第9周作业

1、

由$SLLN$,设事件A发生的概率是p(未知)，进行n次实验成功次数为x,$lim_{n\to+\infty}\overline{x}=lim_{n\to+\infty}\frac{x}{n}=p$,频率解释合理

2、

设$\overline{X}=\frac{1}{n}(X_1+\cdots+X_n)$
因为X_i两两不相关，$Var(\overline{X})=\frac{1}{n^2}\sum_i Var(X_i)=\sum_i\frac{\sigma_i^2}{n^2}\leq\sum_i\frac{c}{n^2}=\frac{c}{n}$
由$Chebyshev$,$P(|\overline{X}-E(\overline{X})|\geq\varepsilon)\leq\frac{\frac{c}{n}}{\varepsilon^2}=\frac{c}{n\varepsilon^2}\to 0,n\to 0$
所以$\lim_{n\to\infty}P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i-\frac{1}{n}\sum_{i=1}^n\mu_i\right|\geq\varepsilon\right)=0$

3、

对$X_1,X_2,\cdots,X_n(iid),E(X_i)=\mu,Var(X_i)=\sigma^2$,记$\overline{X}=\frac{X_1+\cdots+X_n}{n}$

$$\text{由}CLT,lim_{n\to\infty}P(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\leq x)=\Phi(x)$$
$$\lim_{n\to\infty}P(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}>x)=1-\Phi(x)$$
$$\lim_{n\to\infty}P(\left|\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right|>x)=2(1-\Phi(x))$$
$$\lim_{n\to\infty}P(\left|\overline{X}-\mu\right|>\frac{x\sigma}{\sqrt{n}})=2(1-\Phi(x))$$
$$\lim_{n\to\infty}P(\left|\overline{X}-\mu\right|>\varepsilon)=2(1-\Phi(\varepsilon\frac{\sqrt{n}}{\sigma}))$$
$$\text{对}\forall\varepsilon>0,lim_{n\to\infty}2(1-\Phi(\varepsilon\frac{\sqrt{n}}{\sigma}))=0$$

4、

$$E(S^2)=\sigma^2$$
$$\text{由}chebyshev,P(|S^2-E(S^2)|\geq\varepsilon)\leq\frac{Var(S^2)}{\varepsilon^2}$$
$$Var(S^2)=\frac{1}{n-1}Var(\sum_i(X_i-\mu)^2-n(\overline{X}-\mu)^2)$$

5、

$A=\{X_n$不收敛于$a\}, B=\{Y_n$不收敛于$b\}, C=\{\frac{X_n}{Y_n}$不收敛于$\frac{a}{b}\}$
$C\subset(A+B), A, B$的测度为0，C的测度也为0
所以$\frac{X_n}{Y_n}=\frac{a}{b}\ a.s.$

6、

$$\frac{X_1+\cdots+X_n}{Y_1+\cdots+Y_n}=\frac{(X_1+\cdots+X_n)/n}{(Y_1+\cdots+Y_n)/n}$$
$$\text{由}SLLN,=\frac{\mu_X}{\mu_Y}=\frac{2}{5}$$

7、

$$X\sim B(40,0.5)$$

$$\begin{aligned}P(X=20)&=C_{40}^{20}p^{20}(1-p)^{20}\\&\approx0.125\end{aligned}$$
$$y_1=\frac{20-40\times0.5-0.5}{\sqrt{40\times0.5\times0.5}}=-0.156$$
$$y_2=\frac{20-40\times0.5+0.5}{\sqrt{40\times0.5\times0.5}}=0.156$$
$$\text{令}\Phi\text{为正态分布的}cdf,\Phi(y_2)-\Phi(y_1)=0.124$$

8、

(1)

$$\text{设}X\text{为一个买保险人的收益，}E(X)=-0.001\times1000+2=1$$

(2)

$$10000\text{人的保费为}10000\times2=20000$$
$$20\%\text{毛利润下赔付}20000\times(1-20\%)=16000,\text{可理赔}16\text{人}$$
$$\text{设}X\text{为出事故的人数，}X\sim B(10000,0.001)$$
$$P(X\le16)\approx0.973$$

(3)

$$\text{设}f(X)\text{为}X\sim B(10000,0.001)\text{的}cdf$$
$$P(X\le m)\ge0.95,m\text{的最大值为}15$$

$$以95\%的概率可以保证还有20000-15\times 1000=5000利润$$

9、

(1)

$$
\begin{gathered}
P(\left|\frac{X_1+X_2+\cdots+X_n}{n}-0\right|\leq\varepsilon)\\
X_i\sim U(-1,1),Var(X_i)=\frac{(1-(-1))^2}{12}=\frac{1}{3}\\
\sigma^2=\frac{1}{3},\sigma=\frac{1}{\sqrt{3}}\\
P(\frac{\overline{X}}{\sigma/\sqrt{n}}<x)=\Phi(x)\\
P(\overline{X}<\frac{\sigma x}{\sqrt{n}})=\Phi(x)\\
令\varepsilon=\frac{\sigma x}{\sqrt{n}},P(\overline{X}<\varepsilon)=\Phi(\frac{\sqrt{n}\varepsilon}{\sigma})=\Phi(\sqrt{3})=0.958\\
P(|\overline{X}|<\varepsilon)=1-2\times(1-0.958)=0.916
\end{gathered}
$$

(2)

$$
\begin{gathered}
P(\overline{X}<0.2)=\Phi(\sqrt{3n}\times 0.2)\\
P(|\overline{X}|<0.2)=1-2(1-\Phi(0.2\sqrt{3n}))>0.95\\
\Phi(0.2\sqrt{3n})>0.975\\
0.2\sqrt{3n}>1.96\\
n>32.01,n\geq 33
\end{gathered}
$$

(3)

$$
\begin{gathered}
E(\overline{X})=0\\
Var(\overline{X})=\frac{\sigma^2}{n}=\frac{1}{3n}\\
P(|\overline{X}|\geq\varepsilon)\leq\frac{Var(\overline{X})}{\varepsilon^2}\\
P(|\overline{X}|\geq 0.2)\leq\frac{1/3n}{0.04}=0.05\\
n=3.75,n\geq 4\\
Chebyshev相对CLT误差更大
\end{gathered}
$$

10、

$$
设X_i=\begin{cases}0,不合格品\\1,合格品\end{cases}
$$

$$\overline{X}=\frac{X_1+\cdots+X_n}{n},n=5000$$
$$X_1+\cdots+X_n\sim B(n,0.8),\mu=0.8,\sigma^2=p(1-p)=0.16$$
$$\text{由}CLT,Y=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}=\frac{\overline{X}-0.8}{\sqrt{0.16}/\sqrt{5000}}=176.8(\overline{X}-0.8)$$
$$Y\sim N(0,1)$$
$$P(Y\leq x)=\Phi(x)$$
$$P(\overline{X}-0.8\leq\frac{x}{176.8})=\Phi(x)$$
$$\text{令}\varepsilon=\frac{x}{176.8},P(\overline{X}-0.8\leq\varepsilon)=\Phi(176.8\varepsilon)\geq 1-\frac{\alpha}{2}=0.995$$
$$176.8\varepsilon>2.576$$
$$\varepsilon>1.46\%$$
$$\text{合格品数量}=5000\times[80\%-1.46\%,80\%+1.46\%]=[3927,4073]$$

11、

“2015年度西南大学财经学院40000户抽样调查显示，中国基尼系数达到了0.62”
由于CLT,相对理论真值的误差仅近取决于测量次数，而与总数无关，因此在全国14亿中尽管只采样40000户，依然能获得极高的精度

12、

(1)

$$\text{设}X_i=\begin{cases}0,\text{下降}50\%\\1,\text{上涨}70\%\end{cases}$$
$$\text{设}\overline{X}=\frac{X_1+\cdots+X_n}{n}$$
$$X=X_1+\cdots+X_n=n\overline{X}$$
$$Y_n=1.7^X0.5^{n-X}$$
$$\begin{aligned}log(Y_n)&=Xlog3.4+nlog0.5\\&=n\overline{X}log3.4-nlog2\end{aligned}$$
$$E(\overline{X})=0.5,Var(\bar{(X)})=0.5(1-0.5)=0.25$$
$$\text{由}CLT,\frac{\overline{X}-0.5}{0.5/\sqrt{n}}\sim N(0,1)$$
$$\overline{X}\sim N(0.5,\frac{0.25}{n})$$
$$log(Y_n)\sim N(0.5nlog3.4-nlog2,0.25nlog^23.4)$$
$$log(Y_n)\sim N(-0.0813n,0.3744n)$$

(2)

$$
\text{当} n \to \infty, E(log(Y_n)) \to -\infty \\
E(Y_n) = e^{\mu+\sigma^2/2} = e^{0.1059n} \to +\infty
$$

(3)

$$
\begin{aligned}
\forall \varepsilon > 0, Y_n \le \varepsilon &\iff log(Y_n) \le log(\varepsilon) \\
&\iff \frac{log(Y_n) + 0.0813n}{\sqrt{0.3744n}} \le \frac{log(\varepsilon) + 0.0813n}{\sqrt{0.3744n}} \\
&\iff Z \le \frac{log(\varepsilon)}{\sqrt{0.3744n}} + 0.1329\sqrt{n}, Z \sim N(0,1)
\end{aligned}
$$

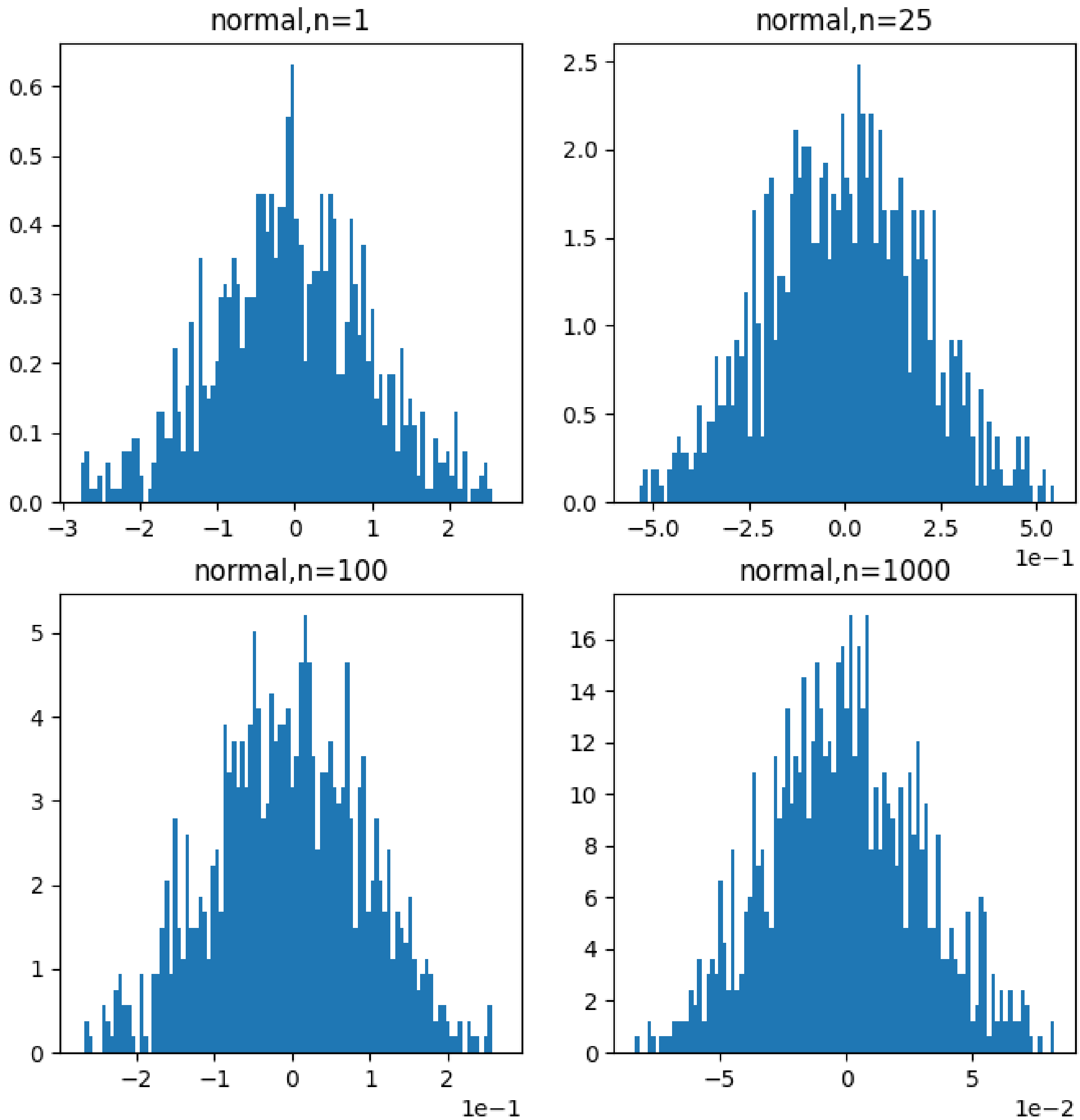
$$
P(Z \le \frac{log(\varepsilon)}{\sqrt{0.3744n}} + 0.1329\sqrt{n}) = \Phi(\frac{log(\varepsilon)}{\sqrt{0.3744n}} + 0.1329\sqrt{n}) \\
\text{当} n \to \infty, P(Y_n \le 0) = \Phi(+\infty) = 1
$$

(4)

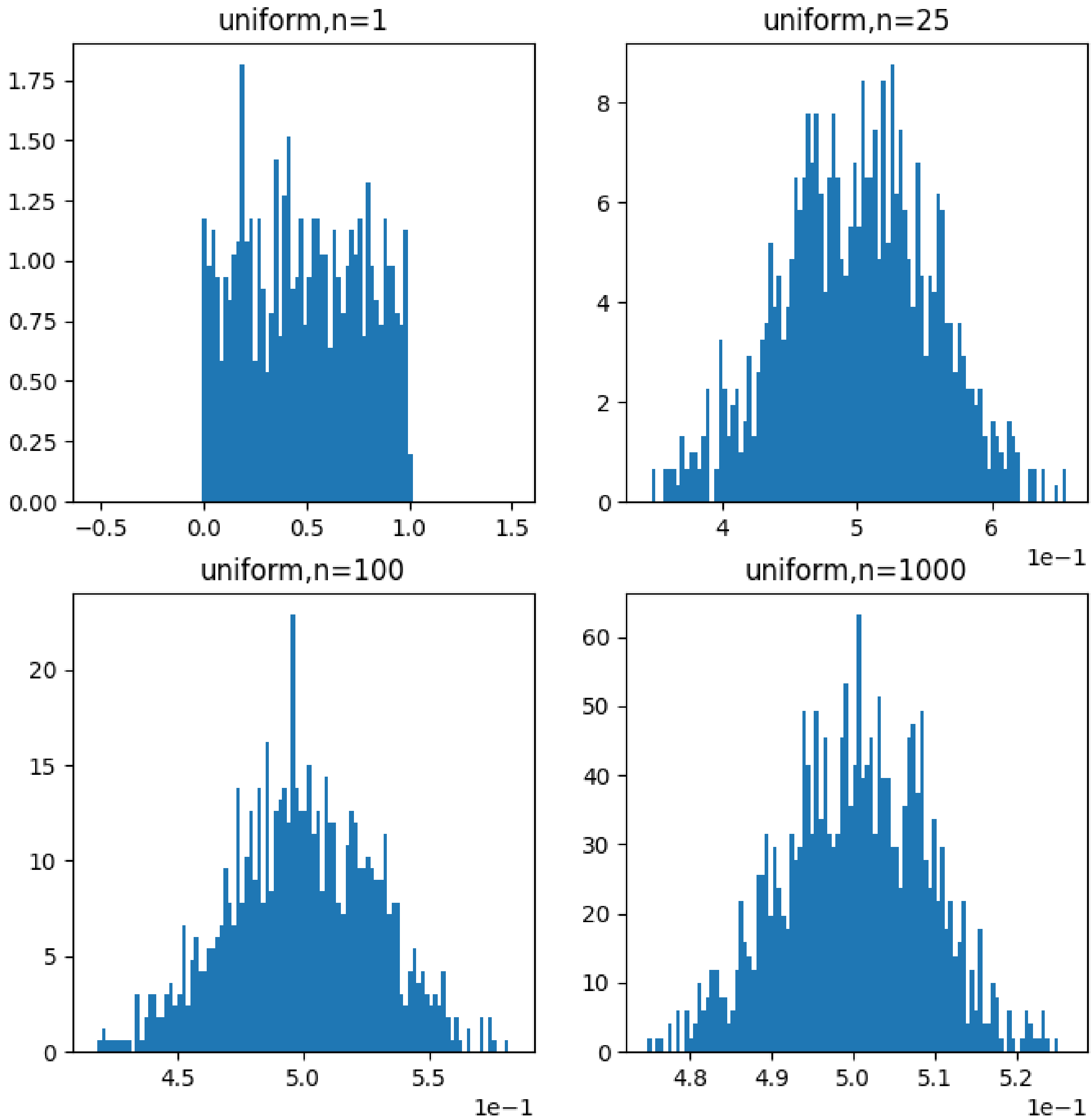
$$
\text{尽管} P\left(\lim_{n\to\infty} Y_n = 0\right) = 1\text{，但对于} Y_n \ne 0 \text{的部分由于} Y_n \text{过大，导致期望不为} 0
$$

14、

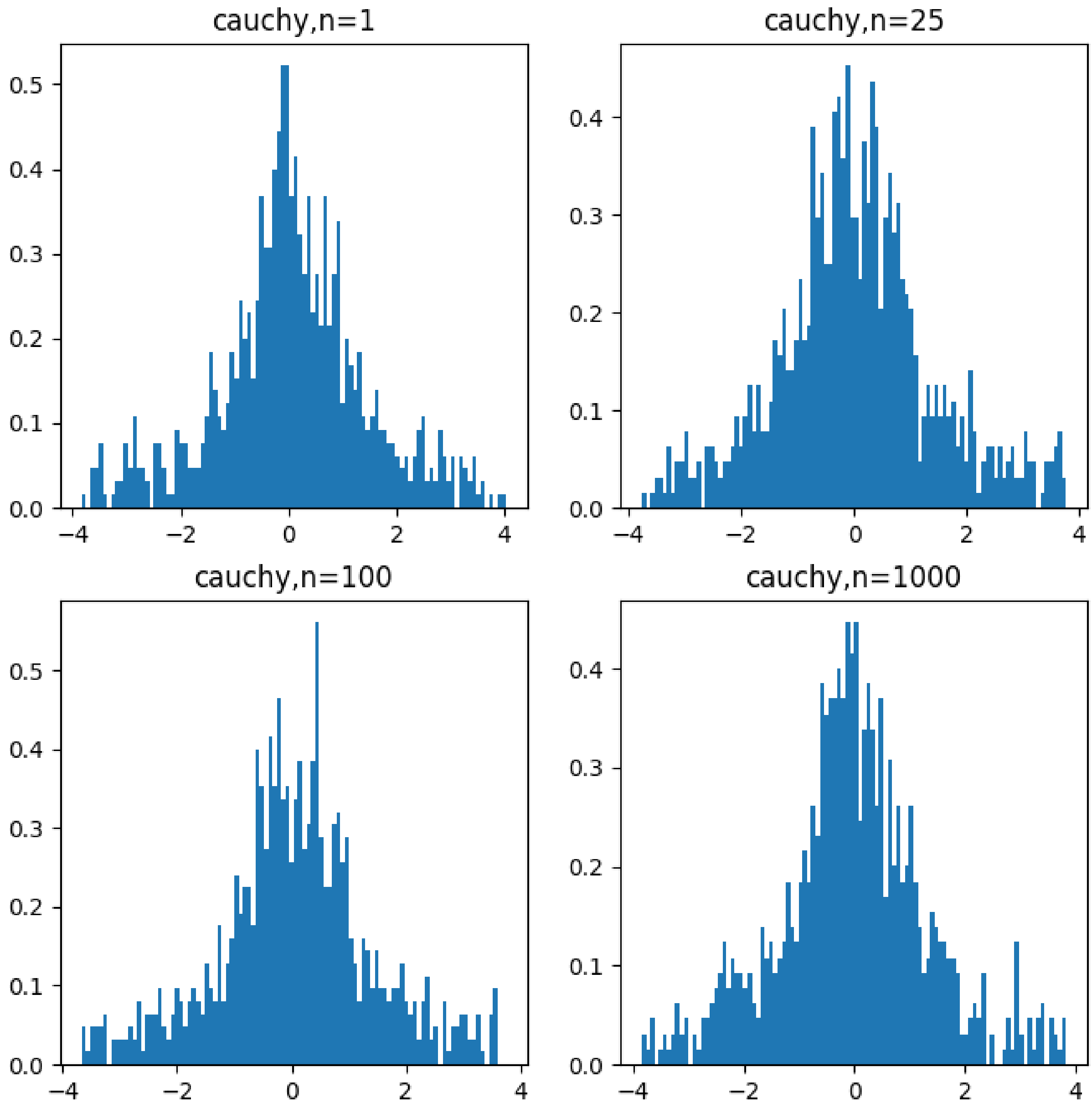
(1)



(2)


uniform,n=1
uniform,n=25
uniform,n=100
uniform,n=1000
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
−0.5
0.0
0.5
1.0
1.5
8
6
4
2
0
4
5
6
1e−1
20
15
10
5
0
4.5
5.0
5.5
1e−1
60
50
40
30
20
10
0
4.8
4.9
5.0
5.1
5.2
1e−1

(3)



(4)

相符，随着n的增大，$\overline{X}$愈发集中

```python
import numpy as np
from scipy.stats import cauchy
import matplotlib.pyplot as plt

def Gen(type='normal', N=1000):
    X1 = []
    X25 = []
    X100 = []
    X1000 = []
    for i in range(N):
        if type=='normal':
            X = np.random.normal(0, 1, N)
        elif type=='uniform':
            X = np.random.uniform(0, 1, N)
        elif type=='cauchy':
            X = cauchy.rvs(loc=0, scale=1, size=N)
        else:
            raise TypeError('Invalid distribution type')
        X1.append(X[0])
        X25.append(np.mean(X[0:25]))
        X100.append(np.mean(X[0:100]))
        X1000.append(np.mean(X[0:1000]))

    X1 = np.array(X1)
    X25 = np.array(X25)
    X100 = np.array(X100)
    X1000 = np.array(X1000)

    q1 = [0 for _ in range(4)]
    q3 = [0 for _ in range(4)]
    iqr = [0 for _ in range(4)]

    q1[0], q3[0] = np.percentile(X1, [25, 75])
    iqr[0] = q3[0] - q1[0]

    q1[1], q3[1] = np.percentile(X25, [25, 75])
    iqr[1] = q3[1] - q1[1]

    q1[2], q3[2] = np.percentile(X100, [25, 75])
    iqr[2] = q3[2] - q1[2]

    q1[3], q3[3] = np.percentile(X1000, [25, 75])
    iqr[3] = q3[3] - q1[3]

    q1 = np.array(q1)
    q3 = np.array(q3)
    iqr = np.array(iqr)

    lower_bound = q1 - 1.5 * iqr
    upper_bound = q3 + 1.5 * iqr
```

```
    fig, axs = plt.subplots(2, 2, figsize=(8, 8))
    axes = axs.flatten()

    axs[0, 0].hist(X1, bins=100, density=True, range=(lower_bound[0], upper_bound[0]))
    axs[0, 0].set_title(f'{type},n=1')

    axs[0, 1].hist(X25, bins=100, density=True, range=(lower_bound[1], upper_bound[1]))
    axs[0, 1].set_title(f'{type},n=25')

    axs[1, 0].hist(X100, bins=100, density=True, range=(lower_bound[2], upper_bound[2])
    axs[1, 0].set_title(f'{type},n=100')

    axs[1, 1].hist(X1000, bins=100, density=True, range=(lower_bound[3], upper_bound[3]
    axs[1, 1].set_title(f'{type},n=1000')

    for i in range(4):
        axes[i].ticklabel_format(style='sci', axis='x', scilimits=(0, 0))

    plt.savefig(f'{type}.png')
    plt.clf()

Gen('normal')
Gen('uniform')
Gen('cauchy')
```