to\infty}\mathrm{P}\left(\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right)=1$。

证明： 利用切比雪夫不等式

若随机变量X的期望、方差存在，则对任意 $\varepsilon>0$，$P\left(\left|X-E(X)\right|\geq\varepsilon\right)\leq\frac{Var(X)}{\varepsilon^2}$。

$\mu_n\sim B(n,p)$，$E(\mu_n)=np$，$Var(\mu_n)=np(1-p)$，则

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

$$P\left(\left|\frac{\mu_A}{n}-p\right|<\varepsilon\right)\geq 1-\frac{Var\left(\frac{\mu_n}{n}\right)}{\varepsilon^2}=1-\frac{\frac{p(1-p)}{n}}{\varepsilon^2}\xrightarrow{n\to\infty}1.$$

****

大数定律的更一般的形式

定理 2：设$\{X_k,k=1,2,\cdots\}$是相互独立同分布的随机变量序列，且其数学期望为μ，

方差为 σ^2。则对于任意给定的 $\varepsilon>0$，$\lim\limits_{n\to\infty}P\left(\left|\frac{1}{n}(X_1+\cdots+X_n)-\mu\right|\geq\varepsilon\right)=0$.

此时，我们称随机变量序列 $\{X_k,k=1,2,\cdots\}$ 服从大数定律。

证明：由于 $X_1,X_2,\cdots,X_n,\cdots$ 相互独立（实际上只需两两不相关即可）且同分布，故由切比雪夫（Chebyshev）不等式知，对任意 $\varepsilon>0$，有

$$P\left(\left|\frac{1}{n}(X_1+\cdots+X_n)-\mu\right|\geq\varepsilon\right)\leq\frac{Var\left(\frac{1}{n}(X_1+\cdots+X_n)\right)}{\varepsilon^2}=\frac{\sigma^2}{n\varepsilon^2}\to 0,\qquad (n\to\infty)$$

****

依概率收敛

$\{X_n,n=1,2,\cdots\}$ 为一个随机变量序列，X 为一随机变量，如果对任意的 $\varepsilon>0$ $\lim\limits_{n\to\infty}P(|X_n-X|<\varepsilon)=1$，则称 $\{X_n,n=1,2,\cdots\}$ 依概率收敛于 X，记作 $X_n\xrightarrow{P}X$。

大数定律的一般形式还有另一中表述方法

定理 3：设 $\{X_n,n=1,2,\cdots\}$ 是相互独立同分布的随机变量序列，且其数学期望为 μ，方差为 σ^2。则 $\lim\limits_{n\to\infty}P\left(\frac{1}{n}(X_1+\cdots+X_n)\leq x\right)=\begin{cases}0, & 若x<\mu,\\ 1, & 若x>\mu.\end{cases}$

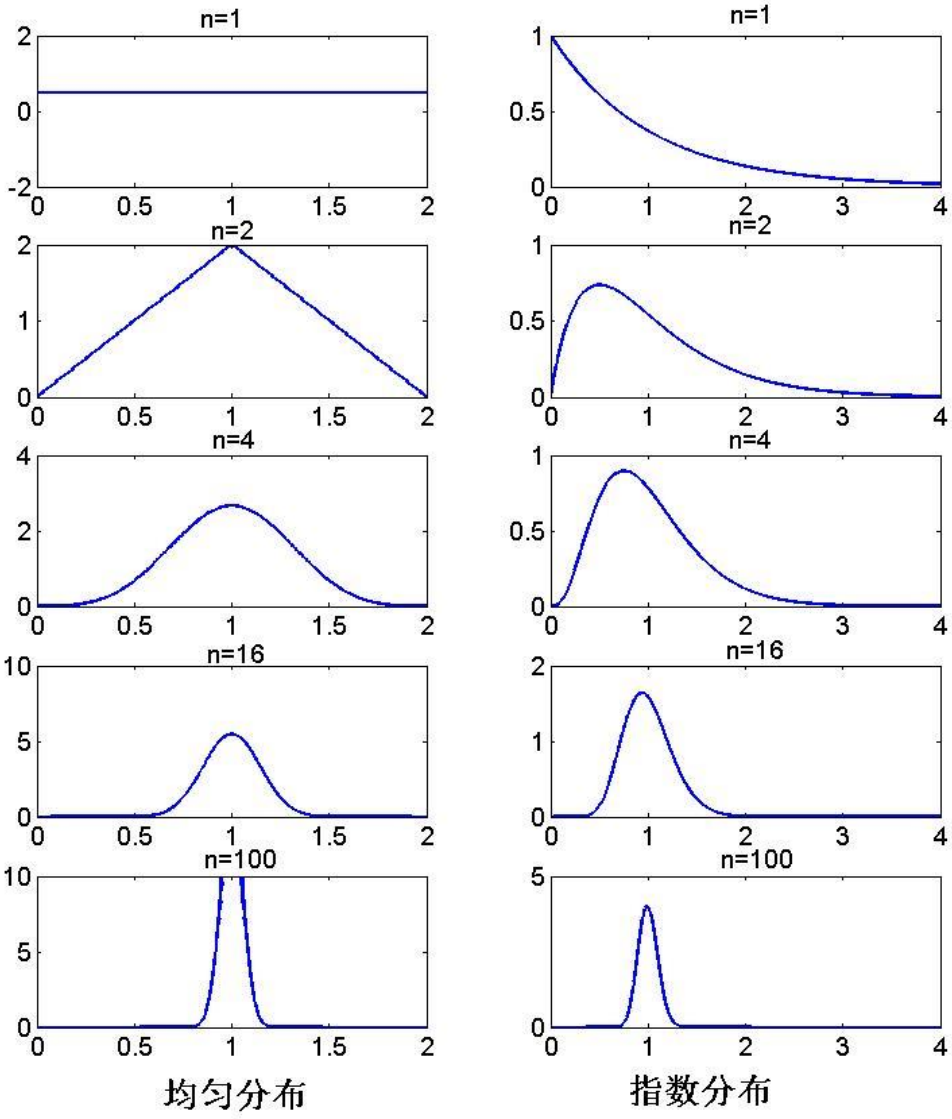
****

为了直观的理解定理2和3，我们分别考虑两个独立同分布随机变量序列 $\{X_k,k=1,2,\cdots\}$ 和 $\{Y_k,k=1,2,\cdots\}$，满足 $X_k\sim U[0,2]$，$Y_k\sim Exp(1)$，并分别记

$$\overline{X}=\frac{1}{n}(X_1+X_2+\cdots X_n),\quad \overline{Y}=\frac{1}{n}(Y_1+Y_2+\cdots Y_n)$$

显然 $E(\overline{X})=1,Var(\overline{X})=\frac{1}{3n},\quad E(\overline{Y})=1,Var(\overline{Y})=\frac{1}{n}$。

这里我们仅画出了 n 分别取1，2，4，16 ，100 时，$\overline{X}$ 及 $\overline{Y}$ 的分布密度的图像。可以看出，随着 n 的增大，$\overline{X}$ 和 $\overline{Y}$ 的取值越来越集中在它们各自的均值1的周围。



由定理 2 所给出的大数定律中，随机变量序列 $\{X_k, k=1,2,\cdots\}$ 需要满足独立同分布且期望和方差均存在的条件。人们又进一步研究这些条件可以得到什么样程度的减弱，其中比较常用的结论有切比雪夫大数定律，马尔科夫大数定律和辛钦大数定律等。其中，辛钦大数定律只要求独立同分布且存在期望，对方差没有限制。目前，各种形式的大数定律的研究仍然远远没有达到完善的程度，仍然在继续。

****

切比雪夫大数定律

设$X_1,X_2,\cdots,X_k\cdots$是两两不相关的随机变量序列，方差有界，则对于任意给定的正数ε，有 $\lim\limits_{n\to\infty}P\left(\left|\frac{1}{n}\sum\limits_{k=1}^{n}X_k-\frac{1}{n}\sum\limits_{k=1}^{n}E\left(X_k\right)\right|<\varepsilon\right)=1.$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}E\left(X_i\right)\right|<\varepsilon\right)=1-P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}E\left(X_i\right)\right|\geq\varepsilon\right)\geq1-\frac{Var\left(\frac{1}{n}\sum\limits_{i=1}^{n}X_i\right)}{\varepsilon^2}$$

$$Var\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\left[\sum_{i=1}^{n}Var\left(X_i\right)+\sum_{1\leq i<j\leq n}Cov\left(X_iX_j\right)\right]\to0$$

马尔科夫大数定律

$\frac{1}{n^2}Var\left(\sum\limits_{k=1}^{n}X_k\right)\to0$（马尔科夫条件），则$\{X_k,k=1,2,\cdots\}$服从大数律，

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

$$\forall\varepsilon>0,\quad P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\frac{1}{n}\sum_{k=1}^{n}E\left(X_k\right)\right|<\varepsilon\right\}\geq1-\frac{Var\left(\frac{1}{n}\sum\limits_{i=1}^{n}X_k\right)}{\varepsilon^2}\to1$$

辛钦大数定律

设$\{X_k,k=1,2,\cdots\}$是独立同分布的随机变量序列。如果其期望$E\left(X_1\right)=\mu$，则对任意$\varepsilon>0$，当$n\to\infty$时，都有$P\left(\left|\frac{X_1+X_2+\cdots+X_n}{n}-\mu\right|<\varepsilon\right)\to1$。

****

12.2 中心极限定理

大数定律说明平均值与期望之间的偏差小于任意给定正的常数的概率趋于1。如果

有方差有限的条件，有下面更精细的结果。

中心极限定理（林德伯格-勒维Lindeberg-Levy）：设$\{X_n\}$是独立同分布的随机变量序列。如果其期望$E(X_1)=\mu$，方差$Var(X_1)=\sigma^2$，则对每一个固定的y有

$\lim\limits_{n\to\infty}\mathrm{P}\left(\dfrac{X_1+X_2+\cdots+X_n-n\mu}{\sigma\cdot\sqrt{n}}\le y\right)=\Phi(y)=\dfrac{1}{\sqrt{2\pi}}\int_{-\infty}^{y}e^{-\frac{t^2}{2}}dt$。

实际上，对独立同分布的随机变量序列$\{X_n\}$，则

$X_1+X_2+\cdots+X_n\sim N\left(E(X_1+X_2+\cdots+X_n),Var(X_1+X_2+\cdots+X_n)\right)$

$E(X_1+X_2+\cdots+X_n)=n\mu$，$Var(X_1+X_2+\cdots+X_n)=n\sigma^2$

$X_1+X_2+\cdots+X_n\sim N(n\mu,n\sigma^2)$，标准化即得：$\dfrac{X_1+X_2+\cdots+X_n-n\mu}{\sqrt{n}\cdot\sigma}\sim N(0,1)$

****

例12.2.1　一个生产线生产的产品成箱包装，每箱的重量是随机的，假设每箱平均重量50kg，标准差为5kg，若用最大载重量5吨的汽车运输产品，试用中心极限定理估计每辆车最多可以装多少箱，才能保障不超载的概率大于0.977（$\Phi(2)=0.977$）。

解：设每箱的重量为随机变量X，则$E(X)=50$，$Var(X)=5^2$；

设一辆车所装各箱产品的重量依次为随机变量$X_1,X_2,\cdots X_n$，则$X_1,X_2,\cdots X_n$相互独立且与X分布相同，所以 $X_1+X_2+\cdots+X_n\sim N(50n,25n)$，

$P(X_1+X_2+\cdots+X_n\le 5000)=P\left(\dfrac{X_1+X_2+\cdots+X_n-50n}{5\sqrt{n}}\le\dfrac{5000-50n}{5\sqrt{n}}\right)\approx\Phi\left(\dfrac{5000-50n}{5\sqrt{n}}\right)$

$\Phi\left(\dfrac{5000-50n}{5\sqrt{n}}\right)\ge 0.977\Rightarrow\dfrac{5000-50n}{5\sqrt{n}}\ge 2\Rightarrow n\le 98.02$。

估计每辆车最多可以装98箱，才能保障不超载的概率大于0.977。

****

例12.2.2（独立和的近似）计算机在进行加法运算时，对每个被加数取整，设所有

的取整误差是相互独立的，且它们都在 $(-0.5,0.5)$ 上服从均匀分布，问大约多少个数相加时，误差总和绝对值小于 10 的概率在 0.90 左右？

解：设随机变量 X_k 表示第 k 个被加数的取整误差，则由题设条件知，$X_1, X_2, \cdots, X_n$ 相互独立，且均服从 $(-0.5,0.5)$ 上的均匀分布，

所以期望 $\mu = E\left(X_k\right) = 0$，方差 $\sigma^2 = Var\left(X_k\right) = \dfrac{1}{12}$。

由题意，要确定满足 $P\left(\left|X_1 + \cdots + X_n\right| < 10\right) \approx 0.90$ 的 n.

由中心极限定理，$\dfrac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}\sigma} = \dfrac{X_1 + \cdots + X_n}{\sqrt{n} \times \sqrt{1/12}}$ 近似服从 $N\left(0,1\right)$，

因此，$P\left(\left|X_1 + \cdots + X_n\right| < 10\right) = P\left(\left|\dfrac{X_1 + \cdots + X_n}{\sqrt{n/12}}\right| < \dfrac{10}{\sqrt{n/12}}\right)$

$$\approx \Phi\left(20\sqrt{3/n}\right) - \Phi\left(-20\sqrt{3/n}\right) = 2\Phi\left(20\sqrt{3/n}\right) - 1$$

$P\left(\left|X_1 + \cdots + X_n\right| < 10\right) \approx 0.90 \Rightarrow \Phi\left(20\sqrt{3/n}\right) \approx 0.95$，

查表得 $20\sqrt{3/n} \approx 1.645$，所以 $n \approx 443$。

大约 443 个数相加时，误差总和的绝对值小于 10 的概率在 0.90 左右。也就是说，被加数不超过 443 个时，可以有不小于 0.90 的概率保证取整误差总和的绝对值小于 10。

****

第 11 周课中二项分布的正态近似，是最早发现的中心极限定理的结果。棣莫弗-拉普拉斯定理是一个特殊的中心极限定理。

棣莫弗-拉普拉斯定理 若 $X \sim B\left(n,p\right)$，则对任何两个常数 a 和 b，$-\infty < a < b < +\infty$，

$\lim\limits_{n\to\infty} P\left(a < \dfrac{X - np}{\sqrt{np\left(1-p\right)}} < b\right) = \int_a^b \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$，即 $n \to \infty$ 时，$X \dot{\sim} N\left(np, np\left(1-p\right)\right)$。

考虑n个独立的$0-1$随机变量$X_k \sim B(1,p)$，$k=1,\cdots,n$，则$X=X_1+X_2+\cdots+X_n$

$E(X_k)=p$，$Var(X_k)=p(1-p)$，所以 $X \sim N\left(np,np(1-p)\right)$

****

例 12.2.3（二项分布的正态逼近） 在某一寿险公司中的一个项目有 3000 个同一年龄的人参加人寿保险，在 1 年里，这些人的死亡率为 0.1%，参加保险的人在年初交纳保险费 10 元，若被保人在 1 年内死亡，保险受益人可以从保险公司领取 2000 元，求保险公司的这个项目 1 年中获利不小于 10000 元的概率。

解：设 1 年内死亡的人数为 X，死亡的概率为 0.001，则 $X \sim B(3000,0.001)$。

保险公司这 1 年的收入为 30000 元，赔付$2000X$ 元。

$$P\left(\text{保险公司1年中获利不小于10000元}\right)=P(30000-2000X \geq 10000)=P(0 \leq X \leq 10)$$

$X \sim B(3000,0.001)$，

$$\frac{X-np}{\sqrt{np(1-p)}}=\frac{X-3000\times 0.001}{\sqrt{3000\times 0.001(1-0.001)}}=\frac{X-3}{1.7312}\text{ 近似服从 } N(0,1)$$

从而 $P(0 \leq X \leq 10)=P\left(\frac{-3}{1.7312} \leq \frac{X-3}{1.7312} \leq \frac{10-3}{1.7312}\right)$

$$\approx \Phi(4.043)-\Phi(-1.733) \approx 1-1+\Phi(1.733) \approx 0.96$$

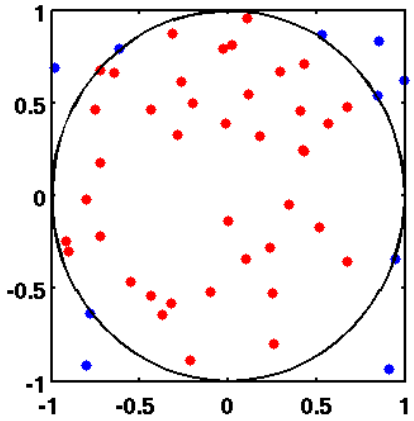
即保险公司的这个项目 1 年中获利不小于 10000 元的概率大约可达到 0.96。

****

12.3 蒙特卡洛（Monte Carlo）算法

蒙特卡洛（Monte Carlo）方法是计算机出现之后，利用概率模型近似计算的方法。

例如右图中单位圆的面积是 π，在 $[-1,1]\times[-1,1]$ 区域内均匀地撒点，落在单位圆内的点标为红色，落在圆外的点标为蓝色。如果共抛了 n 个点，落在单位圆内的红色点有 m 个，则 $\frac{S_{单位圆}}{S_{正方形}}\approx\frac{m}{n}$，已知 $S_{正方形}=4$，则得到 $S_{单位圆}=\pi\approx 4\cdot\frac{m}{n}$，其理论基础是大数定律。



****

设第 k 次撒点落入单位圆内时，随机变量 $X_k=1$，落到单位圆外，则 $X_k=0$。则

$X_k\sim\begin{pmatrix}0 & 1\\ 1-\frac{\pi}{4} & \frac{\pi}{4}\end{pmatrix}$，$k=1,2,\cdots,n$，$E(X_k)=\frac{\pi}{4}$。而 $m=X_1+X_2+\cdots+X_n$，

根据大数定律，对任意的 $\varepsilon>0$，

$$\lim_{n\to\infty}P\left(\left|\frac{1}{n}(X_1+\cdots+X_n)-\frac{\pi}{4}\right|\geq\varepsilon\right)=\lim_{n\to\infty}P\left(\left|\frac{m}{n}-\frac{\pi}{4}\right|\geq\varepsilon\right)=0。$$

Monte Carlo方法的基本想法是构造一个随机变量，使得所希望计算的量是这个随机变量的某个数字特征（通常这个数字特征是数学期望）。然后通过随机模拟的方法得到这个数字特征的估计，从而得到所希望计算的量的估计。可利用中心极限定理对Monte Carlo方法的精度作进一步的分析。

****

例12.3.1 $X_1,X_2,\cdots,X_{2n}$ 相互独立，且均服从(0,1)内的均匀分布

$$Y_k=\begin{cases}4, X_{2k-1}^2+X_{2k}^2<1\\ 0, 其他\end{cases}，k=1,2,\cdots,n，$$

(1) 对任意给定的正整数 n，证明 $\overline{Y}=\frac{Y_1+Y_2+\cdots+Y_n}{n}$ 的期望为 π；

(2) 用中心极限定理估计 $n=100$ 时， $P\left(\left|\bar{Y}-\pi\right|<0.1\right)$；

(3) 用切比雪夫不等式估计， n 取多大时，可保证 $P\left(\left|\bar{Y}-\pi\right|<0.1\right)\geq 0.9$。

(1) 证明： 对所有 $k=1,2,\cdots,n$， Y_k 服从两点分布

$$P\left(Y_k=4\right)=\frac{\pi}{4},\quad P\left(Y_k=0\right)=1-\frac{\pi}{4},\quad E\left(Y_k\right)=\pi。$$

$$E\left(\bar{Y}\right)=E\left(\frac{Y_1+Y_2+\cdots+Y_n}{n}\right)=\frac{E\left(Y_1\right)+\cdots+E\left(Y_n\right)}{n}=\pi。$$

(2) 解： $E\left(\bar{Y}\right)=\pi$， $Var\left(Y_1\right)=E\left(Y_1^{\,2}\right)-E\left(Y_1\right)^2=\pi\left(4-\pi\right)$，

$$Var\left(\bar{Y}\right)=\frac{Var\left(Y_1\right)}{n}=\frac{\pi\left(4-\pi\right)}{n}=\frac{\pi\left(4-\pi\right)}{100},$$

$$\bar{Y}\sim N\left(\pi,\frac{\pi\left(4-\pi\right)}{100}\right),\quad \frac{\bar{Y}-\pi}{\sqrt{\frac{\pi\left(4-\pi\right)}{100}}}=\frac{10\cdot\left(\bar{Y}-\pi\right)}{\sqrt{\pi\left(4-\pi\right)}}\sim N\left(0,1\right)。$$

$$P\left(\left|\bar{Y}-\pi\right|<0.1\right)=P\left(\left|\frac{10\cdot\left(\bar{Y}-\pi\right)}{\sqrt{\pi\left(4-\pi\right)}}\right|<\frac{1}{\sqrt{\pi\left(4-\pi\right)}}\right)$$

$$\approx\Phi\left(\frac{1}{\sqrt{\pi\left(4-\pi\right)}}\right)-\Phi\left(\frac{-1}{\sqrt{\pi\left(4-\pi\right)}}\right)=2\Phi\left(\frac{1}{\sqrt{\pi\left(4-\pi\right)}}\right)-1=0.46$$

可以看到 Monte Carlo方法的精度非常低，用100个样本平均，得到 π 的绝对误差小于0.1的估计的概率不足1/2。 Monte Carlo 方法相比于传统计算方法的优势并不在于二、三维空间面积与体积的近似。科学、工程计算中常常需要计算非常高维空间的体积的问题，维数可能达到几千万甚至上亿的规模，这时传统的方法都不再有效，只有借助于Monte Carlo方法才可能得到有效的结果。

(3) 解： 利用切比雪夫不等式 $P\left(\left|\bar{Y}-\pi\right|<0.1\right)=P\left(\left|\bar{Y}-E\left(\bar{Y}\right)\right|<0.1\right)\geq 1-\frac{Var\left(\bar{Y}\right)}{0.1^2}$

$$1-\frac{Var\left(\bar{Y}\right)}{0.1^2}=1-\frac{100\cdot\pi\cdot\left(4-\pi\right)}{n}\ \geq 0.9$$

$n\geq 1000\cdot\pi\cdot\left(4-\pi\right)\approx 2697$，可保证 $P\left(\left|\bar{Y}-\pi\right|<0.1\right)\geq 0.9$。

****

12.4 伪随机数和随机模拟

随机数是随机算法实现的先决条件。随机数的质量对于随机算法的效果起着极为关键的作用。从 20 世纪二、三十年代开始，人们编制过一些随机数表。但是，这些随机数表远远达不到现代的仿真模拟，Monte Carlo 计算等诸多领域的实际应用。这些实际应用需要大量、快速地生成的随机数。这必须借助计算机程序实现。但计算机上却无法生成真正的随机数。因为在计算机上，一切事情均是确定的。在实际应用中，人们通常以一些简单的算术操作实现某种确定性的规则，以此产生一列看起来很像随机数的数字作为随机数使用。这样的数字序列叫做伪随机数列，它们仅仅在有限的意义下是随机的。最为常用，也最为基本的是生成 0,1 区间上的均匀分布的伪随机数列。

从 1948 年开始，人们对如何用计算机生成好的伪随机数进行了大量的研究，其中包括 Von Neumann，Knuth 这样的著名学者。经过几十年的不断改进，目前为人们广泛采用的伪随机数生成算法大约是 2000 年左右提出的。$(0,1)$区间的均匀随机数很容易得到,除了 Matlab 等专业计算软件,Excel 中也有 rand 命令可生成$(0,1)$区间的均匀伪随机数。

设随机变量U服从$[0,1]$上的均匀分布，函数F为定义在实数集合R的连续单调递增函数，且对任何 $x\in R$ 有 $F(-\infty)=0\leq F(x)\leq 1=F(+\infty)$。则随机变量 $X=F^{-1}(U)$的概率分布函数为$F(x)$。

****

例12.4.1 利用$U(0,1)$分布下的伪随机数生成服从参数为 λ（即期望为$\frac{1}{\lambda}$）的指数

分布的伪随机数。

解： 设随机变量 $X \sim U(0,1)$，考虑随机变量 $Y = \dfrac{-\ln X}{\lambda}$；

当 $y \le 0$ 时， $F_Y(y) = P(Y \le y) = 0$，

当 $y > 0$ 时， $F_Y(y) = P(Y \le y) = P\left(\dfrac{-\ln X}{\lambda} \le y\right) = P\left(X \le e^{-\lambda y}\right) = e^{-\lambda y}$

所以， $Y = \dfrac{-\ln X}{\lambda}$ 服从参数为 λ （即期望为 $\dfrac{1}{\lambda}$）的指数分布。

生成一列服从 $(0,1)$ 区间内均匀分布的伪随机数 $x_1, x_2, \cdots, x_n$，

则 $\dfrac{-\ln x_1}{\lambda}$， $\dfrac{-\ln x_2}{\lambda}$， …， $\dfrac{-\ln x_n}{\lambda}$ 即为服从参数为 λ 的指数分布伪随机数。

****

模拟实例： 独立抛掷一个均匀的色子，设 X_k 是第 k 次掷出的点数，

则 $E(X_k) = 3.5$， $Var(X_k) = \dfrac{1^2 + 2^2 + \cdots + 6^2}{6} - 3.5^2 = \dfrac{35}{12}$。

大数定律的结论是：当 n 很大时，平均点趋于3.5，即平均点数将无限接近于3.5。

而中心极限定理的结论是当 n 很大时平均点数 $\dfrac{X_1 + X_2 + \cdots + X_n}{n}$ 作为一个随机变量

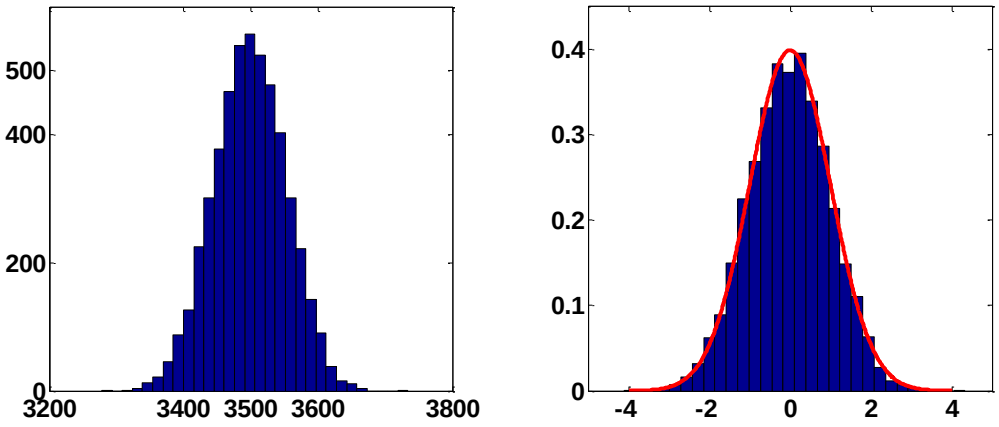
趋于一个期望为3.5，方差为 $\dfrac{Var(X_1)}{n} = \dfrac{35}{12n}$ 的正态分布，设

$Y_n = \dfrac{\dfrac{X_1 + X_2 + \cdots + X_n}{n} - 3.5}{\sqrt{\dfrac{35}{12n}}} = \dfrac{X_1 + X_2 + \cdots + X_n - 3.5n}{\sqrt{\dfrac{35n}{12}}}$， 则 $Y_n \sim N(0,1)$。

当 $n = 1000$ 时， $Y_{1000} = \dfrac{X_1 + X_2 + \cdots + X_{1000} - 3500}{\sqrt{\dfrac{35 \times 1000}{12}}}$ 近似服从标准正态分布。

模拟 $X_1, X_2, \cdots, X_{1000}$，$[0.4549, 0.3325, 0.9437, 0.3031, \cdots] \to [3, 3, 6, 2, \cdots]$

对 $X_1, X_2, \cdots, X_{1000}$ 求和，做5000次模拟，得到下面的左边的直方图。直方图显示的纵坐标显示的是落入对应横坐标区间的样本数。右图显示5000个 Y_{1000} 模拟值的直方图，直方图进行了归一化，即使得直方图围成面积等于1。右图中红色曲线为标准正态分布的密度函数曲线。



$P\{|Y_{1000}| \le y\} \approx 2\Phi(y) - 1$，取 $y = 1$，有 $P\{|Y_{1000}| \le 1\} \approx 2\Phi(1) - 1 = 0.6827$，

在5000次试验中，Y_{1000} 实际落入 $[-1,1]$ 范围的次数是3417，$\frac{3417}{5000} = 0.6834$。

$|Y_{1000}| \le 1$ 对应于点数和的范围，应该是：$[3446, 3554]$，即

$$3500 - \sqrt{\frac{35000}{12}} \le X_1 + X_2 + \cdots + X_{1000} \le 3500 + \sqrt{\frac{35000}{12}}$$。

****

Matlab参考程序（模拟掷n次色子）

```
m=5000; n=1000;                    % m为试验次数，n为每次试验投掷色子的次数
for k=1:m
    b(k)=sum(ceil(rand(1,n)*6));
end
```

```
subplot(1,2,1); hist(b,30);      % m次投掷点数和的直方图
bn=(b-3.5*n)/sqrt(35*n/12); [c1,c2]=hist(bn,30);
c1=c1/(c2(2)-c2(1))/m; subplot(1,2,2);
bar(c2,c1,1);                          % 归一化后的直方图（经验密度函数）
c=-4:0.01:4; hold on;
plot(c,1/sqrt(2*pi)*exp(-c.^2/2)); % 标准正态曲线
**********************************************************
```