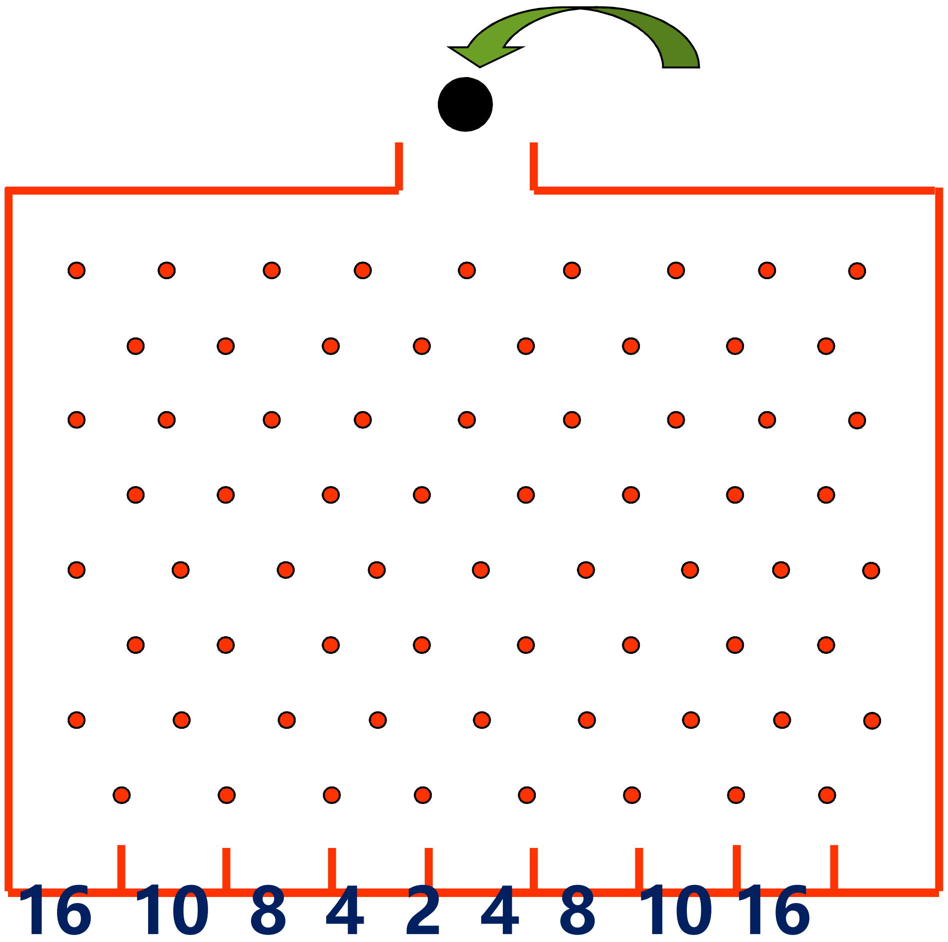
独立同分布中心极限定理

一. 引例

奖券设置问题：

将小球从游戏板上方入口投入，碰到板上钉子后逐层下落，落入底层格中对应数字则为游戏者获得奖券数



恭喜你获得4张奖券！

奖券设置合理吗

符合哪方利益

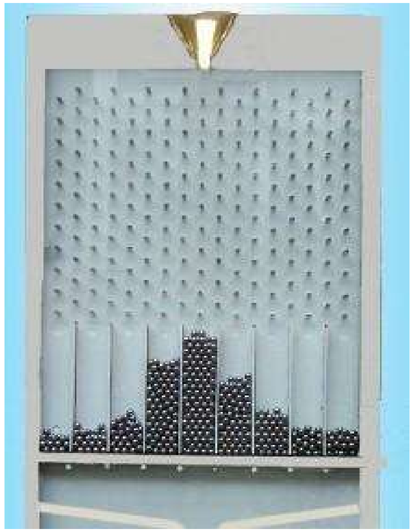
独立同分布中心极限定理

猜测：小球落到中间的机会大，两头小

如何验证？

建立数学模型，描述小球运动过程：

模拟试验



假设：

1. 共有n层钉子;

2. 小球入口处水平位置为坐标原点0;

3. 小球在每层碰到钉子后，向左或向右等可能位移一格，不会出现跳格（位移2格以上）的情况;

4. 小球在不同层向左或向右是相互独立的.

则随机变量序列：$X_k = \begin{cases} 1, & \text{第}k\text{层向右} \\ -1, & \text{第}k\text{层向左} \end{cases}$ $\quad k = 1,\ldots,n$

独立同分布中心极限定理

$$X_k=\begin{cases}1, & \text{第}k\text{层向右}\\ -1, & \text{第}k\text{层向左}\end{cases}\qquad k=1,\ldots,n$$

X_k	-1	1
p	$1/2$	$1/2$

独立同分布

完整描述了小球在n层钉板的运动过程．

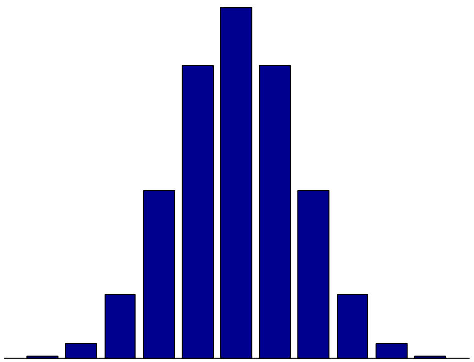
关注：小球经n次碰撞后在钉板底层所处位置

$$Y_n=\sum_{k=1}^{n}X_k$$

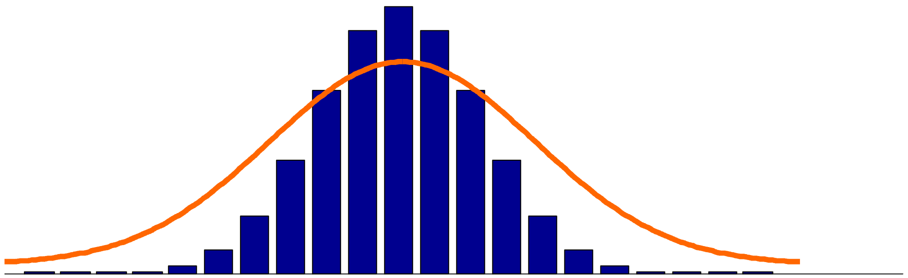
分布情况?

直接计算可得

独立同分布中心极限定理



Y_5的分布律



Y_{10}的分布律

设想：小球在无穷大的钉板中碰撞运动($n\to\infty$)时，

Y_n将呈现怎样的分布规律?

猜测： 正态分布？

事实上，算出 $E(Y_n)=0, D(Y_n)=n$

考虑小球位置的标准化随机变量序列

$$Y_n^* = \frac{Y_n - E(Y_n)}{\sqrt{D(Y_n)}} = \frac{Y_n}{\sqrt{n}}$$

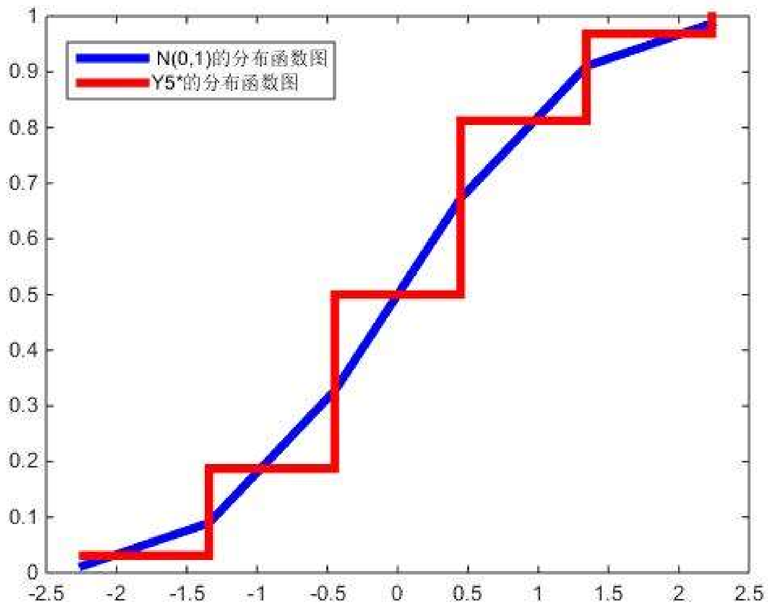
独立同分布中心极限定理

小球位置的标准化随机变量序列

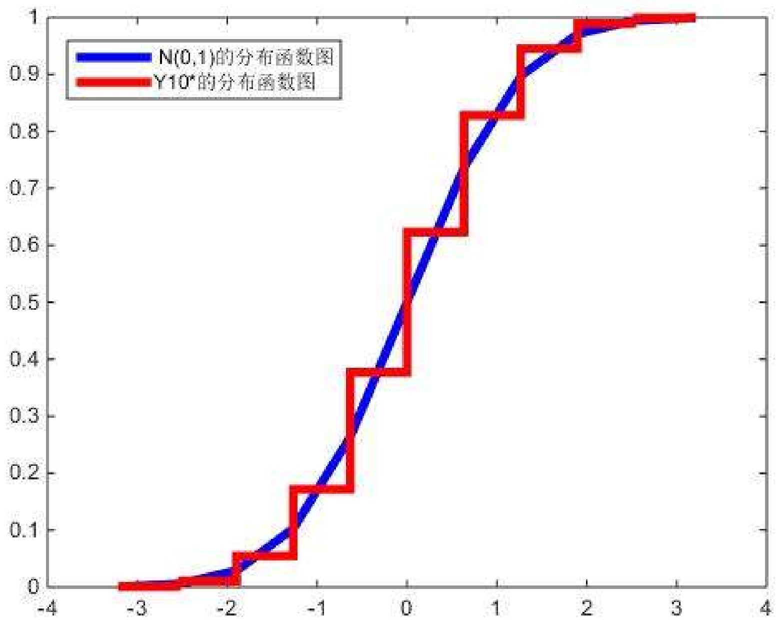
$$Y_n^* = \frac{Y_n - E(Y_n)}{\sqrt{D(Y_n)}} = \frac{Y_n}{\sqrt{n}}$$

基于钉板试验中X_k所满足条件：独立，同两点分布

计算出 Y_n^* 的分布函数， 与标准正态分布函数对比：

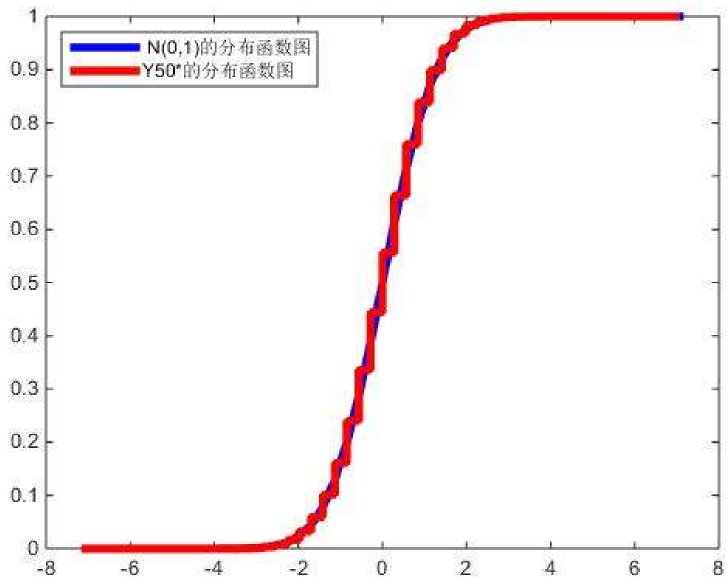


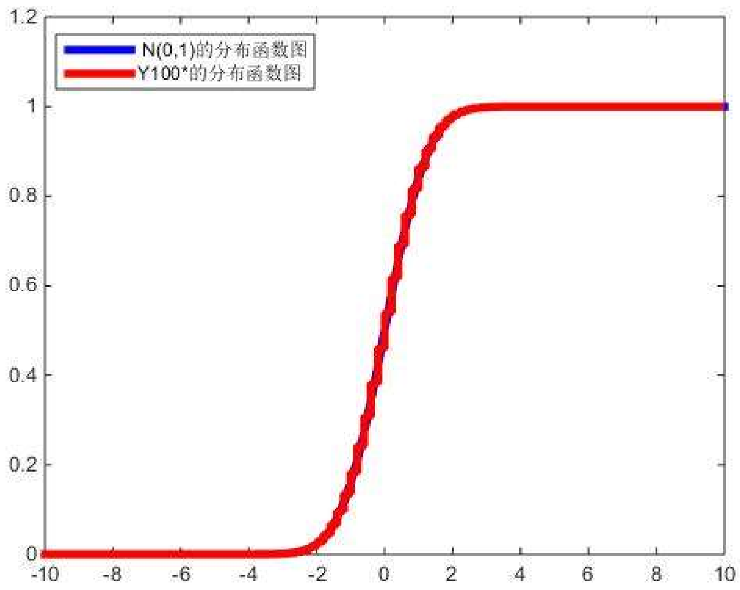
n=5



n=10

独立同分布中心极限定理





n=50　　　　n=100

可从理论上严格证明：

Y_n^* **依分布收敛于标准正态分布随机变量**

更一般的，成立：

三．独立同分布中心极限定理

设$\{X_k\}$, $k=1,2\ldots$为相互独立,具有相同分布的随机变量序列, 且 $E(X_k)=\mu$, $D(X_k)=\sigma^2\neq 0$,（$k=1,2,\ldots$）则有

$$\lim_{n\to\infty} P\left\{\frac{\sum_{k=1}^{n} X_k - E\left(\sum_{k=1}^{n} X_k\right)}{\sqrt{D\left(\sum_{k=1}^{n} X_k\right)}} \le x\right\} = \lim_{n\to\infty} P\left\{\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \le x\right\} = \Phi(x), \quad x\in R$$

称$\{X_k\}$服从中心极限定理

$\{X_k\}$服从中心极限定理的含义：

$\{X_k\}$前n项和的标准化随机变量序列依分布收敛于标准正态分布随机变量

思考: 如何理解定理中要求 $\{X_k\}$ 同分布,期望方差存在?

理解：多个“均匀地小”的独立随机变量的叠加，其分布近似正态分布

解释了实际中什么样的随机变量服从正态分布

四. 独立同分布中心极限定理的应用

1. 概率近似计算

当 n 足够大时，可认为

$$\frac{\sum_{k=1}^{n} X_k - E\left(\sum_{k=1}^{n} X_k\right)}{\sqrt{D\left(\sum_{k=1}^{n} X_k\right)}} \sim N(0,1)$$

近似成立

独立同分布中心极限定理

例1 某保险公司有10000个同阶层的人参加人寿保险，每人每年付600元保险费. 若一年内一个人死亡的概率为0.004，死亡时，其家属可向保险公司领得5万元赔偿金。试问：保险公司每年利润大于300万的概率是多少?

$$P\left\{\sum_{i=1}^{10000} X_i \le 300\right\} = ?$$

解：设 X_i 表示保险公司支付给第 i 户的赔偿金（单位：万元），则 X_i 均服从两点分布

独立同分布

X_i	0	5
p	0.996	0.004

$i=1,\ldots,10000$

且在一定实际条件下，可认为各 X_i 是相互独立的.

则保险公司的全部赔偿金为 $\sum_{i=1}^{10000} X_i$

独立同分布中心极限定理

$$E\left(\sum_{i=1}^{10000} X_i\right)=200,\quad D\left(\sum_{i=1}^{10000} X_i\right)=996,$$

保险公司每年利润大于300万的概率

$$P\left\{\sum_{i=1}^{10000} X_i \le 300\right\} = P\left\{\frac{\sum_{i=1}^{10000} X_i - 200}{\sqrt{996}} \le \frac{300-200}{\sqrt{996}}\right\}$$

近似服从$N(0,1)$（独立同分布中心极限定理）

$$\approx \Phi(3.167) \approx 0.9992$$

结论：保险公司每年利润大于300万元的概率接近100%

2. 正态分布随机数模拟

课后思考题

例2 在工程上，可先产生相互独立的服从区间[0,1]上均匀分布的随机数X_i (i=1,…,60)，再用 $\left(\sum_{i=1}^{60} X_i - 30\right)\Big/\sqrt{5}$ 来作为标准正态分布随机数，为什么？

matlab模拟程序

```
clear
X=rand(60,10000);
Y=(sum(X)-30)/sqrt(5);
hist(Y,40)
```