Probability 2024 Final Project - Question 1

組名：鐘於 pdf
組員：鐘奇恩(b12902013)、劉蕃熙(b12902031)、李瑞恩(b12902067)

```
In the realm of machine learning, we inquire,
Between Normal and Xavier, which to admire?
But what defines "better" in this algorithmic choir?
Let's explore, let's aspire.
```

程式碼與數據資料

git repo: https://github.com/LiuFelicity/prob_final.git
其包含我們修改後的程式碼、跑出來的數據，以及分析數據所用的程式碼

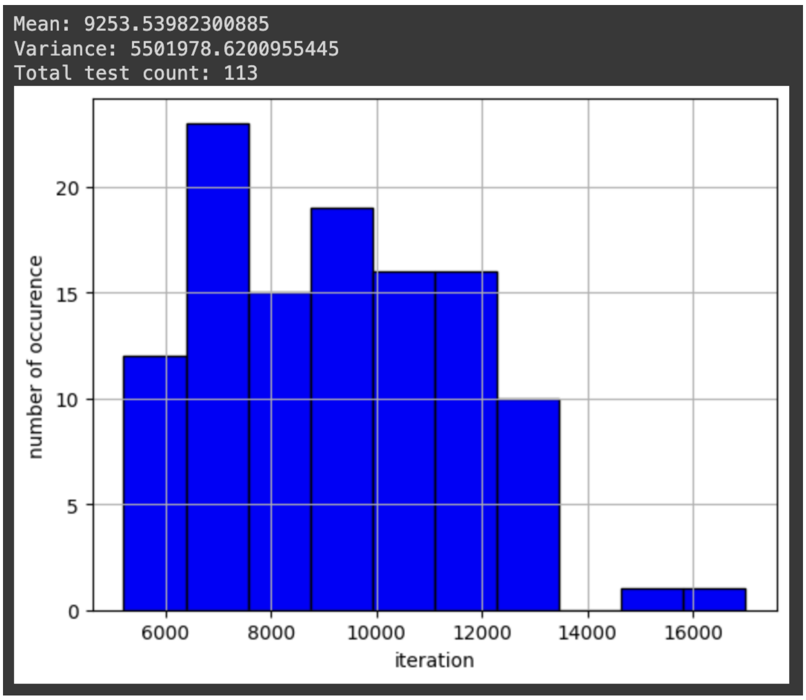
實驗數據

1-1 Does different initialization affect performance

Below are the results of *iteration*:

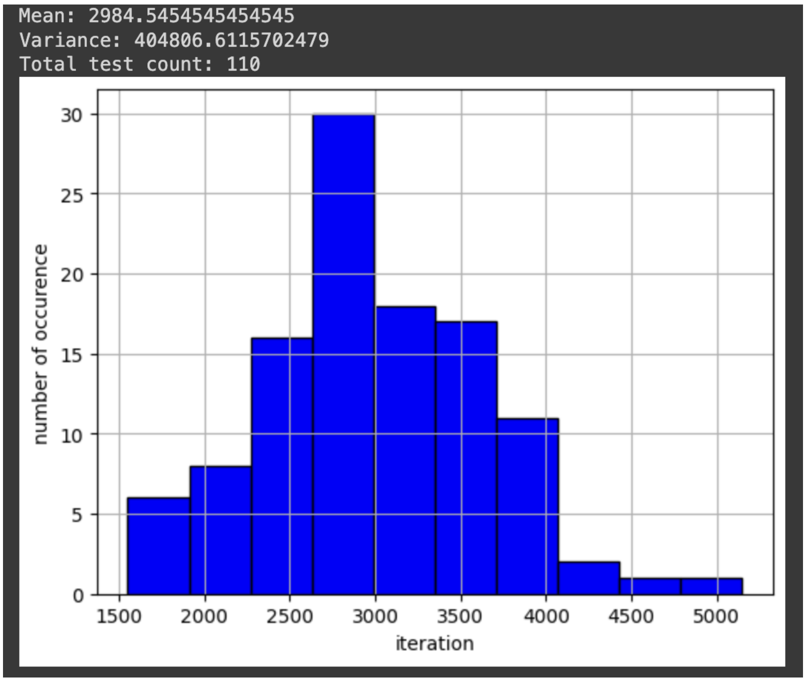
Normal - ChickenRabbit

mean: 9253.53982300885
variance:5501978.6200955445
standard deviation: 2345.6296852
Total test count: 113
histogram:



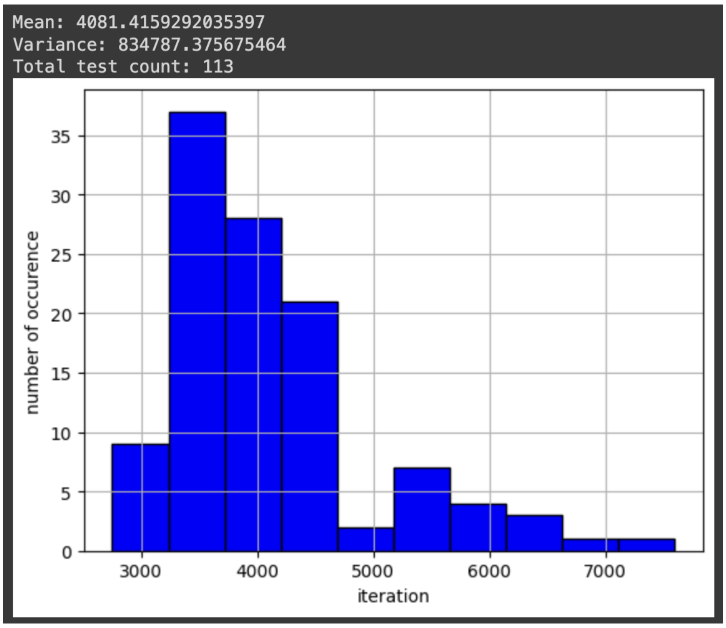
Normal - GCD

Mean: 2984.5454545454545
Variance: 404806.6115702479
standard deviation: 636.244144626
Total test count: 110
histogram:



Xavier - ChickenRabbit

Mean: 4081.4159292035397
Variance: 834787.375675464
standard deviation: 913.666993863
Total test count: 113
histogram:



Xavier - GCD

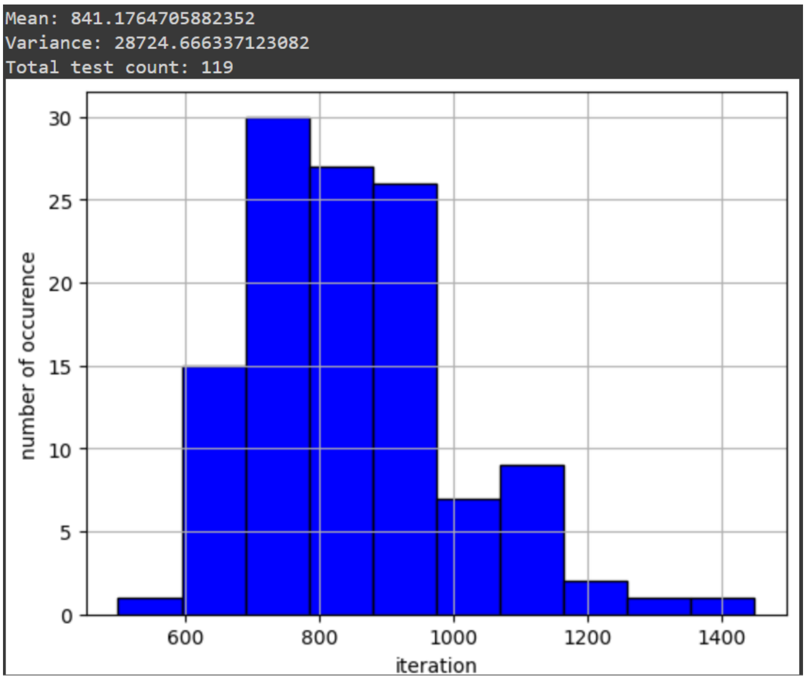
mean: 841.1764705882352

variance: 28724.666337123082

standard deviation: 169.483528218

Total test count: 119

histogram:



1-2

Below are the results of *iteration*:

ChickenRabbit

weight initialization seed = 62

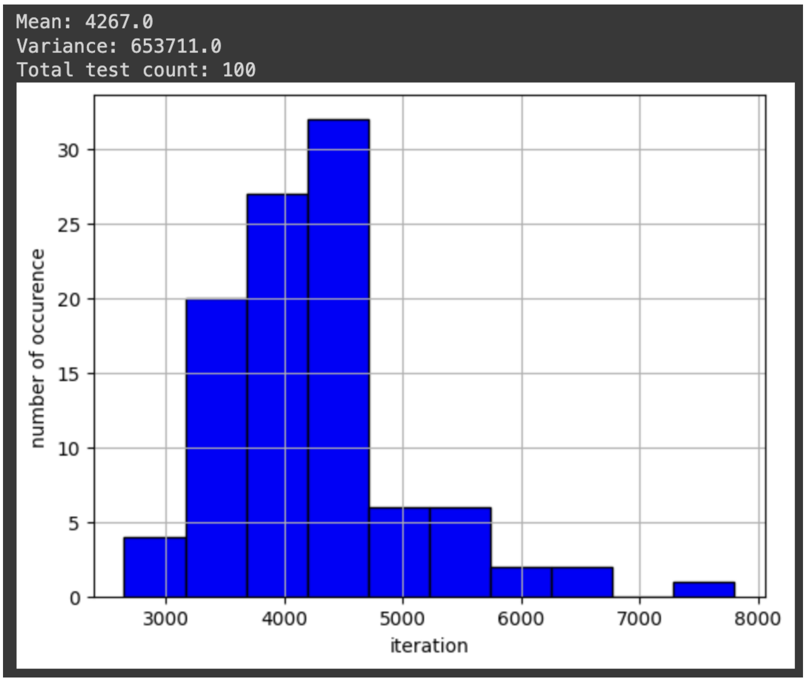
Mean: 4267.0

Variance: 653711.0

standard deviation: 808.523963776

Total test count: 100

histogram:



weight initialization seed = 300

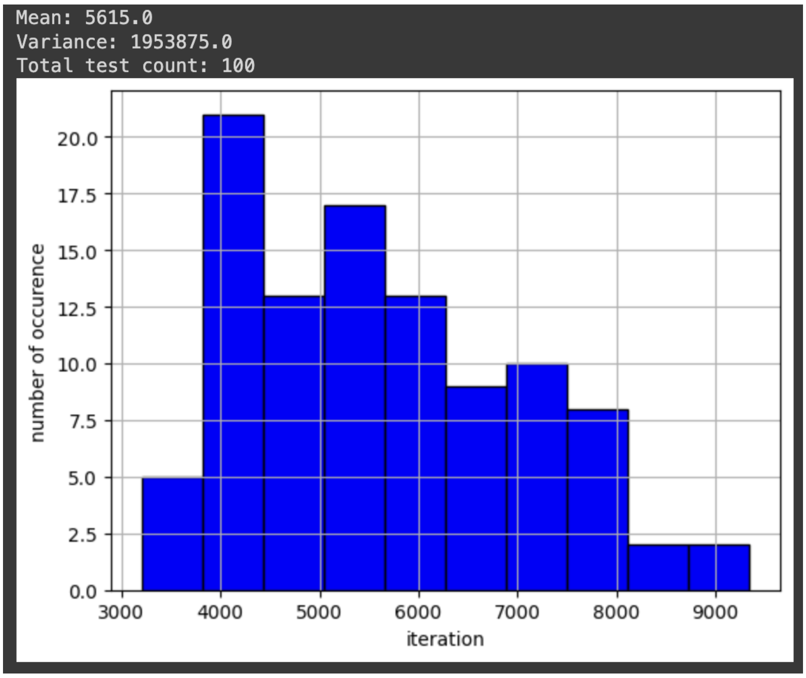
Mean: 5615.0

Variance: 1953875.0

standard deviation: 1397.81078834

Total test count: 100

histogram:



weight initialization seed = 1000

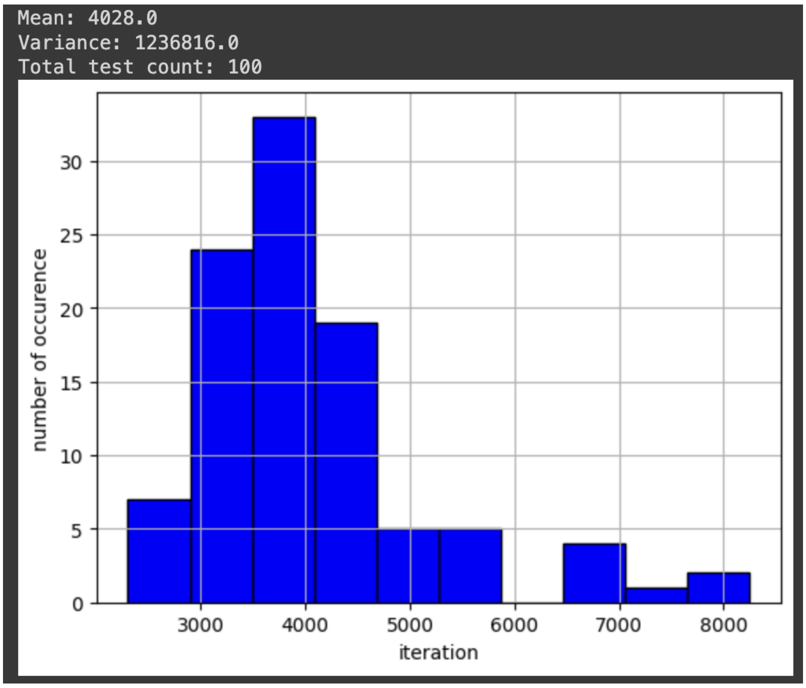
Mean: 4028.0

Variance: 1236816.0

standard deviation: 1112.12229543

Total test count: 100

histogram:



GCD

weight initialization seed = 62

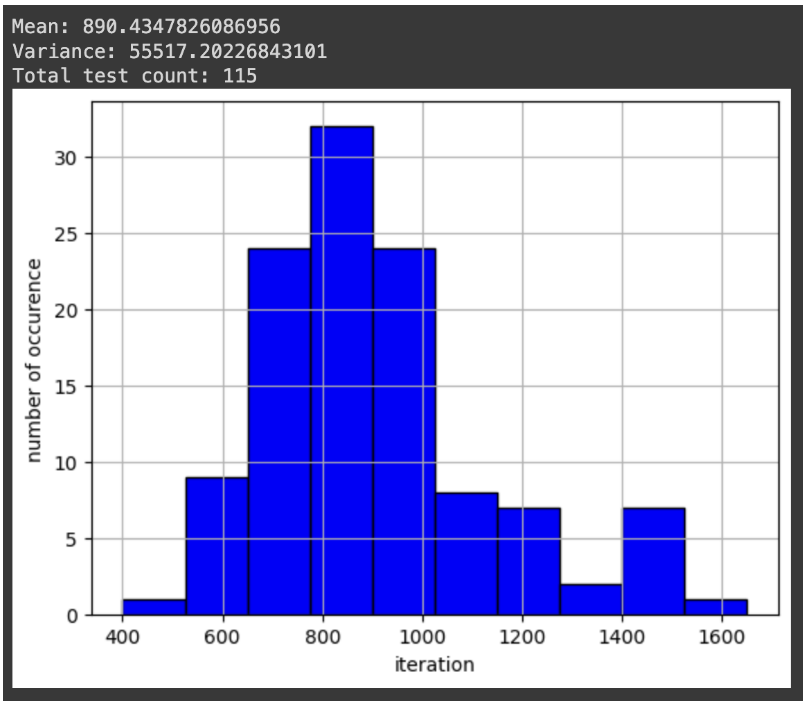
Mean: 890.4347826086956

Variance: 55517.20226843101

standard deviation: 235.620886741

Total test count: 115

histogram:



weight initialization seed = 300

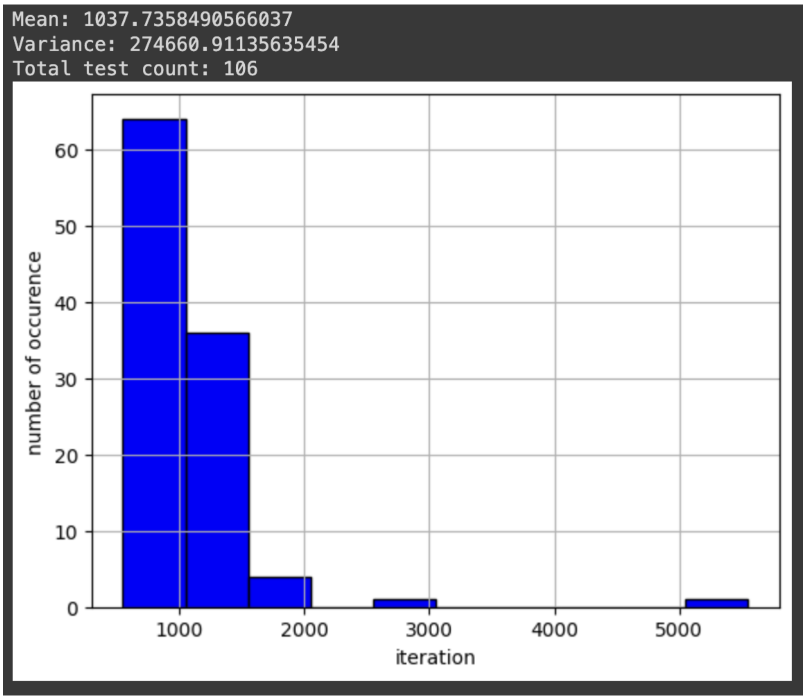
Mean: 1037.7358490566037

Variance: 274660.91135635454

standard deviation: 524.081016024

Total test count: 106

histogram:



weight initialization seed = 1000

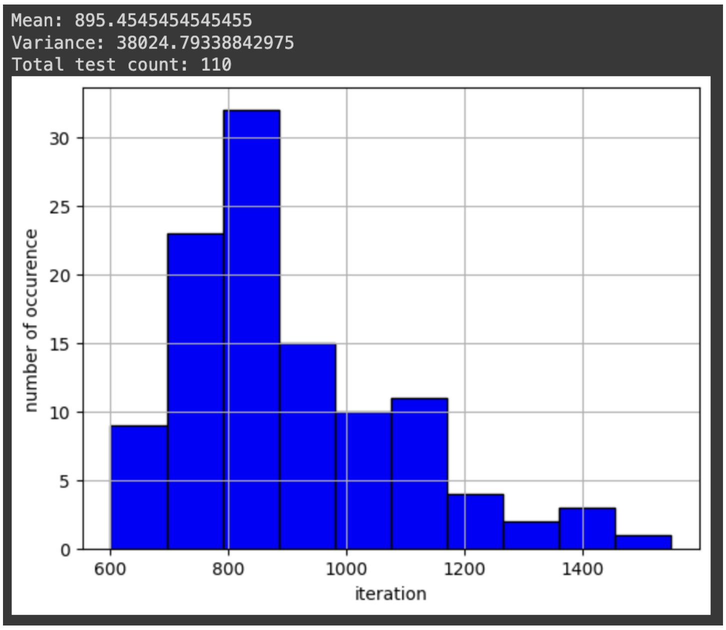
Mean: 895.4545454545455

Variance: 38024.79338842975

standard deviation: 194.999470226

Total test count: 110

histogram:



P-value 分析

1-1

欲對兩種未知分佈之 μ 比較大小順序，可以用以下幾種檢定方法：

- 知道變異數：Z 檢定
- 不知道變異數且變異數一樣：two sample T test
- 不知道變異數且變異數不一樣：Welch test

在這題，我們分別用 Welch's test 在課本中之公式 (註一)，以及 python 內建 *scripy* 函式庫進行驗算，兩者得到的答案相似到小數點後第九位，我們相信其為精度導致的差異，可視為兩者結果一樣。

在此題我們設定 *null hypothesis* (H_0) 為 "Normal is better" ，意思是用 Normal initialize 權重時 ，模型較佳，跑出的 iteration 的平均會較小 。H_1 則是"Xavier is better"，表示 Noraml 跑出的 iteration 平均會比 Xavier 大。

經過計算，我們發現 p-value 都遠小於 0.01，由上課簡報 L8_test_hypothesis 第九頁，可得出 "very strong presumption against neutral hypothesis" 之結論，因此我們可以推翻 H_0，接受 H_1: Xavier is better.

註一 [Welch's test 之公式]

$$t\ score = \frac{\mu_x - \mu_y}{\sqrt{(\frac{s_x^2}{n} + \frac{s_y^2}{m})}}$$

$$自由度\ r = \lfloor \frac{(\frac{s_x^2}{n} + \frac{s_y^2}{m})^2}{\frac{1}{n-1}(\frac{s_x^2}{n})^2 + \frac{1}{m-1}(\frac{s_y^2}{m})^2} \rfloor$$

- μ_x 是 Normal 的平均值
- μ_y 是 Xavier 的平均值
- s_x 是 Normal 的標準差
- s_y 是 Xavier 的標準差
- n 是 Normal 的資料數
- m 是 Xavier 的資料數

ChickenRabbit

t score: 21.841089893860516
degree of freedom: 145
t score 對應的 p-value: 5.295568261719669e-48

GCD

t score: 34.22740243060551
degree of freedom: 123
t score 對應的 p-value: 5.140049420625265e-65

1-2

在此題中，我們定義 H_0 為 "order doesn't affect performance"，H_1 為 "order matters"。

根據 iteration 的 sample variance，我們訂定 當 σ 大於某個數字時，可以說是 order matters ，因此若 sample σ 大於這個門檻，就代表我們做出的實驗結果是支持 order matters 的。

利用變異數假設檢定公式 (註二)，我們可以算出 sample σ 小於 σ 的機率 (即為p-value)，p-value 越小，代表 order matter 越顯著，我們越有足夠大的信心能推翻 H_0.

另外，跑完 weight initialization seed = 62 後，我們還跑了 weight initialalation seed = 300 和 1000，目的是確認 "order matters" 這件事對每一種 initialization 都是正確的。

經過計算我們發現，無論 weight initialization seed，p-value 皆非常小，表示我們有足夠大的信心推翻 H_0，而接受 H_1: order matters。

註二 [變異數假設檢定公式與後續推導]
給定一個分布，和給定的標準差門檻 σ，檢定門檻 α（發生機率要多小我們才會拒絕 H_0），我們接受 H_0 的條件為：

$$\sigma \geq \sqrt{\frac{(n-1)\, s^2}{\chi^2_{1-\alpha,\, n-1}}}$$

$$df = n - 1$$

其中，

- σ 是我們認為 "order matters" 的標準差門檻。
- s^2 是 sample variance
- n 是資料數
- df 是自由度

根據定義，滿足上式的最小 α 即是 p-value。因此，在未給定 α 的情況下，要從給定的標準差門檻 σ 回推 α 對應的 p-value，會滿足：

$$\sigma = \sqrt{\frac{(n-1)\, s^2}{\chi^2_{1-p,\, n-1}}}$$

整理得

$$\chi^2_{1-p,\,n-1} = \frac{(n-1)s^2}{\sigma^2}$$

由此式，我們使用 χ^2 的反函數即得出 p-value（χ^2 的反函數由 python scirpy 函式庫提供）。

ChickenRabbit

weight_seed = 62

sigma=500
Mean: 4267.0
Variance: 653711.0
χ^2: 258.869556
p-value: 3.3306690738754696e-16

weight_seed = 300

sigma=500
Mean: 5615.0
Variance: 1953875.0
χ^2: 773.7345
p-value: 0.0
(p-value 已經小到 python 無法表示)

weight_seed = 1000

sigma=500
Mean: 4028.0
Variance: 1236816.0
χ^2: 489.779136
p-value: 0.0
(p-value 已經小到 python 無法表示)

GCD

weight_seed = 62

sigma=150
Mean: 890.4347826086956
Variance: 55517.20226843101
χ^2: 281.2871581600504
p-value: 3.3306690738754696e-16

weight_seed = 300

sigma=150
Mean: 1037.7358490566037
Variance: 274660.91135635454
χ^2: 1281.7509196629878
p-value: 0.0

weight_seed = 1000

sigma=150
Mean: 895.4545454545455
Variance: 38024.79338842975
χ^2: 184.20899908172635
p-value: 9.042074660348 12e-06

參考資料

https://blog.csdn.net/luoxuexiong/article/details/95772045
http://www.math.ncu.edu.tw/~yu/ms100/boards/lec31_ms_100.pdf
https://www.itl.nist.gov/div898/handbook/eda/section3/eda358.htm