概率论与数理统计 第8次作业

Name: 宋昊原　　Student ID: 2022010755

November 12, 2023

1　概率的频率解释

这种说法几乎是正确的.
令I_i为A的特征变量，当第i次试验结果为A发生时值为1，否则值为0.
则$\frac{m}{n}$是以下随机变量的实测值：

$$\frac{\sum\limits_{i=1}^{n} I_i}{n} = \bar{I}$$

I_i服从均值为$P(A)$的两点分布，根据Kolmogrov强大数定律

$$P(\lim_{n\to\infty} \bar{I} = P(A)) = 1$$

这说明，对于几乎所有的试验情况，$\lim\limits_{n\to\infty} \frac{m}{n} = P(A)$是成立的.
这里"试验情况"指的是$n \to \infty$时的成功或失败组成的无穷序列，而"几乎所有"指不满足上述极限的序列的测度是0. 在一般意义下，概率的频率解释是成立的.

2　Chebyshev弱大数定律

根据Chebyshev不等式

$$P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mu_i| \geq \epsilon) \leq \frac{Var(\frac{1}{n}\sum\limits_{i=1}^{n} X_i)}{\epsilon^2}$$

$$= \frac{E((\sum\limits_{i=1}^{n} X_i)^2) - E^2(\sum\limits_{i=1}^{n} X_i)}{n^2\epsilon^2}$$

$$= \frac{\sum\limits_{i=1}^{n} E(X_i^2) + 2\sum\limits_{1\leq i<j\leq n} E(X_iX_j) - \sum\limits_{i=1}^{n} E^2(X_i) - 2\sum\limits_{1\leq i<j\leq n} E(X_i)E(X_j)}{n^2\epsilon^2}$$

由于

$$E(X_iX_j) - E(X_i)E(X_j) = Cov(X_i, X_j) = 0$$

上式右边

$$= \frac{\sum\limits_{i=1}^{n} \sigma_i^2}{n^2\epsilon^2} \to 0$$

故

$$\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mu_i| \geq \epsilon) = 0$$

3 中心极限定理证明Khinchin弱大数定律

两定理条件相同，根据中心极限定理，我们有

$$\lim_{n\to\infty} P(\frac{(\bar{X}-\mu)\sqrt{n}}{\sigma} \leq x) = \Phi(x)$$

故

$$\lim_{n\to\infty} P(|\bar{X}-\mu| \leq \epsilon) = 2\lim_{n\to\infty} \Phi(-\frac{\sqrt{n}\epsilon}{\sigma}) = 0$$

4 样本方差

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

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

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

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

由于

$$E(X_i^2) = \mu^2 + \sigma^2$$

由Khinchin大数定律，第一项依概率收敛到$\mu^2+\sigma^2$，第二项依概率收敛到μ^2，而第三项是期望有界的量的$\frac{1}{n-1}$，依概率收敛到0.
根据依概率收敛与和、差可交换，S^2依概率收敛到σ^2.

5 以概率1收敛的除法不变性

考虑事件$A=\{X_n$不收敛于$a\}$，$B=\{Y_n$不收敛于$b\}$，$C=\{\frac{X_n}{Y_n}$不收敛于$\frac{a}{b}\}$.
则

$$A^c B^c \subset C^c$$

故

$$(1-P(A))(1-P(B)) \leq 1-P(C)$$

由题设

$$P(A)=P(B)=0$$

于是

$$1-P(C) \geq 1$$

这说明

$$P(C)=0$$

6 5的推论

由Kolmogrov强大数定律，$\bar{X}=\frac{1}{n}(X_1+...+X_n)$以概率1收敛于2，$\bar{Y}=\frac{1}{n}(Y_1+...+Y_n)$以概率1收敛于5，由第5题，$\frac{\bar{X}}{\bar{Y}}=\frac{X_1+...+X_n}{Y_1+...+Y_n}$以概率1收敛于$\frac{2}{5}$.

7 二项分布的正态近似

$$X \sim B(40, \frac{1}{2})$$

$$P(X = 20) = 0.1254$$

$$X \to N(20, 10)$$

$$P(X = 20) \approx \Phi(\frac{0.5}{sqrt10}) - \Phi(\frac{-0.5}{sqrt10}) = 0.1256$$

8 保险公司

8.1

一年内发生事故的次数$X \sim B(10000, 0.001)$，故

$$E(X) = 10$$

保险公司净利润的期望为

$$E(20000 - 1000X) = 10000 > 0$$

合理.

8.2

$$20000 - 1000X \geq 4000 \Leftrightarrow X \leq 16$$

利用正态分布近似，$X \to N(10, 9.99)$，则

$$P(X \leq 16) \approx \Phi(\frac{6}{\sqrt{9.99}}) = 0.9712$$

8.3

这相当于解方程

$$\Phi(\frac{x - 10}{\sqrt{9.99}}) = 0.95$$

解得

$$x \approx 15.2$$

故至少有毛利润

$$20000 - 1000 \times 15 = 5000$$

9 随机误差

9.1

根据中心极限定理，设该随机变量的均值为X，则有

$$X \to N(0, \frac{1}{75})$$

$$P(|X| < 0.2) = \Phi(\sqrt{75} \times 0.2) - \Phi(-\sqrt{75} \times 0.2) = 0.9167$$

9.2

这相当于

$$\Phi(0.2\sqrt{3n})-\Phi(-0.2\sqrt{3n})>0.95$$

即

$$\Phi(0.2\sqrt{3n})>0.975$$

查表得

$$0.2\sqrt{3n}>1.96$$

即

$$n>32$$

9.3

根据Chebyshev不等式

$$1-P(|X|<\epsilon)\leq\frac{Var(X)}{\epsilon^2}$$

即

$$P(|X|<\epsilon)\geq 1-\frac{1}{3n\epsilon^2}$$

故

$$n>\frac{1}{3\alpha\epsilon^2}$$

代入得

$$n>166$$

可见用Chebyshev不等式进行放缩的误差非常大，会严重高估平均误差.

10 合格率

设样本合格率为X，则X为5000个$B(0.8)$变量的平均，根据中心极限定理，近似地有

$$X\sim N(0.8,3.2\times10^{-5})$$

于是ϵ满足

$$\Phi(\frac{\epsilon}{\sqrt{3.2\times10^{-5}}})=0.995$$

查表知

$$\frac{\epsilon}{\sqrt{3.2\times10^{-5}}}=2.58$$

即

$$\epsilon=0.0146$$

此时合格品数范围是3927∼4073个.

11 抽样调查实例

2023年11月11日，清新时报发表文章《调查丨我在清华不恋爱》，选择了学校143位同学进行恋爱现状及观念的调查，其中有60%的同学处于单身状态，可以计算，若要求精度为0.05，则，此结果的置信水平为

$$\alpha=2(1-\Phi(\frac{\sqrt{143}\times0.05}{\sqrt{0.6\times0.4}}))=0.22$$

若要求置信水平达到0.05以下，精度应不低于0.08，这说明可以以95%的置信度认为校内单身率在52%∼68%，但如果要认为单身率在55%∼65%，置信度就会降至78%，基本没有参考价值.

12 股票涨跌

12.1

设n天内有X_n天涨价，则

$$X_n \sim B(n, \frac{1}{2})$$

且

$$Y_n = 1.7^{X_n} \times 0.5^{n-X_n}$$

故

$$\ln Y_n = X_n \ln 1.7 + (n - X_n) \ln 0.5 = \ln \frac{1.7}{0.5} X_n + n \ln 0.5$$

根据中心极限定理，$\frac{X_n}{n}$近似服从$N(\frac{1}{2}, \frac{1}{4n})$，即$X_n$近似服从$N(\frac{n}{2}, \frac{n}{4})$.
于是，近似地有

$$\ln Y_n \sim N((\frac{\ln 3.4}{2} + \ln 0.5)n, (\ln 3.4)^2 \frac{n}{4})$$

经计算得

$$\ln Y_n \sim N(-0.08126n, 0.3744n)$$

12.2

$n \to \infty$时，$E(Y_n) \approx -0.08126n \to -\infty$.

12.3

不会了

12.4

$1.7^{0.5} \times 0.5^{0.5} = 0.922 < 1$，这意味着平均每天的股价增幅为$-7.8\%$.

13 经验分布函数

13.1

考虑随机变量$I_i(x)$，则$F_n(x) = \frac{1}{n} \sum\limits_{i=1}^{n} I_i(x) = \bar{I}(x)$，$I_i(x)$独立同分布. 则

$$E(F_n(x)) = E(I_i(X)) = F(x)$$

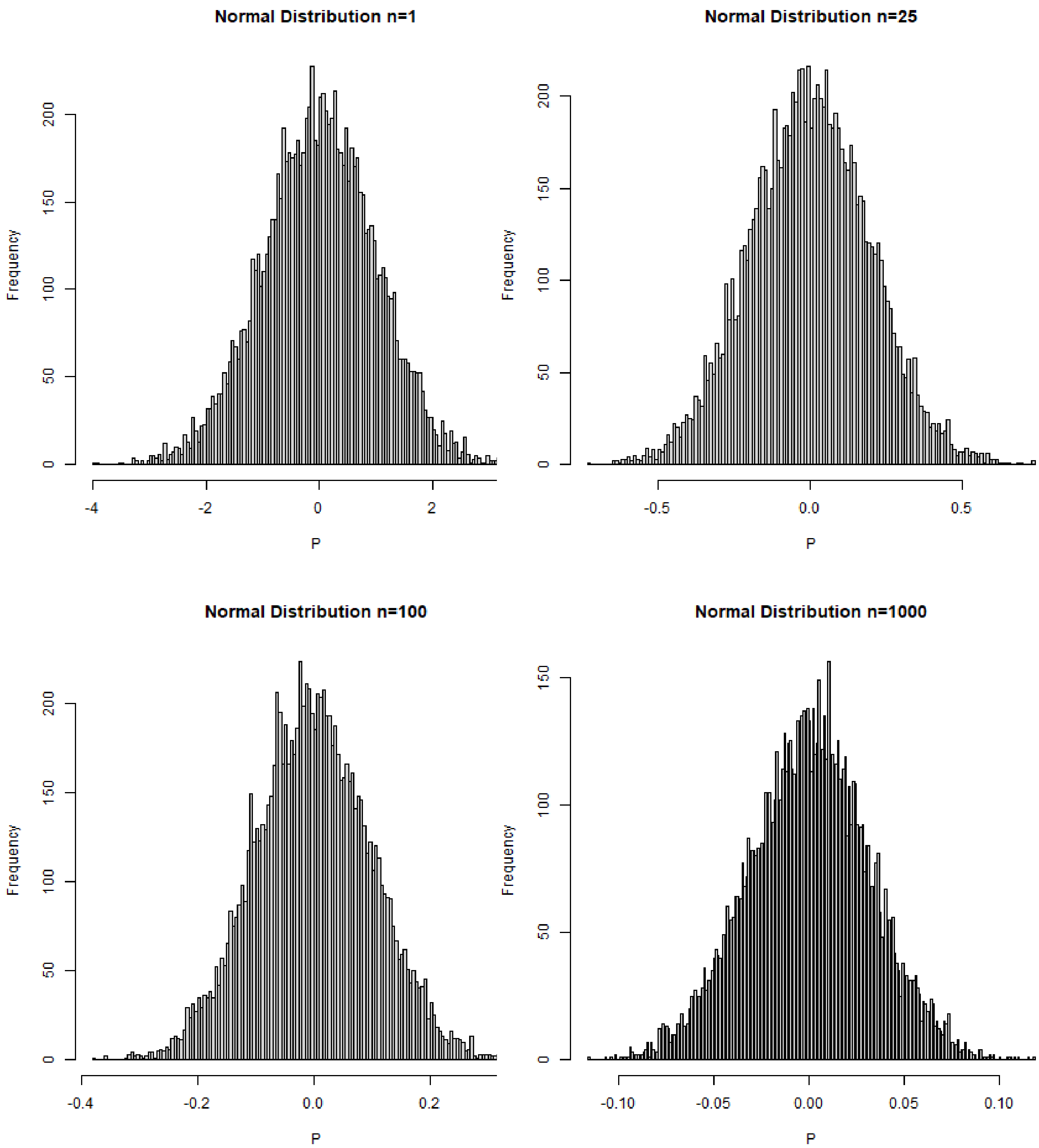
$$Var(F_n(x)) = \frac{Var(I_i(X))}{n} = \frac{F(x)(1 - F(x))}{n}$$

13.2

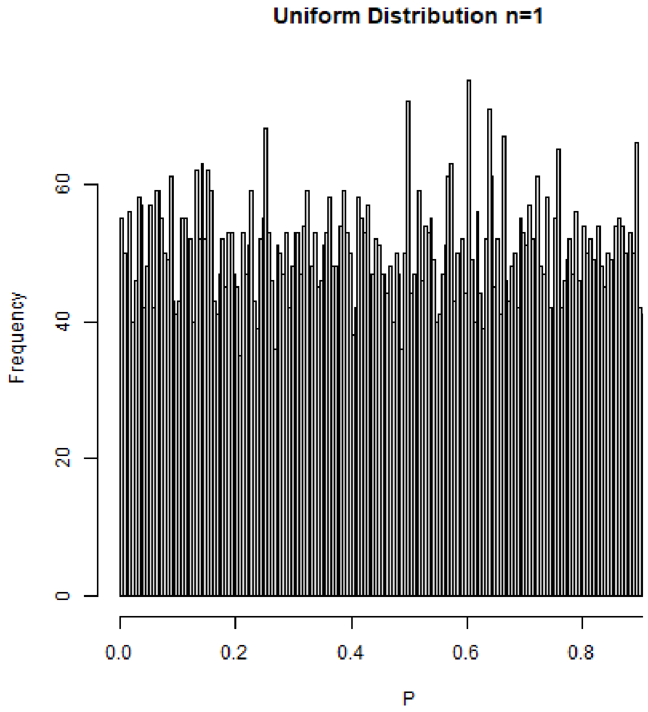
根据Kolmogrov强大数定律，对每个x都有$F_n(x)$以概率1收敛于$F(x)$.

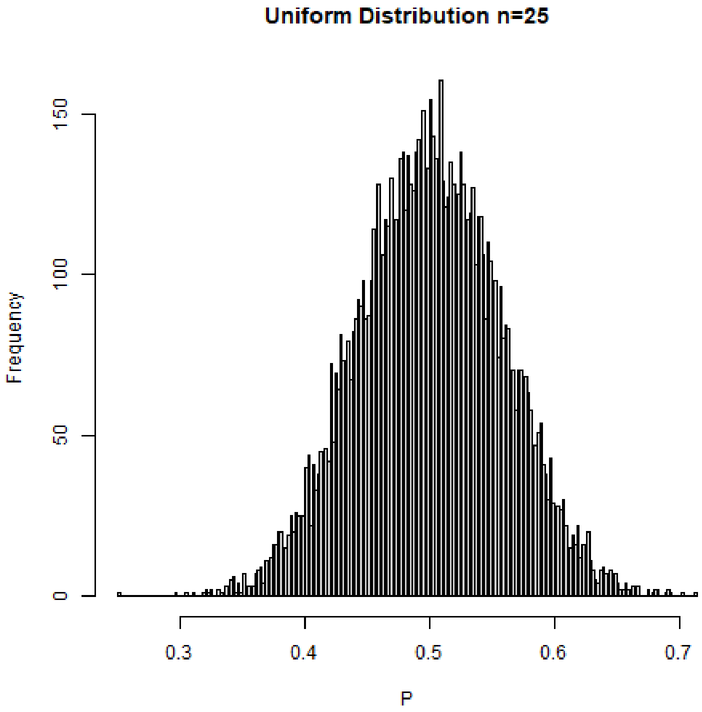
14 计算机试验：中心极限定理

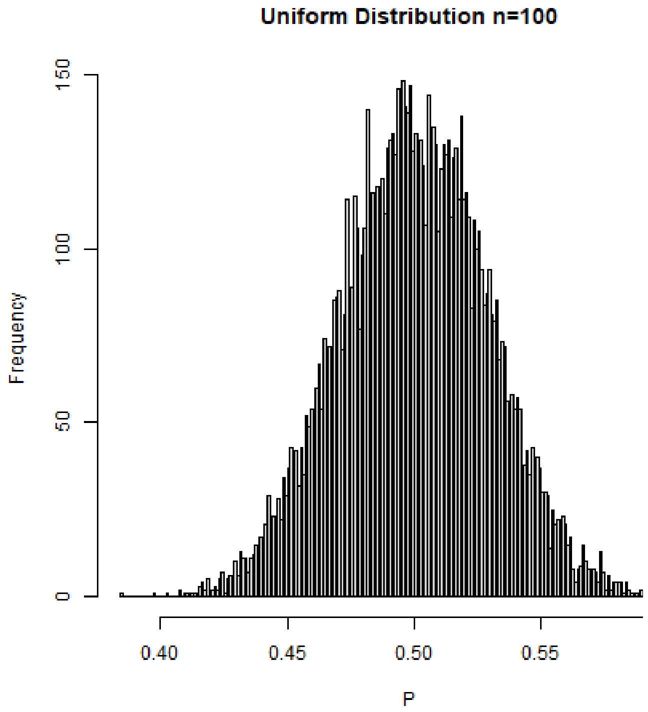
14.1 正态分布

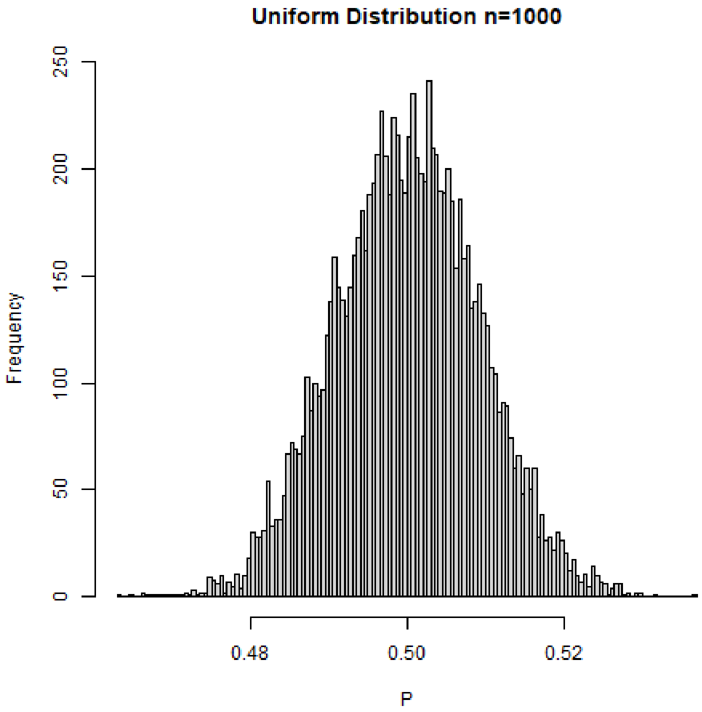


14.2 均匀分布

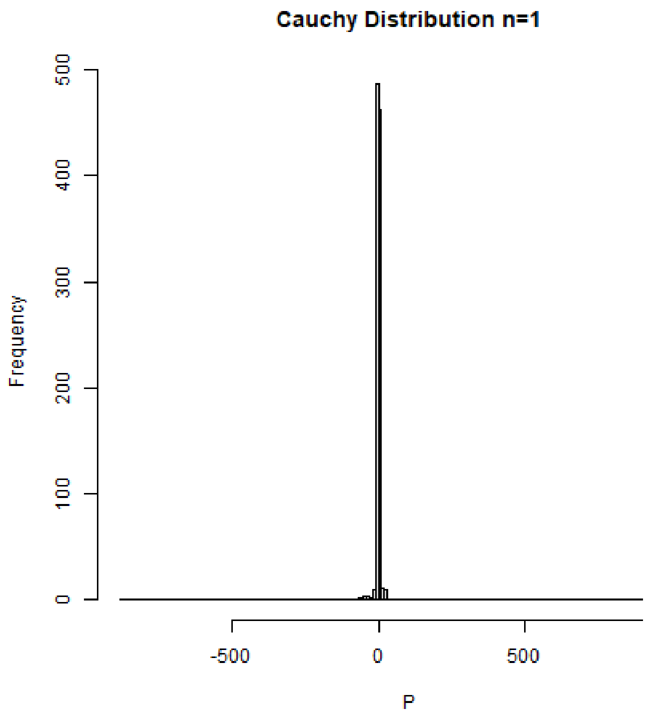

Uniform Distribution n=1
Frequency
0
20
40
60
0.0
0.2
0.4
0.6
0.8
P

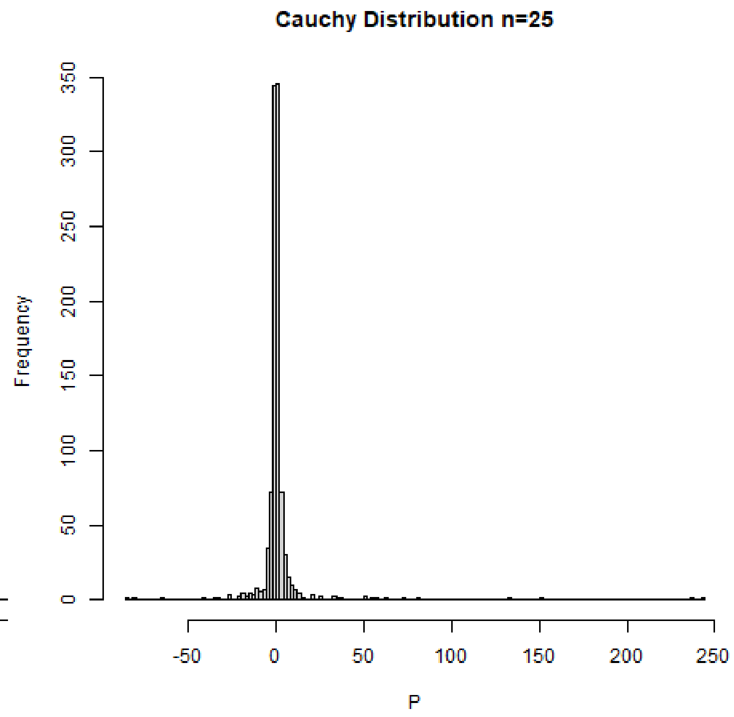

Uniform Distribution n=25
Frequency
0
50
100
150
0.3
0.4
0.5
0.6
0.7
P

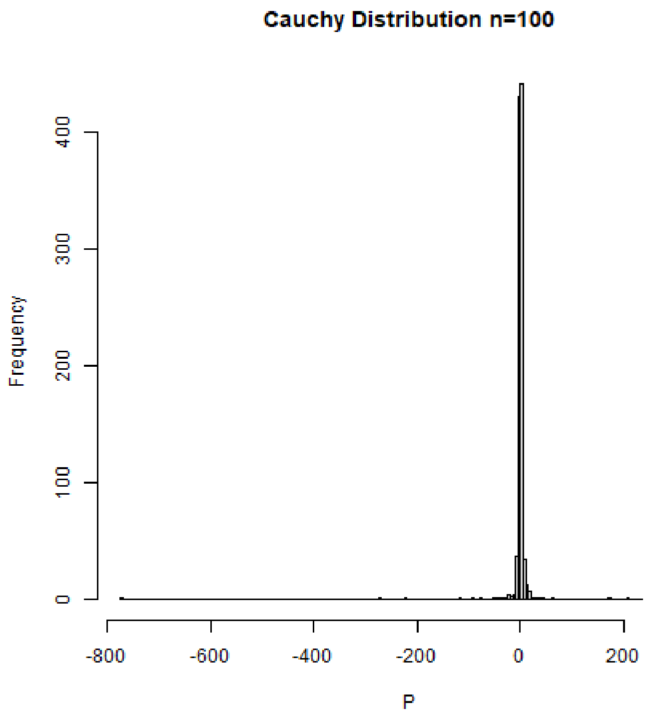

Uniform Distribution n=100
Frequency
0
50
100
150
0.40
0.45
0.50
0.55
P

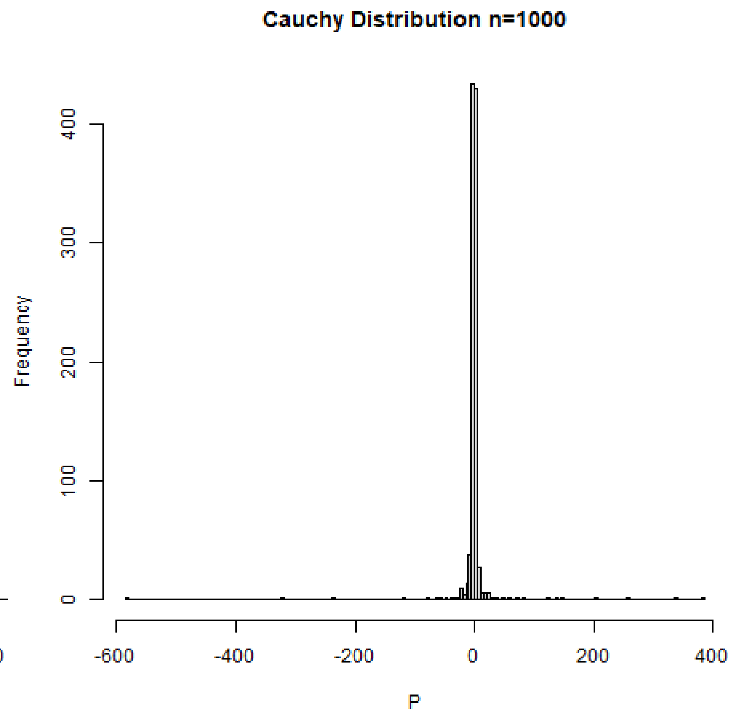

Uniform Distribution n=1000
Frequency
0
50
100
150
200
250
0.48
0.50
0.52
P

14.3 Cauchy分布









14.4 结论

正态分布和均匀分布符合中心极限定理，而Cauchy分布不符合，因为Cauchy分布的方差不存在.