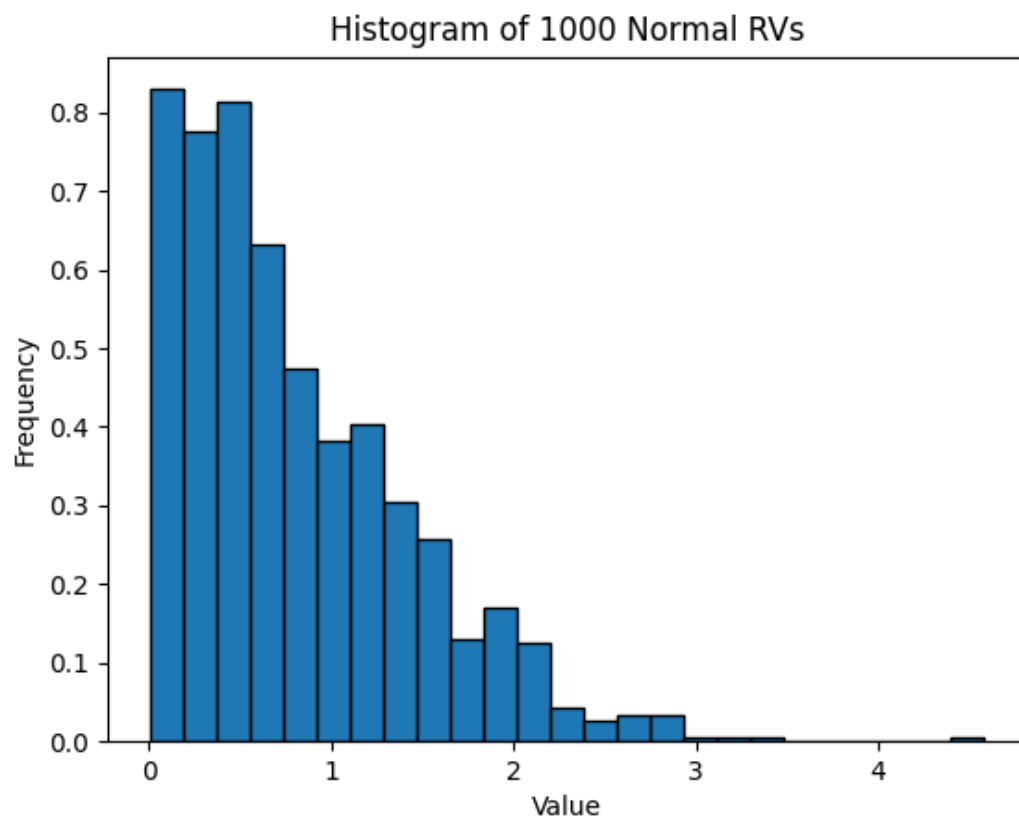


1121_SIMULATION AND STATISTICAL COMPUTING

#HW5

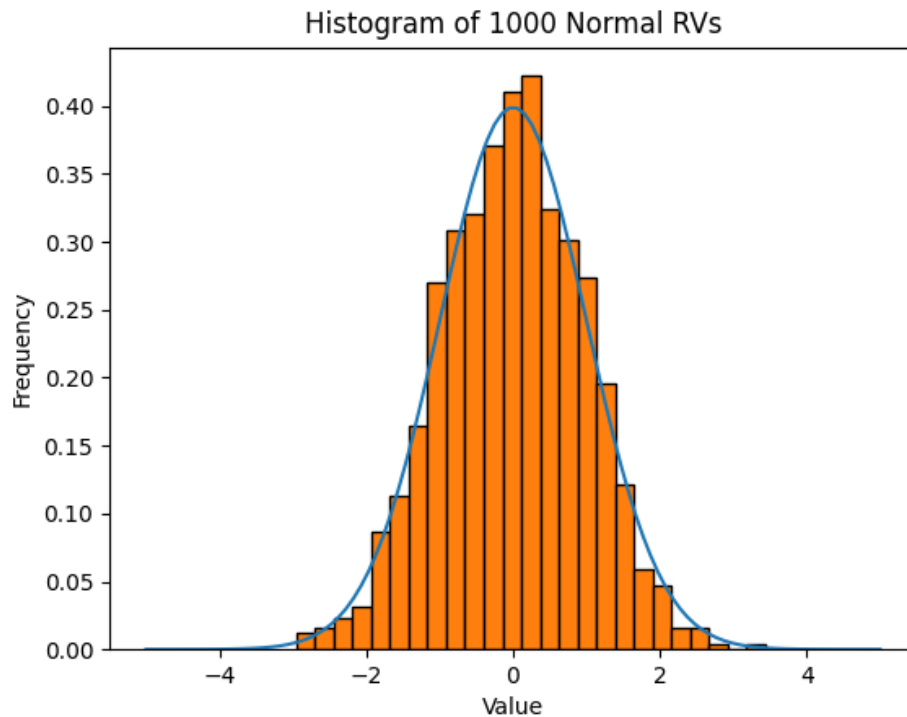
1. Change Example 5f to generate a Standard Normal Distribution



圖(一) 5f所產生 1000 個隨機變數分布

若單純重現 5f 之方法產生出的隨機變數並沒有負數。為了產生 Standard Normal Random Variable，該分佈以零為中心，我修改了演算法，以確保它以 50% 的機率生成正數和 50% 的機率生成負數。

```
if U <= math.exp(-(Y-1)**2 / 2):  
    if(random.randint(0,1) == 1):  
        return -Y  
    return Y
```



圖(二) 修改後符合常態分佈的結果

2.Generate a Normal Random Variable by Rejection Method to Represent Student Test Scores

```
def target_pdf(x):  
    return 1 / (stddev * np.sqrt(2*np.pi)) * np.exp(-0.5 * ((x - mean) / stddev)**2)  
  
def Uniform_pdf(x):  
    return 1/101  
  
C = target_pdf(mean) / Uniform_pdf(mean)
```

為了透過 Reject Method 產生 Normal Distribution Random Variable，我將 $g(x)$ 設置為一個均勻分佈（0~100）的隨機變數。為了找到常數 C ，需要尋找 $f(x) / g(x)$ 的最大值。由於 Normal Distribution 的機率密度在 Mean 處最大，因此我將 $mean$ 參數代入以求得最大值。

```

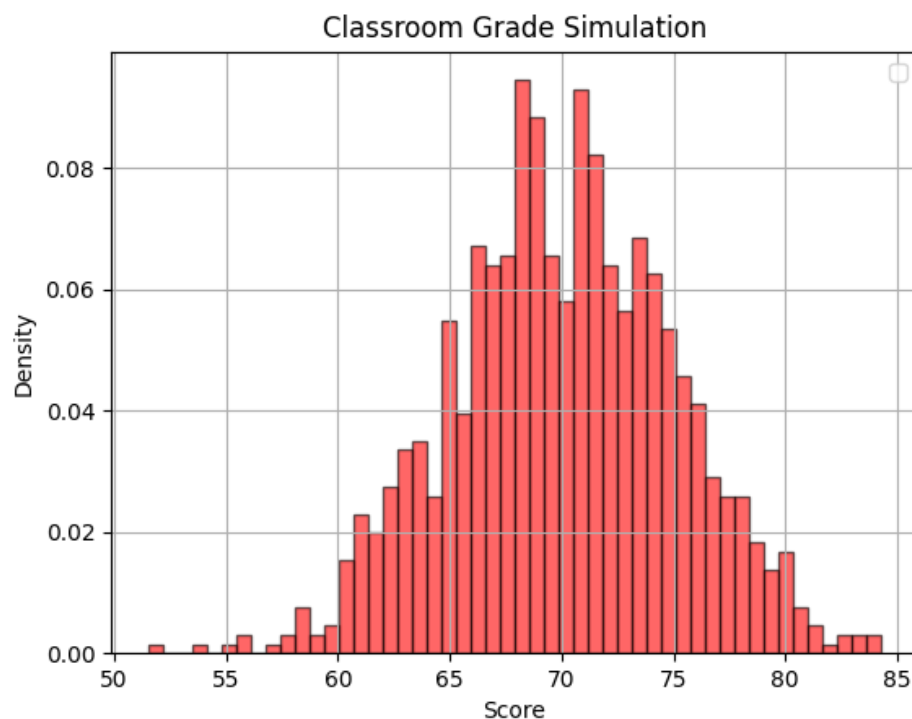
x = random.uniform(0, 100)

accept_prob = target_pdf(x) / (C * Uniform_pdf(x))

if random.random() < accept_prob:
    samples.append(x)
    break

```

接著，我隨機生成了 1000 個介於 0 到 100 之間的均勻分佈隨機數，這些隨機數代表學生成績。然後，我使用拒絕法（Reject Method）將它們轉換為符合常態分佈的隨機數。



圖(三) 透過 Reject Method 產生的全校 1000 位同學成績分布圖

我想知道這樣的成績分布是否符合常態分布的情形，因此我將計算在標準差內的同學數量並取得其數量所佔據全部同學的人數比例。

```

# 1個標準差內的學生數量
in1dev = 0
for i in range(len(samples)):
    if(samples[i]<=75 and samples[i] >= 65):
        in1dev += 1

```

```
# 2個標準差內的學生數量
in1dev = 0
for i in range(len(samples)):
    if(samples[i]<=80 and samples[i] >= 60):
        in1dev += 1
```

```
# 3個標準差內的學生數量
in1dev = 0
for i in range(len(samples)):
    if(samples[i]<=85 and samples[i] >= 55):
        in1dev += 1
```

```
The proportion of students within one standard deviation from the mean in relation to the total number of students.0.698
The proportion of students within two standard deviation from the mean in relation to the total number of students.0.969
The proportion of students within three standard deviation from the mean in relation to the total number of students.0.997
```

其結果分別為：

在一個標準差範圍內，有 69.8%的學生的成績落在平均值附近。

在兩個標準差範圍內，有 96.9%的學生的成績落在平均值附近。

在三個標準差範圍內，有 99.7%的學生的成績落在平均值附近。

與常態分佈：

在一個標準差範圍內，約有 68%的學生的成績落在平均值附近。

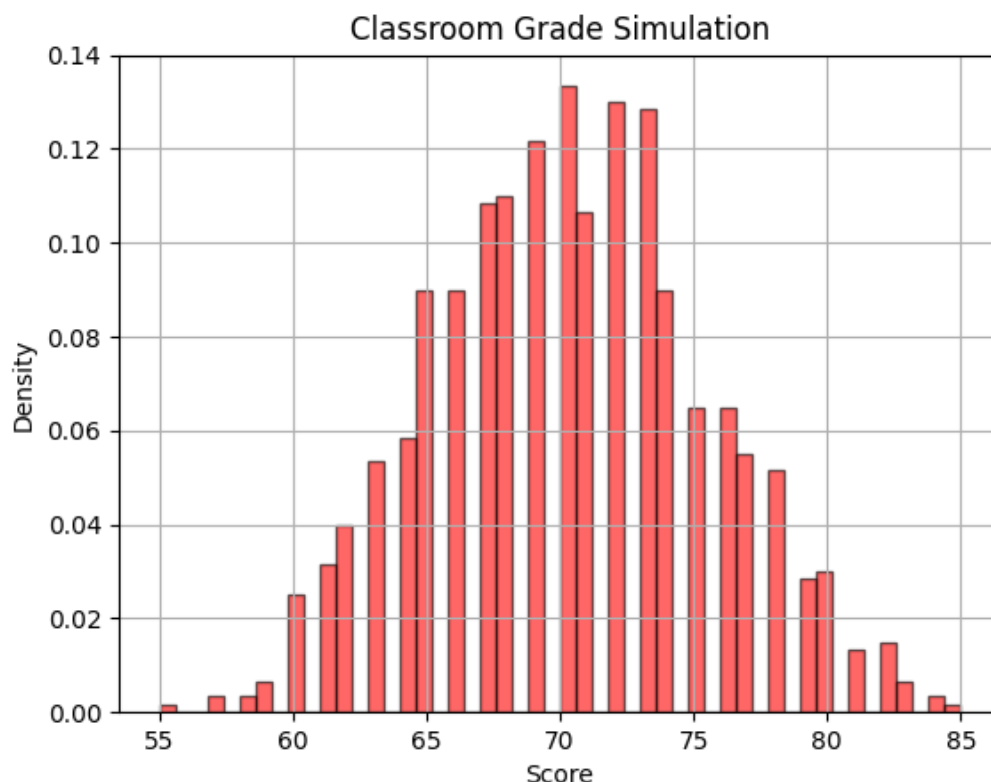
在兩個標準差範圍內，約有 95%的學生的成績落在平均值附近。

在三個標準差範圍內，約有 99.7%的學生的成績落在平均值附近。

接近。

雖然這樣的結果確實接近常態分佈，然而這樣的成績並不符合現況，由於大部分的學生成績通常為整數，而我所產生的隨機數皆為浮點數，因此我將所有的隨機數改為 0-100 的整數並觀察結果。

```
# x = random.uniform(0, 100)
x = random.randint(0, 100)
```



圖(四) 透過 Reject Method 產生的全校 1000 位同學整數型態成績

```
The proportion of students within one standard deviation from the mean in relation to the total number of students.0.704
The proportion of students within two standard deviation from the mean in relation to the total number of students.0.967
The proportion of students within three standard deviation from the mean in relation to the total number of students.1.0
```

這樣的模擬方法分布結果變為為：

在一個標準差範圍內，有 70.4%的學生的成績落在平均值附近。

在兩個標準差範圍內，有 96.7%的學生的成績落在平均值附近。

在三個標準差範圍內，有 100%的學生的成績落在平均值附近。

與常態分佈：

在一個標準差範圍內，約有 68%的學生的成績落在平均值附近。

在兩個標準差範圍內，約有 95%的學生的成績落在平均值附近。

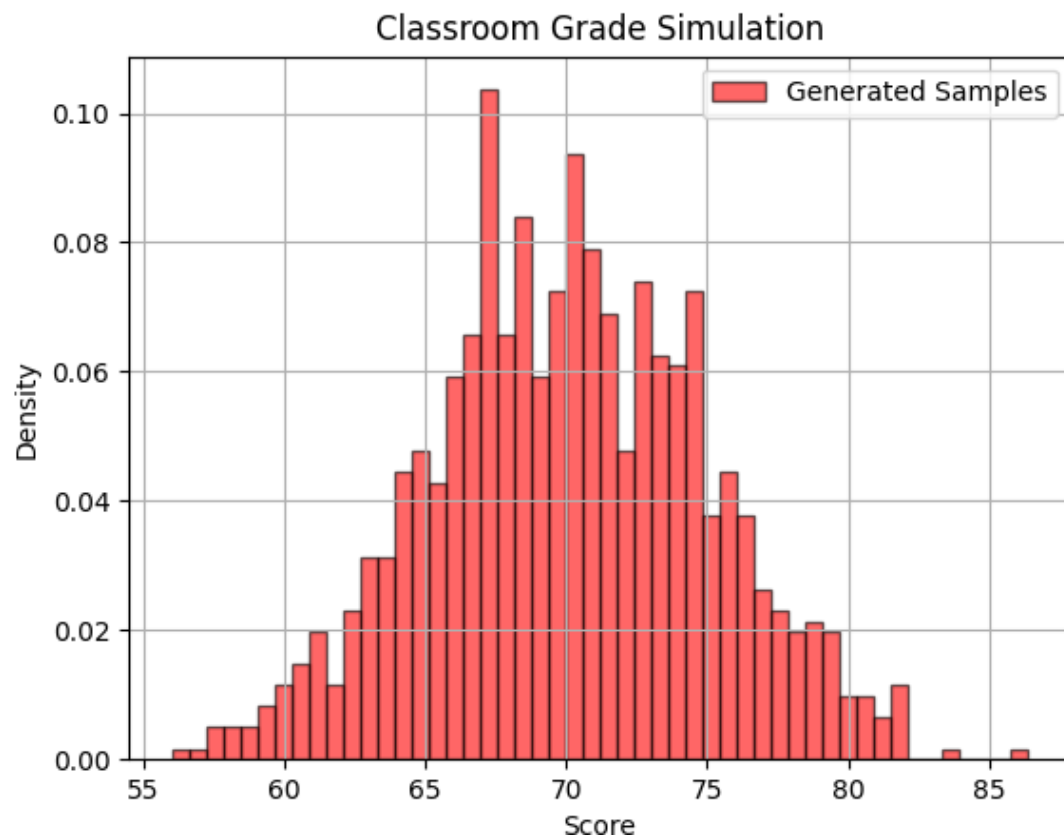
在三個標準差範圍內，約有 99.7%的學生的成績落在平均值附近。

差距較大。

我的結論是，當生成的隨機數為整數類型時，透過拒絕法可能無法有效地轉換為另一種隨機變數。這可能是因為原始隨機變數是連續型的。透過整數型態的隨機變數進行轉換可能會產生誤差，無法精確地反映連續分布的特性。

3.Generate a Normal Random Variable by Linear Transformation to Represent Student Test Scores

另外我透過將 5f 產生的標準常態分佈隨機變數的結果透過線性轉換，轉換成符合學生成績的分布並觀察結果。



圖(五) 透過 Linear Transformation 產生的全校 1000 位同學成績分布

The proportion of students within one standard deviation from the mean in relation to the total number of students.0.69
The proportion of students within two standard deviation from the mean in relation to the total number of students.0.957
The proportion of students within three standard deviation from the mean in relation to the total number of students.0.999

這樣的模擬方法分布結果變為為：

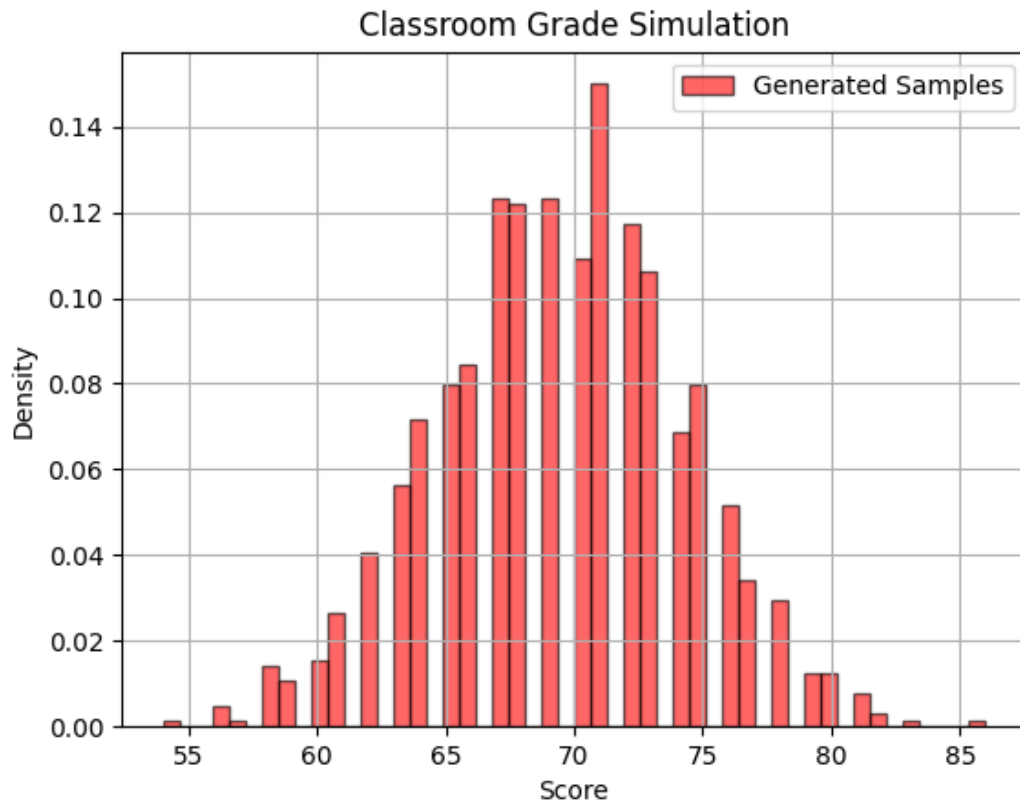
在一個標準差範圍內，有 69%的學生的成績落在平均值附近。

在兩個標準差範圍內，有 95.7%的學生的成績落在平均值附近。

在三個標準差範圍內，有 999%的學生的成績落在平均值附近。

與常態分佈接近。

同樣的，若我把同學成績改為整數型態觀察結果。



圖(四) 透過 Linear Transformation 產生的全校 1000 位同學整數型態成績

```
The proportion of students within one standard deviation from the mean in relation to the total number of students.0.745
The proportion of students within two standard deviation from the mean in relation to the total number of students.0.97
The proportion of students within three standard deviation from the mean in relation to the total number of students.0.998
```

這樣的模擬方法分布結果變為為：

在一個標準差範圍內，有 **74.5%** 的學生的成績落在平均值附近。

在兩個標準差範圍內，有 **99%** 的學生的成績落在平均值附近。

在三個標準差範圍內，有 **998%** 的學生的成績落在平均值附近。

與常態分佈差距較大，並且與 Rejection Method 結論一致。