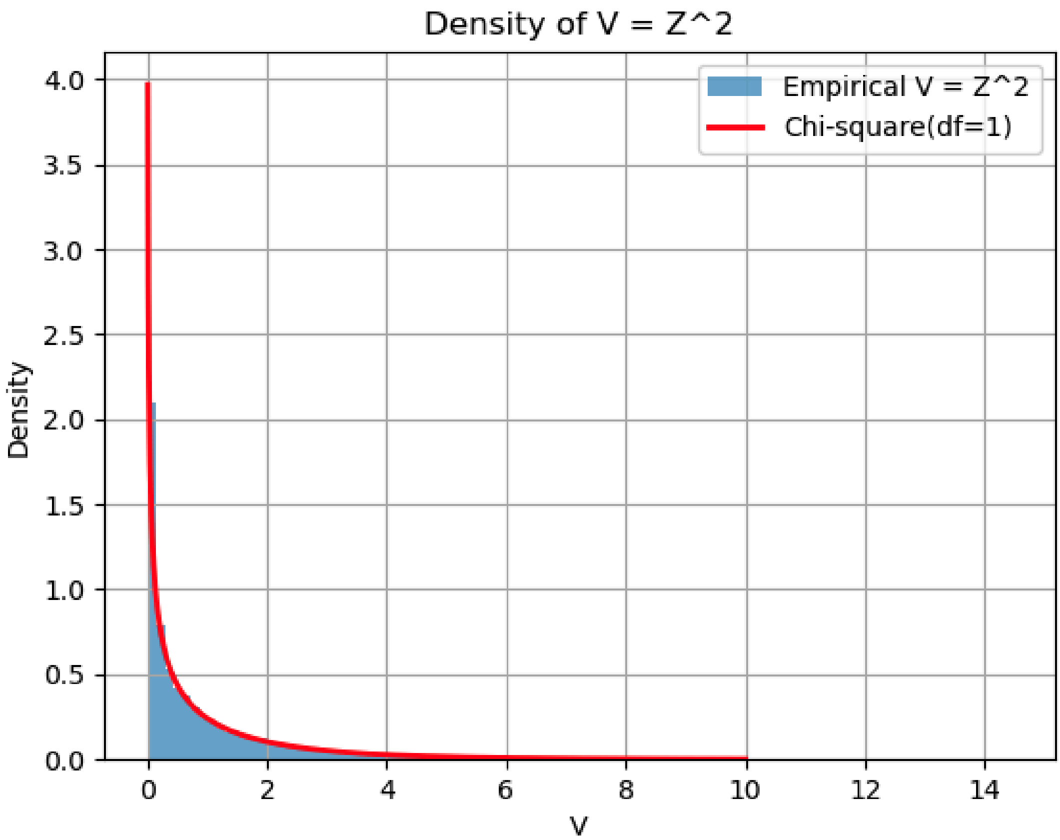
문제 **9-1. R**을 이용해서 정규분포를 따르는 확률변수 **Z**의 샘플을 **10000**개 뽑고, **V=Z^2**를 만들어보세요.

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import chi2

# 1. 정규분포 Z에서 10000개 샘플 생성
np.random.seed(0)
Z = np.random.normal(loc=0, scale=1, size=10000)

# 2. V = Z^2
V = Z**2

# 3. 밀도 그래프
plt.figure()
plt.hist(V, bins=100, density=True, alpha=0.6, label='Empirical V = Z^2')
x = np.linspace(0, 10, 1000)
plt.plot(x, chi2.pdf(x, df=1), 'r-', lw=2, label='Chi-square(df=1)')
plt.title('Density of V = Z^2')
plt.xlabel('V')
plt.ylabel('Density')
plt.legend()
plt.grid(True)
plt.show()

# 4. V < 1, V < 4의 비율 계산
prob_v_lt_1_empirical = np.mean(V < 1)
prob_v_lt_4_empirical = np.mean(V < 4)

# 5. 자유도 1인 chi-square 분포의 이론값과 비교
prob_v_lt_1_theoretical = chi2.cdf(1, df=1)
prob_v_lt_4_theoretical = chi2.cdf(4, df=1)

# 결과 출력
print(f"Empirical P(V < 1): {prob_v_lt_1_empirical:.4f}")
print(f"Theoretical P(V < 1) (Chi2 df=1): {prob_v_lt_1_theoretical:.4f}")
print()
print(f"Empirical P(V < 4): {prob_v_lt_4_empirical:.4f}")
print(f"Theoretical P(V < 4) (Chi2 df=1): {prob_v_lt_4_theoretical:.4f}")
```

1) **R**을 이용해서 **V**의 **density plot**을 그리세요.



2) **10000**개의 샘플로 구한 **V**가 **1**보다 작은 비율은 얼마인가요**?** 자유도가 **1**인 **Chi-square** 분포가 **1**보다 작은 값을 가질 확률과 유사한가요**?**

```
Empirical P(V < 1): 0.6899
Theoretical P(V < 1) (Chi2 df=1): 0.6827
```

비율은 **10000**개 중 **6899**개가 작다.
유사하다.

3) **10000**개의 샘플로 구한 **V**가 **4**보다 작은 비율은 얼마인가요**?** 자유도가 **1**인 **Chi-square** 분포가 **4**보다 작은 값을 가질 확률과 유사한가요**?**

```
Empirical P(V < 4): 0.9574
Theoretical P(V < 4) (Chi2 df=1): 0.9545
```

비율은 **10000**개 중 **9574**개가 작다.
유사하다.

문제 **9-2.** 확률변수 **Z_1, Z_2, ..., Z_6** 를 표준정규분포로부터 얻은 무작위표본일 때**,** 아래 식이 **t-**분포를 따르기 위한 상수 **c**는 얼마인가요**?**

$$c\left(\frac{Z_4+2Z_5+3Z_6}{\sqrt{Z_1^2+Z_2^2+Z_3^2}}\right) \sim t(\cdot)$$

t-분포는 다음과 같은 형태를 가진다.

$$T = \frac{X}{\sqrt{Y/\nu}} \sim t(\nu)$$

$X \sim N(0, 1)$: 표준 정규
$Y \sim \chi^2(\nu)$: 자유도 ν의 카이 제곱
X와 Y는 독립

분자: $Z_4 + 2Z_5 + 3Z_6$

선형 결합이므로, $X \sim N(0, \sigma^2)$ 형태.
분산은 다음과 같습니다.

$$Var(X) = Var(Z_4 + 2Z_5 + 3Z_6) = 1^2 + 2^2 + 3^2 = 1 + 4 + 9 = 14$$

즉,

$$X \sim N(0, 14)$$

분모:

$$\sqrt{Z_1^2 + Z_2^2 + Z_3^2} \Rightarrow \chi^2(3)\text{의 루트}$$

$$\frac{X}{\sqrt{14}} = \frac{Z_4+2Z_5+3Z_6}{\sqrt{14}} \sim N(0, 1)$$

따라서 전체 **t-**분포 형태는 다음과 같아야 한다.

$$T = \frac{Z_4+2Z_5+3Z_6}{\sqrt{14}} / \sqrt{\frac{Z_1^2+Z_2^2+Z_3^2}{3}} \sim t(3)$$

이를 정리하면:

$$\frac{1}{\sqrt{\frac{14}{3}}} \cdot \frac{Z_4+2Z_5+3Z_6}{\sqrt{Z_1^2+Z_2^2+Z_3^2}}$$

따라서

$$c = \frac{1}{\sqrt{\frac{14}{3}}} = \sqrt{\frac{3}{14}}$$

문제 **9-3.** 평균 μ_x, 분산 σ^2인 정규분포로부터 표본크기 **10**인 무작위표본을 얻고**,** 이와 독립적으로 평균 μ_y, 공통된 분산 σ^2을 가지는 정규분포로부터 표본크기 **10**인 무작위표본을 얻었을 때**,** S_x^2이 S_y^2보다 **2**배 이상 클 확률은**?**

두 모집단은 아래와 같다.

$N(\mu_x, \sigma^2)$, $N(\mu_y, \sigma^2)$를 따른다.

각 표본의 크기는 $n = 10$
표본 분산: S_x^2, S_y^2

조건: $S_x^2 > 2 \cdot S_y^2$

즉, 다음 확률을 구하라는 뜻이다.

$$P(\frac{S_x^2}{S_y^2} > 2)$$

표본 분산의 비는 다음과 같은 자유도 $n - 1$인 F 분포를 따른다.

$$F = \frac{S_x^2}{S_y^2} \sim F(n_x - 1, n_y - 1)$$

여기서는:

$n_x = n_y = 10 \Rightarrow$ 자유도: $(9, 9)$
따라서: $F \sim F(9, 9)$

구하고자하는 확률은

$$P(F > 2) = 1 - F_{cdf}(2; 9, 9)$$

```python
from scipy.stats import f

# 자유도
d1, d2 = 9, 9

# 누적분포함수 값
cdf_val = f.cdf(2, d1, d2)
prob = 1 - cdf_val

print(f"P(F > 2) = {prob:.4f}")
```

```
P(F > 2) = 0.1582
```

15.82%이다.