篇名：

Python 的伽碼分布的驗證

作者：

潘建蒼。國立台北科技大學。電資二。106820003。

陳品睿。國立台北科技大學。電資二。106820026。

壹●前文

一、研究背景

伽瑪分布是統計學的一種連續機率函數。隨機變數x，若其 p.d.f.為

$$f(x) = \frac{x^{(\alpha-1)}\beta^{\alpha}e^{-\beta x}}{\Gamma(\alpha)} \quad , \quad x > 0 \quad \text{……①}$$

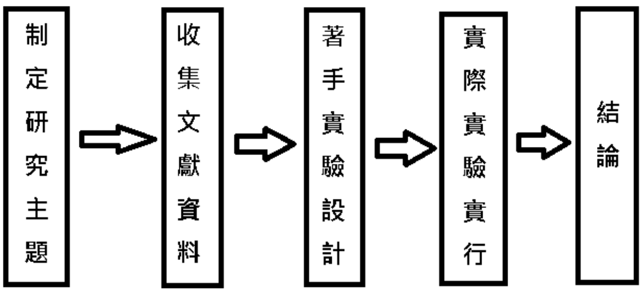
其中$\alpha, \beta > 0$為二正數，便稱為有參數α, β之伽瑪分布(gamma distribution), 以$\Gamma(\alpha, \beta)$表之，在此伽瑪函數之定義為

$$\Gamma(\alpha) = \int_0^{\infty} t^{(\alpha-1)}\, e^{-t}\, \mathrm{dt} \quad , \quad \alpha > 0 \quad \text{……②}$$

二、研究動機

在實作上，我們常常利用 python 中、已經寫好的伽瑪分布的函式；然而我們卻常常忽略一種可能性——函式有可能是錯誤的。所以我們這一組想透過實驗的方式來驗證 python 函式庫的正確性。

三、研究流程



圖一 實驗流程圖

貳●正文

一、實驗設計

（一）前提

固定公式①中的變數$\beta=0.3$和$\alpha=20$

（二）流程

在本次實驗中我們只有一個變數因子 У 為樣本數。У 為一就是總共取一次樣本，二就是總共取二次樣本，以此類推……。故 X 是為大於等於 0 的整數。

為了防止電腦的運算超過它的負荷，我們還設了一個變數 Ψ 為取值的一個上屆。而 Ψ 的值為：

$$\mathbf{E}[\mathbf{X}] = \frac{\alpha}{\beta}$$

$$\boldsymbol{\sigma}(\mathbf{X}) = \sqrt{\frac{\alpha}{\beta^2}}$$

$$\Psi = \mathbf{E}[\mathbf{X}] + \mathbf{10}\boldsymbol{\sigma}(\mathbf{X})$$

在此實驗中$\Psi = 215.73786$。$t = X$。

在取第一次的樣本中，當 $t = i$ 時，我們使用隨機變數函數隨機產生一個 0 到 1 的一個數字(為了方便，稱呼這個數字為 R_i)，藉由已知的 p.d.f.，我們可以知道對應到 R_i 機率為 P_i。

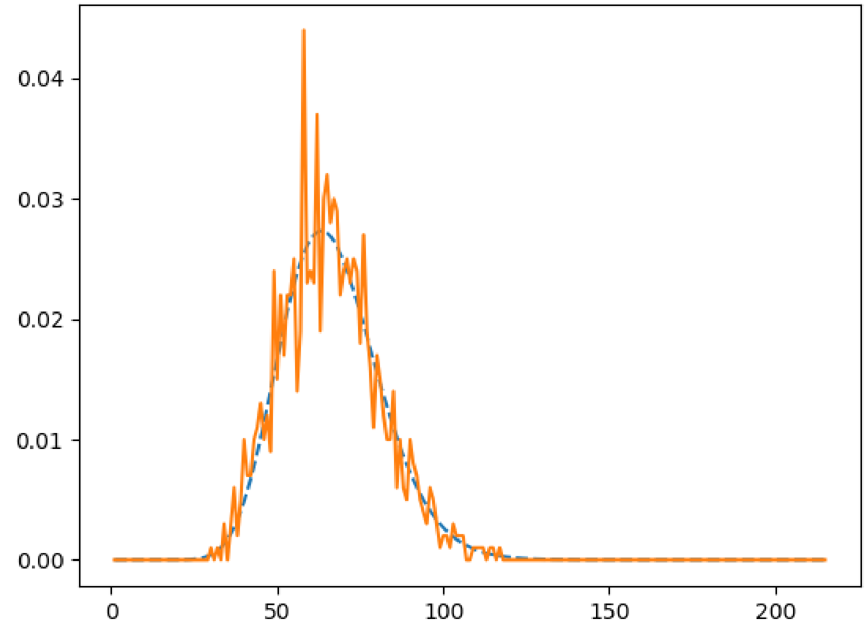
若 $R_i <= P_i$ 則代表此樣本在 $t = i$ 發生，此時 $t = i$ 的成功樣本數加一，重新從 $t = 1$ 開始跑下一個樣本。

若 $R_i > P_i$ 則代表此樣本在 $t = i$ 時不發生，此樣本繼續執行 $t = i + 1$ 時的判斷。若到 $t = \mathrm{int}(\Psi)$ 前皆不發生，則此樣本結束，重新從 $t = 1$ 開始跑下一個樣本。

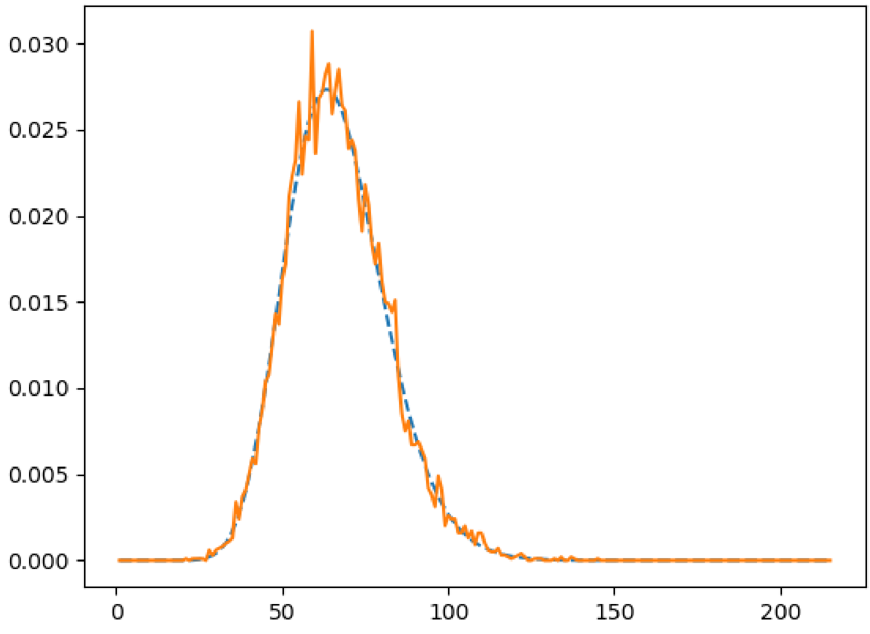
在之後的取 y 次樣本中，我們重複同樣的流程。最後畫出圖表：橫軸為 t，縱軸為成功次數除以總樣本數。

二、實驗結果

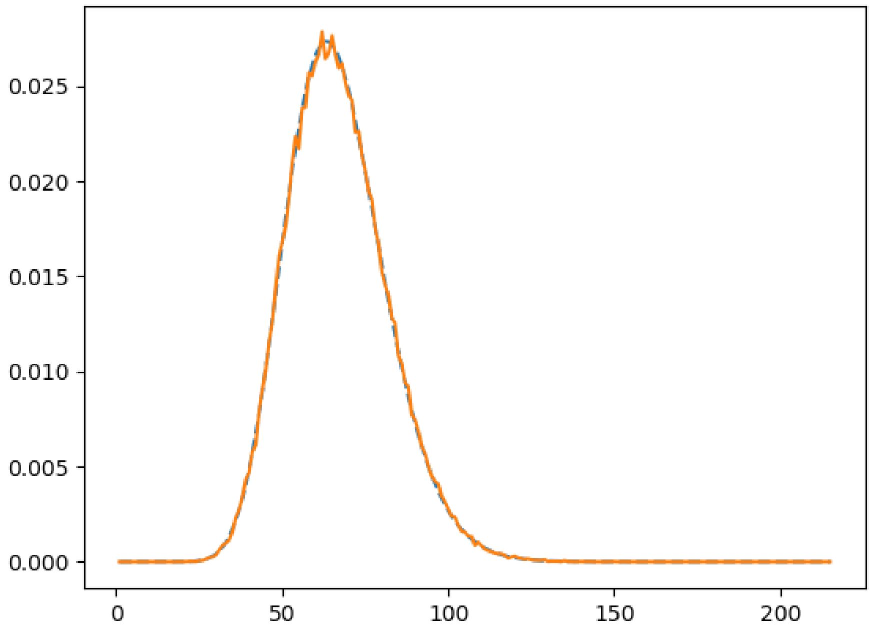
橘色的圖是實驗組，藍色虛線是用 python 的 gamma 分佈函式直接描繪的對照組



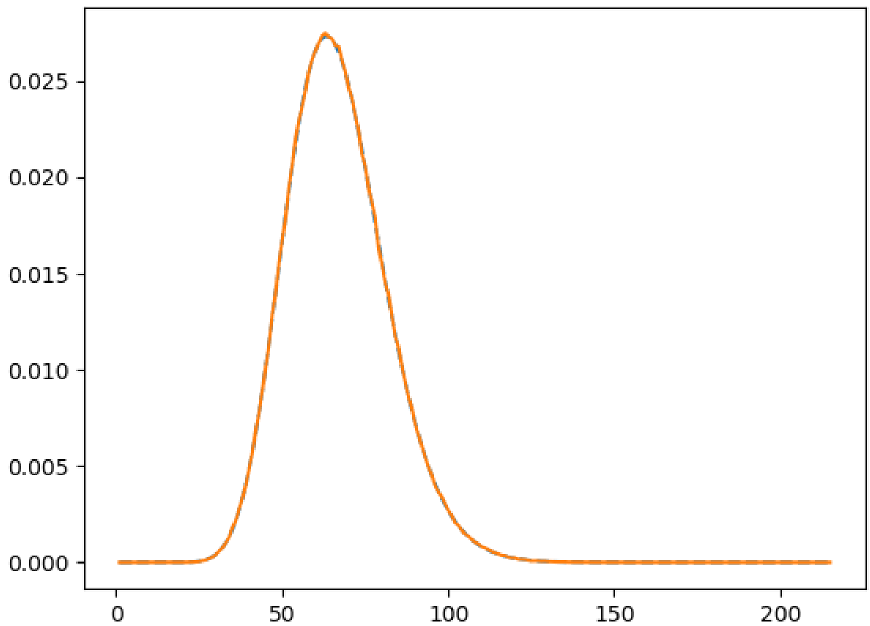
圖(二) $y = 1000$時的實驗機率分布



圖(三) $y = 10000$時的實驗機率



圖(四) $y = 100000$時的實驗機率分布



圖(五) $y = 1000000$時的實驗機率分布

參●結論

一、研究結果

根據以上的圖表，我們發現當變因 y 的值越大，也就是操作的越多次，實驗所描繪出來的圖會越來越像用函式所描繪的結果。

二、問題與討論

1. Did you use a library function found somewhere or did you write the function yourself? If it is a library function found somewhere, please write a short description of the function, including where to find it, how to use it, what are the parameters for the function. If you wrote the function yourself, you should describe the method you used.

Answer：我們是用自己寫出的函式產生實驗組，用 python 的函式產生對照組。自己寫的函式：

已知在 x 為整數時，$\Gamma(\alpha) = (\alpha - 1)!$

因此$f(x) = \frac{x^{(\alpha-1)}\beta^{\alpha}e^{-\beta x}}{\Gamma(\alpha)}$可改寫成

$$f(x) = \frac{x^{(\alpha-1)}\beta^{\alpha}e^{-\beta x}}{(\alpha-1)!} = \frac{\alpha x^{\alpha}\beta^{\alpha}e^{-\beta x}}{\alpha! x}$$ 。

為了防止數字過大無法運算，可再改寫成

$$f(x) = \frac{\alpha e^{-\beta x}}{x}\prod_{i=1}^{\alpha}\left(\frac{x\beta}{i}\right)$$

肆●參考資料

http://www.stat.nuk.edu.tw/prost/Web/pdf9.htm
維基百科：
https://zh.wikipedia.org/wiki/%E4%BC%BD%E7%8E%9B%E5%88%86%E5%B8%83

伍●附件

一、程式碼

```python
import matplotlib.pyplot as plt
import math
import random
import scipy.stats as stats

def gamma_distribute(alpha,beta,t):
    sum = math.exp(-beta*t)
    for i in range(1,alpha+1):
        sum = sum*beta*t/i
    sum *= alpha/t
    return sum

beta = 0.3
alpha = 20
sigma = math.sqrt(alpha)/beta
limit = alpha/beta + 10*sigma
t = []
y = []
y_py = []
y_random = [0]*int(limit)
samples = 100000
for i in range(1, int(limit)+1):
    t.append(i)
    y.append(gamma_distribute(alpha,beta,i))
    y_py.append(stats.gamma.pdf(i, a = alpha, scale = 1/beta))
for j in range(samples):
    for i in range(int(limit)):
        if random.random() <= y[i]:
            y_random[i] += 1/samples
#plt.plot(t, y, '--')
plt.plot(t, y_py, '--')
plt.plot(t, y_random, '-')
plt.show()

```