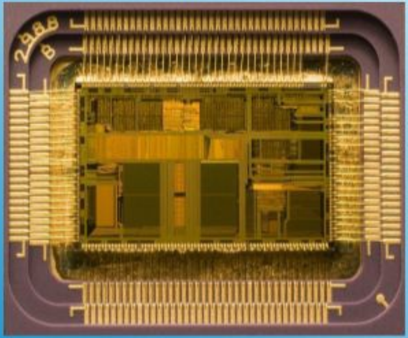
機率統計

Probability & Statistics

（使用 R 軟體）



- *R 軟體*
- *機率統計*
- *中央極限定理*
- *檢定*
- *回歸*
- *時間序列*
- *品質管制*
- *貝氏網路*
- *HMM 學習*
- *EM 算法*

作者：陳鍾誠

機率統計 -- 使用 R 軟體

陳鍾誠

2013 年 3 月

機率統計 -- 使用 R 軟體

6. 均匀分布 (Uniform distribution)
7. 常態分布 (Normal Distribution)

機率統計 -- 使用 R 軟體

序

機率統計是「自然科學」與「社會科學」都共同需要的數學語言，傳統上許多學校都會採用商用的 SPSS 或 SAS 等軟體作為課程實作的工具。但是這些軟體都是商用軟體，學生在家中很難合法安裝。

事實上，開放原始碼領域的「R軟體」比起 SPSS 或 SAS 毫不遜色，而且具有相當多的開放資源，已經成為學習機率統計的重要工具。在本書中，我們將使用 「R 軟體」進行機率統計的實作，讓學習者能以「實作印證理論」，加深學習的效果。

R 軟體當中預設就包含了各式各樣的機率模型，以及各種統計工具，可以讓讀者一邊學習，一邊用簡單的指令進行機率統計的實驗。透過這種學習方式，讀者可以透過程式實際體會理論的意義。舉例而言，當我們看到常態分配的機率模型時，我們可以直接使用下列指令來畫出常態分布，並且用程式產生符合常態分布模型的樣本，以便進行某種互動式的學習，用實驗體會常態分布的意義。

我很喜歡用武俠小說中的「氣宗與劍宗」與學術中的「理論與實務」相對比。氣宗的人一開始就學習一堆理論，直到內力充足以後才慢慢了解實務操作；而劍宗的人則是直接拿起劍就找人比試，在不了解理論的情況下直接出招，一開始進步很快，但到後來就會發現內功不夠深厚，遇到高手時總要吃虧的。

雖然我的求學過程比較像是氣宗的教育方式，但是最後寫出來的書卻往往像是劍宗的教材，或許這就是一種內心矛盾的反射，因為理論念得多了，就發現自己在實務上一竅不通，所以寫出來之後往往從實務出

發，最後才回到理論進行印證。

在本書中，我們將採用「理論=>實務=>理論=>實務...」的循環，交替的說明原理與實務，我們會先給大家看一大堆「圖片」、「影片」甚至是「程式」，然後再回過頭來講解原理，透過這種方式，希望大家都能達到「氣劍合一」的境界。

陳鍾誠 2012/10/23 於 金門大學 資訊工程系

授權聲明

機率與統計

簡介

在現實的生活當中，有許多我們無法準確描述的現象，這些現象的出現包含了某種程度的隨機性。舉例而言，我們無法精確的預知明天是否會下雨、股票會漲或者會跌、匯率會如何波動，人會不會生病等等。

但是當這些現象出現的總體量很多的時候，我們就可以「統計」出該事件發生的「機率」，於是我們的天氣預報可以預測明天下雨的機率、我們也可以統計出股票漲跌的機率、人們生病的機率、或者某人買了一張彩券後中獎的機率等等。

因此，機率與統計可以說是一體的兩面，當我們知道某個基本事件的先驗機率1時，我們可以根據此一機率計算某個組合事件發生的機率。例如我們可以知道兩顆公平的骰子同時出現 6 點的機率是 1/36 ，而連續投擲公平的銅版五次，每次都是正面的機率為 1/32等。

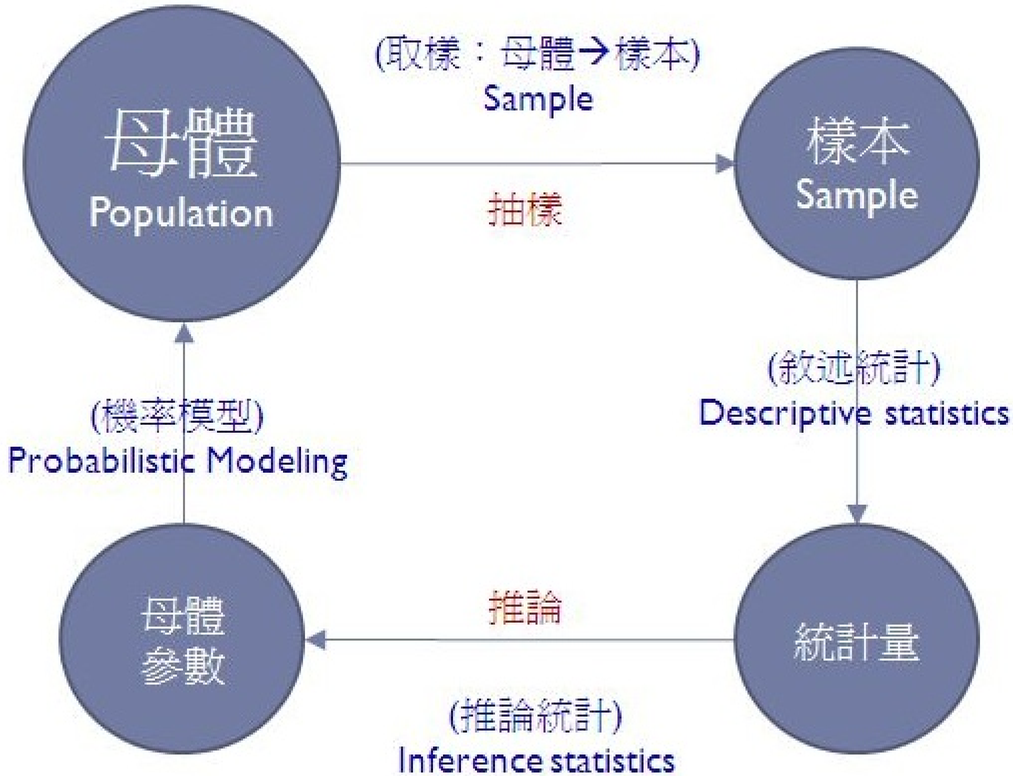
但是當我們不知道基本事件的發生機率時，我們該怎麼辦呢？此時統計的價值就顯現出來了，我們可以先進行很多次的實驗，以便透過計算的方式算出某個「事件」的機率，這種統計方法就稱為「敘述統計」。接著我們可以根據這個統計的結果，推論出某些衍生事件的機率，或者推算出此一事件是否「可信」，這樣的統計就稱為「推論統計」。

如果用更數學化的語言來說，我們可以透過已知的「母體模型」，以及某些「母體參數」計算某個事件的機率，或者用電腦隨機產生這些事件。這種用參數與模型產生隨機事件的過程，可以用電腦的方是透過過程

式模擬產生，此時電腦其實是以機率的角度在模擬母體的運作方式，這樣的電腦模擬方法稱為「蒙地卡羅法」。

但是如果我們是在是先取得一群樣本之後，開始計算樣本的某個統計量是多少，這種計算就稱為「敘述統計」。如果我們進一步透過「敘述統計」的「統計量」去推估「母體的某個參數值應該是多少？」，這樣的推估方法就稱為「推論統計」。

下圖顯示了「母體、參數、樣本與統計量」之間的關係。



機率與統計之關係

如果我們精確知道母體的機率模型與參數，我們就完全掌握了母體的機率分布，這就是從機率角度的看

法。但是在很多情況下我們無法清楚的知道母體的模型與參數，此時我們就可以採用「抽樣的方式」，從母體中取得或觀察到某些「樣本」，再透過這些樣本去計算出某些「敘述統計量」，接著再用這些統計量去推估母體的參數。

機率理論

舉例而言，假如我們已經某個銅版是公平的，也就是兩面的機率都是 1/2 ，那麼我們就可以直接透過「機率法則」計算某個序列，例如連續五次都投出正面的機率為 $\frac{1}{2^5} = \frac{1}{32}$ 。

但是如果我們不知到銅版「正反面的機率」，那麼我們就必須改用「統計的方法」，例如連續投擲該銅版一千次，然後計算「正面與反面各為多少次」，接著再透過這些「正反面次數的統計量」，去推估某事件的出現機率。

假如我們投擲該銅版一千次的結果，發現正面出現 508 次，反面出現 492 次，那麼我們就可以推估「正面的機率為 0.508」，而「反面出現的機率為 0.492」，接著再去推估連續出現五次正面的機率為 $0.508^5 = 0.03383$ 。

當然，機率的模型並不是都像擲骰子或銅版那樣簡單的，有時機率模型本身就有點困難。舉例而言，布瓦松 (Poisson) 分布是用來描述「一段連續時間內」某個隨機事件發生的次數，其離散機率密度函數如下所示。

$$pois(\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

如果沒有學過機率的話，相信一般人很難看懂該機率分布的意義。因此在本書的後序章節中，我們將會先介紹機率的法則與模型，以便讓讀者能先對這些機率模型所代表的意義有清楚的認識，然後再進入「敘述統計」與「推論統計」的數學世界，希望透過這樣的方法，讀者能更清楚的理解整個機率統計的數學理論，並且能用 R 程式的實作來印證這些理論。

統計理論

同樣的，統計理論也不只是算算出現次數這麼簡單的。舉例而言，當我們想知道某一組統計量是否「合理」時，我們會採用「信賴區間」的方式描述該組統計量的合理的母體參數範圍，以下是一個範例。

請計算出以下列樣本序列的「平均值之 95% 信賴區間」：

```
3.6146570 4.1440593 2.5726955 5.2325581 2.0635500
2.6294660 2.8541827 2.4816312 1.5836851 3.2193062
2.8205306 3.5037204 2.6107131 4.1870588 2.4506509
2.4849244 4.5343839 0.7606934 3.5219675 1.7019120
```

這樣的計算顯然不是簡單的「計算出現次數」而已。要能進行「信賴區間」的計算，顯然我們必須學習更多的數學理論，才能知道如何計算，也才能清楚的掌握計算結果的意義。

應用

機率統計的應用涵蓋面非常的廣，從社會科學到自然科學都會用到，這是一門有著極強實用性的數學，很少數學像機率統計一樣有著如此強大的實用性。

在社會科學當中，我們會用機率統計來檢驗某個抽樣調查是否可信，某個抽樣調查顯示了何種意義等等？甚至像是社會科學領域的經典，塗爾幹的自殺論當中，即是採用機率統計的方法檢驗哪些因子會造成自殺現象的增加或減少等等，這些都是機率統計在社會科學上很明顯的應用案例。

在自然科學當中，學習生物或醫學的研究者也會透過機率統計來計算並研究某個藥物是否對特定疾病具有療效，或者某個檢測結果是否顯示該病人已經得到某種疾病。而學習電腦的程式設計者則可以透過機率統計模型進行「蒙地卡羅式的隨機模擬」，以便計算某個現象的機率。或者透過像「貝氏網路」這樣的機率模型以進行事件的機率計算，甚至是透過像「隱馬可夫鏈算法」(Hidden Markov Model) 或 EM 學習法 (Expectation-Maximization Algorithm) 等方法來學習某個機率模型與參數，以便讓程式能根據輸入樣本得到預測某些事件的能力。這些都是機率統計在自然科學上典型的應用案例。

R 軟體實作：簡介與基本操作

簡介

R 軟體是專門為了機率統計而設計的一種開放原始碼軟體，是免費的自由軟體。

市面上有許多與 R 類似的商用軟體，像是 SPSS, SAS, MINITAB, S-PLUS 等，但是這些軟體是要花錢買的。

R 軟體所使用的程式語言，被稱為 R 語言。

R 語言 與 S-PLUS 所使用的語言很類似，兩者都衍生自貝爾實驗室 Rick Becker, Allan Wilks, John Chambers 所創造的 S 語言，R 語言基本上是 GNU 所實作的 S 語言版本。

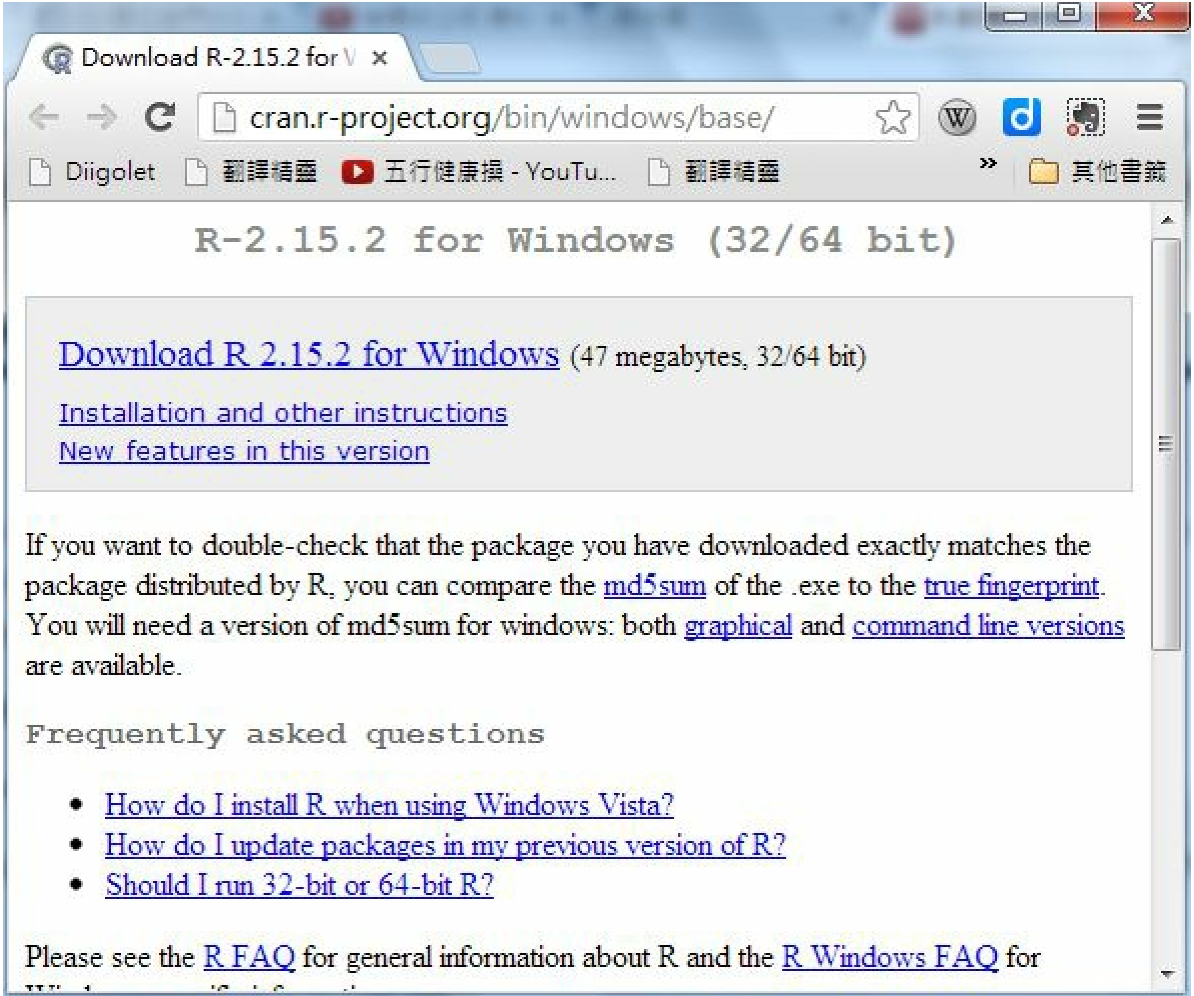
筆者篡寫此文時，R 所採用的 S 語言演化到了第四版，因此稱為 S4。

安裝

R 軟體的官方網站為 http://www.r-project.org/ ，其中有個相當重要的子網站稱為 CRAN (Comprehensive R Archive Network)，其網址為 http://cran.r-project.org/ ，您可以從這個網站中下載 R 軟體。

舉例而言，筆者使用的是 Windows作業系統，因此可以從以下網址下載到最新版的 R 軟體。

- http://cran.r-project.org/bin/windows/base/


Download R-2.15.2 for W
cran.r-project.org/bin/windows/base/
Diigolet
翻譯精靈
五行健康操 - YouTu...
翻譯精靈
其他書籤
R-2.15.2 for Windows (32/64 bit)
Download R 2.15.2 for Windows (47 megabytes, 32/64 bit)
Installation and other instructions
New features in this version
If you want to double-check that the package you have downloaded exactly matches the package distributed by R, you can compare the md5sum of the .exe to the true fingerprint. You will need a version of md5sum for windows: both graphical and command line versions are available.
Frequently asked questions
How do I install R when using Windows Vista?
How do I update packages in my previous version of R?
Should I run 32-bit or 64-bit R?
Please see the R FAQ for general information about R and the R Windows FAQ for

R 軟體下載頁面

舉例而言，筆者點選時為 Download R 2.15.2 for Windows 這個連結， 這會下載位於下列網址的檔案：

- http://cran.r-project.org/bin/windows/base/R-2.15.2-win.exe

下載完畢後，請啟動該安裝檔，然後不斷按「下一步」就可以完成安裝了，過程非常簡單。

以下網址中的 Youtube 影片介紹了 R 軟體的下載、安裝、套件、網站、電子書等等，有興趣的朋友可以看看。

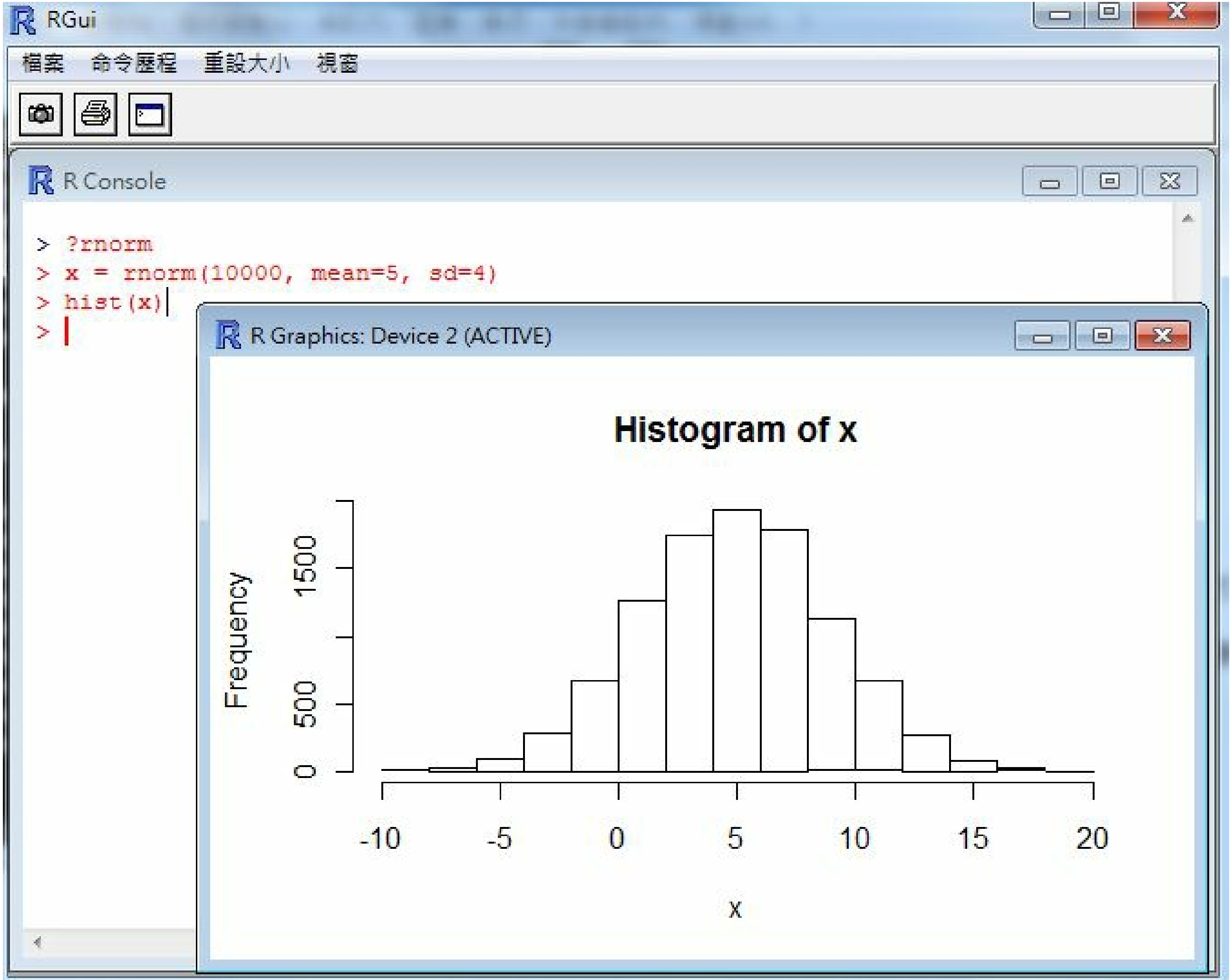
- http://www.youtube.com/watch?v=AipnE4s8sKk

基本操作

為了說明 R 軟體的用法，並用以學習機率統計的概念，本系列文章將運用 R 來說明機率統計的理論，讓程式人可以透過實作學會機率統計，並且學會 R 軟體中的 S 語言。

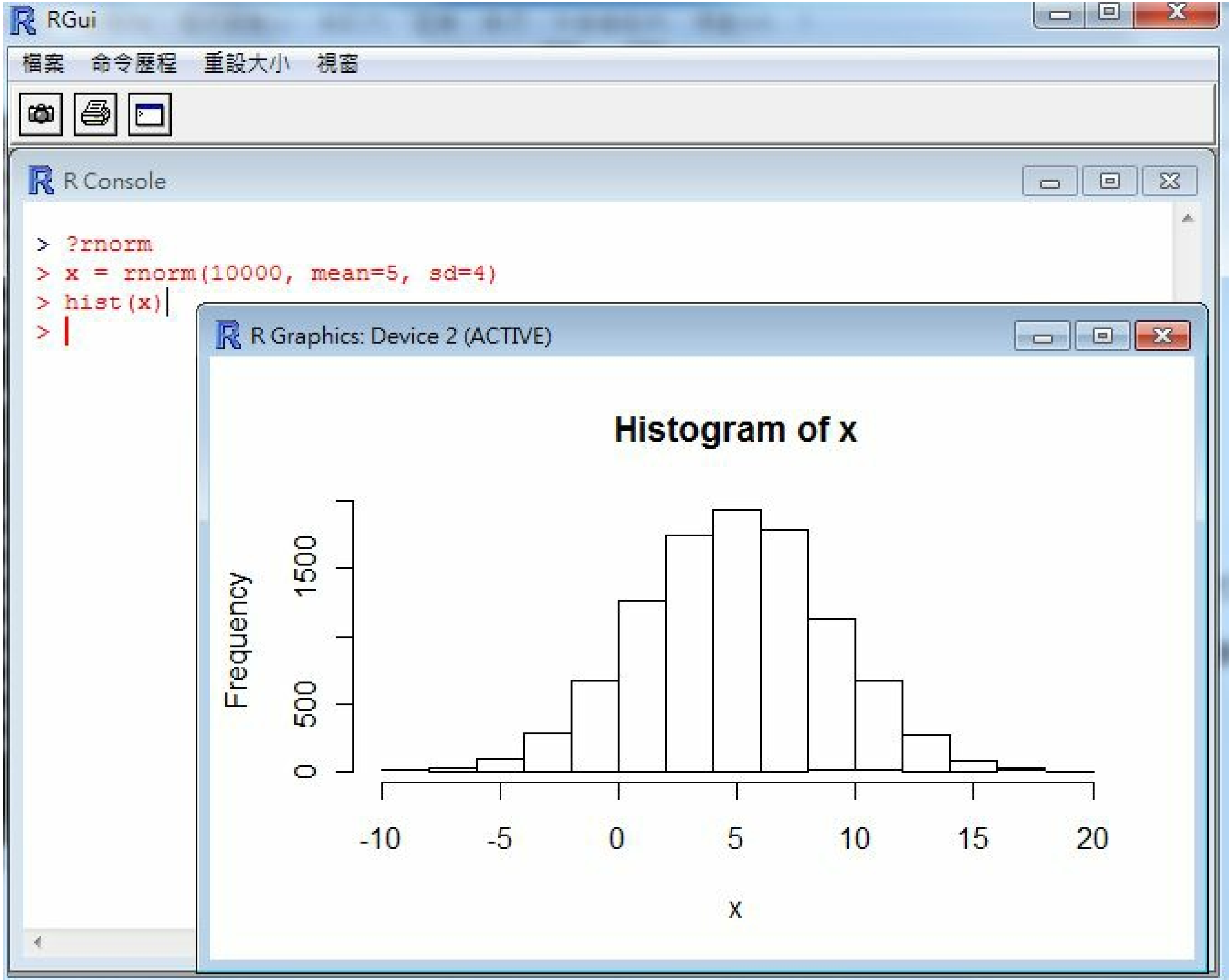
為了避免太過枯燥，我們將不會先介紹 R 的基本語法，而是先用一系列的操作，讓讀者體會 R 的能力，然後再慢慢回到語言的教學上面。

以下是筆者用 R 軟體取樣後會出樣本統計圖的畫面，簡單的幾個指令就可以得到統計結果，是不是很棒呢？


RGui
檔案 命令歷程 重設大小 視窗
R Console
> ?rnorm
> x = rnorm(10000, mean=5, sd=4)
> hist(x)
>
R Graphics: Device 2 (ACTIVE)
Histogram of x
Frequency
0
500
1500
-10
-5
0
5
10
15
20
x

圖、R 軟體執行畫面

第一個指令 ?rnorm 是要求 R 軟體查詢 rnorm 這個指令，R 軟體會顯示以下的說明網頁，您可以看到 rnorm 指令是與常態分部 (The Normal Distribution) 有關的。


RGui
檔案 命令歷程 重設大小 視窗
R Console
> ?rnorm
> x = rnorm(10000, mean=5, sd=4)
> hist(x)
>
R Graphics: Device 2 (ACTIVE)
Histogram of x
Frequency
0
500
1500
-10
-5
0
5
10
15
20
x

圖、R 軟體的說明網頁

在 R 軟體中，對於任何一個機率分布 xxxx，都會實作出以 d, p, q, r 為字首的四種函數，例如對於常態分布 Normal Distribution (簡寫為 norm) 而言，就有 dnorm, pnorm, qnorm, rnorm 等四個函數，功能分別如下所示：

函數	說明	語法
dnorm	常態分布的機密度函數	dnorm(x, mean = 0, sd = 1, log = FALSE)
pnorm	常態分布的機分布函數	pnorm(q, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
qnorm	常態分布的分位數函數	qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
rnorm	常態分布隨機樣本函數	rnorm(n, mean = 0, sd = 1)

上表中的 mean 代表平均數，sd 代表 Standard Deviation (標準差)，n 是隨機產生的樣本個數，x 是隨機變數值，q 是累積值，p 是機率值，n 則是產生的樣本數。

您可以發現函數中，有些參數後面有 = 的指定 (像是 mean=0, sd=1, log=FALSE, ...)，有些卻沒有 (像是 x, q, p, n) 等，這些指定代表預設值，也就是如果您不指定這些參數的值，那麼將會自動代入預設值。

所以 rnorm(100) 代表 rnorm(100, mean = 0, sd = 1) 的意思，也就是該函數會產生平均數為 mean=0，標準差

為 sd=1 的隨機樣本共 100 個。

關於這些函數的更詳細的說明如下表所示。

字首	函數意義	範例	說明
d	機率密度函數	dnorm(1.96)	P(X=x)
p	累積機率函數 (CDF)	pnorm(1.96)=0.975	P(X≤x)
q	計算百分位數	qnorm(0.975)=1.96	q 系列為 p 系列的反函數; 所以 qnorm(pnorm(1.96)) = 1.96
r	抽樣函數	rnorm(100)	傳回 100 個標準常態分布的樣本向量

看懂這些函數之後，讓我們再度列出上圖的操作指令，仔細觀察看看每一個指令的意義。

```
?rnorm
```

```
x = rnorm(10000, mean=5, sd=4)


hist(x)
```

指令 x = rnorm(10000, mean=5, sd=4) 代表我們要用平均值為 5, 標準差為 4 的常態分布，隨機產生 10000 個樣本，然後將這些樣本存到 x 陣列當中。

指令 hist(x) 代表要用這些樣本畫出統計的直方圖 (Histogram)，於是就畫出了圖中的那個長條狀圖形。

現在、請讀者試著看看下列操作，看看您是否能夠讀懂這些操作的意義。

```
rnorm(10, 3, 2)
> x
 [1] 2.5810213 0.5399127 5.0005020 5.3402693 2.7900723 3.9638088 5.2119685
 [8] 2.2209882 2.9935943 7.0308419
> a=dnorm(1.96)
> a
[1] 0.05844094
> b=pnorm(1.96)
> b
[1] 0.9750021
> c=qnorm(b)
```

```
> c
[1] 1.96
> d=rnorm(10)
> d
 [1] -0.32913677  0.77788306 -1.80862496  0.16694598 -0.65656254 -1.76305925
 [7]  1.18237502  0.19651748 -0.07898685  0.73970933
>
```

參考文獻

* 維基百科:機率論
* Wikipedia:Probability Theory
* 機率密度函數 PDF (連續) -- http://en.wikipedia.org/wiki/Probability_density_function
* 機率質量函數 PMF (離散) -- http://en.wikipedia.org/wiki/Probability_mass_function
* http://en.wikipedia.org/wiki/Statistics
* http://en.wikipedia.org/wiki/Descriptive_statistics
* http://en.wikipedia.org/wiki/Inferential_statistics
* http://en.wikipedia.org/wiki/Bayesian_Inference
* http://en.wikipedia.org/wiki/Correlation
* http://en.wikipedia.org/wiki/Analysis_of_variance
* http://en.wikipedia.org/wiki/Design_of_experiments

- http://en.wikipedia.org/wiki/Regression_analysis
- http://en.wikipedia.org/wiki/Student%27s_t-test

機率的概念

樣本空間

機率論中，樣本空間是一個實驗或隨機試驗所有可能結果的集合，而隨機試驗中的每個可能結果稱為樣本點。通常用 S、Ω 或 U 表示。例如，如果拋擲一枚硬幣，那麼樣本空間就是集合{正面，反面}。如果投擲一個骰子，那麼樣本空間就是 {1,2,3,4,5,6} 。

事件

一個事件是由樣本空間中的一個子集合，例如令 A 為骰子的點數為 {1, 3, 5} 的事件，其機率可以寫為 P(A) = P({1,3,5})。

機率的詮釋方法

某個事件的發生率，機率很低代表該事件不太可能出現 (很罕見，但是並非不會出現)，機率很高代表該事件非常可能發生。

機率的詮釋方式可以分為下列三種方式。

詮釋 1. 個人方式：(Personal Approach) : 完全按照個人直覺的解釋方式 (不客觀)。

詮釋 2. 相對頻率方式 (Relative Frequency Approach) ： $P[A]=\frac{f}{n}$

- 說明：f 為實驗中事件 A 出現的次數，n 為實驗進行的次數。 此方法乃是基於實驗觀察的結果的方式。

詮釋 3. 古典方式 (Classical Approach) ： $P[A]=\frac{n(A)}{n(S)}$

- 說明： n(A) 為事件 A 可能出現的次數 N(S) 為實驗可能進行的次數。此方法乃是將將實驗的可能出像 (outcome) 假設為等可能發生 (equaly likely)。

機 率 公 理

一般人學習數學的時候都是從直覺概念開始的，例如我們小學的時候透過算幾個蘋果學到加法，然後用好幾排的蘋果學到乘法，接著就會背誦九九乘法表，然後在中學的時候導入變數的觀念，於是學會了聯立方程式的解法。

但是數學家們看數學往往是從公理系統開始的，透過公理系統進行推論以建立定理，然後推論出整個數學體系。讓我們學習一下數學家的想法，先來看看機率的公理系統有何特色。

以下三條法則是機率的基本公理：

公理 (1). $P(S)=1$

公理 (2). $P(A)\geq 0$

公理 (3). $P(A1\cup A2)=P(A1)+P(A2)$; $if\ \ A1\cap A2=\emptyset$

公理 (1) 中的 S 代表機率的樣本空間，也就是所有可能發生的事件所形成的集合，這個集合的發生機率為 1，意義是沒有任何事件落在樣本空間之外。

公理 (2) 中的 A 代表任一事件，而 $P(A)\geq 0$ 則代表任何事件的發生機率必須是正的，沒有負的機率值。

公理 (3) 中的 A1, A2 代表任兩個事件，如果 A1 與 A2 沒有交集，那麼其聯集發生的機率將會是其機率的總和，也就是 $P(A1\cup A2)=P(A1)+P(A2)$ 。

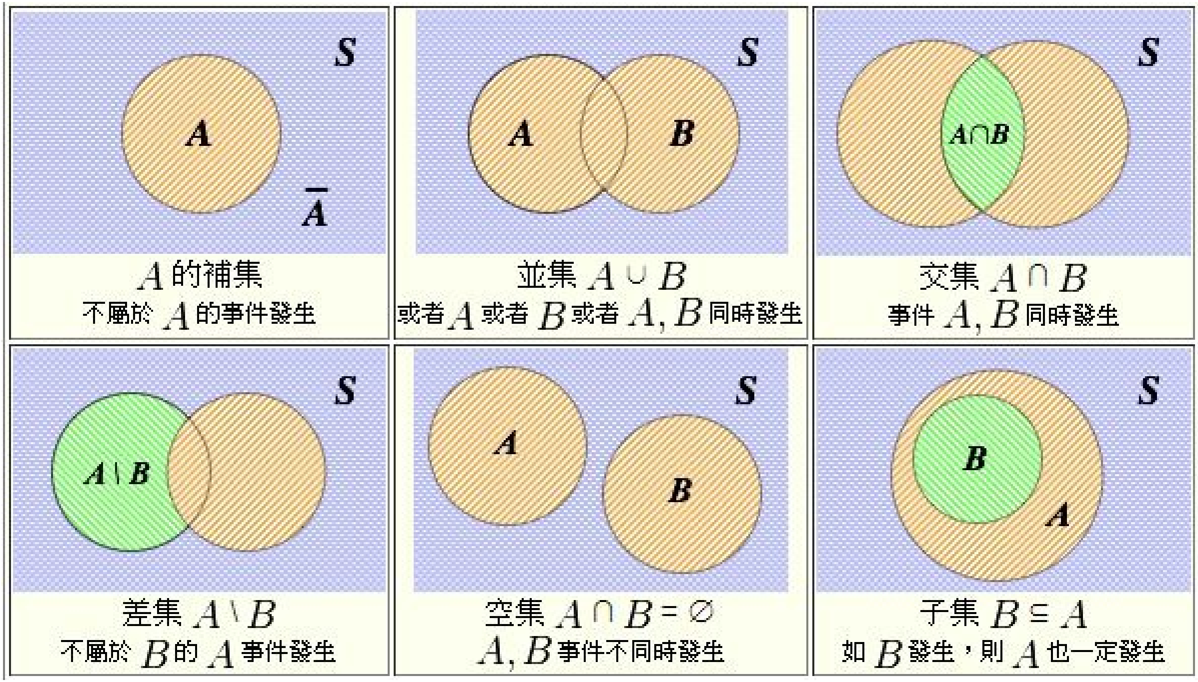
當這個公理系統確定下來之後，我們就可以透過這些法則進行一些基本的推論，舉例而言，我們應該可以很容易的證明以下這些定理。

定理 1. $P(\emptyset)=0$

定理 2. $P(A')=1-P(A)$

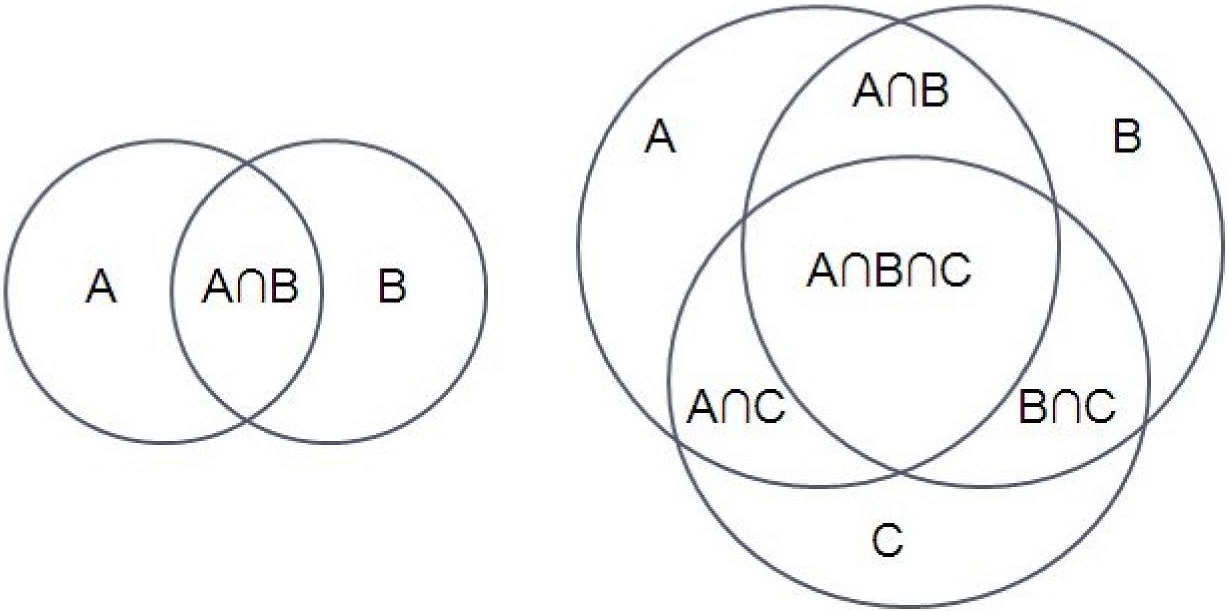
定理 3. $P(A1 \cup A2) = P(A1) + P(A2) - P(A1 \cap A2)$

但是，這些定理又代表甚麼意義呢？其實從下列凡氏圖上可以很清楚的看得出來這些定理的直覺意義。



圖、兩個集合 A, B 的凡氏圖

基本上，機率系統是建構在集合論之上的一門數學系統，所以我們可以用集合論的凡氏圖來理解這些公理與定理的意義。下圖左方是兩個集合 A,B 所形成的凡氏圖，而右方則是三個集合 A, B, C 所形成的凡氏圖。



圖、三個集合 A, B, C 的凡氏圖

習題：機率定理的證明

習題 1 :

定理：證明 $P(\emptyset)=0$

證明：

$P(S\cup\emptyset)=P(S)+P(\emptyset)$; 根據公理 (3)

$S=S\cup\emptyset$; 根據集合論

$P(S)=1$; 根據公理 1

$1=P(S)=P(S\cup\emptyset)=P(S)+P(\emptyset)$; 根據集合論與公理 (3)

所以 $P(\emptyset)=1-P(S)=1-1=0$

習題 2 :

定理：證明 P(A') = 1-P(A) ; 其中的 A' 代表 A 的補集，也就是 $A\cup A'=S; A\cap A'=\emptyset$

證明：

因為 $A \cup A' = S; A \cap A' = \emptyset$; 根據 A' 的定義

$P(A' \cup A) = P(A') + P(A)$; 根據公理 3

$P(A') + P(A) = P(A' \cup A) = P(S) = 1$; 根據公理 3 與公理1

所以 $P(A') = 1 - P(A)$

習題 3 :

定理：證明 $P(A1 \cup A2) = P(A1) + P(A2) - P(A1 \cap A2)$

證明：

因為 $A1 \cup A2 = (A1 - (A1 \cap A2)) \cup A2$; 根據集合論 (用文氏圖可以理解其直覺意義)

$(A1 - (A1 \cap A2)) \cap A2 = \emptyset$; 根據集合論 (用文氏圖理解，只是為了方便)

$P(A1) = P((A1 - (A1 \cap A2)) \cup (A1 \cap A2)) = P(A1 - (A1 \cap A2)) + P(A1 \cap A2)$
; 根據公理 3

所以 $P(A1 - (A1 \cap A2)) = P(A1) - P(A1 \cap A2)$

推論

$$P(A1 \cup A2) = P((A1-(A1 \cap A2)) \cup A2) = P(A1-(A1 \cap A2)) + P(A2)$$

$$= P(A1) - P(A1 \cap A2) + P(A2) = P(A1) + P(A2) - P(A1 \cap A2);$$

所以 $P(A1 \cup A2) = P(A1) + P(A2) - P(A1 \cap A2)$

機率模型

因此、只要指定了所有可能事件的發生率，我們就可以完整的描述一個機率模型，舉例而言，日常生活中最常見的機率模型，大概就是丟銅板和擲骰子了， 以下是我們對這兩個機率系統的描述。

範例 1：丟銅板

在投擲銅板的機率過程中，其樣本空間 S={正, 反} ，

而其中一個常見的隨機變數 X ，是用來計算銅板的正面數量，

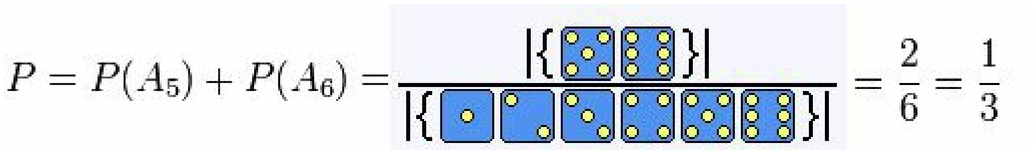
此時，P(正) = 0.5，而 P(反) = 0.5

範例 2：擲骰子

在投擲骰子的機率過程中，其樣本空間 S={1點,2點,3點,4點,5點,6點}，

此時，P(1點) = P(2點) = ... = P(6點) = 1/6。

所以，在一次擲骰子中，得到 5 點或者 6 點的機率，可以圖示如下。



圖、擲一次骰子 5 點或 6 點的機率

練習：R 軟體與機率密度函數

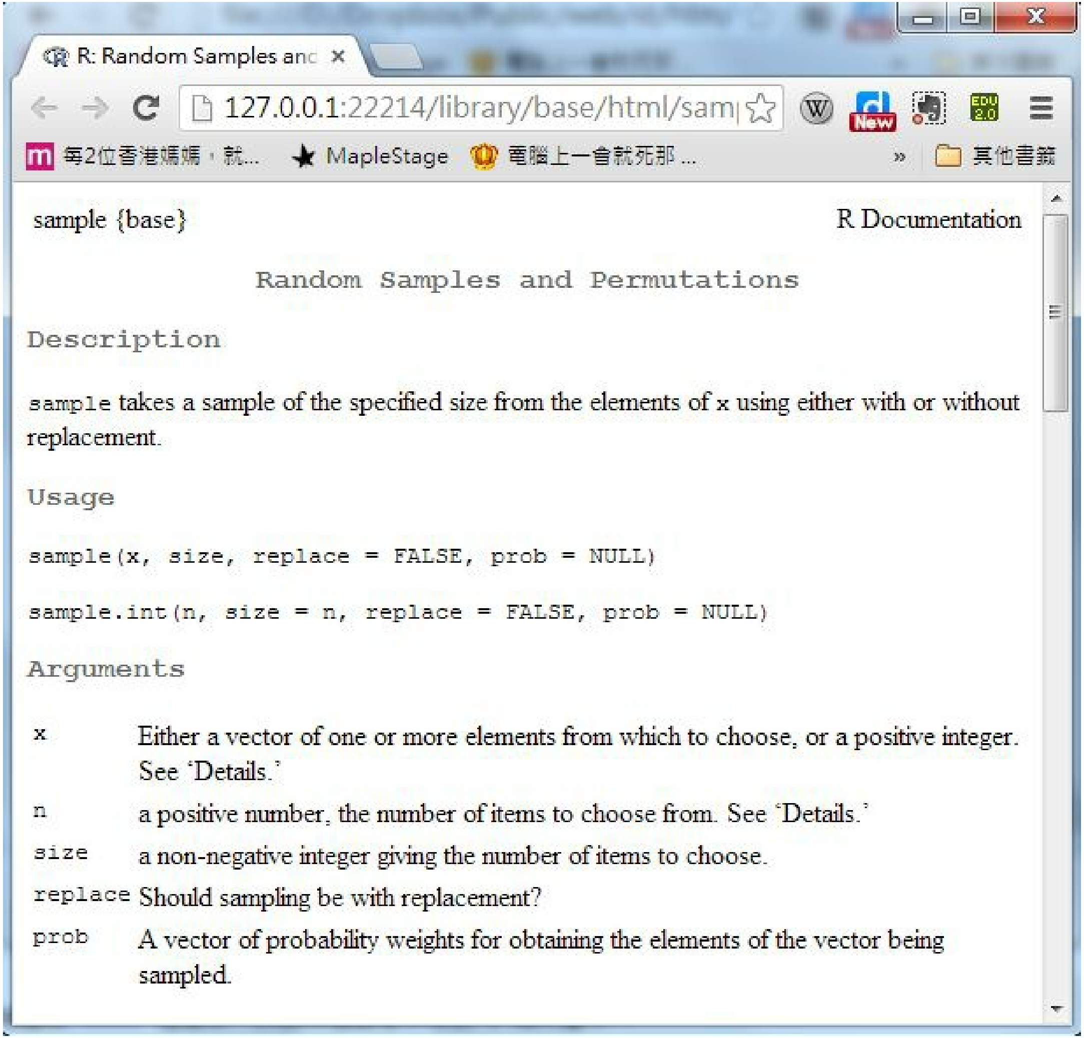
我們可以透過 R 軟體進一步瞭解機率密度函數的意義，舉例而言， R 當中有個 sample() 函數，我們只要使用該函數就可以模擬擲骰子或銅板的過程。

您可以用 「?函數」 的方式查詢某函數的功能，因此當我們在 R 軟體中鍵入 ?sample 時， R 軟體會輸出下列訊息：

```
> ?sample
```

```
starting httpd help server ... done
```

然後就開啟下列的網頁畫面



sample {base} R Documentation

Random Samples and Permutations

Description

`sample` takes a sample of the specified size from the elements of `x` using either with or without replacement.

Usage

```
sample(x, size, replace = FALSE, prob = NULL)

sample.int(n, size = n, replace = FALSE, prob = NULL)
```

Arguments

`x`	Either a vector of one or more elements from which to choose, or a positive integer. See 'Details.'
`n`	a positive number, the number of items to choose from. See 'Details.'
`size`	a non-negative integer giving the number of items to choose.
`replace`	Should sampling be with replacement?
`prob`	A vector of probability weights for obtaining the elements of the vector being sampled.

圖、sample 函數的 Help 畫面

您可以看到 sample 函數的原型為 sample(x, size, replace = FALSE, prob = NULL)，於是可以用下列指令模擬投擲骰子十次的行為。

```
> sample(1:6, 10)
錯誤在sample(1:6, 10) :
  cannot take a sample larger than the population when 'replace = FALSE'
> sample(1:6, 10, replace=TRUE)
 [1] 3 2 4 4 4 2 6 3 3 3
>
```

您可以看到當我們用 sample(1:6, 10, replace=TRUE) 的指令時，可以正確的模擬出投擲骰子十次的隨機過程，得到 3 2 4 4 4 2 6 3 3 3 這個序列， 但是若我們沒有指定 replace=T (TRUE)，的時候，代表要採用取樣後不放回的方式，但是這種方式最多只能做六次，所以就得到失敗的結果。

不過如果我們指定的樣本數 k 在六個以下，那麼 sample(1:6, k) 是會成功的，以下是一個範例：

```
> sample(1:6, 6)
[1] 2 6 4 1 5 3
```

同樣的，我們也可以用 sample 函數模擬投擲銅板的過程，只是由於同板只有兩個面 (正面與反面)，因此我們可以用以下的方式模擬：

```
> face = c("正", "反")
> sample(face, 10, replace=TRUE)
 [1] "正" "反" "正" "反" "反" "正" "反" "正" "正" "反"
```

上述模擬中的第一個指令 face = c("正", "反")，代表我們要建立一個有兩個字串元素 [正, 反] 的陣列。然後第二個指令 sample(face, 10, replace=TRUE) 是用這樣的陣列去產生 10 個樣本 (取後放回的方式)。

有時候，我們希望模擬的事物，其機率並非平均的，舉例而言，像是灌過鉛的骰子，或者是像台灣的廟裏面常見的「擲茭」， 其機率可能是不平均的，對這種情況我們就可以指定 sample(x, size, replace = FALSE, prob = NULL) 這個函數的第四個參數， 也就是 prob 來模擬。

舉例而言，假如「擲茭」的正面機率是 0.6，而反面機率是 0.4，那麼我們就可以用下列方式模擬「擲茭」十次的過程。

```
> sample(face, 10, replace=TRUE, c(0.6, 0.4))
 [1] "反" "正" "反" "反" "反" "正" "正" "正" "正" "正"
```

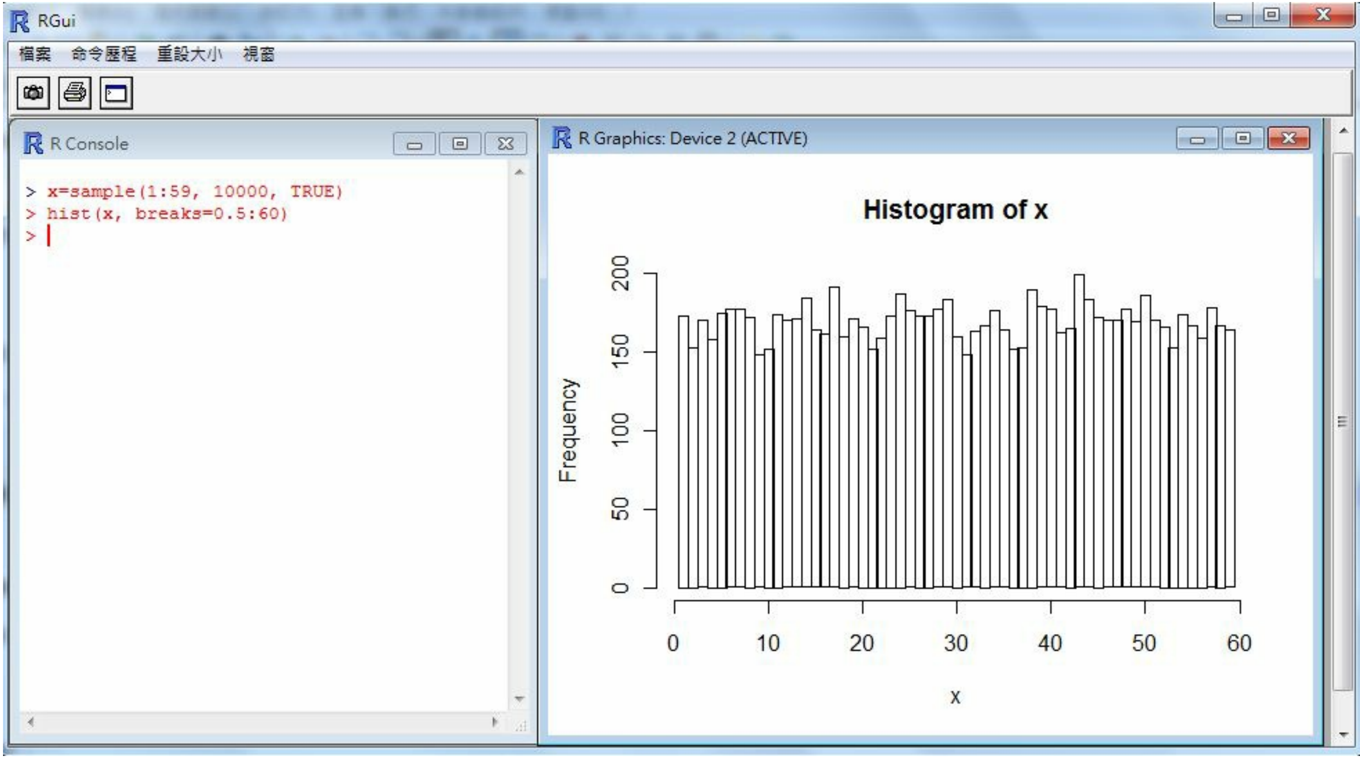
習 題

習題 1 : 請模擬從班上隨機抽學生一萬次，看看誰與你最有緣 (抽到次數最多)

解答：假設班上有 59 人，那麼編為 1 到 59 號，於是我們可以用下列程式，進行 1 萬次抽樣， 並繪出統計圖。

```
> x=sample(1:59, 10000, TRUE)
> hist(x, breaks=0.5:60)
```

執行結果



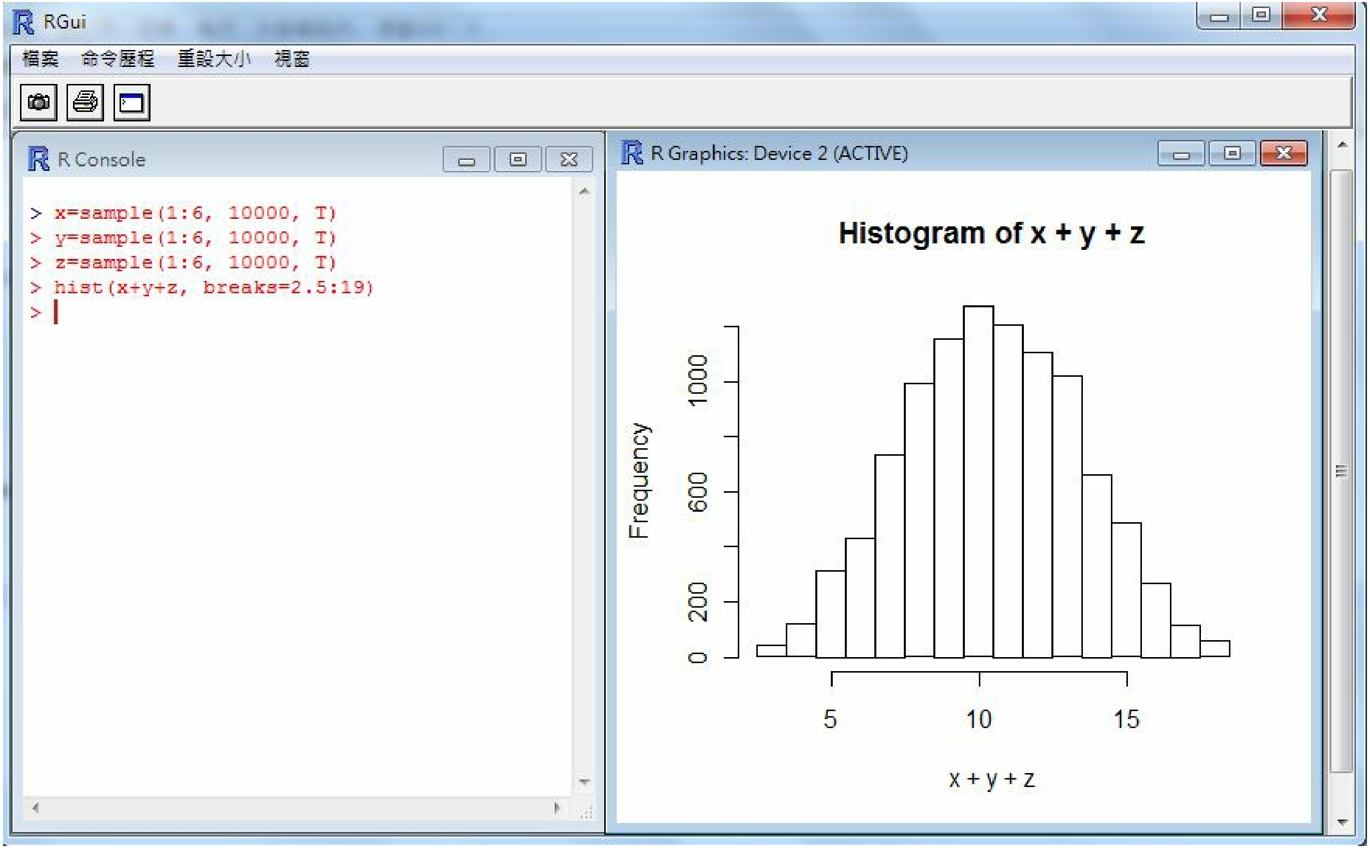
圖、最有緣的同學統計圖

習題 2 : 感受中央極限定理

```
> x=sample(1:6, 10000, T)
```

```
> y=sample(1:6, 10000, T)
> z=sample(1:6, 10000, T)
> hist(x, breaks=0.5:7)
> hist(y, breaks=0.5:7)
> hist(z, breaks=0.5:7)
> hist(x+y, breaks=1.5:13)
> hist(x+y+z, breaks=2.5:19)
```

執行結果


RGui
檔案 命令歷程 重設大小 視窗
R Console
> x=sample(1:6, 10000, T)
> y=sample(1:6, 10000, T)
> z=sample(1:6, 10000, T)
> hist(x+y+z, breaks=2.5:19)
>
R Graphics: Device 2 (ACTIVE)
Histogram of x + y + z
Frequency
0
200
600
1000
5
10
15
x + y + z

圖、x+y+z 的分布圖

條件機率

條件機率的定義：

在 A 事件出現的情況下，B 事件出現的機率，稱為 P(B|A)。

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

範例 1：

舉例而言，假如我們已知某投擲骰子的結果為偶數 (事件 A=偶數)，那麼結果為 3 點 (事件 B=3點) 的機率為多少？

這個條件機率可以用下列算式表示。

P(B|A) = P(3點|偶數)

範例 2：

當然、B 不一定要是 A 的子集合，舉例而言，假如 B 為「不大於 3 點」的事件，那麼我們就可以將條件機率表示如下：

P(B|A) = P(不大於3點|偶數)

獨立事件

獨立事件的定義：

事件 A 與 B 彼此獨立，則 A, B 兩事件同時出現的機率為

$$P(A \cap B) = P(A)P(B)$$

請注意數學中定義的意義，定義代表某種規定，是不需要證明的，只要不符合這種規定的，就不能用此一名詞描述，也就是不符合此定義。

因此、並非所有的事件 A, B 都會是獨立的，但若事件 A, B 符合上述規定的話，我們就稱這兩個事件彼此獨立。

舉例而言，假如對於一個公平的骰子而言，請問下列的 A,B 事件之間是否彼此獨立。

範例 1. 兩事件不獨立的情況

問題：請問 「A=偶數, B=3點」這兩個事件是否獨立

解答：

$P(A) = 3/6 = 1/2$

$P(B) = 1/6$

$P(A \cap B) = 0$

$P(A)\ P(B) = 1/2 * 1/6 =1/12$

由於 P(A B) P(A) P(B) ，所以這兩個事件彼此不獨立。

範例 2. 兩事件獨立的情況

問題：請問 「A=偶數, B=不大於 4 點」這兩個事件是否獨立

解答：

P(A) = 3/6 = 1/2

P(B) = 4/6 = 2/3

P(A ∩ B) = P({2點, 4點}) = 2/6 = 1/3

P(A) P(B) = 1/2 * 2/3 = 1/3

由於 P(A ∩ B) = P(A) P(B) ，所以這兩個事件彼此獨立。

習題：請證明以下定理：

定理 1. 若 A , B 彼此獨立，則 $P(A|B) = P(B|A) * \frac{P(A)}{P(B)}$

定理 2. $A_1, A_2, \ldots, A_k$ 彼此獨立 <=> $P(A_1 \cap A_2 \ldots \cap A_k) = P(A_1) P(A_2) \ldots P(A_k)$

定理 3. 乘法規則： $P(A \cap B) = P(B|A) P(A)$

習題 : 請舉出一組獨立事件的範例

貝氏定理

貝氏定理： $P(A|B)=P(B|A)\frac{P(A)}{P(B)}$

證明：

由條件機率的定義可得 $P(B|A)=\frac{P(A\cap B)}{P(A)}$ ，也可以得到 $P(A|B)=\frac{P(A\cap B)}{P(B)}$

所以 $P(A\cap B)=P(B|A)P(A)=P(A|B)P(B)$

於是得到 $P(A|B)=P(B|A)\frac{P(A)}{P(B)}$

習題：

習題 1. 請驗證 「A=偶數, B=3點」這兩個事件是否符合貝氏定理

習題 2. 請驗證 「A=偶數, B=不大於 4 點」這兩個事件是否符合貝氏定理

條件獨立

條件獨立的定義：

假如 A 與 B 在給定 C 的情況下條件獨立，那麼以下算式成立：

$$P(A,B|C)=P(A|C)*P(B|C);$$

習題：請證明以下定理：

定理： $$P(A,B|C)=P(C|A)*P(C|B)*\frac{P(A)P(B)}{P(C)^2}$$

習題：牙疼的診斷問題

本問題來自人工智慧的經典教科書 「Artificial Intelligence: A Modern Approach」第三版，475 頁。

問題描述：當病人來看牙醫時，該病人可能有蛀牙或沒蛀牙，也可能有牙痛或沒有牙痛，而牙醫可能會找到牙痛的原因或找不到。

因此有下列三個隨機變數

X:(蛀) 蛀牙與否 (Cavity) Y:(痛) 牙痛與否 (Toothache) Z:(找) 是否找到痛的牙 (Catch)

假如這個問題個統計機率都已經知道了，如下表所示。

	牙痛 (Y=1)		不牙痛 (Y=0)	
	找到 (Z=1)	找不到 (Z=0)	找到 (Z=1)	找不到 (Z=0)
蛀牙(X=1)	0.108	0.012	0.072	0.008
沒蛀牙 (X=0)	0.016	0.064	0.144	0.576

請回答下列問題

- 問題 1：請計算 P(沒痛) = ?
- 問題 2：請計算 P(找到 | 牙痛) = ?
- 問題 3：請問這是一個合理的機率分布嗎？
- 問題 4：請計算 P(找到 | 蛀牙) = ?
- 問題 5：請計算 P(找到, 牙痛) = ?
- 問題 6：請計算 P(蛀 | 找到), P(蛀), P(找到), P(找到 | 蛀) ，然後驗證下列貝氏定理是否成立。

    - P(找到|蛀) = P(蛀|找到) P(找到)/P(蛀)

解答

R 的 陣列是用以行為主的順序 (Column Major Order)，請看下列檔案中的說明：

- http://cran.r-project.org/doc/manuals/R-lang.pdf

> 2.2.2 The dim attribute is used to implement arrays. The content of the array is stored in a vector in column-major order and the dim attribute is a vector of integers specifying the respective extents of the array. R ensures that the length of the vector is the product of the lengths of the dimensions. The length of one or more dimensions may be zero.

所以我們必須用以行為主的順序 (Column Major Order) 將機率列舉出來，如下表所示：

蛀 X	痛 Y	找 Z	P(X,Y,Z)
0	0	0	0.576
1	0	0	0.008

0	1	0	0.064
1	1	0	0.012
0	0	1	0.144
1	0	1	0.072
0	1	1	0.016
1	1	1	0.108

而且 R 的陣列是從 1 開始算的，不像 C 語言是從 0 開始算的，因此還必須將上表修改如下：

蛀 X	痛 Y	找 Z	P(X,Y,Z)
1	1	1	0.576
2	1	1	0.008
1	2	1	0.064
2	2	1	0.012

1	1	2	0.144
2	1	2	0.072
1	2	2	0.016
2	2	2	0.108

```
> p <- array(c(0.576, 0.008, 0.064, 0.012, 0.144, 0.072, 0.016, 0.108),c(2,2,2))
> p
, , 1

      [,1]  [,2]
[1,] 0.576 0.064
[2,] 0.008 0.012

, , 2

      [,1]  [,2]
[1,] 0.144 0.016
```

```
[2,] 0.072 0.108

> p[1,1,1]
[1] 0.576
> p[2,1,1]
[1] 0.008
> p[1,2,1]
[1] 0.064
> p[2,2,1]
[1] 0.012
> p[1,1,2]
[1] 0.144
> p[2,1,2]
[1] 0.072
> p[1,2,2]
[1] 0.016
> p[2,2,2]
[1] 0.108
> dimnames(p)[[1]] = c("沒蛀", "蛀")
> dimnames(p)[[2]] = c("沒痛", "痛")
> dimnames(p)[[3]] = c("沒找", "找")
```

```
> p
, , 沒找

     沒痛    痛
沒蛀 0.576 0.064
蛀   0.008 0.012

, , 找

     沒痛    痛
沒蛀 0.144 0.016
蛀   0.072 0.108
```

解答1：P(沒痛) = 0.8 計算過程：

```
> p[,"沒痛",]
     沒找    找
沒蛀 0.576 0.144
蛀   0.008 0.072
> sum(p[,"沒痛",])
[1] 0.8
```

解答2：P(找到 | 牙痛) = 0.62

```
> p[,,"找"]
      沒痛   痛
沒蛀 0.144 0.016
蛀   0.072 0.108
> sum(p[,,"找"])
[1] 0.34
> sum(p[,"痛","找"])
[1] 0.124
> sum(p[,"痛","找"])/sum(sum(p[,"痛",]))
[1] 0.62
```

解答3：請問這是一個合理的機率分布嗎？ (是的，因為總和為 1，而且每個機率直都介於 0 到1之間)

```
> sum(p)
[1] 1
> 0<=p & p <=1
, , 沒找

     沒痛   痛
```

```
沒蛀 TRUE TRUE
蛀   TRUE TRUE

, , 找

     沒痛   痛
沒蛀 TRUE TRUE
蛀   TRUE TRUE
```

問題 4：請計算 P(找到 | 蛀牙) = ?

```
>  sum(p["蛀",,"找"])/sum(p["蛀",,])
[1] 0.9
```

問題 5：請計算 P(找到, 牙痛) = ?

```
> sum(p[,"痛","找"])
[1] 0.124
```

解答6：請計算 P(蛀 | 找到), P(蛀), P(找到), P(找到 | 蛀) ，然後驗證下列貝氏定理是否成立。

P(蛀 | 找到) = p(找到|蛀) * p(蛀)/p(找到)

說明：

P(蛀 | 找到) = 0.5294118, P(蛀)=0.2, P(找到)=0.34, P(找到 | 蛀)=0.9

P(蛀 | 找到) = 0.5294118 = 0.9 * 0.2/0.34 = = p(找到|蛀) * p(蛀)/p(找到)

```
> pab = sum(p["蛀","找"])/sum(p[,"找"]) # pab = P(蛀 | 找到)
> pba = sum(p["蛀","找"])/sum(p["蛀",]) # pba = P(找到 | 蛀)
> pa = sum(p["蛀",]) # pa = P(蛀)
> pb = sum(p[,"找"]) # pb = P(找到)
> pab
[1] 0.5294118
> pba
[1] 0.9
> pa
[1] 0.2
> pb
[1] 0.34
```

```
> pba*pa/pb
[1] 0.5294118
> pab-pba*pa/pb
[1] 0
```

所以

```
p(蛀|找)
= sum(p["蛀","找"])/sum(p[,,"找"])
= pab
= pba * pa / pb
= p(找|蛀) * p(蛀)/p(找)
= sum(p["蛀","找"])/sum(p[,,"蛀"])* sum(p[,,"蛀"])/ sum(p["找",,])
```

完整的操作過程

```
> p <- array(c(0.576, 0.008, 0.064, 0.012, 0.144, 0.072, 0.016, 0.108),c(2,2,2))
> p
, , 1


     [,1]  [,2]
```

```
[1,] 0.576 0.064
[2,] 0.008 0.012

, , 2

      [,1]  [,2]
[1,] 0.144 0.016
[2,] 0.072 0.108

> p[1,1,1]
[1] 0.576
> p[2,1,1]
[1] 0.008
> p[1,2,1]
[1] 0.064
> p[2,2,1]
[1] 0.012
> p[1,1,2]
[1] 0.144
> p[2,1,2]
[1] 0.072
```

```
> p[1,2,2]
[1] 0.016
> p[2,2,2]
[1] 0.108
> dimnames(p)[[1]] = c("沒蛀", "蛀")
> dimnames(p)[[2]] = c("沒痛", "痛")
> dimnames(p)[[3]] = c("沒找", "找")
> p
, , 沒找

      沒痛    痛
沒蛀 0.576 0.064
蛀   0.008 0.012

, , 找

      沒痛    痛
沒蛀 0.144 0.016
蛀   0.072 0.108

> p[,"沒痛",]
```

```
      沒找    找
沒蛀 0.576 0.144
蛀   0.008 0.072
> p[,,"找"]
      沒痛    痛
沒蛀 0.144 0.016
蛀   0.072 0.108
> sum(p[,,"找"])
[1] 0.34
> sum(p[,"痛","找"])
[1] 0.124
> sum(p[,"痛","找"])/sum(sum(p[,"痛",]))
[1] 0.62
> sum(p)
[1] 1
> 0<=p & p <=1
, , 沒找


     沒痛   痛
沒蛀 TRUE TRUE
蛀   TRUE TRUE
```

```
, , 找

    沒痛  痛
沒蛀 TRUE TRUE
蛀   TRUE TRUE

> sum(p["蛀",,"找"])/sum(p["蛀",,])
[1] 0.9
> sum(p["蛀",,"找"])/sum(p["蛀",,])
[1] 0.9
> sum(p[,"痛","找"])/sum(p[,"痛",])
[1] 0.62
> sum(p[,"痛","找"])
[1] 0.124
> pab = sum(p["蛀",,"找"])/sum(p[,,"找"]) # pab = P(蛀 | 找到)
> pba = sum(p["蛀",,"找"])/sum(p["蛀",,]) # pba = P(找到 | 蛀)
> pa = sum(p["蛀",,]) # pa = P(蛀)
> pb = sum(p[,,"找"]) # pb = P(找到)
> pab
[1] 0.5294118
```

```
> pba
[1] 0.9
> pa
[1] 0.2
> pb
[1] 0.34
> pba*pa/pb
[1] 0.5294118
> pab-pba*pa/pb
[1] 0
>
```

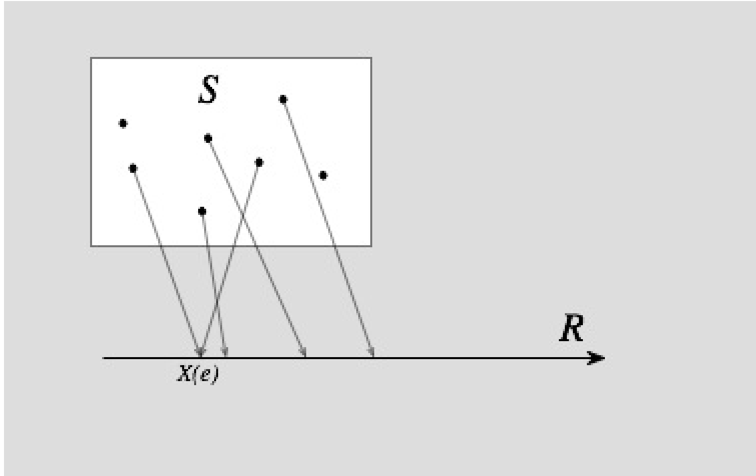
隨機變數

隨機變數簡介

如果 X 指定給機率空間 S 中每一個事件 e 一個實數 X(e)，同時針對每一個實數 r 都有一個事件集合 A_r 與其相對應，其中 $A_r = e{:}X(e) \le r$ ，那麼 X 被稱作隨機變數。

隨機變數是機率理論當中非常重要的一個概念，但是卻也非常容易被誤解，因為隨機變數其實是一種函數，而非只是簡單的變數，以下是機變數的定義。

> 隨機變數 **(Random Variable)**
>
> 定義：隨機變數是以樣本空間 S 為定義域的實數值函數，可以寫為 X(s)，其中 $s \in S, X(s) \in R$
>
> 換言之，隨機變數 X 是一個機率空間 (probability space) 中的函數， 可以寫為 $X(S) \to R$ ，該函數將 S 的某一子集合映射到實數領域 R。



圖、實數坐標軸上的隨機變數示意圖

舉例而言，投擲一個銅板時，可能出現正面或反面，此時的樣本空間 S = {正面、反面}。

假如這是一個公平的銅板，兩面的機率各為 1/2，那麼我們就可以寫為 P(正面) = 1/2， P(反面) = 1/2。

但是在這樣的描述當中，並沒有函數的概念，因此不符合隨機變數的定義。

如果我們用一個函數 X，代表銅板正面出現的次數，那麼 X 會將 {正面、反面} 映設到 {1,0} ，這樣的函數才符合隨機變數的定義。我們可以寫為 $X(S) \rightarrow R$ ，其中的樣本空間 S={正面、反面}，且 X(正面)=1，X(反面) = 0。

為何要這麼麻煩呢？為何我們不直接指定樣本空間中每一元素的機率就好了呢？

原因之一是，採用隨機變數概念的描述，才能將函數引入到機率模型中，這樣也才能更方便的描述一系列的隨機試驗。

舉例而言，假如我們投擲兩個銅版，出現正面的個數為一個隨機變數，假如這個隨機變數稱為 X2，那麼 X2 的定義域 (樣本空間) 就是 S2={正正、正反、反正、反反}，那麼隨機變數 X2 就會將 S2 空間中的元素映射到 {2, 1, 0} 這些實數值上，如下所示：

X2(正正) = 2
X2(正反) = 1
X2(反正) = 1
X2(反反) = 0

這樣我們就可以用「機率密度函數」來描述各個事件出現的機率，例如用 P[X2=2] 代表出現兩次正面的機率，P[X2=1] 代表出現一次正面的機率，而P[X2=0] 代表沒有出現正面的機率。

範例：

隨機擲兩個骰子，整個事件空間可以由 36 個元素組成：

$$S=\{(i,j)|i=1,...,6;j=1,...,6\}$$

這裡可以構成多個隨機變數，比如隨機變數 X (獲得的兩個骰子的點數和) 或者隨機變數 Y (獲得的兩個骰子的點數差)，隨機變數 X 可以有 11 個整數值，而隨機變數 Y 只有 6 個。

$X(i,j) := i + j, x = 2,3,...,12$；

$Y(i,j) := |i - j|, y = 0,1,2,3,4,5.$；

在此，我們引入了一個新的概念，稱為「機率密度函數」，讓我們更仔細的看看這個概念的意義。

習題 1：

問題： 請定義擲筊(博杯)的隨機變數

說明：

a. 有杯映射到 1，沒杯映射到 0
b. 樣本空間為 {正正、正反、反正、反反}
c. { 正反、反正} 稱為有杯

解答：

X({正正}) = 0

X({正反}) = 1

X({反正}) = 1

X({反反}) = 0

補充：假如博杯正面積率為 0.6，反面機率為 0.4，而且兩個杯之間互相獨立，那麼假如根據機率公理第三條，可以算出：

P({正正}) = P(正) * P(正) = 0.6*0.6 = 0.36

P({正反}) = P(正) * P(反) = 0.6*0.4 = 0.24

P({反正}) = P(反) * P(正) = 0.4*0.6 = 0.24

P({反反}) = P(反) * P(反) = 0.4*0.4 = 0.16

所以

P(X=1) = P({正反, 反正}) = P({正反}) + P({反正}) = 0.24 + 0.24 = 0.48

$P(X=0) = P(\{正正, 反反\}) = P(\{正正\}) + P(\{反反\}) = 0.36 + 0.16 = 0.52$

習題 2：

問題： 假如現在從你身上抽一滴血，請回答下列兩個問題。

1. 請定義一個隨機變數 X 代表那滴血中的白血球數量。

提示： 樣本空間 S = 此時此刻你身上的所有白血球 = {w1,w2,......,wn}

2. 請算出一滴血液中有三顆白血球的機率，假設該滴血液占你總血量的 1/1000。

解答 1：

$X(A) = |A|$

說明：

A 是一個事件，也就是白血球的樣本空間 S 的子集合，例如： A = {w1, w5, w9}

|A| 代表 A 集合的大小，也就是元素個數，舉例而言：

如果 A = {w1, w5, w9} ，那麼 |A| = 3

如果 B = {w2, w8}，那麼 |B| = 2

如果 C = {}，那麼 |C| = 0

如果 D = S，那麼 |D| = n

解答 2：

P(X=3) = P({A | X(A) = 3}) = P({{w1, w2, w3}) + P({w1, w2, w4}) + ...

假如任一顆白血球被抽到的機率等於該滴血液佔全身血液的比率，由於該滴血液佔總血量的 1/1000，所以給顆白血球被抽到的機率為 1/1000。

而且假設這些白血球沒有智慧，也不會聚合在一起，因此相互之間獨立，那麼由於每顆白血球被抽到的機率為 1/1000，因此 P(w1) = P(w2) = ... P(wn) = 1/1000。

那麼初步想法是 P(w1w3) = P(w1) * P(w3) = 1/1000 * 1/1000 。

但是上述的想法有個小問題，那就是該情況代表其它白血球都沒被抽到，因此所謂的 P(w1w3) 真正的意思應該是

$$P(w_1\overline{w_2}w_3\overline{w_4}....\overline{w_n})=(\frac{1}{1000})^2(\frac{999}{1000})^{n-2}$$

所以 P(X=3) 應該算法如下：

$$P(X=3)=P(\{A|X(A)=3\})=(\frac{1}{1000})^3(\frac{999}{1000})^{n-3}\binom{n}{3}$$

推而廣之，P(X=k) 的機率之算法如下：

$$P(X=k)=P(\{A|X(A)=k\})=(\frac{1}{1000})^k(\frac{999}{1000})^{n-k}\binom{n}{k}$$

事實上，這個題目的機率分布就是下一章的二項分布，如下所示：

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

而且、當 n 趨近於無限大時，這個分布將會趨近於布瓦松分布，如下所示：

$$P(X=k)=\lambda^k e^{-\lambda}/k!$$

其中的 λ 之意義為，在單位時間 (或單位面積、體積) 內，事件的出現次數平均為 λ 次。

離散與連續

如果隨機變數 X 的取值是有限的或者是可數無窮盡的值，則稱 X 為離散隨機變數，如下所示：

$$X = \{x_1, x_2, x_3, \ldots,\}$$

如果 X 由全部實數或者由一部分區間組成，則稱 X 為連續隨機變數，連續隨機變數的值是不可數及無窮盡的。

$$X = \{x | a \le x \le b\}$$

例如：擲骰子和丟銅版，都是離散型的隨機變數。而常態分布或均等分布，則是連續型的隨機變數之分布。

機率密度函數 (Probabilistic Density Function)

機率密度函數 **(Probabilistic Density Function, PDF)**

定義：機率密度函數則是一個符合機率公理的的函數 P，當我們寫 P[X=x] 時，意味著 x 是一個特定實數，其機率定義如下：

$$P[X=x]=P(S_x)=P(\{s:X(s)=x\})=\sum_{s\in S_x}P(s)$$

其中的 S_x 乃是一個 S 的子集合，定義為 $S_x=\{s:X(s)=x\}$ 。

舉例而言，P[X=2] 代表 $P(\{s\in S:X(s)=2\})$ 的機率。

讓我們來看看更多的機率密度函數的範例。

範例 1：

在投擲銅板的機率過程中，其樣本空間 S={正, 反} ，

而其中一個常見的隨機變數 X ，是用來計算銅板的正面數量，

也就是 X(正) =1, X(反) = 0。

此時，P[X=1] = P({正}) = 0.5，而 P[X=0] = P({反}) = 0.5

範例 2：

在投擲兩個銅板的機率過程中，其樣本空間 S={正正, 正反, 反正, 反反} ，

而其中一個常見的隨機變數 X ，是用來計算銅板的正面數量，

也就是 X(正正) =2, X(正反) = X(反正) = 1, X(反反) = 0。

P[X=2] = P({正正}) = 0.25 P[X=1] = P({正反,反正}) = 0.5 P[X=0] = P({反反}) = 0.25

範例 3：

在投擲骰子的機率過程中，其樣本空間 S={1點,2點,3點,4點,5點,6點}，

而其中一個常見的隨機變數 X ，是用來計算點數的，

也就是 X(1點) =1, X(2點) = 2, ... X(6點) = 6。

此時，P[X=1] = P[X=2] = ... = P[X=6] = 1/6。

範例 4：

在投擲骰子的機率過程中，其樣本空間 S={1點,2點,3點,4點,5點,6點}，

而其中一個不常見的隨機變數 Y ，是用來辨認偶數點的，

也就是 Y(1點) =0, Y(2點) = 1, Y(3點) = 0, Y(4點) = 1, Y(5點) = 0, Y(6點) = 1 。

此時，P[Y=1] = P[Y=0]= 1/2。

累加分配函數 (Cumulative Distribution Function)

有了上述的「隨機變數」與「機率密度函數」之後，我們就可以很容易的定義「累加分配函數」這種在「實數值」上的概念了。

累加分配函數 **(Cumulative Distribution Function, CDF)**

定義：累加分配函數 F(x) 代表所有小於 x 的機率密度函數之累加值 $F(x) = P[X \le x]$

離散情況：$F(x_0) = \sum_{x \le x_0} P[X = x] = \sum_{x \le x_0} P(x)$

連續情況：$F(x_0) = \int_{-\infty}^{x_0} P[X = x] dx = \int_{-\infty}^{x_0} f(x) dx$

為了方便起見，我們經常會將 P[X=1] 簡寫成 P(1) 或 f(1)，P[X=x] 簡寫成 P(x) 或 f(x)。 所以上面公式中的 P(x) 是離散情況中機率密度函數 P[X=x] 的簡寫，而 f(x) 是連續情況中 P[X=x] 的簡寫。

隨機變數的代數運算

在單一個樣本空間 S 中，可以有很多不同的隨機變數 X, Y, ... ，因為將 S 映射到 R 的函數有很多，例如：

1. X(s) 其中 $s \in S, X(s) \in R$
2. Y(s) 其中 $s \in S, Y(s) \in R$
3. ...

在機率統計中，我們經常看到隨機變數可以像數值一樣進行 +, -, * 等運算。舉例而言，假如 X, Y 均為隨機變數，那麼 「X+Y」,「X-Y」,「X*Y」 等都是隨機變數。

但是在前文中，我們看到了隨機變數其實被定義為一個「實數值函數」 $X(S) \to R$ ，那麼這些 +, -, * 等運算就是在函數上所進行的運算，這些運算的函意到底是甚麼呢？

3X 的意義

隨機變數 3X 代表的是一個函數 Z=3X，其中 Z 函數對每一個元素 s 的映射值均為 X 的 3 倍，也就是：

Z(s) = 3 * X(s)

範例：

問題：令 X 為擲骰子點數的隨機變數，也就是 X(k點)=k (k=1..6)，那麼隨機變數 3X 代表的是 Z(k點)=3*X(k點)=3k 這個函數。

根據這樣的表示方法，如果 Z = 3X，那麼請計算下列機率值。

1. 請問 P[Z=3] = ?, (答案為1/6)
2. 請問 P[Z=1] = ?, (答案為0)
3. 請問 P[Z=18] = ?, (答案為1/6)
4. 請問 P[Z=5] = ?, (答案為0)

範例：

問題：令 X 為丟銅板所得正面次數的隨機變數，也就是 X(正)=1, X(反)=0，那麼隨機變數 Z=3X 代表

的是 Z(正)=3，Z(反)=0 這個函數。

X+Y 的意義

隨機變數 X+Y 代表的是一個函數 Z=X+Y，其中 Z 函數對每一個元素 s 的映射值均為 X + Y 的映射值總和，也就是：Z(s) = X(s)+Y(s)

範例：

令 X, Y 均為為擲骰子點數的隨機變數，也就是 X(k點)=Y(k點)=k (k=1..6)，那麼 X+Y 代表的是隨機變數 Z(k點)=2k 這個隨機變數。

範例：

問題：令 X 為擲骰子點數的隨機變數，Y 為丟銅板所得正面次數的隨機變數，那麼 X+Y 這個隨機變數代表甚麼意義呢？

解答：這兩個隨機變數的定義域不同，因此不能相加，但是若我們將定義域擴展為聯合分布，那麼就可以相加。

在這個範例中，X 與 Y 兩者的定義域 SX, SY 並不相同，因此必須用聯合隨機分布的概念，也就是同

時投擲一顆骰子與一個銅板，才能有效說明 X+Y 的意義。

對於定義域不同的兩個隨機變數而言，其樣本空間可用兩者的「笛卡兒」乘積代表，也就是 SX = {1點,...., 6點} , 而 SY = {正, 反}。

此時 X+Y 所在的樣本空間，必須解釋為 {1點,...., 6點} 與 {正, 反} 兩者的笛卡兒乘積，總共有 12 種可能，聯合分布的樣本空間 S 如下所示。

S = (S_X, S_Y) = { (1點, 正), (1點,反), (2點, 正), (2點,反), ...(6點, 正), (6點,反)}

因此，Z = X+Y 所代表的隨機變數，其實是一個 Z 函數，該函數將 (S_X, S_Y) 映射到實數 R 中，其中的 X 作用在 S_X 上，而 Y 作用在 S_Y 上，也就是： Z(s) = Z(x, y) = X(x)+Y(y)

所以，P(Z=2) 也可以寫成 P(X+Y = 2) ，也就是 P({(1點, 正), (2點,反)})，因此 P(Z=2) 的機率為 2/12 = 1/6。

X Y 的意義

隨機變數 X Y 代表的是一個函數 Z=X Y，其中 Z 函數對每一個元素 s 的映射值均為 X Y 的映射值乘積，也就是：

Z(s) = X(s) Y(s)

範例：

問題：令 X 為擲骰子點數的隨機變數，Y 為丟銅板所得正面次數的隨機變數，那麼 X Y 這個隨機變數代表甚麼意義呢？

解答：這兩個隨機變數的定義域不同，因此不能相加，但是若我們將定義域擴展為聯合分布，那麼就可以相加。

同上一個範例，X 與 Y 兩者的定義域 (S_X, S_Y) 並不相同，樣本空間仍然用其「笛卡兒」乘積代表。

S = (S_X, S_Y) = { (1點, 正), (1點,反), (2點, 正), (2點,反), ...(6點, 正), (6點,反)}

因此，Z = X Y 所代表的隨機變數，其實是一個 Z 函數，該函數將 (S_X, S_Y) 映射到實數 R 中，其中的 X 作用在 S_X 上，而 Y 作用在 S_Y 上。

所以，P(Z=2) 也可以寫成 P(X Y = 2) ，也就是 P({(2點, 正)}) ，因此 P(Z=2) 的機率為 1/12。

X^k 的意義

隨機變數 X^k 代表的是一個函數 $Z = X^k$ ，其中 Z 函數對每一個元素 s 的映射值均為 X(s) 的 k 次方，也就是： $Z(s) = X^k(s)$

範例：X 為投擲 1 顆骰子點數的隨機變數，且定義 $Z = X^2$，請問隨機變數 P(Z=4) 的機率為何？

解答：

$$Z(s) = 4 = X^2(s) = X(s) * X(s) \rightarrow X(s) = 2$$

所以 P(Z=4) 相當於 P(X=2) = P({2點}) = 1/6

但必須注意的是 Z 的定義域雖仍然為 ({1點,...., 6點})，但是值域卻為 1,4,9,16,25,36。

結語

隨機變數 X, Y, Z, ... 乃是一種作用於樣本空間 S 的實函數，此種函數會將樣本點映射到實數中，例如：$X(S) \rightarrow R$ 代表函數 X 將樣本空間中的元素 s 映射到某個實數值 x。

利用隨機變數映射完成之後，就可以比較大小，因此可以計算「機率密度函數」與「累加分配函數」，這樣就能利用加總或積分去計算某個區間內的機率，讓機率模型得以進行數學性的運算。

我想這是為甚麼數學家要將隨機變數定義成實函數的原因之一吧！

機率分布

簡介

在程式設計領域，「設計模式」是一些經常被使用到的物件樣式，而在數學領域，也同樣存在著某些「常見模式」，在機率統計領域，這些「常見模式」就是機率分布。

機率分布可以分為「離散型」與「連續型」兩類，離散型的機率分布通常只會有整數型的值，而連續型的機率分布則在整個實數軸上都可能產生樣本。

伯努力試驗 (Bernoulli trial)

所有的離散型機率分布，幾乎都是從「伯努力試驗」這個概念開始的，讓我們先來瞭解一下何謂「伯努力試驗」。

伯努利試驗是一項只有兩種可能結果的隨機試驗，可以用下列機率分布描述：

$$P[X=1]=p$$
$$P[X=0]=1-p$$

換句話說、伯努力試驗是一種 YES or NO (1 or 0) 的試驗。舉例而言，像是「丟銅版、生男生女、一地區

某天最高溫是否超過 30 度、擲骰子是否超過 2 點」等等，都可以用伯努力實驗描述。

伯努力試驗的概念很簡單，以下是一些範例：

範例 1 :

丟一個公正銅板，用隨機變數 X 將正面映射為 1 ，反面映射為 0，那麼就可以用 P[X=1]=0.5, P[X=0]=0.5 表示這個機率模型。

在 R 軟體中， Sample 函數可以用來模仿柏努力試驗。

舉例而言，以下是範例一的丟銅板試驗，指令 sample(0:1, 10, replace=T, prob=c(0.5,0.5)) 代表連續進行 10 次柏努力試驗，成功失敗機率各為 0.5。

```
> sample(0:1, 10, replace=TRUE, prob=c(0.5,0.5))
 [1] 1 0 1 1 0 1 0 1 0 1
> sample(0:1, 10, replace=T)
 [1] 0 1 1 1 0 0 1 1 1 0
```

說明：X(反面)=0, X(正面)=1, 第一個參數 0:1 分別代表 {反面、正面} 的映射結果，而第二個參數代表只投擲一次，第三個參數 replace 代表樣本取後是否放回，這在頭銅板的範例必須用 replace=TRUE, 因為這次

投正面之後不代表下次不能再出現正面，而 prob 則是指定的機率分布，如果不指定則代表採用平分的機率分布，以這個例子就是各為 0.5 的方式。

範例 2 :

假如用機率描述生男生女這件事， X({生男})=1, X({生女})=0, 且生男生的機率為 0.53, 生女生的機率為 0.47，那麼就可以用 P[X=1]=0.53, P[X=0]=0.47 表示這個機率模型。

```
> sample(0:1, 10, replace=T, prob=c(0.47, 0.53))
 [1] 0 1 1 0 1 1 0 1 1 0
```

二項分布 (Binomial distribution)

如果我們進行 n 次的伯努力試驗，每一次的實驗都可以用隨機變數描述， P(ti=1) = p, P(ti=0)=1-p ，而且這些試驗 {t1, t2, ..., tn} 之間是獨立的，那麼我們就可以用二項分布來描述 n 次實驗的可能機率分布。

由於這 n 次實驗相互獨立，假如 (t1 t2 ... tn) 代表這個實驗的一個可能出像，因此 P(t1 t2 .... tn) = P(t1) P(t2) .... P(tn)。

令 X 代表一個可以將 (t1 t2 ... tn) 映射到伯努力試驗成功 (Yes) 次數的函數，那麼、n 次實驗中出現 k 次 1 的機會，可以用以下算式表示。

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

舉例而言，投擲公正銅板 5 次，得到 3 次正面的機率為 $P(X=3)=\binom{5}{3}p^3(1-p)^{5-3}$ ，其中 p=0.5。

範例：

假如生男生的機率為 0.53, 生女生的機率為 0.47，而且每位母親生男生女的事件之間都是獨立的。

某母親 A 想要生 3 個小孩，請問至少有一個男生的機會為多少。

用機率描述生男生女這件事， X({生男})=1, X({生女})=0, 那麼可以計算至少生一個男生的機率如下：

$P(X>=1)=P(X=2,3)=\sum_{k in \{2,3\}}\binom{3}{k}p^k(1-p)^{n-k}$ ，其中 p = 0.53, (1-p) = 0.47。

讓我們用 R 軟體計算一下

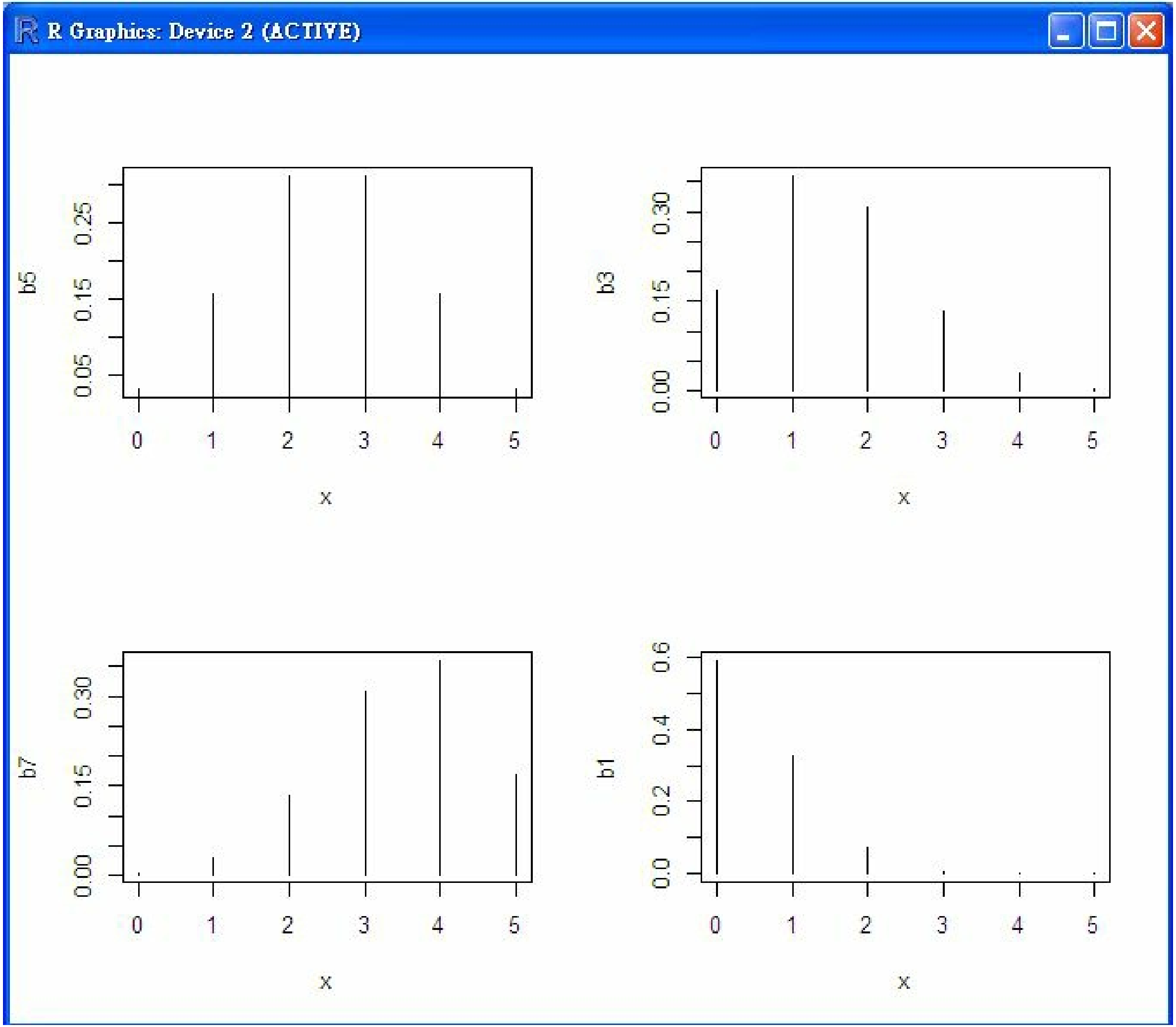
```
> dbinom(1, 3, 0.53)+dbinom(2,3, 0.53)+dbinom(3,3,0.53)
```

```
[1] 0.896177
> sum(dbinom(c(1,2,3), 3, 0.53))
[1] 0.896177
> x=c(1,2,3)
> x
[1] 1 2 3
> p=dbinom(x, 3, 0.53)
> p
[1] 0.351231 0.396069 0.148877
> sum(p)
[1] 0.896177
```

二項分布的圖形

```
> par(mfrow=c(2,2))
> x = 0:5
> b5 = dbinom(x, 5, 0.5)
> plot(x, b5, type="h")
> b3 = dbinom(x, 5, 0.3)
> plot(x, b3, type="h")
> b7 = dbinom(x, 5, 0.7)
> plot(x, b7, type="h")
```

```
> b1 = dbinom(x, 5, 0.1)
> plot(x, b1, type="h")
```


R Graphics: Device 2 (ACTIVE)
b5
0.05
0.15
0.25
0 1 2 3 4 5
x
b3
0.00
0.15
0.30
0 1 2 3 4 5
x
b7
0.00
0.15
0.30
0 1 2 3 4 5
x
b1
0.0
0.2
0.4
0.6
0 1 2 3 4 5
x

習題

1. 請問丟 10 個公平的銅板，有三個正面的機會是多少？
2. 請問丟 n 個公平的銅板，正面次數 <= k 的機率是多少？
3. 請問丟 10 個公平的銅板，得到正面次數的期望值為何？

幾何分布 (Geometric distribution)

如果我們連續進行一系列的伯努力試驗，直到成功才停止，那麼我們需要進行多少次實驗呢？

關於這種「直到成功才停止」的問題，可以用幾何分布來描述，以下是幾何分布的定義。

$$P(X=k)=(1-p)^{k-1}p=q^{k-1}p$$

舉例而言，假如我們連續投擲公正銅版，直到出現正面才停止，那麼我們需要投擲 k 次才會得到第一個正面的機率，就會是 $(1-p)^{k-1}p$ ，其中的 p=0.5。

範例：

假如生男生的機率為 0.53, 生女生的機率為 0.47，而且每位母親生男生女的事件之間都是獨立的。

某位母親決定要一直生小孩，直到有一個女孩為止，請問她在生小孩個數不大於三個就能完成任務的機率為多少？

用機率描述生男生女這件事， X({生女})=1, X({生男})=0, 那麼就可以累加下列算式以計算結果。

$$\sum_{k=1}^{3} P(X=k) = \sum_{k=1}^{3} (1-p)^{k-1} p = \sum_{k=1}^{3} (1-0.53)^{k-1} 0.47$$

讓我們用 R 軟體計算一下，必須注意的是，R 軟體中的幾何分布 dgeom 的定義為 $p(x)=p(1-p)^x$，也就是用失敗次數當 x，因此其公式與上面的有所不同，必須修改如下：(其中的 x 代表失敗次數)。

$$P(X=x)=(1-p)^x p = q^x p$$

```
> dgeom(0, 0.47)
[1] 0.47
> dgeom(1, 0.47)
[1] 0.2491
> sum(dgeom(c(0,1,2), 0.47))
[1] 0.851123
```

習題

1. 請問丟公平的銅板時，得到第 1 次正面時投擲次數 k 的機率分布為何？該分布的期望值為何？
2. 請問丟公正的骰子時，得到第 1 次 6 點時投擲次數 k 的機率分布為何？該分布的期望值為何？

負二項分布

如果我們對「幾何分布」進行擴充，改成「持續進行試驗直到取得 r 次成功為止」，那麼其機率分布又該如何描述呢？

這樣的機率分布就稱為負二項分布，其公式如下：

$$P(X=k)=\binom{k-1}{r-1}(1-p)^{k-r}p^{r}$$

舉例而言，假如我們連續投擲公正銅版，直到出現三次正面才停止，那麼我們需要投擲 k 次才會得到第一個正面的機率，就會是 $\binom{k-1}{r-1}(1-p)^{k-r}p^{r}$ ，其中的 p=0.5。

讓我們用 R 軟體計算一下，必須注意的是，R 軟體中的負二項分布 dbinom 的定義為 Γ(x+n)/(Γ(n) x!) p^n (1-p)^x ，也就是用 n=r, x=k-r-1 的代換方式。

其中的 x 同樣代表失敗次數，而 n 代表成功次數，Γ(n) 代表排列數，所以 Γ(x+n)/(Γ(n) x!) 其實也就是

(x+n-1)!/((n-1)! x!) ，也就是 $\binom{x+n-1}{x}$ 的意思。

R 的操作範例

```
> dnbinom(0, 3, 0.5)
[1] 0.125
> dnbinom(1, 3, 0.5)
[1] 0.1875
> dnbinom(0:10, 3, 0.5)
 [1] 0.125000000 0.187500000 0.187500000 0.156250000 0.117187500 0.082031250
 [7] 0.054687500 0.035156250 0.021972656 0.013427734 0.008056641
> n=3
> x=1
> p=0.5
> gamma(x+n)/(gamma(n)*prod(1:x)) * p^n * (1-p)^x
[1] 0.1875
> choose(x+n, n) * p^n * (1-p)^x
[1] 0.25
> choose(x+n-1, x) * p^n * (1-p)^x
[1] 0.1875
```

範例：

假如生男生的機率為 0.53, 生女生的機率為 0.47，而且每位母親生男生女的事件之間都是獨立的。

某位母親決定要一直生小孩，直到有三個女孩為止，請問她在生小孩個數不大於 5 個就能完成任務的機率為多少？

用機率描述生男生女這件事， X({生女})=1, X({生男})=0, 那麼就可以累加下列算式以計算結果。

$$\sum_{k=1}^{5} P(X=k) = \sum_{k=1}^{5} \binom{k-1}{r-1} (1-p)^{k-r} p^{r}$$, 其中 p=0.47, r=3。

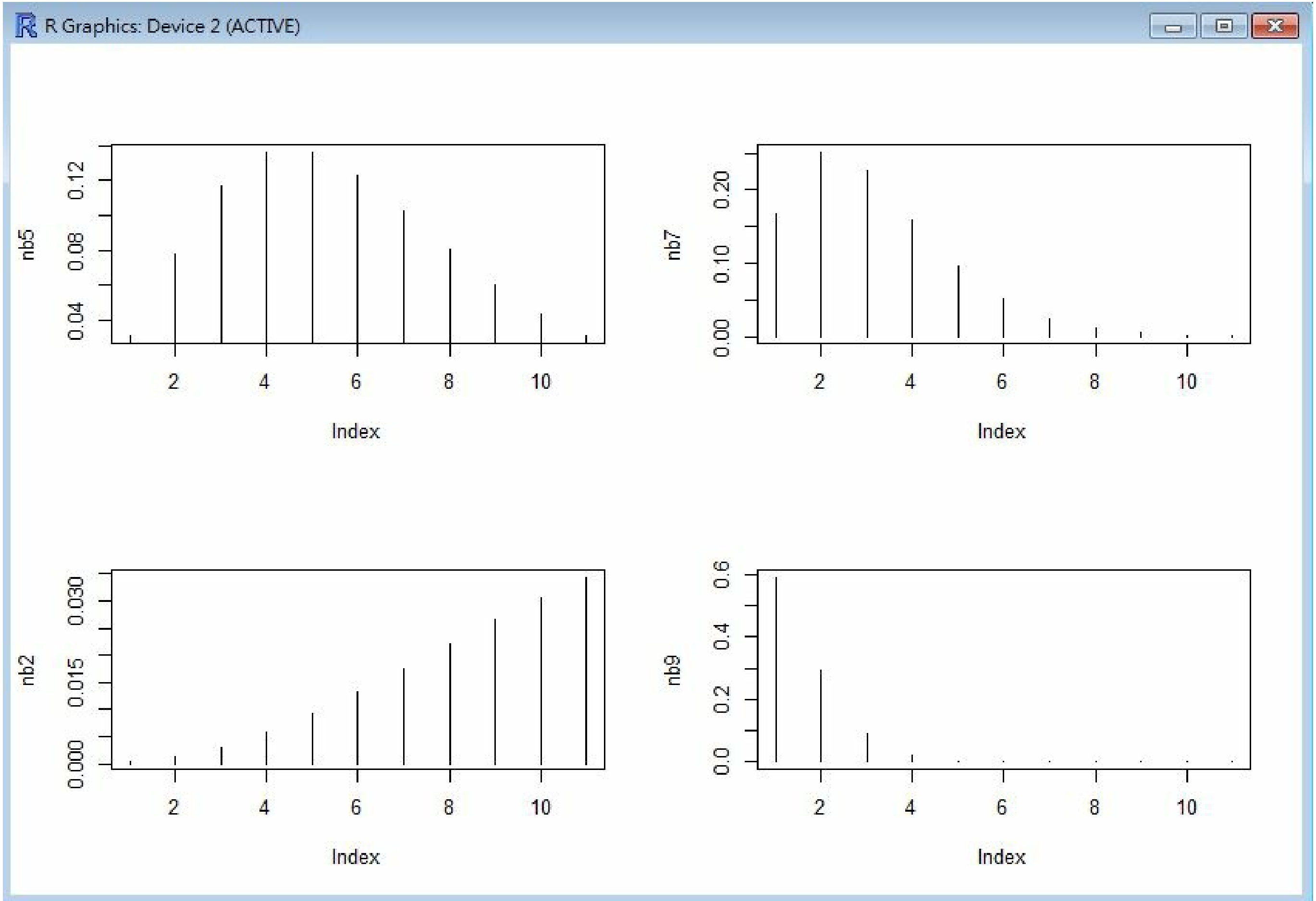
但是由於 R 是用失敗次數

```
> dnbinom(3, 3, 0.47)
[1] 0.1545686
> dnbinom(4, 3, 0.47)
[1] 0.122882
> p=dnbinom(c(3,4,5), 3, 0.47)
> p
[1] 0.15456857 0.12288201 0.09117845
```

```
> sum(p)
[1] 0.368629
```

負二項分布的圖形

```
> par(mfrow=c(2,2))
> nb5 = dnbinom(x, 5, 0.5)
> plot(nb5, type="h")
> nb7 = dnbinom(x, 5, 0.7)
> plot(nb7, type="h")
> nb2 = dnbinom(x, 5, 0.2)
> plot(nb2, type="h")
> nb9 = dnbinom(x, 5, 0.9)
> plot(nb9, type="h")
```


R Graphics: Device 2 (ACTIVE)
nb5
0.04
0.08
0.12
nb7
0.00
0.10
0.20
nb2
0.000
0.015
0.030
nb9
0.0
0.2
0.4
0.6
2
4
6
8
10
Index

布瓦松分布（Poisson distribution）

在離散機率分布當中，布瓦松分布算是相當特別的一個，因為「布瓦松分布」是描述「連續區域內出現幾個樣本」的分布。舉例而言，像是舀一瓢水會撈到的草履蟲數量，或者抽一滴血會抽到的白血球數量等等。

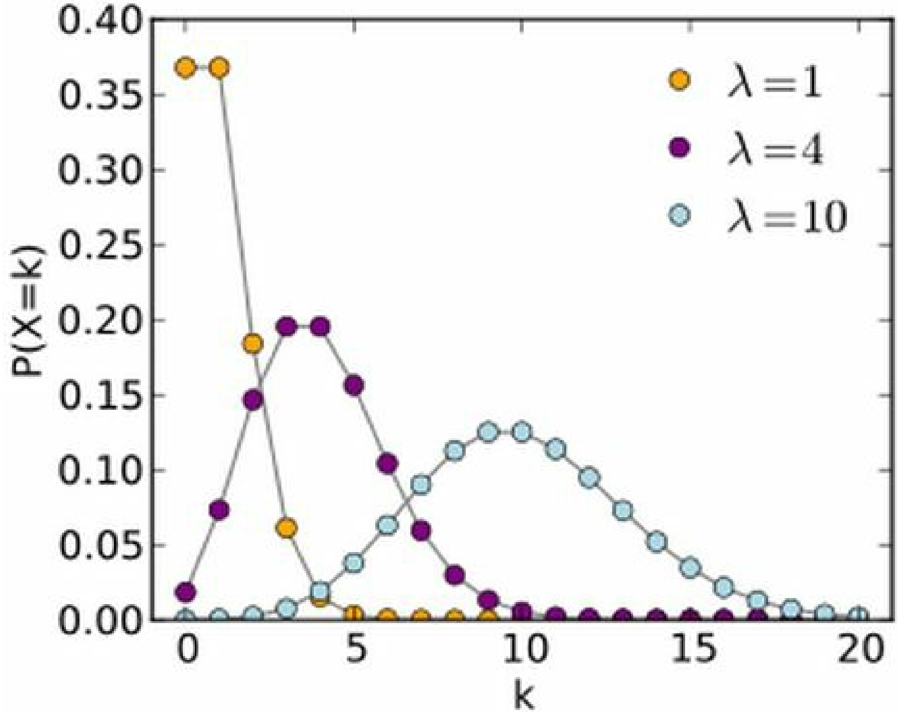
布瓦松分布的公式如下所示，其中的 λ 代表每單位區域內會出現的樣本平均數。

$$P(X=k)=\lambda^k e^{-\lambda}/k!$$

要瞭解布瓦松分布，得從二項分布的極限開始想起，以下是三種不同 λ 參數的布瓦松分布圖：

Poisson

Probability mass function

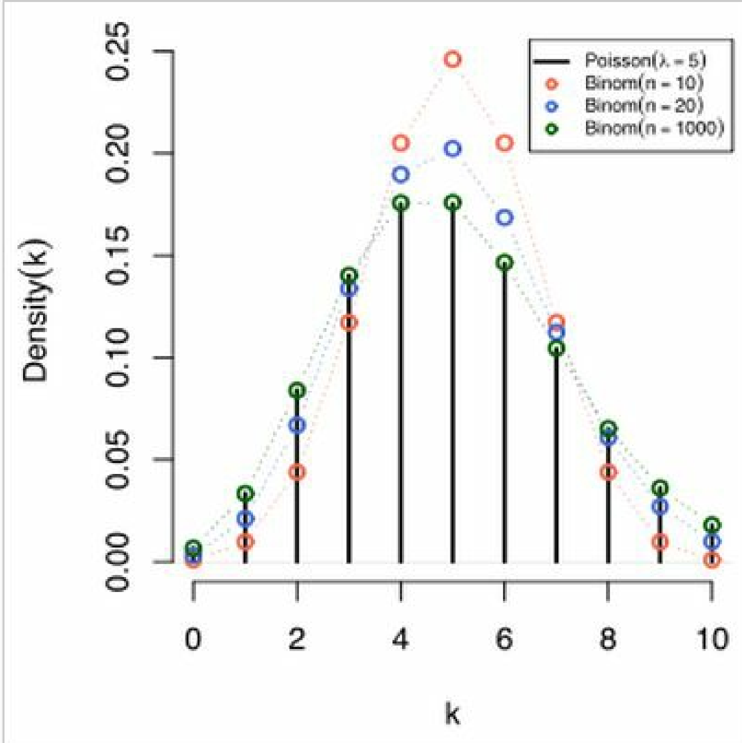


The horizontal axis is the index *k*, the number of occurrences. The function is only defined at integer values of *k*. The connecting lines are only guides for the eye.

圖、布瓦松分布

來源：http://en.wikipedia.org/wiki/Poisson_distribution

事實上、布瓦松分布是二項分布在 n 趨近無限大的極限情況。



Comparison of the Poisson distribution (black lines) and the binomial distribution with n=10 (red circles), n=20 (blue circles), n=1000 (green circles). All distributions have a mean of 5. The horizontal axis shows the number of events *k*. Notice that as n gets larger, the Poisson distribution becomes an increasingly better approximation for the binomial distribution with the same mean.

圖、布瓦松分布是二項分布 n 趨近無限大的極限情況

來源：http://en.wikipedia.org/wiki/File:Binomial_versus_poisson.svg

為了說明布瓦松分布與二項分布之間的關係，讓我們用以下的白血球範例來看看隱藏在這兩個分布背後的關係：

範例：抽血時白血球數量的問題

問題： 假如現在從你身上抽一滴血，請回答下列兩個問題。

1. 請定義一個隨機變數 X 代表那滴血中的白血球數量。

提示： 樣本空間 S = 此時此刻你身上的所有白血球 = {w1,w2,......,wn}

2. 請算出一滴血液中有三顆白血球的機率，假設該滴血液占你總血量的 1/1000。

解答 1：

X(A) = |A|

說明：

A 是一個事件，也就是白血球的樣本空間 S 的子集合，例如： A = {w1, w5, w9}

|A| 代表 A 集合的大小，也就是元素個數，舉例而言：

如果 A = {w1, w5, w9} ，那麼 |A| = 3

如果 B = {w2, w8}，那麼 |B| = 2

如果 C = {}，那麼 |C| = 0

如果 D = S，那麼 |D| = n

解答 2：

P(X=3) = P({A | X(A) = 3}) = P({{w1, w2, w3}) + P({w1, w2, w4}) + ...

假如任一顆白血球被抽到的機率等於該滴血液佔全身血液的比率，由於該滴血液佔總血量的 1/1000，所以給顆白血球被抽到的機率為 1/1000。

而且假設這些白血球沒有智慧，也不會聚合在一起，因此相互之間獨立，那麼由於每顆白血球被抽到的機率為 1/1000，因此 P(w1) = P(w2) = ... P(wn) = 1/1000。

那麼初步想法是 P(w1w3) = P(w1) * P(w3) = 1/1000 * 1/1000 。

但是上述的想法有個小問題，那就是該情況代表其它白血球都沒被抽到，因此所謂的 P(w1w3) 真正的意思應該是

$$P(w_1\overline{w_2}w_3\overline{w_4}....\overline{w_n})=(\frac{1}{1000})^2(\frac{999}{1000})^{n-2}$$

所以 P(X=3) 應該算法如下：

$$P(X=3)=P(\{A|X(A)=3\})=(\frac{1}{1000})^3(\frac{999}{1000})^{n-3}\binom{n}{3}$$

推而廣之，P(X=k) 的機率之算法如下：

$$P(X=k)=P(\{A|X(A)=k\})=(\frac{1}{1000})^k(\frac{999}{1000})^{n-k}\binom{n}{k}$$

事實上，這個題目的機率分布就是下一章的二項分布，如下所示：

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

而且、當 n 趨近於無限大時，這個分布將會趨近於布瓦松分布，如下所示：

$$P(X=k)=\lambda^k e^{-\lambda}/k!$$

其中的 λ 之意義為，在單位時間 (或單位面積、體積) 內，事件的出現次數平均為 λ 次。

習 題

習題：假設每 1CC 的血所含的白血球平均為 10 顆，那麼請問你抽 1CC 的血時，抽到 8 顆白血球的機率是多少。

解答：

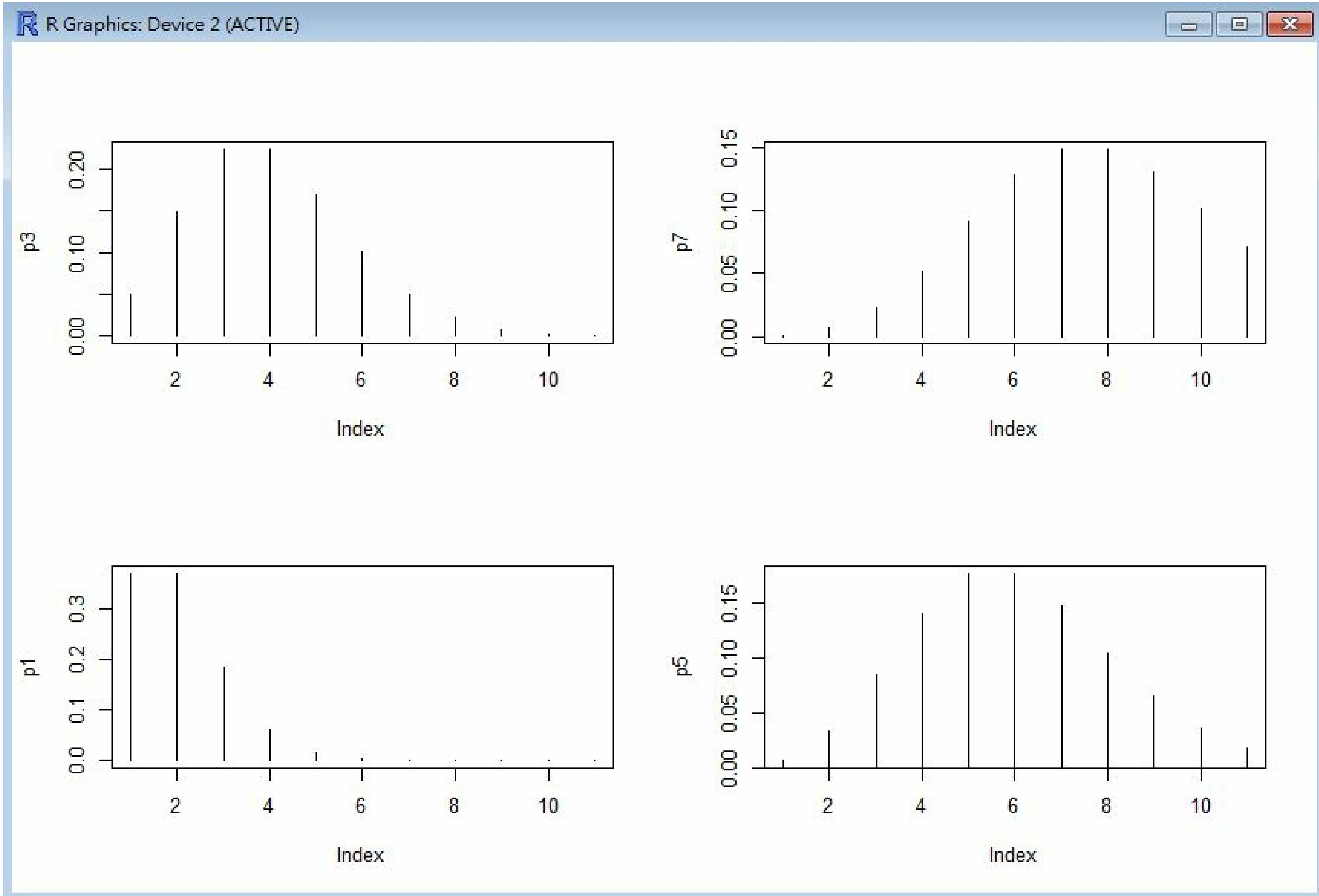
$\lambda = 10$，因此布瓦松分布為 $p(x) = 10^x e^{-10} / x!$ ，將 x=8 代入，得到 $p(8) = 10^8 e^{-10} / 8!$

其數值可以用 R 軟體計算，如下所示：

```
> ?dpois
> dpois(8, 10)
[1] 0.112599
> 10^8*exp(-10)/prod(1:8)
[1] 0.112599
```

布瓦松分布的圖形

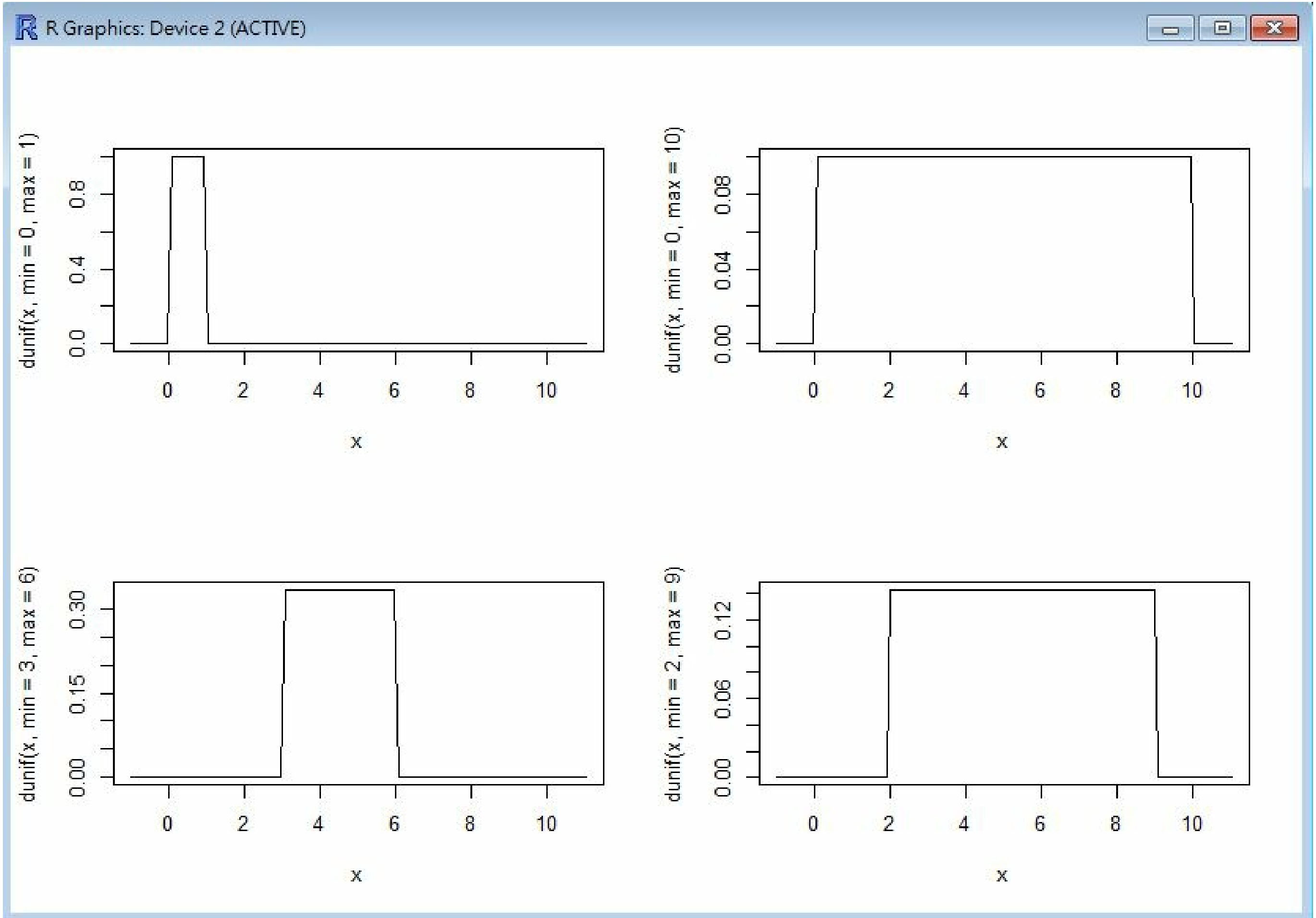
```
> par(mfrow=c(2,2))
> x = 0:10
> p3 = dpois(x, lambda=3)
> plot(p3, type="h")
> p7 = dpois(x, lambda=7)
> plot(p7, type="h")
> p1 = dpois(x, lambda=1)
> plot(p1, type="h")
> p5 = dpois(x, lambda=5)
> plot(p5, type="h")
```



均匀分布 (Uniform distribution)

均勻分布的圖形

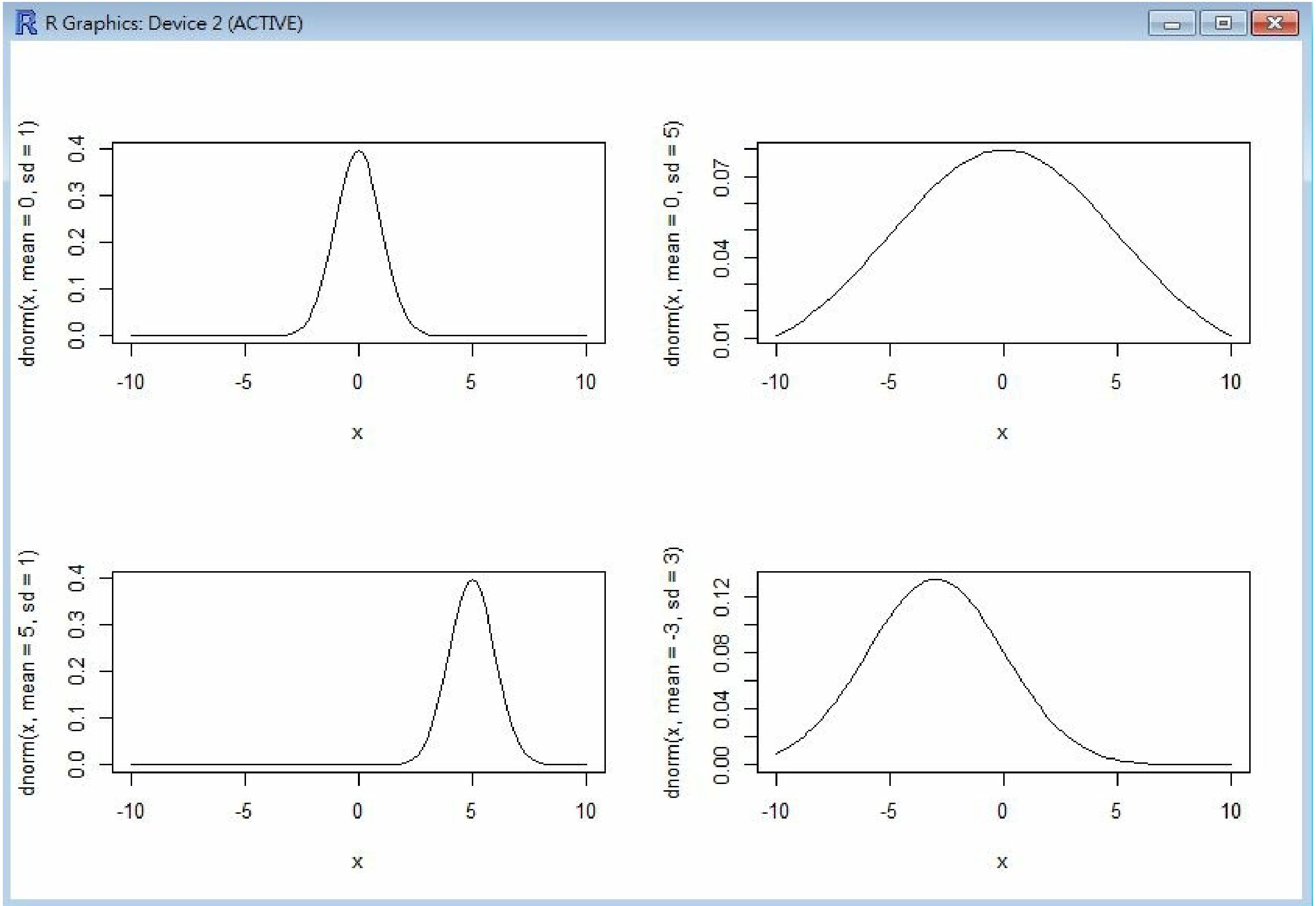
```
> dunif(0.5)
[1] 1
> dunif(0.9)
[1] 1
> dunif(2)
[1] 0
> dunif(-1)
[1] 0
> par(mfrow=c(2,2))
> x=0:10
> curve(dunif(x, min=0, max=1), from=-1, to=11)
> curve(dunif(x, min=0, max=10), from=-1, to=11)
> curve(dunif(x, min=3, max=6), from=-1, to=11)
> curve(dunif(x, min=2, max=9), from=-1, to=11)
```


R Graphics: Device 2 (ACTIVE)
dunif(x, min = 0, max = 1)
0.0
0.4
0.8
0
2
4
6
8
10
x
dunif(x, min = 0, max = 10)
0.00
0.04
0.08
0
2
4
6
8
10
x
dunif(x, min = 3, max = 6)
0.00
0.15
0.30
0
2
4
6
8
10
x
dunif(x, min = 2, max = 9)
0.00
0.06
0.12
0
2
4
6
8
10
x

常態分布 (Normal Distribution)

常態分布的圖形

```
> dnorm(0)
[1] 0.3989423
> dnorm(0.5)
[1] 0.3520653
> dnorm(2.5)
[1] 0.0175283
> par(mfrow=c(2,2))
> curve(dnorm(x, mean=0, sd=1), from=-10, to=10)
> curve(dnorm(x, mean=0, sd=5), from=-10, to=10)
> curve(dnorm(x, mean=5, sd=1), from=-10, to=10)
> curve(dnorm(x, mean=-3, sd=3), from=-10, to=10)
```


R Graphics: Device 2 (ACTIVE)
dnorm(x, mean = 0, sd = 1)
0.0 0.1 0.2 0.3 0.4
-10 -5 0 5 10
x
dnorm(x, mean = 0, sd = 5)
0.01 0.04 0.07
-10 -5 0 5 10
x
dnorm(x, mean = 5, sd = 1)
0.0 0.1 0.2 0.3 0.4
-10 -5 0 5 10
x
dnorm(x, mean = -3, sd = 3)
0.00 0.04 0.08 0.12
-10 -5 0 5 10
x

附件：離散型機率分布表格整理

以下是離散型機率分布的匯總表格，讀者現在還看不懂的話沒有關係，我們在後文中會解說其中較重要的幾個分布。

離散機率模型	密度函數	R 函數名稱	說明
二項分布	$\binom{n}{x} p^x (1-p)^{n-x}$	binom(n:size, p:prob)	n:樣本數, p:正面機率, n 次試驗中有 x 個成功的機率
多項分布	$\frac{n!}{x_1!...x_k!} p_1^{x_1} ... p_k^{x_k}$	multinom(n:size, p(1..k):prob)	n:樣本數, p[1..n]:各項的機率
負二項分布	$\binom{x-1}{r-1} (1-p)^{x-r} p^r$	nbinom(size, prob)	x:樣本數, , p:正面機率, 要得到第 r 次成功所需要的試驗次數
幾何分布	$(1-p)^{x-1} p$	geom(p:prob)	p: 成功機率, 第一次成功所需要的試驗次數

超幾何分布	$\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$	hyper(N:m,n:n,r:k)	m:白球數量, n:黑球數量, k:抽出球數, 同二項分布，但取樣後不放回
布瓦松分布	$\frac{e^{-\lambda s}\lambda s^{x}}{x!}$	pois(lambda)	k:期望值, $\lambda = \frac{k}{s}$, 在 s 時間內，事件出現平均 k 次

附件：連續型機率分布表格整理

以下是連續型機率分布的匯總表格，讀者現在還看不懂的話沒有關係，我們在後文中會解說其中較重要的幾個分布。

連續機率模型	密度函數	R 函數	說明
均勻分布 (Uniform)	$\frac{1}{b-a}$	unif(a:min, b:max)	a:範圍下限, b: 上限 出現機會均等
常態分布 (Normal)	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}[(x-\mu)/\sigma]^2}$	norm(mean, sd)	中央極限定理：x1+x2+...+xk; 當 k 越大就越接近常態分布

伽瑪分布 (Gamma)	$\frac{1}{\Gamma(a)b^a}x^{a-1}e^{-x/b}$	gamma(shape, rate = 1, scale = 1/rate)	$\Gamma(k)=\int_0^{\infty} z^{k-1}e^{-z}dz$ 指數分布與卡方分布都是 Gamma 分布的特例
指數分布 (Exponential)	$\frac{1}{b}e^{-x/b}$	exp(rate)	伽瑪分布($a=1,b=\frac{1}{\lambda}$) 布瓦松過程中，第一次事件 出現的時間 W
卡方分布 (Chi-Square)	$\frac{1}{2^{\gamma/2}\Gamma(\gamma/2)}\ x^{\gamma/2-1}e^{-x/2}$	chisq(df, ncp)	伽瑪分布($b=2,a=\gamma/2$) 利用樣本推斷母體變異數
柯西分布 (Cauchy)	$\frac{1}{\pi}\frac{a}{a^2+(x-b)^2}$	cauchy(b:location, a:scale)	
威布爾分布 (Weibull)	$abx^{b-1}e^{-ax^b}$	weibull(a:shape, b:scale)	$\rho(t)=\frac{f(t)}{R(t)}$ 可靠度工程：f(x) 失敗時間,

			R(t) 可靠度, $\rho(t)$ 失敗率
T 分布 (T)	$\frac{z}{\sqrt{x_{\gamma}^{2}/\gamma}}$	t(df, ncp)	估計變異數時使用的分布
F 分布 (F)	$\frac{x_{\gamma_1}^{2}/\gamma_1}{x_{\gamma_2}^{2}/\gamma_2}$	f(df1, df2, ncp)	等變異數 F 檢定時使用
貝塔分布 (Beta)		beta(a:shape1, b:shape2, ncp)	
對數常態分布 (Log Normal)		lnorm(meanlog, sdlog)	
邏輯分布		logis(location, scale)	
Signrank		signrank(n)	
威爾斯		wilcox(m, n)	a,b 為兩組樣本

期望值與動差生成函數

期望值

定義：期望值 E(X) , (通常用符號 μ 代表, $\mu = E(X)$)

離散分布：$E[X] = \sum_{s \in S, x = X(s)} xP(x)$ ；通常簡寫為 $\sum_{x \in X(S)} xP(x)$ 或者直接寫 $\sum_{\forall x} xP(x)$

連續分布：$E[X] = \int_{-\infty}^{\infty} xf(x)dx$

期望值的相關定理：

1. $E[c] = c$;

2. $E[cX]=cE[X]$;

3. $E[X+Y]=E[X]+E[Y]$;

證明

定理 1: E[c] = c

$E[c]=\sum_{x \in X(S)} (c^* P(x))$; 根據期望值定義

$=c \sum_{x \in X(S)} P(x)$; 根據基本算術

$=c$; 因為 P(x) 是機率密度函數

定理 2: E[c X] = c E[X]

$E[cX]=\sum_{x \in X(S)} (c^* x^* P(x))$; 根據期望值定義

$$= c \sum_{x \in X(S)} (x^* P(x))$$; 根據基本算術

$$= c \; E[X]$$; 根據期望值定義

定理 3 : E[X + Y] = E[X] + E[Y]

假如離散隨機變數 X, Y 的機率密度函數分別用 $P(X), P(Y)$ 代表。

$$E[X+Y] = \sum_{s \in S, x = X(S), y = Y(s)} (xP(x) + yP(y))$$; 根據期望值定義

$$= \sum_{x \in X(S)} (xP(X)) + \sum_{y \in Y(S)} (yP(y))$$; 根據乘法對加法的分配率

$$= E(X) + E(Y)$$;

以上證明了離散的情況，連續的情況雷同，請比照上述寫法撰寫。

變異數

定義：變異數 Var(X)

離散隨機變數 X 的變異數 Var(X) 定義如下

$$Var(X) = \sigma^2 = E[(X-\mu)^2] = \sum_{x \in X(S)} (x-\mu)^2 P(x)$$

說明：

1. 上式中的 Var(X) 稱為 X 的變異數，而其平方根 $\sum$ 稱為 X 的標準差。 (μ 為 X 的期望值)
2. 以上算式中 $\sum$ 的下標均為 $x \in X(S)$ ，而非 $x \in S$ ，也就是 x 是實數值，而非樣本點。
3. 這也是為何要將隨機變數定義為實函數的原因，這樣才能對這些「變數」進行 +, -, * > 等代數運算，並且可以進行期望值與變異數的計算。

定理：$Var(X) = E[X^2] - (E[X])^2$

$Var(X) = E[(X-\mu)^2]$；

$$=E[X^2-2\mu X+\mu^2];$$
$$=E[X^2]-2\mu E[X]+\mu^2;$$
$$=E[X^2]-2E[X]E[X]+E[X]^2;$$
$$=E[X^2]-E[X]^2.$$

期望值的函數

有時，我們會想計算某個隨機變數之函數的期望值，像是 E[g(X)]。

某隨機變數 X 之函數 g(X) 的期望值

期望值 **E[g(X)]**：

$$E[g(X)]=\sum_{x\in X(S)} g(x)P(x)$$

舉例而言，$E[X^3+2X^2+3X+2]$ 就是一個隨機變數 X 的函數

$g(x)=X^3+2X^2+3X+2$ 的期望值。

而且、由於上述定理 1, 2, 3 的特性，這些期望值的函數還可以拆開來算，舉例如下：

$E[X^3+2X^2+3X+2]$;

$=E[X^3]+E[2X^2]+E[3X]+E[2]$;

$=E[X^3]+2E[X^2]+3E[X]+2$.

在以上的範例中， $E[X]$ 稱為 X 的 1 級動差， $E[X^2]$ 稱為 X 的 2 級動差， $E[X^3]$ 稱為 X 的 3 級動差 ....

k 階動差 (Kth Ordinary Moment)

定義： $E[X^k]$ 稱為隨機變數 X 的 k 階動差 (Kth ordinary moment)

動差的概念就像是期望值的多項式，我們可以將任何一個多項式的動差寫成 k 個動差的組合，這樣就能將任何的函數的動差給支解。

但是、可惜的是，即使我們將函數分解成動差的組合，其計算上仍然是相當複雜的，但是如果我們只是想變任某個期望值函數對應的原始機率分布為何，那麼可以藉助「動差生成函數」來完成這項任務，以下是動差生成函數的定義。

動差生成函數

定義：隨機變數 X 的動差生成函數 (Moment Generating Function, m.g.f) $m_X(t)$ 為以下函數

$$m_X(t) = E\left[e^{tX}\right] = E\left(1 + tX + \frac{(tX)^2}{2!} + ... + \frac{(tX)^k}{k!} + ...\right)$$

動差存在的條件是期望值 $E\left[e^{tX}\right]$ 在開區間 (-h, h) 內是有限的。

根據以上定義，離散分布與連續分布的動差生成函數分別可以寫成以下算式：

離散分布：$E\left(e^{tX}\right) = \sum_{x \in S} e^{tx} P(x)$

連續分布：$E(e^{tX})=\int_{-\infty}^{\infty}e^{tx}f(x)dx$

那麼、動差生成函數到底有甚麼用呢？

關於這個問題，可以讓我們回到泰勒展開式這個微積分的概念來看，就能理解「動差生成函數」背後的原理了。

根據泰勒展開式，我們可以將函數 e^{tX} 展開如下：

$$e^{tX}=1+tX+\frac{(tX)^2}{2!}+....+\frac{(tX)^k}{k!}+.....$$

您可以看到在上述展開式當中，不管 k 為何，每一項的 X^k 都存在，並不會消失，而且 X^k 的係數為 $\frac{t^k}{k!}$ ， 因此、只要在某個夠小的開區間 (-h, h) 內這個動差生成函數是有限的，那麼隨機變數 X 與函數 e^{tX} 之間將會有對映關係，而機率密度函數 P(X) 與動差生成函數 $E(e^{tX})$ 也可以被證明有一對一的對映關係。

於是、動差生成函數就成了一個機率分布的「指紋」，意思是如果兩個隨機變數 X, Y 的動差生成函數 $E(e^{tX})=E(e^{tY})$ ，則這兩個機率分布也必然相同。

思考 1：

思考：為何動差生成函數可以做為一個機率分布的「指紋」呢？

說明：如果兩個機率分布 P(X) 與 P(Y) 的動差生成函數相同，那麼將意味著 $E(e^{tX}) = E(e^{tX})$，根據泰勒展開式可得到

$$E(1 + tX + ... + \frac{(tX)^k}{k!} + ...) = E(1 + tY + ... + \frac{(tY)^k}{k!} + ...)$$

因此在每一階的動差上，$E[X^k]$ 都與 $E[Y^k]$ 相同，因此這兩個分布也就應該是一樣的了。

回顧 1：

f(x) 在 0 點的泰勒展開式 (麥克羅林級數) 可以作為一個函數的指紋，意思是如果兩個函數的泰勒展開式相同，則這兩個函數必然相同 (這點是高等微積分課程的核心)。

回顧 2：

函數 f(x) 的特徵函數 (Characteristic function) 為

$$E(e^{itX}) = e^{i}E(e^{tX}) = e^{i}E\left(1 + tX + \frac{(tX)^2}{2!} + ... + \frac{(tX)^k}{k!} + ...\right)$$

結 語

為何數學家要將隨機變數定義成一種函數，然後相樣本映射到實數空間上，而不是直接對樣本進行機率運算呢？筆者認為應該是為了期望值而布的局，因為將樣本映射到實數之後，才能用下列算式計算期望值。

$$E[X] = \sum_{x \in X(S)} x P(x)$$

而隨機變數之間的代數運算，像是「3X」,「X+Y」,「X-2Y」,「X*Y」,「X*X*X*X」等 運算的結果，也仍然是一種作用在樣本空間 S 的實函數，只是當 X, Y 兩者的樣本空間有所不同時，我們必須以兩者樣本空間的迪卡兒乘積 $S = (S_X, S_Y)$ 作為樣本空間。

在這種情況下，期望值函數也才能運作在 +, - * 等運算空間中，得到以下的廣義期望值：

$$E[g(X)] = \sum_{x \in X(S)} g(x)P(x)$$

「動差生成函數」可以做為機率分布的指紋，因此如果兩個機率分布的「動差生成函數」相同，那麼其機率分布也會相同。

「動差生成函數」的定義如下：

$$m_X(t) = E[e^{tX}] = E\left(1 + tX + \frac{(tX)^2}{2!} + ... + \frac{(tX)^k}{k!} + ...\right)$$

聯合分布

聯合密度函數

定義：離散聯合密度函數

表示符號： $P_{XY}(x,y)=P[X=x,Y=y]$

必要條件：

1. $P_{XY}(x,y)\geq 0$；
2. $\sum_{\forall x}\sum_{\forall y}P_{XY}(x,y)=1$；

定義 2：連續聯合密度函數

表示符號： $P_{XY}(x,y)$ (範圍： $f(x)=x^2$)

必要條件：

1. $P_{XY}(x,y) \geq 0$；
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_{XY}(x,y)dydx = 1$；
3. $P[a \leq X \leq b, c \leq Y \leq d,] = \int_{a}^{b}\int_{c}^{d} P_{XY}(x,y)dydx$；

邊際密度函數

定義：離散邊際密度函數

1. 只有 X 的情況：$P_X(x) = \sum_{\forall y} P_{XY}(x,y)$
2. 只有 Y 的情況：$P_Y(y) = \sum_{\forall x} P_{XY}(x,y)$

定義：連續邊際密度函數

1. 只有 X 的情況：$P_X(x)=\int_{-\infty}^{\infty}P_{XY}(x,y)dy$
2. 只有 Y 的情況：$P_Y(y)=\int_{-\infty}^{\infty}P_{XY}(x,y)dx$

聯合分配的期望值

定義：聯合分配的期望值 E[H(X,Y)]

1. 離散的情況：$E[H(X,Y)]=\sum_{\forall x}\sum_{\forall y}H(x,y)P_{XY}(x,y)$
2. 連續的情況：$E[H(X,Y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}H(x,y)P_{XY}(x,y)dydx$

定義：聯合分配中單一變數的期望值

1. 離散： $E[X]=\sum_{\forall x}\sum_{\forall y} xP_{XY}(x,y)$

2. 離散： $E[Y]=\sum_{\forall x}\sum_{\forall x} xP_{XY}(x,y)$

3. 連續： $E[X]=\int_{-\infty}^{\infty} xP_{XY}(x,y)dydx$

4. 連續： $E[Y]=\int_{-\infty}^{\infty} yP_{XY}(x,y)dydx$

共變異數（Covariance，協方差）

定義：共變異數 Cov(X,Y)

$$\sigma_{XY}=Cov(X,Y)=E[(X-\mu_X)(Y-\mu_Y)]$$

定理：共變異數與期望值之關係

$$Cov(X,Y) = E[XY] - E[X]E[Y]$$

定理：相互獨立時的期望值

如果 X, Y 相互獨立 ，則 E[X Y] = E[X] E[Y] 。

相關係數 (Correlation)

定義：

相關係數 $Cor(X,Y) = \rho_{XY}$

$$Cor(X,Y) = \rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

定理：

$$f(x) = x^2$$

定理：

$$|\rho_{XY}| = 1 \Leftrightarrow y = \beta_0 + \beta_1 X$$

實作：相關係數的R 程式

```
> x = sample(1:10, 10)
> x
 [1]  1  8 10  5  3  7  9  4  2  6
> cor(x, x+1)
[1] 1
> cor(x, -x)
[1] -1
```

```
> cor(x, 0.5*x)
[1] 1
> cor(x, 0.5*x+1)
[1] 1
> cor(x, -0.5*x+1)
[1] -1
> y=sample(1:100, 10)
> y
 [1]  4 53 20 68 29 74 17 49 78 62
> cor(x,y)
[1] -0.06586336
>
```

多變數聯合分布的情況

聯合分布與條件機率

定義：如果 X,Y 滿足下列條件，則稱 X, Y 兩者之間獨立：

$$P_{XY}(x,y) = P_X(x) P_Y(y)$$

多個變數的貝氏定理

1. $P(A,B|C)=P(C|A,B)*\frac{P(A,B)}{P(C)}$;
2. $P(A|B,C)=P(B,C|A)*\frac{P(A)}{P(B,C)}$;
3. $P(A,B|C,D)=P(C,D|A,B)*\frac{P(A,B)}{P(C,D)}$;

其他情況可以類推，只要能正確改寫 A , B 為任何隨機變數序列都行。

條件獨立與貝氏定理

假如 A 與 B 在給定 C 的情況下條件獨立，那麼以下算式成立：

$$P(A,B|C)=P(A|C)*P(B|C)=P(C|A)*P(C|B)*\frac{P(A)P(B)}{P(C)^2}$$

結語

兩個位於樣本空間 S 的聯合隨機行為，會導致樣本空間變成其迪卡兒乘積 S×S，其樣本空間大小變成 $|S|^2$。

而 k 個位於樣本空間 S 的聯合隨機行為，會導致樣本空間變成其迪卡兒乘積 S×S×...×S ，其樣本空間大小變成 $|S|^n$。

如果兩個位於不同樣本空間 (S_X, S_Y) 的聯合隨機行為，則會導致樣本空間變成 $S_X \times S_Y$，其樣本空間大小變為 $|S_X||S_Y|$。

此時 X, Y 的機率密度函數將會採用以下的「邊際機率密度函數」之算法，以便將聯合樣本空間 (S_X, S_Y) 中的機率與單一樣本空間 S_X 或 S_Y 中的機率關聯起來。

$$P(X=x) = P(x,*) = \sum_{y \in S_Y} P(x,y)$$

$$P(Y=y) = P(*,y) = \sum_{x \in S_X} P(x,y)$$

最後我們必須強調的是，樣本空間的選擇並沒有一定的標準，您可以視問題的需要來定義樣本空間，通常我們會盡量利用獨立的特性，讓樣本空間越小越好，否則將會很難計算。

附錄 A：常見的機率分布

二項分布 (Binomial distribution)

伯努力試驗

伯努利試驗(Bernoulli trial) 一種只有兩種可能結果的隨機試驗，可以用下列機率分布描述：

$$Pr[X=1]=p$$
$$Pr[X=0]=1-p$$

換句話說、伯努力試驗是一種 YES or NO (1 or 0) 的試驗。舉例而言，像是「丟銅版、生男生女、 一地區某天最高溫是否超過 30 度」等等，都可以用伯努力實驗描述。

二項分布的意義

如果我們進行 n 次的伯努力試驗，而且這些試驗之間是獨立的，那麼我們就可以用二項分布來描述 n 次實驗的可能機率分布。

二項分布公式

分布公式：$f(x)=\binom{n}{x}p^{x}(1-p)^{n-x}$

意義：dbinom(x; n, p)：在 n 次柏努力試驗中有 x 次成功的機率 (已知單次試驗成功機率為 p)。

R 的公式：dbinom(x; n, p) = p(x) = choose(n,x) p^x (1-p)^(n-x)

- R 函數：binom(size=n:樣本數, prob=p:成功機率)
- http://stat.ethz.ch/R-manual/R-patched/library/stats/html/Binomial.html

二項定理：$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$

特性

1. $E(X) = \mu = np$
2. $Var(X) = \sigma^2 = np(1-p) = npq$

動差生成函數：$m_x(t) = ((1-p)+pe^t)^n = (q+pe^t)^n$

習題

1. 請問丟 10 個公平的銅板，有三個正面的機會是多少？
2. 請問丟 n 個公平的銅板，正面次數 <= k 的機率是多少？
3. 請問丟 10 個公平的銅板，得到正面次數的期望值為何？

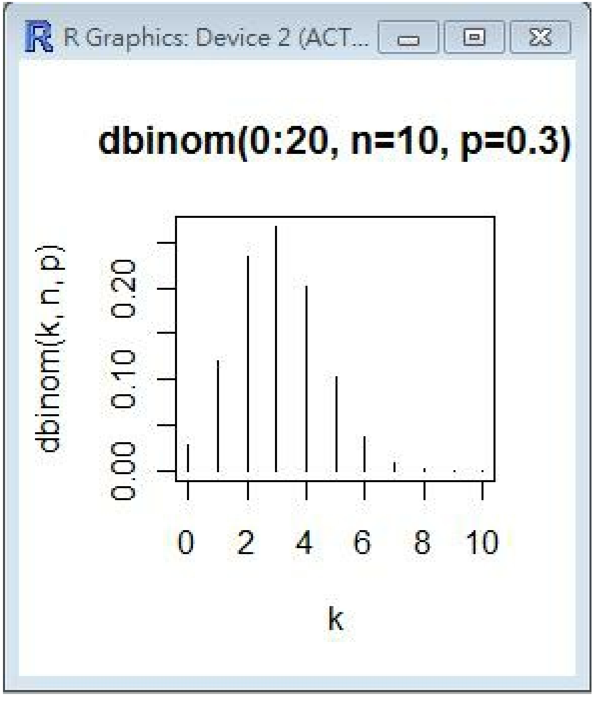
R 程式範例：伯努力試驗

```
> y <- rbinom(50, 25, .4)
> m1 <- mean(y)
> m2 <- sum(y) / 25
> y
 [1] 12  9  9  9 12 11 10 11  5  7  8  7 16  6 12 13  9 12  9 13  7 12 15  8
[25]  9  7 10  4 10 10  9 10 13  8 10 14  8 11 11 10 10  9  7 13  5  5 11 13
[49]  9  8
> m1
[1] 9.72
> m2
[1] 19.44
> m3 <- sum ( (y-m1)^2 ) / 50
> m3
[1] 6.8816
>
```

說明： y 中的每個數字，代表模擬投擲 25 次白努力試驗後，成功的次數有幾次。因此 rbinom(50, 25, .4) 總共進行了 50*25 次白努力試驗。

R 程式範例：二項分布曲線圖

```
> n=10; p=0.3; k=seq(0,n)
> plot(k, dbinom(k,n,p), type='h', main='dbinom(0:20, n=10, p=0.3)', xlab='k')
>
```



二項分布的圖形

R 程式範例：（定理）常態分配可用來逼近二項分布

假如 n 夠大的話，通常只要 n*min(p, 1-p) > 5 就可以採用下列逼近方式

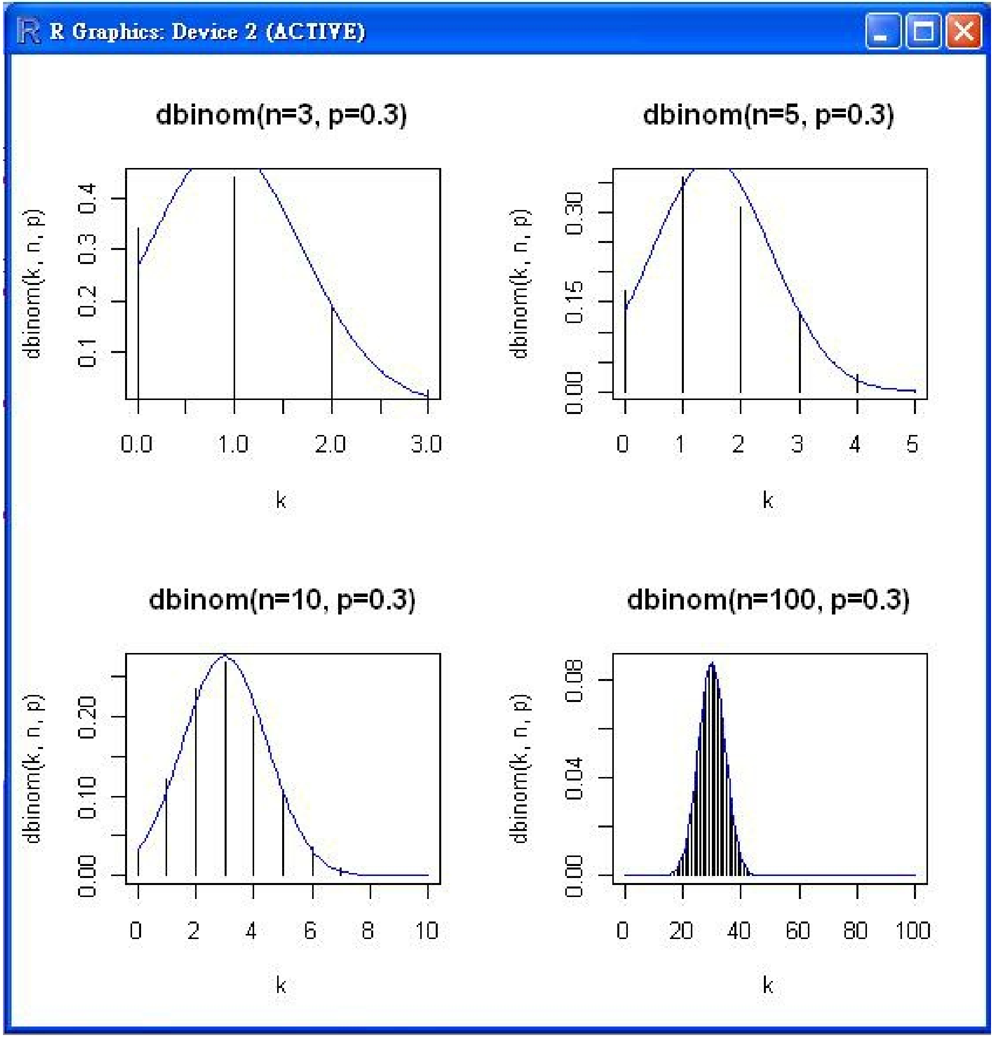
$$binom(n,p) \rightarrow norm(np, np(1-p))$$

原始程式：

```
op=par(mfrow=c(2,2))
n=3; p=0.3; k=seq(0,n)
plot(k, dbinom(k,n,p), type='h', main='dbinom(n=3, p=0.3)', xlab='k')
curve(dnorm(x,n*p,sqrt(n*p*(1-p))), add=T, col='blue')

n=5; p=0.3; k=seq(0,n)
plot(k, dbinom(k,n,p), type='h', main='dbinom(n=5, p=0.3)', xlab='k')
curve(dnorm(x,n*p,sqrt(n*p*(1-p))), add=T, col='blue')

n=10; p=0.3; k=seq(0,n)
plot(k, dbinom(k,n,p), type='h', main='dbinom(n=10, p=0.3)', xlab='k')
curve(dnorm(x,n*p,sqrt(n*p*(1-p))), add=T, col='blue')
```

```r
n=100; p=0.3; k=seq(0,n)
plot(k, dbinom(k,n,p), type='h', main='dbinom(n=100, p=0.3)', xlab='k')
curve(dnorm(x,n*p,sqrt(n*p*(1-p))), add=T, col='blue')
```

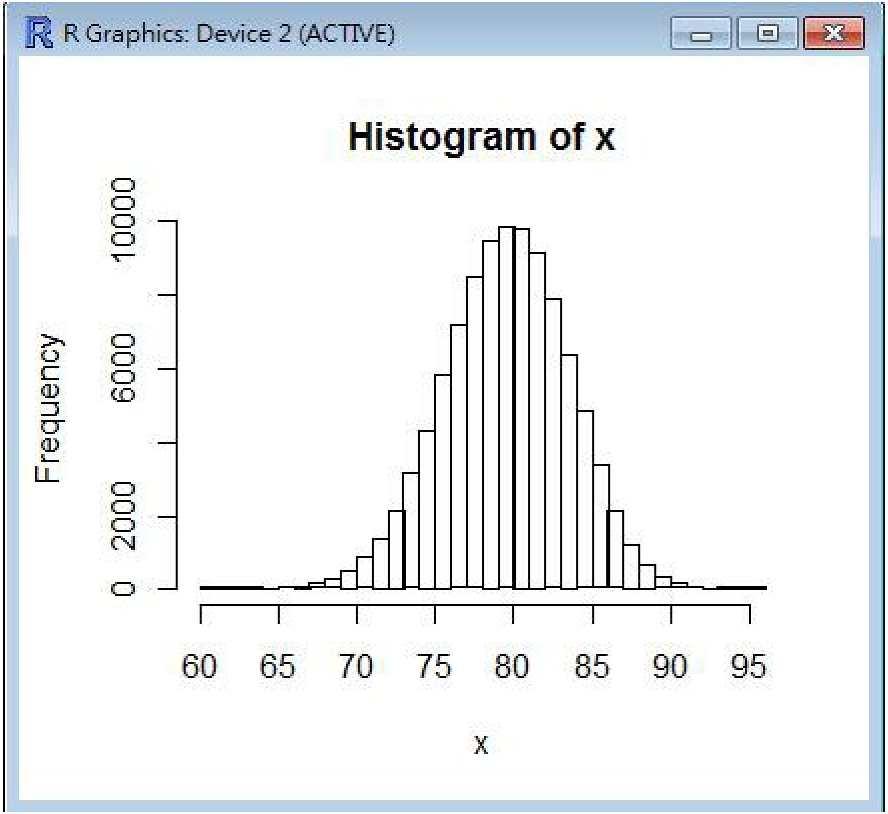
輸出圖形：



R Graphics: Device 2 (ACTIVE)
dbinom(n=3, p=0.3)
dbinom(n=5, p=0.3)
dbinom(n=10, p=0.3)
dbinom(n=100, p=0.3)
dbinom(k, n, p)
k
0.1 0.2 0.3 0.4
0.0 1.0 2.0 3.0
0.00 0.15 0.30
0 1 2 3 4 5
0.00 0.10 0.20
0 2 4 6 8 10
0.00 0.04 0.08
0 20 40 60 80 100

R 程式範例：二項分布統計圖

```
> x = rbinom(100000, 100, 0.8)
> hist(x, nclas=max(x)-min(x)+1)
>
```



參考文獻

- Distributions in the stats package -- http://stat.ethz.ch/R-manual/R-patched/library/stats/html/Distributions.html

- Wikipedia:二項分佈 -- http://zh.wikipedia.org/wiki/%E4%BA%8C%E9%A0%85%E5%88%86%E4%BD%88
- Wikipedia:Binomial_distribution -- http://en.wikipedia.org/wiki/Binomial_distribution

負二項分布（Netative binomial distribution）

公式： $f(x)=\binom{x-1}{r-1}(1-p)^{x-r}p^r$

範圍：r=1,2,3,.... ; x= r, r+1, r+2, ...

意義：要得到第 r 次成功所需要的試驗次數 x;

R 函數： nbinom(size, prob) ; r:size:成功數, p:prob:成功機率

- http://stat.ethz.ch/R-manual/R-patched/library/stats/html/NegBinomial.html

特性

1. $E(X)=r/p$
2. $Var(X)=r(1-p)/p^2=rq/p^2$

動差生成函數： $m_x(t)=\frac{(pe^t)^r}{(1-(1-p)e^t)^r}=\frac{(pe^t)^r}{(1-qe^t)^r}$

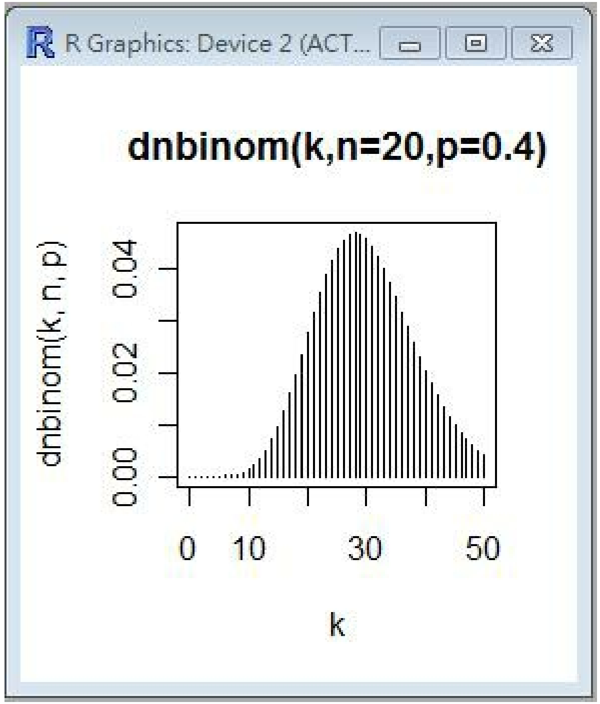
習題

1. 請問丟公平的銅板時，在得到第三次正面的要求下，其投擲次數 x 的機率分布為何？該分布的期望值為何？

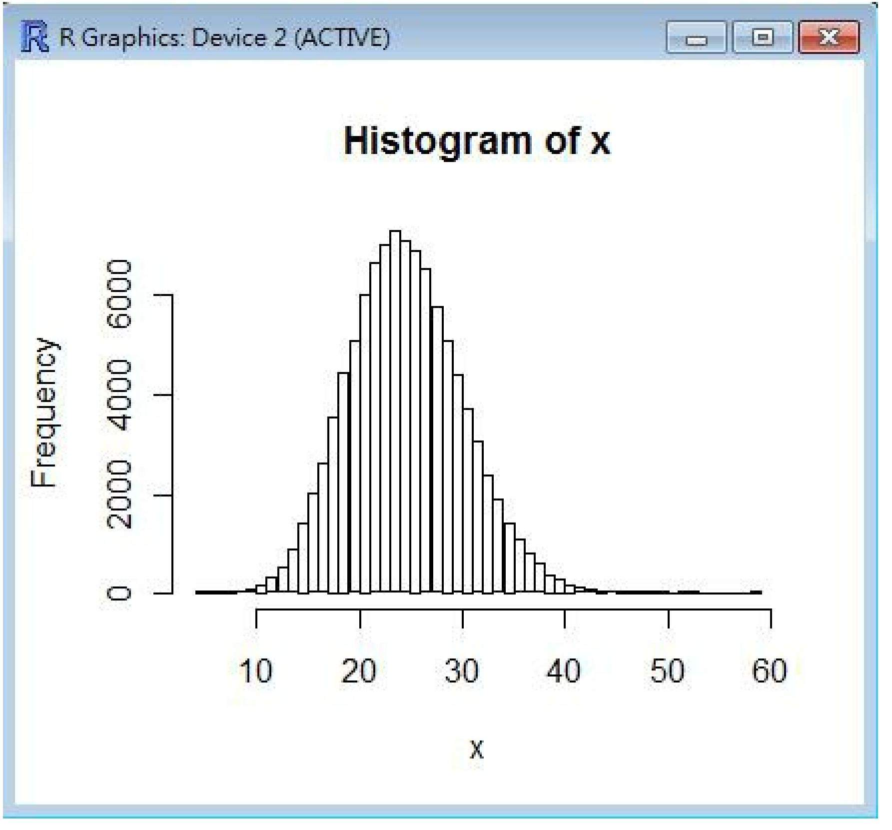
2. 請問丟公平的銅板時，在得到第 r 次正面的要求下，其投擲次數 x 的機率分布為何？該分布的期望值為何？

R 程式範例：負二項分布曲線圖

```
> n=20; p=0.4; k=seq(0,50)
> plot(k, dnbinom(k,n,p), type='h', main='dnbinom(k,n=20,p=0.4)', xlab='k')
>
```




```
> x = rnbinom(100000, 100, 0.8)
> hist(x, nclass=max(x)-min(x)+1)
>
```



R 程式範例（進階）

```
require(graphics)
```

```r
x <- 0:11
dnbinom(x, size = 1, prob = 1/2) * 2^(1 + x) # == 1
126 /  dnbinom(0:8, size  = 2, prob  = 1/2) #- theoretically integer

## Cumulative ('p') = Sum of discrete prob.s ('d');  Relative error :
summary(1 - cumsum(dnbinom(x, size = 2, prob = 1/2)) /
                  pnbinom(x, size  = 2, prob = 1/2))

x <- 0:15
size <- (1:20)/4
persp(x,size, dnb <- outer(x, size, function(x,s) dnbinom(x,s, prob= 0.4)),
      xlab = "x", ylab = "s", zlab="density", theta = 150)
title(tit <- "negative binomial density(x,s, pr = 0.4)  vs.  x & s")

image  (x,size, log10(dnb), main= paste("log [",tit,"]"))
contour(x,size, log10(dnb),add=TRUE)

## Alternative parametrization
x1 <- rnbinom(500, mu = 4, size = 1)
x2 <- rnbinom(500, mu = 4, size = 10)
x3 <- rnbinom(500, mu = 4, size = 100)
```

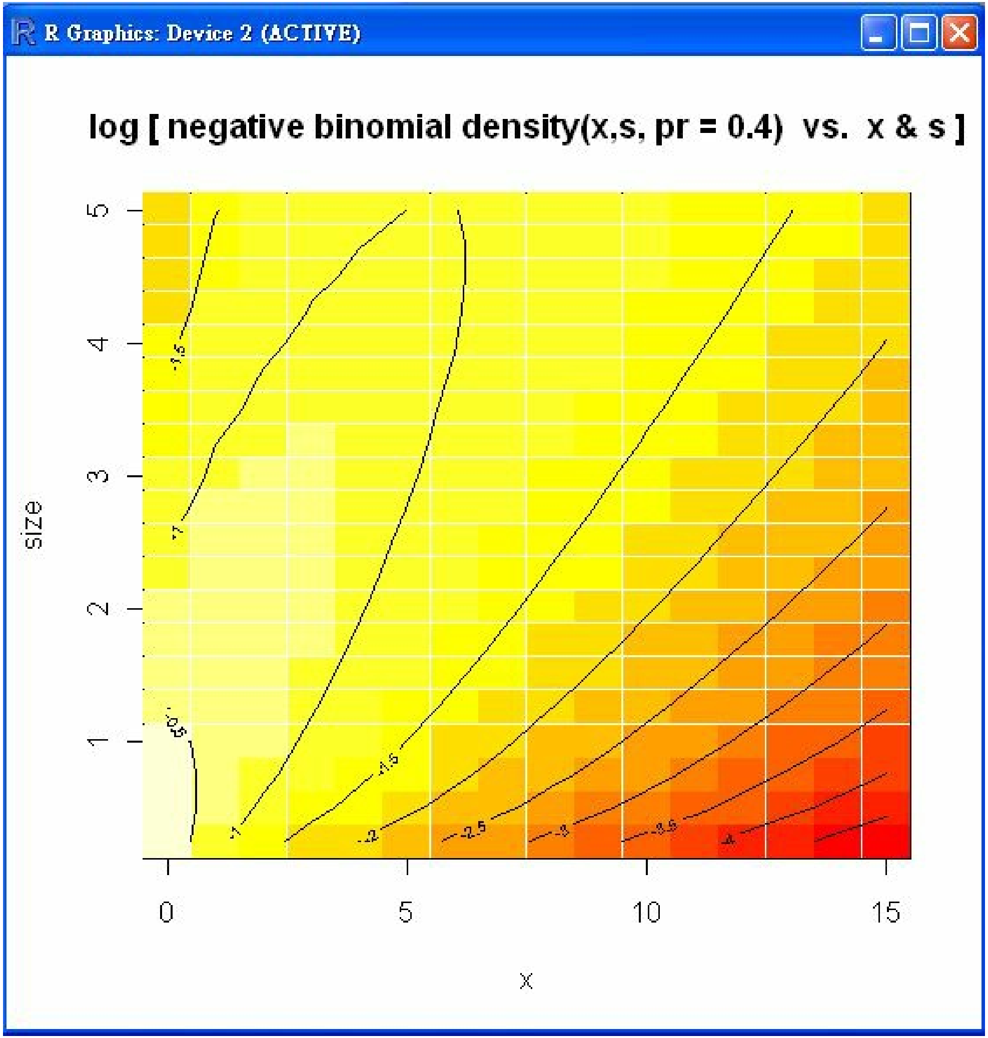
```
h1 <- hist(x1, breaks = 20, plot = FALSE)
h2 <- hist(x2, breaks = h1$breaks, plot = FALSE)
h3 <- hist(x3, breaks = h1$breaks, plot = FALSE)
barplot(rbind(h1 ![](../timg/d7d2ac3ced7f.jpg) counts, h3$counts),
        beside = TRUE, col = c("red","blue","cyan"),
        names.arg = round(h1 ![](../timg/728bbfdb9a46.jpg) breaks)]))
```

執行結果：

```
> require(graphics)
> x <- 0:11
> dnbinom(x, size = 1, prob = 1/2) * 2^(1 + x) # == 1
 [1] 1 1 1 1 1 1 1 1 1 1 1 1
> 126 /  dnbinom(0:8, size  = 2, prob  = 1/2) #- theoretically integer
[1]   504.0   504.0   672.0  1008.0  1612.8  2688.0  4608.0  8064.0 14336.0
>
> ## Cumulative ('p') = Sum of discrete prob.s ('d');  Relative error :
> summary(1 - cumsum(dnbinom(x, size = 2, prob = 1/2)) /
+                   pnbinom(x, size  = 2, prob = 1/2))
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-2.22e-16 -2.22e-16 -2.22e-16 -1.48e-16  0.00e+00  0.00e+00
```

```
>
> x <- 0:15
> size <- (1:20)/4
> persp(x,size, dnb <- outer(x, size, function(x,s) dnbinom(x,s, prob= 0.4)),
+       xlab = "x", ylab = "s", zlab="density", theta = 150)
> title(tit <- "negative binomial density(x,s, pr = 0.4)  vs.  x & s")
>
> image  (x,size, log10(dnb), main= paste("log [",tit,"]"))
> contour(x,size, log10(dnb),add=TRUE)
>
> ## Alternative parametrization
> x1 <- rnbinom(500, mu = 4, size = 1)
> x2 <- rnbinom(500, mu = 4, size = 10)
> x3 <- rnbinom(500, mu = 4, size = 100)
> h1 <- hist(x1, breaks = 20, plot = FALSE)
> h2 <- hist(x2, breaks = h1$breaks, plot = FALSE)
> h3 <- hist(x3, breaks = h1$breaks, plot = FALSE)
> barplot(rbind(h1 ![](../timg/d7d2ac3ced7f.jpg) counts, h3$counts),
+         beside = TRUE, col = c("red","blue","cyan"),
+         names.arg = round(h1 ![](../timg/728bbfdb9a46.jpg) breaks)]))
```

繪圖結果：


R Graphics: Device 2 (ACTIVE)
log [negative binomial density(x,s, pr = 0.4) vs. x & s]
size
x
0
5
10
15
1
2
3
4
5
-1.5
-1
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
-4

參考文獻

- Wikipedia:負二項分布
- Wikipedia:Negative_binomial_distribution

幾何分布 (Geometric distribution)

$$P(X=x)=(1-p)^{x-1}p=q^{x-1}p$$

範圍：r=1,2,3,.... ; x= r, r+1, r+2, ...

意義：第一次成功所需要的試驗次數。

R 函數： geom(prob) ; p:prob:成功機率, x-1:size:失敗次數, q:失敗機率

- R 的公式： $p(x)=p(1-p)^{x}$
- R 當中的 x 代表失敗次數，而非第一次成功的次數，因此 R 當中的 x 相當於上式中的 (x-1)
- http://stat.ethz.ch/R-manual/R-patched/library/stats/html/Geometric.html

特性：

1. $E[X]=1/p$

2. $Var(X)=\frac{(1-p)}{p^2}=\frac{q}{p^2}$

動差生成函數：$m_x(t)=\frac{pe^t}{1-(1-p)e^t}=\frac{pe^t}{1-qe^t}$

習題

1. 請問丟公平的銅板時，得到第 1 次正面時投擲次數 x 的機率分布為何？該分布的期望值為何？
2. 請問丟公正的骰子時，得到第 1 次 6 點時投擲次數 x 的機率分布為何？該分布的期望值為何？

R 程式範例：曲線圖

```
p=0.7; k=seq(0,10)
plot(k, dgeom(k, p), type='h', main='dgeom(p=0.5)', xlab='k')
```

R 程式範例：

```
qgeom((1:9)/10, prob = .2)
Ni <- rgeom(20, prob = 1/4); table(factor(Ni, 0:max(Ni)))
```

執行結果：

```
> qgeom((1:9)/10, prob = .2)
[1]  0  0  1  2  3  4  5  7 10
> Ni <- rgeom(20, prob = 1/4); table(factor(Ni, 0:max(Ni)))


 0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17
 4  5  3  2  0  3  1  0  0  0  1  0  0  0  0  0  0  1
>
```

參考

- Wikipedia:幾何分佈
- Wikipedia:Geometric_distribution

超幾何分布（Hypergeometric distribution）

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

意義：N 個球中有白球有 r 個，黑球 N-r 個，取出 n 個球，其中有 x 個白球的機率; (取後不放回)

R 函數： hyper(m,n,k) = choose(m, x) choose(n, k-x) / choose(m+n, k)

- R 函數的意義：m+n 個球中有白球有 m 個，黑球 n 個，取出 k 個球，其中有 x 個白球的機率; (取後不放回)
- R 的網址：http://stat.ethz.ch/R-manual/R-patched/library/stats/html/Hypergeometric.html
- 課本與 R 之間對應公式：N=>m+n; n=>k; r=>m
- R 的公式：$P(X=x;m,n,k)=\frac{\binom{m}{x}\binom{n}{k-x}}{\binom{m+n}{k}}$

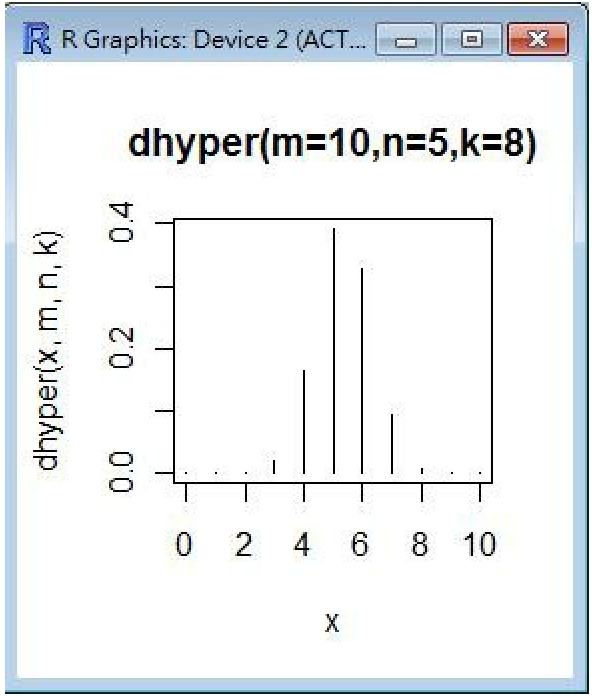
特性

1. $E[X]=k(\frac{m}{m+n})$
2. $Var(X)=k(\frac{n}{m+n})(\frac{m}{m+n})(\frac{m+n-k}{m+n-1})$

動差生成函數：$m_x(t)=???$

R 程式範例：曲線圖

```
m=10; n=5; k=8
x=seq(0,10)
plot(x, dhyper(x, m, n, k), type='h', main='dhyper(m=10,n=5,k=8)', xlab='x')
```



R 程式範例：

```r
m <- 10; n <- 7; k <- 8
x <- 0:(k+1)
rbind(phyper(x, m, n, k), dhyper(x, m, n, k))
all(phyper(x, m, n, k) == cumsum(dhyper(x, m, n, k)))# FALSE
```

```r
## but error is very small:
signif(phyper(x, m, n, k) - cumsum(dhyper(x, m, n, k)), digits=3)
```

執行結果：

```
> m <- 10; n <- 7; k <- 8
> x <- 0:(k+1)
> rbind(phyper(x, m, n, k), dhyper(x, m, n, k))
     [,1]         [,2]       [,3]     [,4]      [,5]      [,6]      [,7]
[1,]    0 0.0004113534 0.01336898 0.117030 0.4193747 0.7821884 0.9635952
[2,]    0 0.0004113534 0.01295763 0.103661 0.3023447 0.3628137 0.1814068
           [,8]       [,9] [,10]
[1,] 0.99814891 1.00000000     1
[2,] 0.03455368 0.00185109     0
> all(phyper(x, m, n, k) == cumsum(dhyper(x, m, n, k)))# FALSE
[1] FALSE
> ## but error is very small:
> signif(phyper(x, m, n, k) - cumsum(dhyper(x, m, n, k)), digits=3)
 [1]  0.00e+00  0.00e+00  1.73e-18  0.00e+00 -5.55e-17  1.11e-16  2.22e-16
 [8]  2.22e-16  2.22e-16  2.22e-16
>
```

參考文獻

- Wikipedia:超幾何分布
- Wikipedia:Hypergeometric_distribution

布瓦松分布 (Poisson distribution)

意義：在單位時間內，事件出現平均 λ 次的機率分布。

公式： $f(x) = \frac{e^{-k} k^x}{x!}$

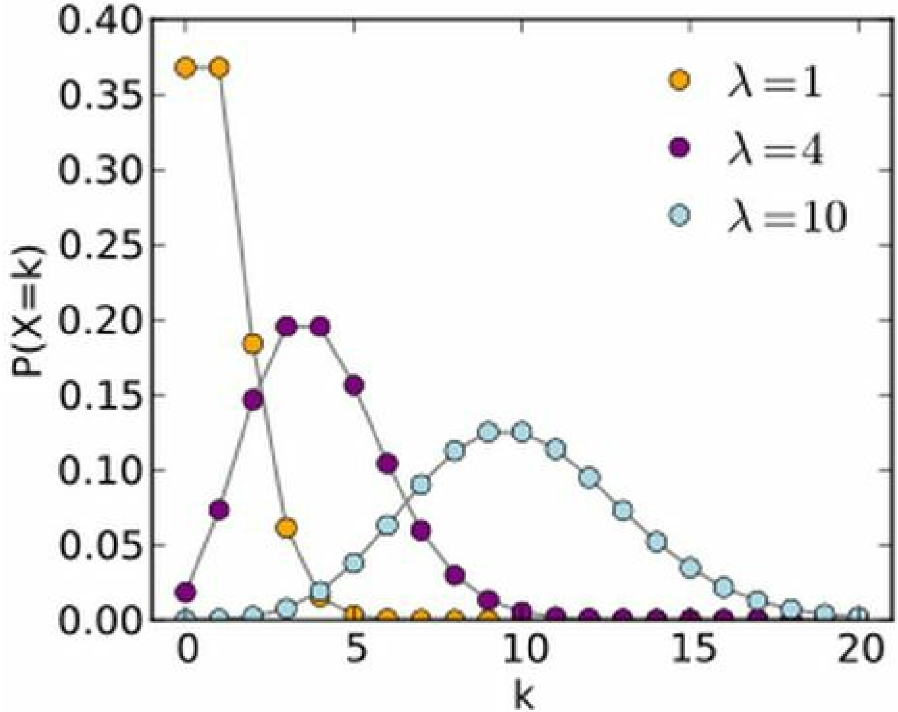
R 的公式： $p(x) = \lambda^x e^{-\lambda} / x!$

R 函數：pois(λ:事件平均出現次數)

變數意義： $k = \lambda$

Poisson

Probability mass function



The horizontal axis is the index *k*, the number of occurrences. The function is only defined at integer values of *k*. The connecting lines are only guides for the eye.

圖、布瓦松分布

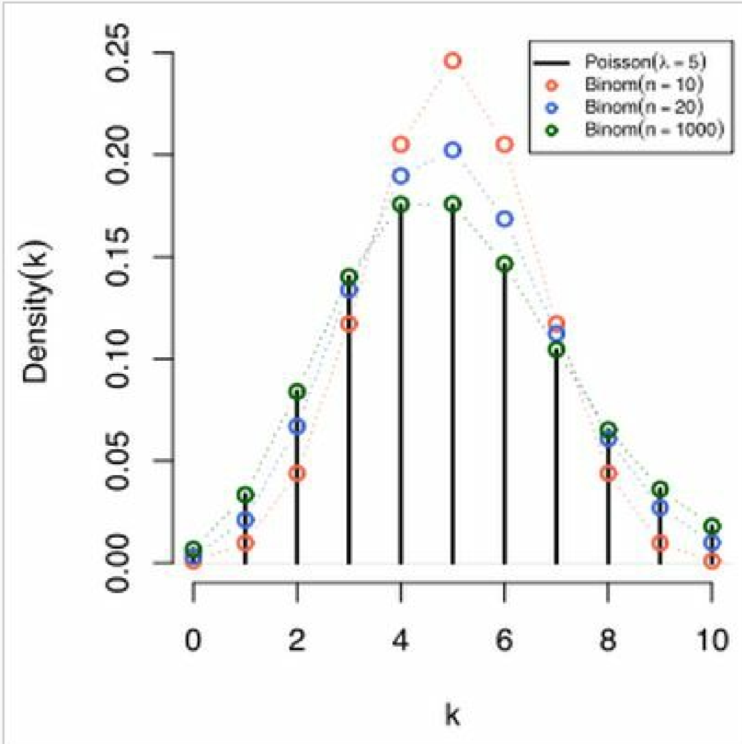
來源：http://en.wikipedia.org/wiki/Poisson_distribution

特性：

布瓦松分布可以與泰勒展開式中的 Maclaurin 級數對映起來，所謂的 Maclaurin 級數就是泰勒展開式在 0 點的展開式。

> If the Taylor series is centered at zero, then that series is also called a Maclaurin series, named after the Scottish mathematician Colin Maclaurin, who made extensive use of this special case of Taylor series in the 18th century.

$$e^x = 1 + x + x^2/2! + x^3/3! + ... + x^k/k! + ...$$



Comparison of the Poisson distribution (black lines) and the binomial distribution with n=10 (red circles), n=20 (blue circles), n=1000 (green circles). All distributions have a mean of 5. The horizontal axis shows the number of events *k*. Notice that as n gets larger, the Poisson distribution becomes an increasingly better approximation for the binomial distribution with the same mean.

圖、布瓦松分布是二項分布 n 趨近無限大的極限情況
來源：http://en.wikipedia.org/wiki/File:Binomial_versus_poisson.svg

布瓦松分配的公式來源

布瓦松分配可視為二項分配的極限形式，當 binom(n, p) 當中 n 趨近於無限大，而 p 非常小的時候，就會趨近布瓦松分配。

關鍵公式：$\lim_{n \to \infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$

證明過程：

$$\lim_{n \to \infty} P(X_n = k) = \lim_{n \to \infty}\binom{n}{k}p^k(1-p)^{n-k} = \lim_{n \to \infty}\frac{n!}{(n-k)!k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}.$$

$$= \lim_{n \to \infty}\underbrace{\left[\frac{n!}{n^k(n-k)!}\right]}_{A_n}\left(\frac{\lambda^k}{k!}\right)\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\to \exp(-\lambda)}\underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{\to 1}.$$

$$= \left[\lim_{n \to \infty} A_n\right]\left(\frac{\lambda^k}{k!}\right)\exp(-\lambda) \to \left(\frac{\lambda^k}{k!}\right)\exp(-\lambda).$$

其中的 An 趨近於 1 ，證明如下：

$$A_n = \frac{n!}{n^k(n-k)!} = \frac{n\cdot(n-1)\cdots(n-(k-1))}{n^k} = 1\cdot(1-\frac{1}{n})\cdots(1-\frac{k-1}{n}) \to 1\cdot 1\cdots 1 = 1 \cdot$$

期望值與變異數

1. $E(X) = k = \lambda$
2. $Var(X) = k = \lambda$

動差生成函數： $m_x(t) = e^{k(e^t-1)} = e^{\lambda(e^t-1)}$

習題：

習題：抽血時白血球數量的問題

問題： 假如現在從你身上抽一滴血，請回答下列兩個問題。

1. 請定義一個隨機變數 X 代表那滴血中的白血球數量。

提示： 樣本空間 S = 此時此刻你身上的所有白血球 = {w1,w2,......,wn}

2. 請算出一滴血液中有三顆白血球的機率，假設該滴血液占你總血量的 1/1000。

解答 1：

X(A) = |A|

說明：

A 是一個事件，也就是白血球的樣本空間 S 的子集合，例如： A = {w1, w5, w9}

|A| 代表 A 集合的大小，也就是元素個數，舉例而言：

如果 A = {w1, w5, w9} ，那麼 |A| = 3

如果 B = {w2, w8}，那麼 |B| = 2

如果 C = {}，那麼 |C| = 0

如果 D = S，那麼 |D| = n

解答 2：

P(X=3) = P({A | X(A) = 3}) = P({{w1, w2, w3}) + P({w1, w2, w4}) + ...

假如任一顆白血球被抽到的機率等於該滴血液佔全身血液的比率，由於該滴血液佔總血量的 1/1000，所以給顆白血球被抽到的機率為 1/1000。

而且假設這些白血球沒有智慧，也不會聚合在一起，因此相互之間獨立，那麼由於每顆白血球被抽到的機率為 1/1000，因此 P(w1) = P(w2) = ... P(wn) = 1/1000。

那麼初步想法是 P(w1w3) = P(w1) * P(w3) = 1/1000 * 1/1000 。

但是上述的想法有個小問題，那就是該情況代表其它白血球都沒被抽到，因此所謂的 P(w1w3) 真正的意思應該是

$$P(w_1\overline{w_2}w_3\overline{w_4}....\overline{w_n}) = (\frac{1}{1000})^2(\frac{999}{1000})^{n-2}$$

所以 P(X=3) 應該算法如下：

$$P(X=3) = P(\{A|X(A)=3\}) = (\frac{1}{1000})^3(\frac{999}{1000})^{n-3}\binom{n}{3}$$

推而廣之，P(X=k) 的機率之算法如下：

$$P(X=k) = P(\{A|X(A)=k\}) = (\frac{1}{1000})^k(\frac{999}{1000})^{n-k}\binom{n}{k}$$

事實上，這個題目的機率分布就是下一章的二項分布，如下所示：

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

而且、當 n 趨近於無限大時，這個分布將會趨近於布瓦松分布，如下所示：

$$P(X=k)=\lambda^k e^{-\lambda}/k!$$

其中的 λ 之意義為，在單位時間 (或單位面積、體積) 內，事件的出現次數平均為 λ 次。

習題：假設每 1CC 的血所含的白血球平均為 10 顆，那麼請問你抽 1CC 的血時，抽到 8 顆白血球的機率是多少。

解答：

λ = 10，因此布瓦松分布為 $p(x)=10^x e^{-10}/x!$ ，將 x=8 代入，得到 $p(8)=10^8 e^{-10}/8!$

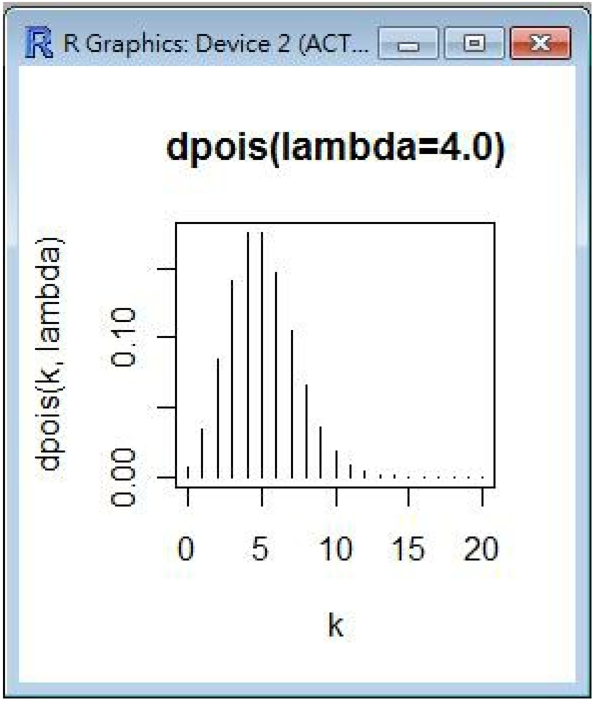
其數值可以用 R 軟體計算，如下所示：

```
> ?dpois
```

```
> dpois(8, 10)
[1] 0.112599
> 10^8*exp(-10)/prod(1:8)
[1] 0.112599
```

R 程式範例：曲線圖

```
lambda=5.0; k=seq(0,20);
plot(k, dpois(k, lambda), type='h', main='dpois(lambda=4.0)', xlab='k')
```



R 程式範例：

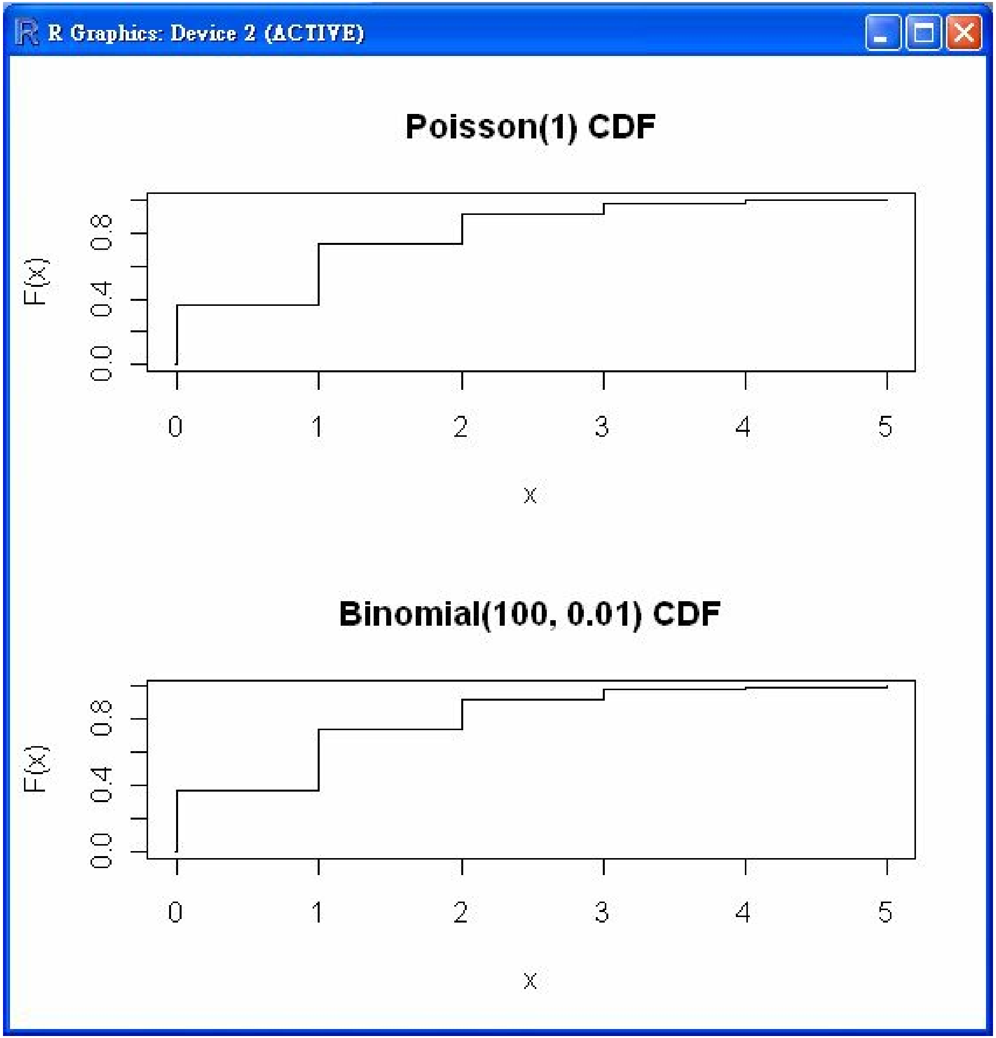
```
require(graphics)


-log(dpois(0:7, lambda=1) * gamma(1+ 0:7)) # == 1
Ni <- rpois(50, lambda = 4); table(factor(Ni, 0:max(Ni)))
```

```
1 - ppois(10*(15:25), lambda=100)  # becomes 0 (cancellation)
    ppois(10*(15:25), lambda=100, lower.tail=FALSE)  # no cancellation

par(mfrow = c(2, 1))
x <- seq(-0.01, 5, 0.01)
plot(x, ppois(x, 1), type="s", ylab="F(x)", main="Poisson(1) CDF")
plot(x, pbinom(x, 100, 0.01),type="s", ylab="F(x)",
     main="Binomial(100, 0.01) CDF")
```

執行結果：

```
> require(graphics)
>
> -log(dpois(0:7, lambda=1) * gamma(1+ 0:7)) # == 1
[1] 1 1 1 1 1 1 1 1
> Ni <- rpois(50, lambda = 4); table(factor(Ni, 0:max(Ni)))

 0  1  2  3  4  5  6  7  8
 1  3  6  8 11 11  4  3  3
>
```

```
> 1 - ppois(10*(15:25), lambda=100)  # becomes 0 (cancellation)
 [1] 1.233094e-06 1.261664e-08 7.085799e-11 2.252643e-13 4.440892e-16
 [6] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
[11] 0.000000e+00
>     ppois(10*(15:25), lambda=100, lower.tail=FALSE)  # no cancellation
 [1] 1.233094e-06 1.261664e-08 7.085800e-11 2.253110e-13 4.174239e-16
 [6] 4.626179e-19 3.142097e-22 1.337219e-25 3.639328e-29 6.453883e-33
[11] 7.587807e-37
>
> par(mfrow = c(2, 1))
> x <- seq(-0.01, 5, 0.01)
> plot(x, ppois(x, 1), type="s", ylab="F(x)", main="Poisson(1) CDF")
> plot(x, pbinom(x, 100, 0.01),type="s", ylab="F(x)",
+      main="Binomial(100, 0.01) CDF")
>
```


R Graphics: Device 2 (ACTIVE)
Poisson(1) CDF
F(x)
0.0 0.4 0.8
0 1 2 3 4 5
x
Binomial(100, 0.01) CDF
F(x)
0.0 0.4 0.8
0 1 2 3 4 5
x

參考文獻

- Wikipedia:卜瓦松分佈
- Wikipedia:Poisson_distribution

均勻分布 (Uniform distribution)

意義：在範圍 (a,b) 內的出現機會均等。

R 函數：unif(a:min, b:max)

課本公式： $f(x) = \frac{1}{b-a}$

R 的公式：f(x) = 1/(max-min)

變數意義：a:min 範圍下限, b:max 上限

R 程式範例

```
> op=par(mfrow=c(2,2))
> curve(dunif(x, 0, 1), -2, 10)
> curve(dunif(x, 1, 5), -2, 10)
```

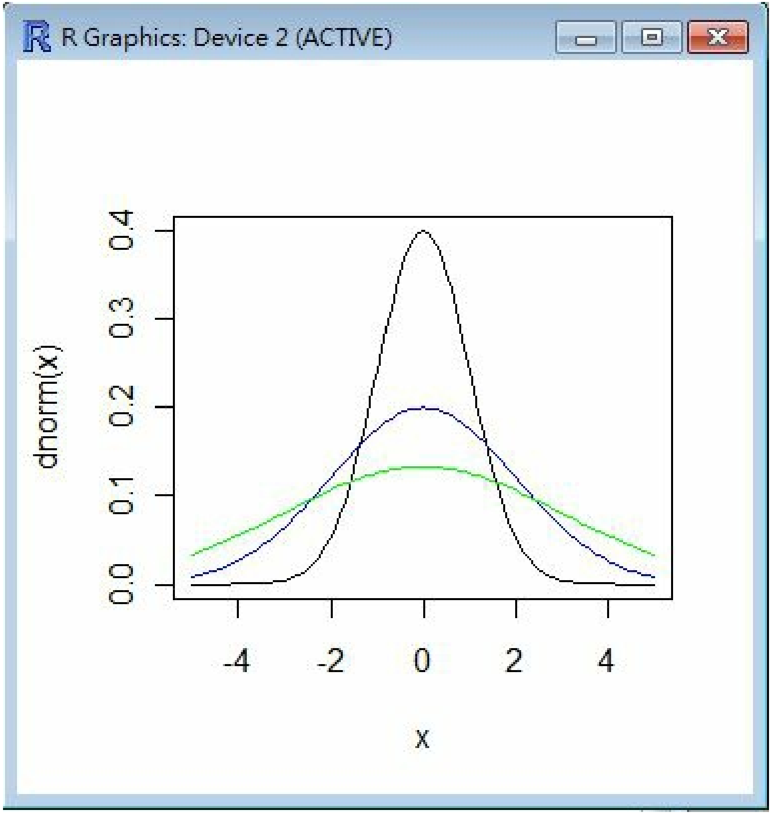
```
> curve(dunif(x, -1, 9), -2, 10)
> curve(dunif(x, 10, 110), 0, 200)
>
```

常態分布 (Normal Distribution)

意義：有誤差的對稱性分布形式，中央高兩邊低的形式。

繪圖：用 R 繪製標準差為 1, 2, 3, 的常態分布。

```
curve(dnorm(x), -5, 5, col="black")
curve(dnorm(x, sd=2), -5, 5, col="blue", add=T)
curve(dnorm(x, sd=3), -5, 5, col="green", add=T)
```



R 函數：norm(mean, sd)

公式： $\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}[(x-\mu)/\sigma]^2}$

R 的公式：f(x) = 1/(√(2 π) σ) e^-((x - μ)^2/(2 σ^2))

重要性：根據中央極限定理，任何 n 個獨立樣本的平均值趨近於常態分布。

中央極限定理： $x_1 + x_2 + ... + x_n \to N(\mu, \delta / \sqrt{n})$

標準差

1 標準差： $P[-1\sigma < X - \mu < 1\sigma] \simeq 0.68$ 。

2 標準差： $P[-2\sigma < X - \mu < 2\sigma] \simeq 0.95$ 。

3 標準差： $P[-3\sigma < X - \mu < 3\sigma] \simeq 0.997$ 。

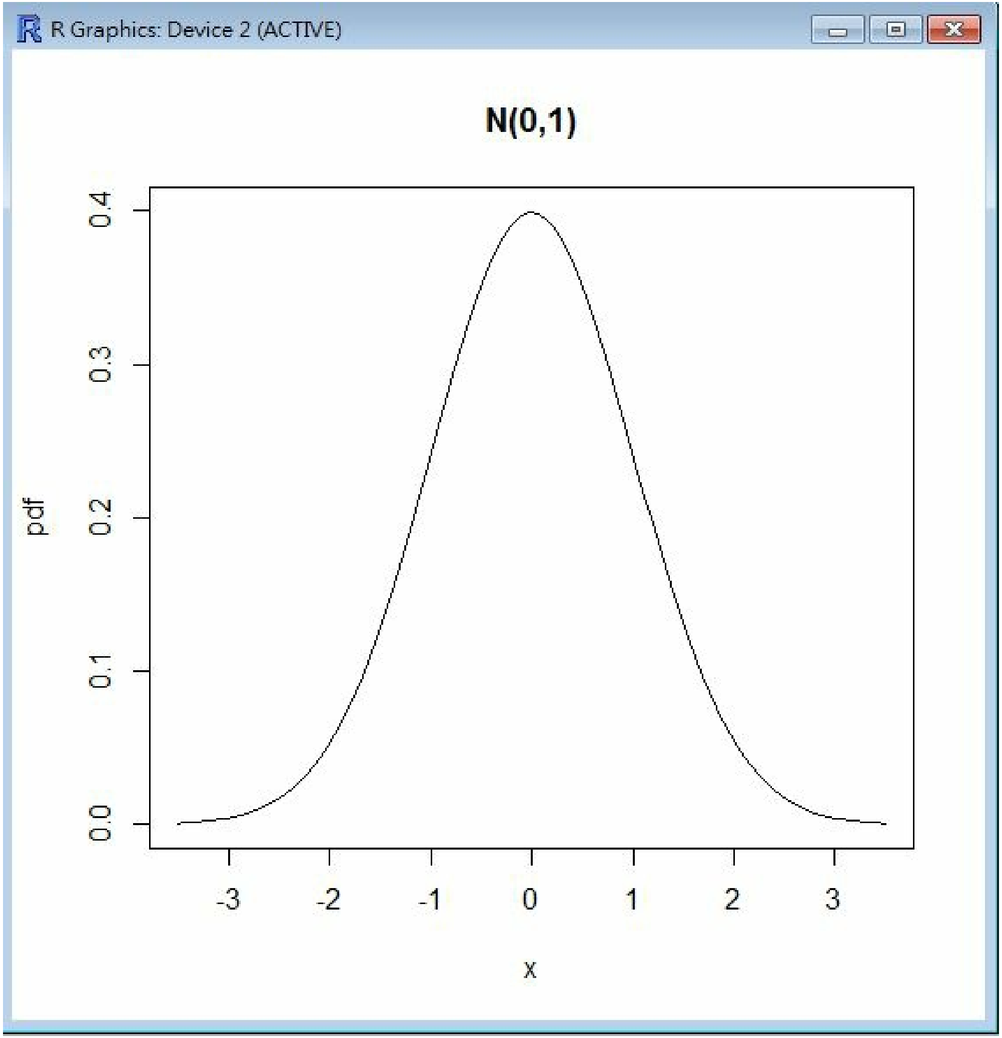
```
> pnorm(1)-pnorm(-1)
[1] 0.6826895
> pnorm(2)-pnorm(-2)
[1] 0.9544997
> pnorm(3)-pnorm(-3)
[1] 0.9973002
> pnorm(4)-pnorm(-4)
[1] 0.9999367
> pnorm(5)-pnorm(-5)
```

```
[1] 0.9999994
> pnorm(6)-pnorm(-6)
[1] 1
```

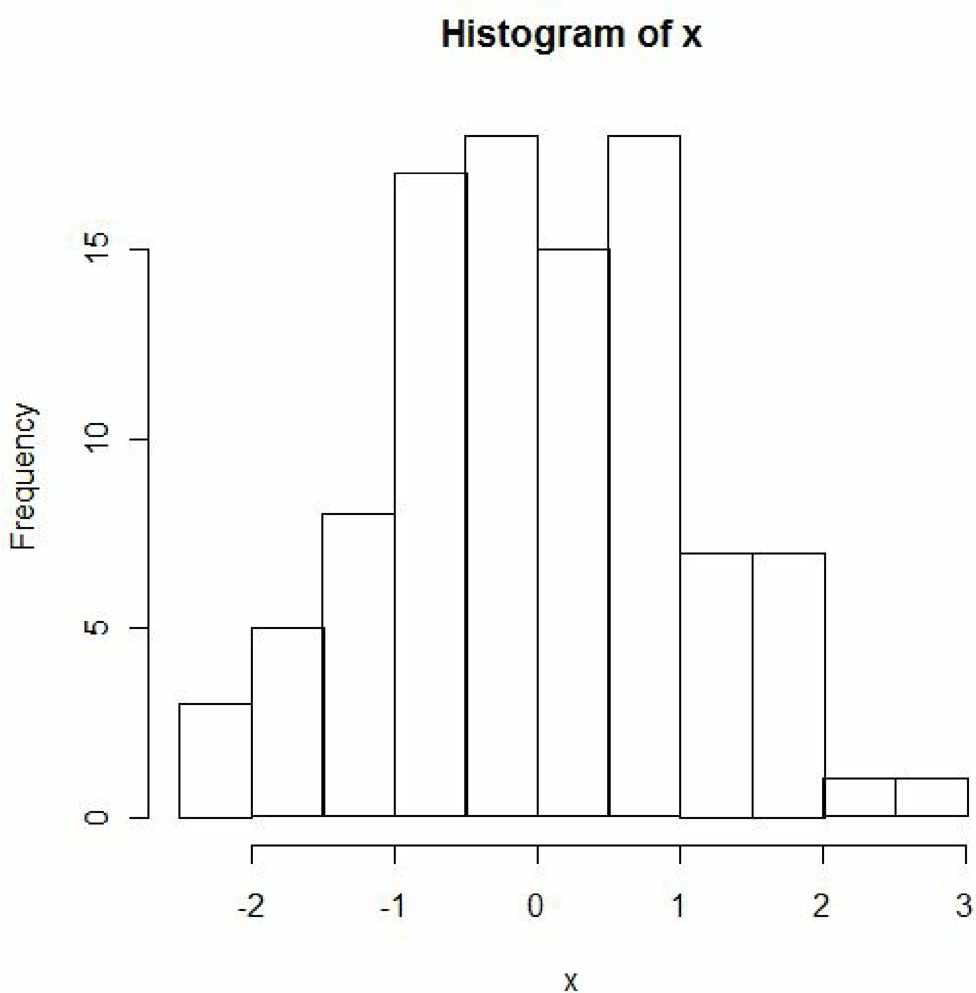
所以現在大家應該知道「工業管理學」上「六標準差」的要求，是很嚴苛的了吧！

R 程式範例

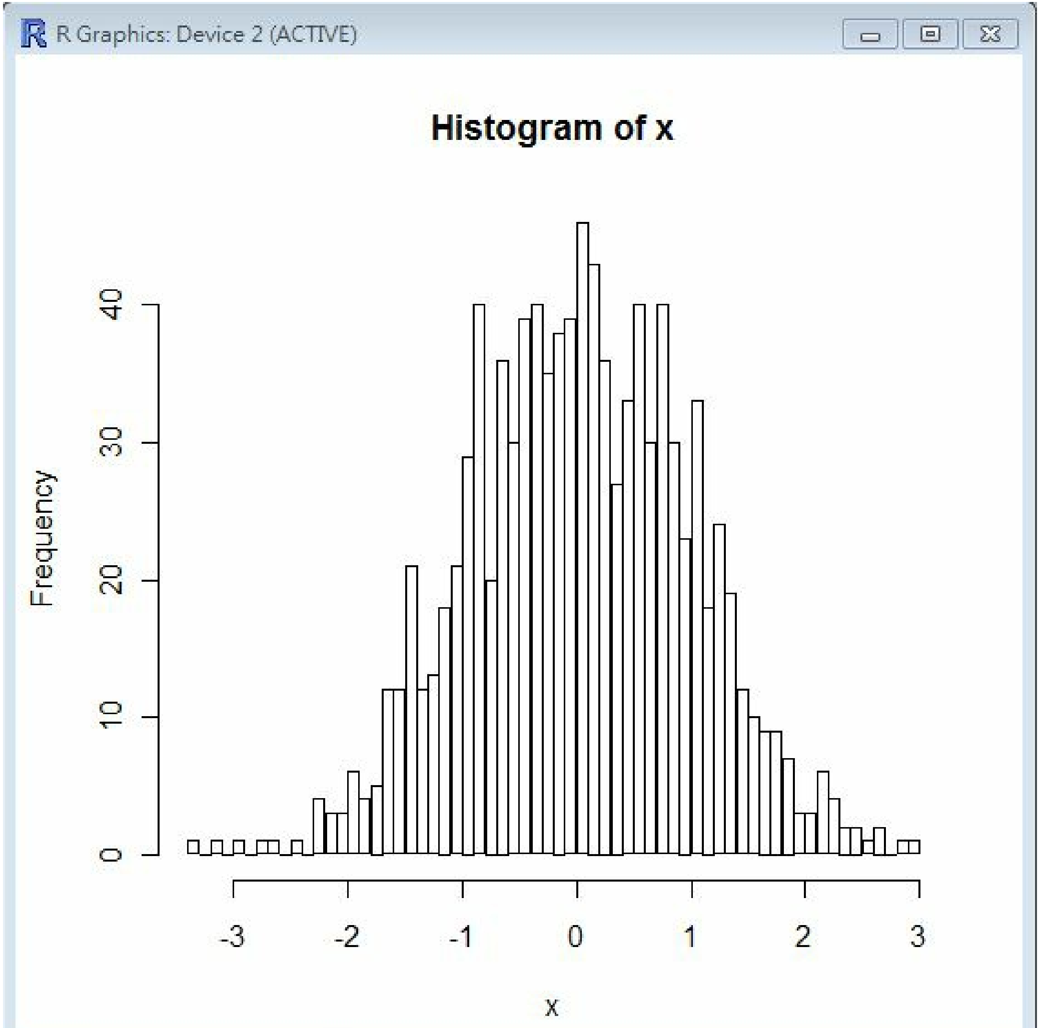
```
> dnorm(0)
[1] 0.3989423
> dnorm(0.5)
[1] 0.3520653
> dnorm(2.5)
[1] 0.0175283
> curve(dnorm(x), from = -3.5, to = 3.5, ylab="pdf", main="N(0,1)")
>
```


R Graphics: Device 2 (ACTIVE)
N(0,1)
pdf
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3
x

```
> x = rnorm(100)
> hist(x, nclass=8)
>
```


Histogram of x
Frequency
0
5
10
15
-2
-1
0
1
2
3
x

```
> x = rnorm(1000)
> hist(x, nclass=50)
>
```


R Graphics: Device 2 (ACTIVE)
Histogram of x
Frequency
0
10
20
30
40
-3
-2
-1
0
1
2
3
x