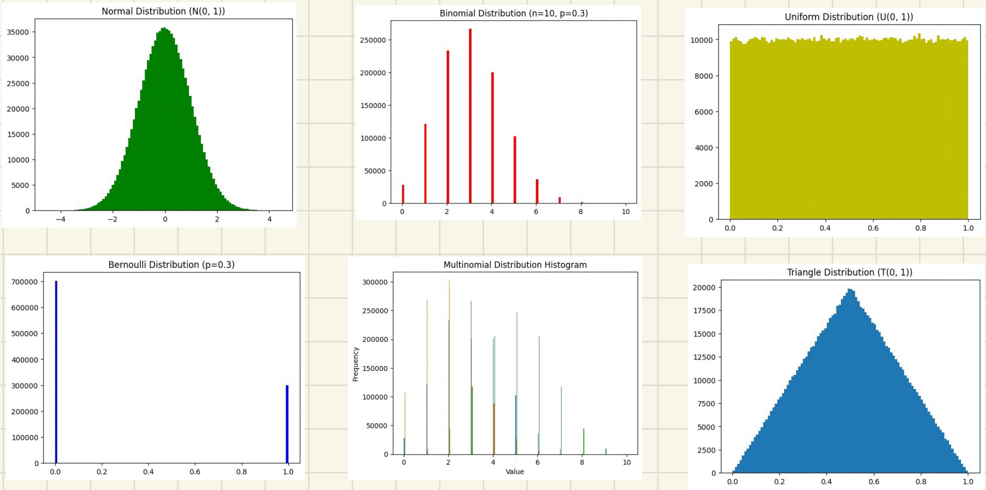
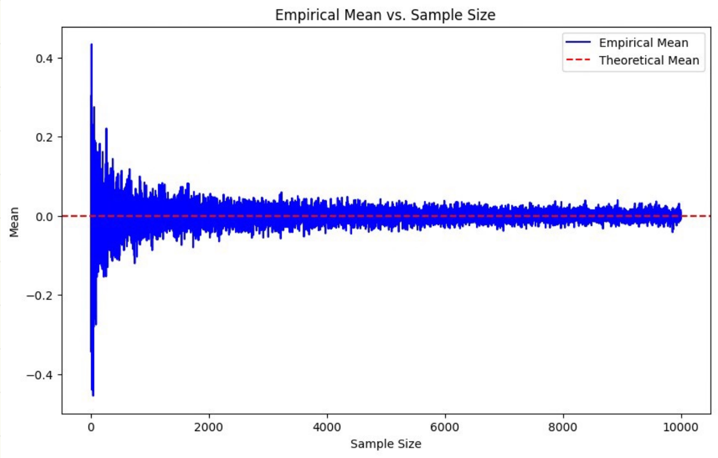
1)

```python
def sample_normal(sample_size=10, mu=0, std=1):
    # TODO#1.2:
    samples = norm.rvs(loc=mu, scale=std, size=sample_size)
    return samples


def sample_bernoulli(sample_size=10, p=0.5):
    # TODO#1.2:
    samples = bernoulli.rvs(p, size=sample_size)
    return samples


def sample_binomial(sample_size=10, n=10, p=0.5):
    # TODO#1.3:
    samples = binom.rvs(n, p, size=sample_size)
    return samples


def sample_multinomial(sample_size=10, n=100, p=[0.3, 0.2, 0.5]):
    # TODO#1.4:
    samples = multinomial.rvs(n, p, size=sample_size)
    return samples


def sample_uniform(sample_size=10, from_x=0, to_x=1):
    # TODO#1.5:
    samples = uniform.rvs(loc=from_x, scale=to_x - from_x, size=sample_size)
    return samples


def sample_triangle(sample_size=10, a=0, b=1):
    # TODO#1.6:
    samples = triang.rvs(c=0.5, loc=a, scale=b - a, size=sample_size)
    return samples
```



2)

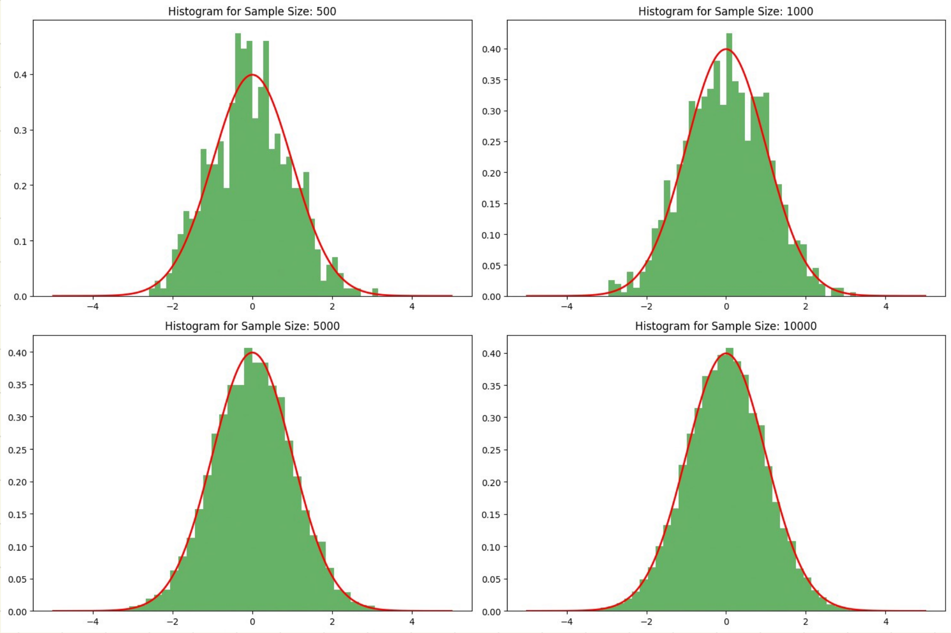
```python
sample_sizes = np.arange(1, 10001)  # Sample sizes from 1 to 10000
empirical_means = []  # List to store empirical means

# Calculate empirical means for different sample sizes
for size in sample_sizes:
    samples = sample_normal(sample_size=size)
    empirical_mean = np.mean(samples)
    empirical_means.append(empirical_mean)

# Theoretical mean of the normal distribution (N(0, 1)) is 0
theoretical_mean = 0

# Plotting the graph
plt.figure(figsize=(10, 6))
plt.plot(sample_sizes, empirical_means, color='b', label='Empirical Mean')
plt.axhline(y=theoretical_mean, color='r', linestyle='--', label='Theoretical Mean')
plt.xlabel('Sample Size')
plt.ylabel('Mean')
plt.title('Empirical Mean vs. Sample Size')
plt.legend()
plt.show()
```



- จากกราฟจะเห็นได้ว่า Empirical mean กับ theoritical mean ของ independent samples จำนวนมากๆ นั้นมีค่าใกล้เคียงกัน

3)

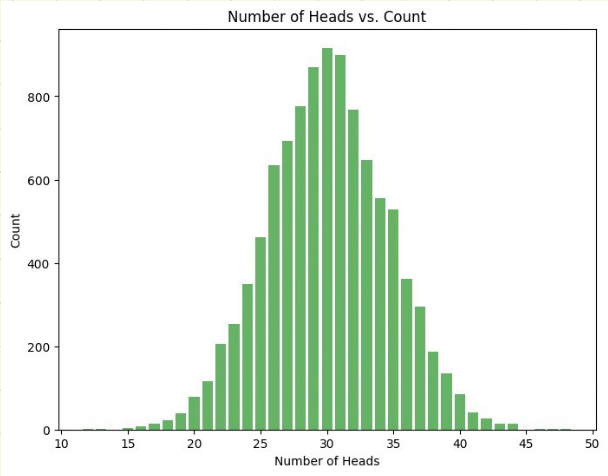
```python
sample_sizes = [500, 1000, 5000, 10000]
num_bins = 40

plt.figure(figsize=(15, 10))

for idx, size in enumerate(sample_sizes, 1):
    # Generate samples
    samples = sample_normal(sample_size=size)

    # Plot histogram
    plt.subplot(2, 2, idx)
    plt.hist(samples, bins=num_bins, density=True, alpha=0.6, color='g')
    plt.title(f'Histogram for Sample Size: {size}')

    # Plot true PDF (normal distribution)
    x = np.linspace(-5, 5, 1000)
    pdf = norm.pdf(x, loc=0, scale=1)
    plt.plot(x, pdf, 'r', linewidth=2)

plt.tight_layout()
plt.show()
```



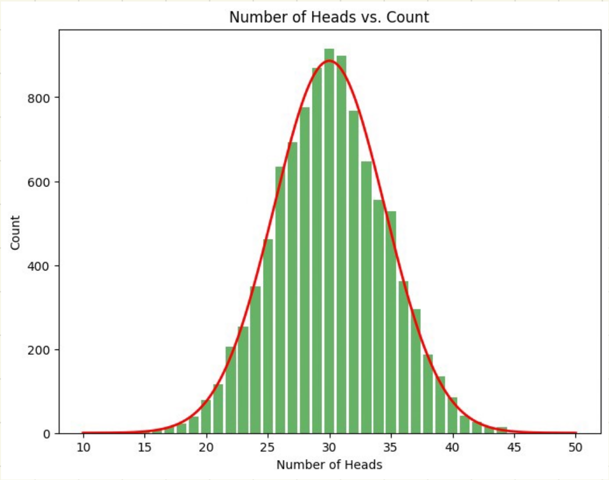
- จากทั้ง 4 กราฟ นำมาเปรียบเทียบกับ pdf จะเห็นได้ว่าทั้ง 4 กราฟมีลักษณะคล้ายกับ pdf
โดยยิ่ง sample size มาก จะยิ่ง fit กับ pdf

4)

```python
np.random.seed(0)  # Set seed for reproducibility
num_simulations = 10000
num_tosses = 100
probability_heads = 0.3

# Simulate coin tosses
tosses = np.random.choice([0, 1], size=(num_simulations, num_tosses), p=[1 - probability_heads, probability_heads])
num_heads = np.sum(tosses, axis=1)

# Count the occurrences of each number of heads
unique_heads, counts = np.unique(num_heads, return_counts=True)

# Plot the relationship between number of heads and count
plt.figure(figsize=(8, 6))
plt.bar(unique_heads, counts, color='g', alpha=0.6)
plt.title('Number of Heads vs. Count')
plt.xlabel('Number of Heads')
plt.ylabel('Count')
```



จากกราฟจะเห็นได้ว่ามีลักษณะใกล้เคียงกับ normal distribution



5) Probability using CLT

$$z = \frac{(x-\mu)}{\sigma} \sim N(\mu, \sigma^2)$$

$\mu = 100 \cdot 0.3 \qquad \sigma^2 = 100 \cdot 0.3 \cdot 0.7 \qquad \sim N(30, 21)$

$\mu = 30 \qquad \sigma^2 = 21$

เปิดตาราง

$$\rightarrow P(x > 40) = P\left(z > \frac{40-30}{\sqrt{21}}\right) = P(z > 2.182) = 0.0143$$

6) Probability using binomial distribution is

$$P(x > 40) = 1 - P(x \le 40)$$

$$= 1 - \sum_{i=0}^{k=40} \binom{100}{i} \cdot 0.3^{i} \cdot 0.7^{100-i}$$

$$= 0.0125$$

use binom.cdf to calculate →

```python
from scipy.stats import binom

actual_probability = 1 - binom.cdf(40, num_tosses, probability_heads)

print(f"Actual probability of getting more than 40 heads: {actual_probability:.4f}")
```

```
Actual probability of getting more than 40 heads: 0.0125
```

and the difference between CLT's and binomial's $= 0.0143 - 0.0125$

$= 0.0018$ → very small different

10.1)

$$P(\text{Fail}) = \int P(\text{Fail} \mid t)\, P(t)\, dt$$

$$= \int_{\mu-1}^{\mu+1} \left(\frac{0.97}{2250}(t-15)^2 + 0.001\right) 0.5\, dt$$

$$= \frac{1}{2}\left(\left[\frac{0.97}{2250}\frac{(t-15)^3}{3} + 0.001t\right]_{\mu-1}^{\mu+1}\right)$$

$$= \frac{1}{2}\left(\frac{0.97}{2250}\frac{(\mu-14)^3}{3} + \cancel{0.001(\mu+1)} - \left(\frac{0.97}{2250}\frac{(\mu-16)^3}{3} + \cancel{0.001(\mu-1)}\right)\right)$$

$$= \frac{1}{2}\left(\frac{0.97}{6750}\left((\mu-14)^3 - (\mu-16)^3\right) + 0.002\right)$$

$$= \frac{1}{2}\left(\frac{0.97}{6750}\left(6\mu^2 - 180\mu + 1352\right) + 0.002\right)$$

$$= 0.00043(\mu^2 - 30\mu + 227.653) = 0$$ ← doesn't want any failure

พิจารณา $b^2 - 4ac = 30^2 - 4 \cdot 1 \cdot 227.653 = -10.612$ $\therefore P(\text{Fail}) = 0$ ไม่มีคำตอบเป็น จน.จริง

find minimum $P(\text{fail})$ → $\dfrac{dP(\text{fail})}{d\mu} = 0$

$$\frac{1}{2} \cdot \frac{0.97}{6750}\left(3(\mu-14)^2 - 3(\mu-16)^2\right) = 0$$

$$(\mu-14)^2 - (\mu-16)^2 = 0$$

$$(\mu-14)^2 = (\mu-16)^2$$

$$\mu = 15_{\#}$$

10.2)

$$P(\text{Fail}) = \frac{1}{2}\left(\frac{0.97}{6750}\left((15-14)^3 - (15-16)^3\right) + 0.002\right) \approx 0.001144$$

↑

for each disk

10.3) From 10.2), P(Fail) for each disk = 1.144×10^{-3}

increase disk_num until it satisfies the condition

The code can be provided as follows:

```
from scipy.stats import binom

sample_size = 1000000
final_answer = 0
disk_num = 0

while True:
    prob = (1.144 * (10 ** (-3))) ** disk_num
    binomm = binom.rvs(n=10000, p=prob, size=sample_size)
    answer = 0

    for e in binomm:
        if e > 1:
            answer += 1

    prob_answer = answer / len(binomm)

    if prob_answer < 0.01 / 100:
        final_answer = disk_num
        break

    disk_num += 1

print("Final Answer:", final_answer)
```

Ans 2 disks

11)

11.1)

Σ	a	b	c	d
a	10×10^{-3}	0	4×10^{-3}	5×10^{-3}
b	0	3×10^{-3}	0	0
c	4×10^{-3}	0	12×10^{-3}	2×10^{-3}
d	5×10^{-3}	0	2×10^{-3}	15×10^{-3}

From the table, the pairs of coin that independent are (a,b), (b,c), (b,d)

11.2)

```
Expected Return for Coin a at T = 30: 0.967569883068716
Expected Return for Coin b at T = 30: 0.6008413072308074
Expected Return for Coin c at T = 30: 1.3102683838584235
Expected Return for Coin d at T = 30: 1.1617379044936387
Expected Return for Coin a at T = 180: 6.962440467911151
Expected Return for Coin b at T = 180: 4.306062121049301
Expected Return for Coin c at T = 180: 10.710457851882566
Expected Return for Coin d at T = 180: 10.390751434059398
```

11.3)

Σ	a	b	c	d
a	10×10^{-3}	0	4×10^{-3}	5×10^{-3}
b	0	3×10^{-3}	0	0
c	4×10^{-3}	0	12×10^{-3}	2×10^{-3}
d	5×10^{-3}	0	2×10^{-3}	15×10^{-3}

$\mathrm{Var}[x] = E[(x - E[x])^2] = \sigma^2$

* variance of each coin

$\mathrm{Var}[a] = 10 \times 10^{-3}$

$\mathrm{Var}[b] = 3 \times 10^{-3}$ ← lowest variance

Ans coin b

$\mathrm{Var}[c] = 12 \times 10^{-3}$

$\mathrm{Var}[d] = 15 \times 10^{-3}$

11.4) Because they can only lose their entire investment if they loss
but if they gain, they have probability to get more tha 100% of thier invesment.
This make the expected value >0

11.5)

Strategy	Expected Return (T=30)	Expected Return (T=180)	Variance (T=30)	Variance (T=180)	Probability of Profit (T=30)	Probability of Profit (T=180)
1	0.9199	7.0140	42.0449	1239.0167	0.4569	0.3983
2	0.6540	4.5906	10.7169	155.9407	0.5265	0.5544
3	1.2676	10.5016	52.3667	2375.8565	0.4630	0.4129
4	1.3040	9.3178	73.6345	3185.5091	0.4398	0.3518
5	0.8642	5.6899	13.3928	353.6537	0.5259	0.5499
6	1.0544	8.5897	31.8701	1168.5708	0.4902	0.4529
7	1.1072	7.7461	37.6973	1162.4972	0.4771	0.4279

11.6 strategy 4 พิจารณาจาก Expected(return) ทั้ง T=30, 180

11.7 Strategy 2 เพราะมี variance น้อยที่สุดทั้ง T=30 และ T=180

11.8 variance ของ strategy 7 มีค่ามากกว่า เพราะ:

Σ	a	b	c	d
a	10×10^{-3}	0	4×10^{-3}	5×10^{-3}
b	0	3×10^{-3}	0	0
c	4×10^{-3}	0	12×10^{-3}	2×10^{-3}
d	5×10^{-3}	0	2×10^{-3}	15×10^{-3}

จะเห็นได้ว่า $cov(r_a, r_c) < cov(r_a, r_d)$
ทำให้ directional relationship ระหว่าง
two random variables
ส่งผลให้ a,d มี variance ที่เยอะกว่า (กระจายมากกว่า)

11.9 เป็น general pratice for good investment เพราะมีการคำนวณค่าทางสถิติที่หลากหลาย
ทำให้มีข้อมูลในการตัดสินใจได้ประมาณนึง แต่ จำนวนวัน (T=30, 180) ยังเป็นจำนวนวันที่น้อยมาก
ทำให้ค่ายังดูเพี้ยน แปลก และไม่น่าเชื่อถือมากพอ ที่จะเป็นข้อมูลประกอบตัดสินใจได้