# 附錄 02：統計機率分布函數

## 系統性記憶

R 語言有一系列的函數可以幫使用者產出不同機率分布的數值、機率值或特定值，這一系列函數的外觀很相似，是四個英文字母與機率分布名稱的排列組合，因此在一一展示之前我們先將**英文字母**與**機率分布名稱**拆開成兩個表讓讀者可以更有系統性地瞭解這些函數：

英文字母的部分：

|  |  |
| --- | --- |
| 英文字母 | 意義 |
| d | **density** 的縮寫，回傳機率密度值 |
| p | **probability** 的縮寫，回傳累積機率值 |
| q | **quantile** 的縮寫，回傳分位數 |
| r | **random** 的縮寫，回傳隨機值 |

機率分布的部分：

|  |  |
| --- | --- |
| 機率分布名稱 | 意義 |
| unif | 均勻分布 |
| norm | 常態分布 |
| binom | 二項式分布 |
| pois | Poisson 分布 |
| chisq | 卡方分布 |

上表的每個英文字母都可以下表的機率分布組合，像是 d 加上 unif 的 dunif() 函數就是回傳均勻分布的密度值；因此我們有 4X5 共 20 組靈活的函數能夠應用。

## 均勻分布

預設均勻分布的最小值為 0，最大值為 1，如果想要調整可以分別修改 min 參數與 max 參數。

### dunif() 函數

回傳均勻分布的機率密度值：

> x <- seq(from = -1, to = 2, by = 0.01)  
> y <- dunif(x)  
> plot(x, y, type = "l", ylab = "Density")

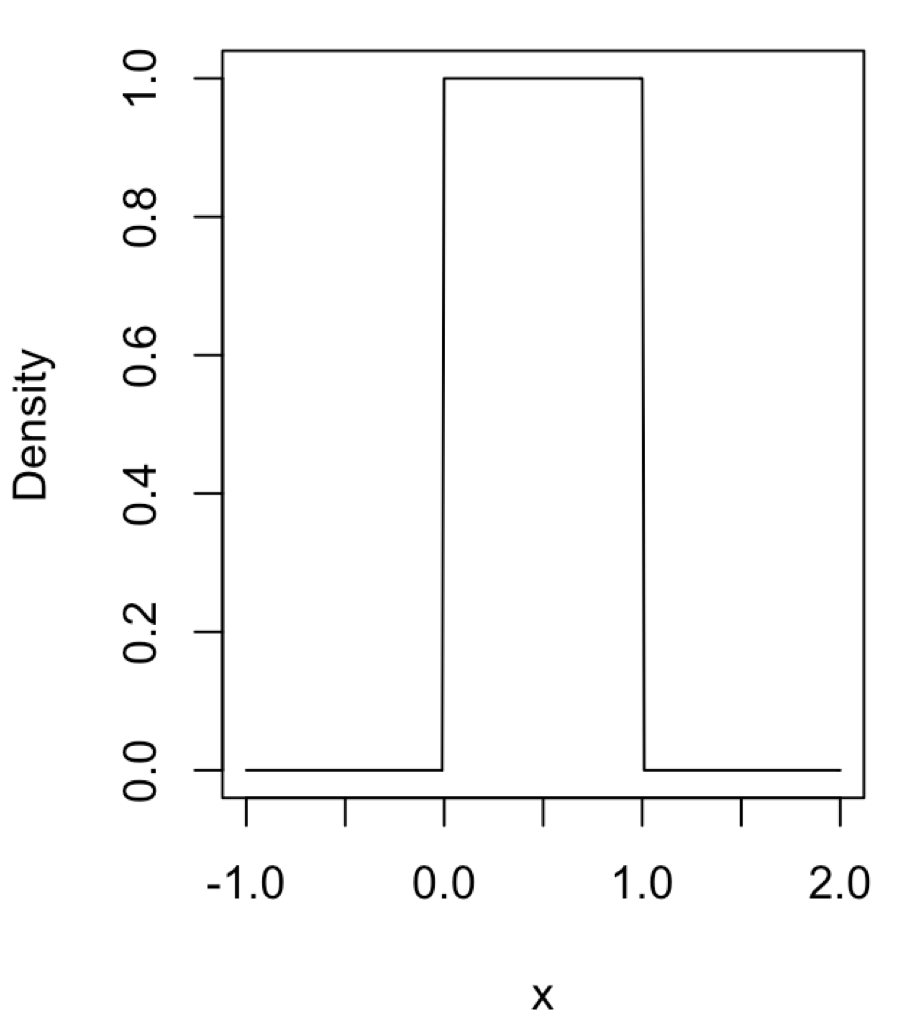


圖 a2-1 均勻分布的機率密度值

### punif() 函數

回傳對應輸入的均勻分布累積機率值：

> punif(0.9)  
[1] 0.9

### qunif() 函數

回傳對應累積機率值的均勻分布輸入：

> qunif(0.9)  
[1] 0.9

### runif() 函數

回傳 n 個符合均勻分布的隨機值：

> x <- runif(1000)  
> hist(x, ylab = "Frequency")

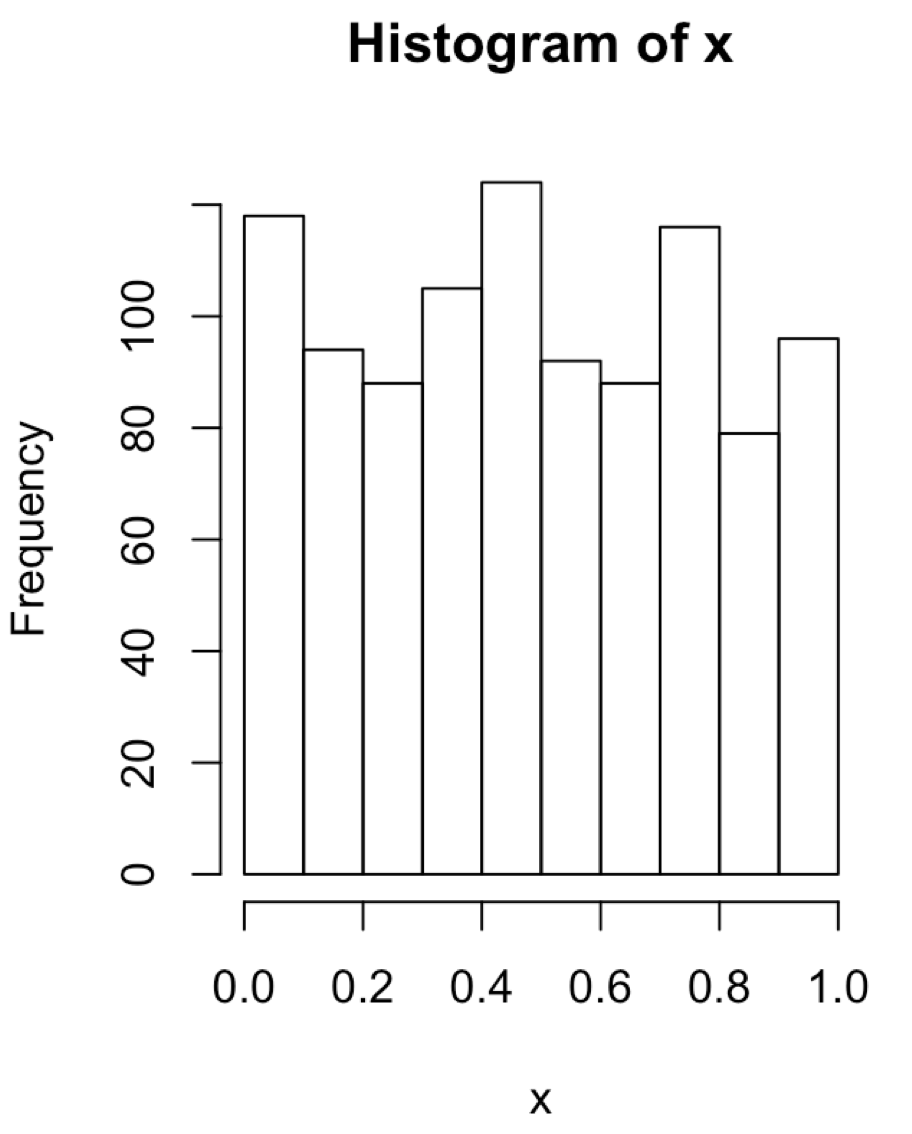


圖 a2-2 均勻分布的隨機值

## 常態分佈

預設常態分布為標準常態分布，平均值為 0，標準差為 1，如果想要調整可以分別修改 mean 參數與 sd 參數。

### dnorm() 函數

回傳常態分布的機率密度值：

> x <- seq(from = -3, to = 3, by = 0.01)  
> y <- dnorm(x)  
> plot(x, y, type = "l", ylab = "Density")

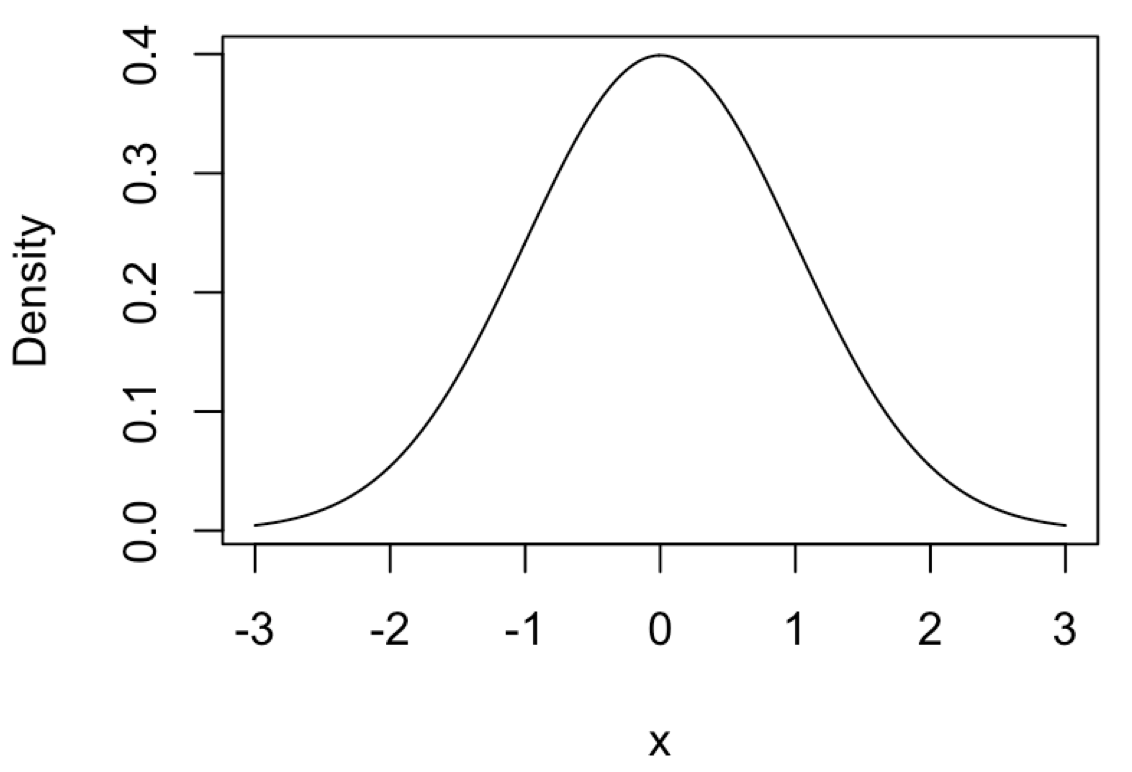


圖 a2-3 常態分布的機率密度值

### pnorm() 函數

回傳對應輸入的常態分布累積機率值：

> pnorm(1.96)  
[1] 0.9750021

### qnorm() 函數

回傳對應累積機率值的常態分布輸入：

> qnorm(0.975)  
[1] 1.959964

### rnorm() 函數

回傳 n 個符合常態分布的隨機值：

> x <- rnorm(1000)  
> hist(x, ylab = "Frequency")

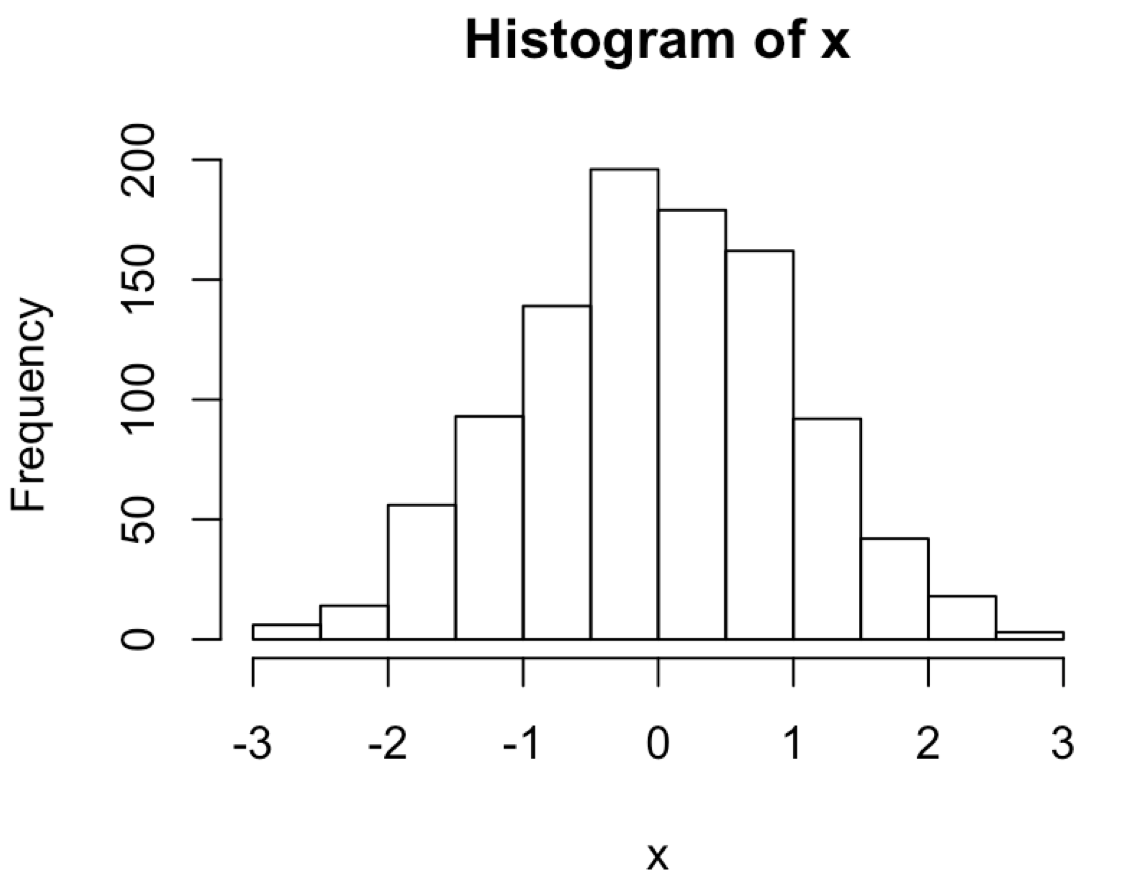


圖 a2-4 常態分布的隨機值

## 二項式分布

二項式分布可以簡單想像成投擲銅板的結果，size 參數代表投擲的次數，prob 參數則代表銅板正面（Head）的機率，一枚公正硬幣意指的就是 prob = 0.5。

### dbinom() 函數

回傳投擲一枚公正硬幣 100 次，出現 0 到 100 次正面的機率分布：

> x <- 0:100  
> y <- dbinom(x, size = 100, prob = 0.5)   
> plot(x, y, type = "l", ylab = "Probability")

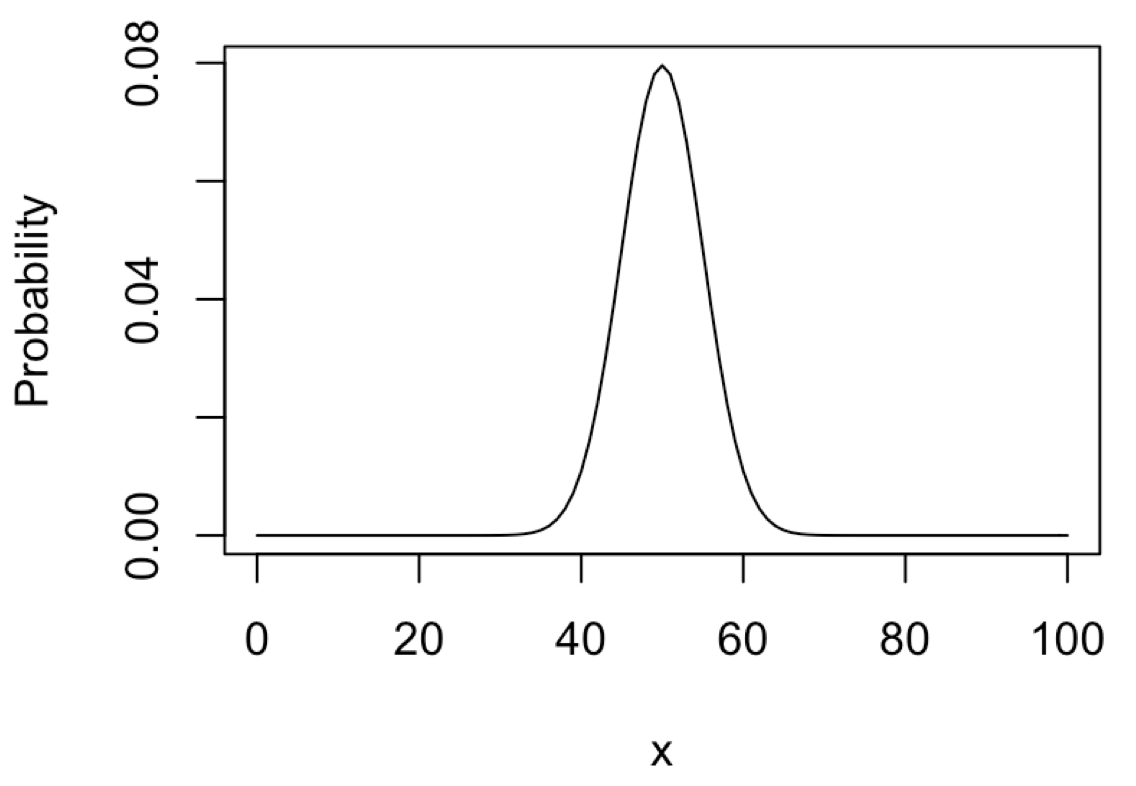


圖 a2-5 二項式分布的機率分布

我們比較三枚硬幣：公正硬幣、擲出正面的機率是 0.7 的硬幣與擲出正面的機率是 0.3 的硬幣，那麼投擲這三枚硬幣 100 次，出現 0 到 100 次正面的機率分布分別是：

> x <- 0:100  
> y1 <- dbinom(x, size = 100, prob = 0.5)  
> y2 <- dbinom(x, size = 100, prob = 0.7)  
> y3 <- dbinom(x, size = 100, prob = 0.3)  
> plot(x, y1, type = "l", ylab = "Probability", ylim = c(0, max(y1, y2, y3)))  
> lines(y2, col = "red")  
> lines(y3, col = "green")

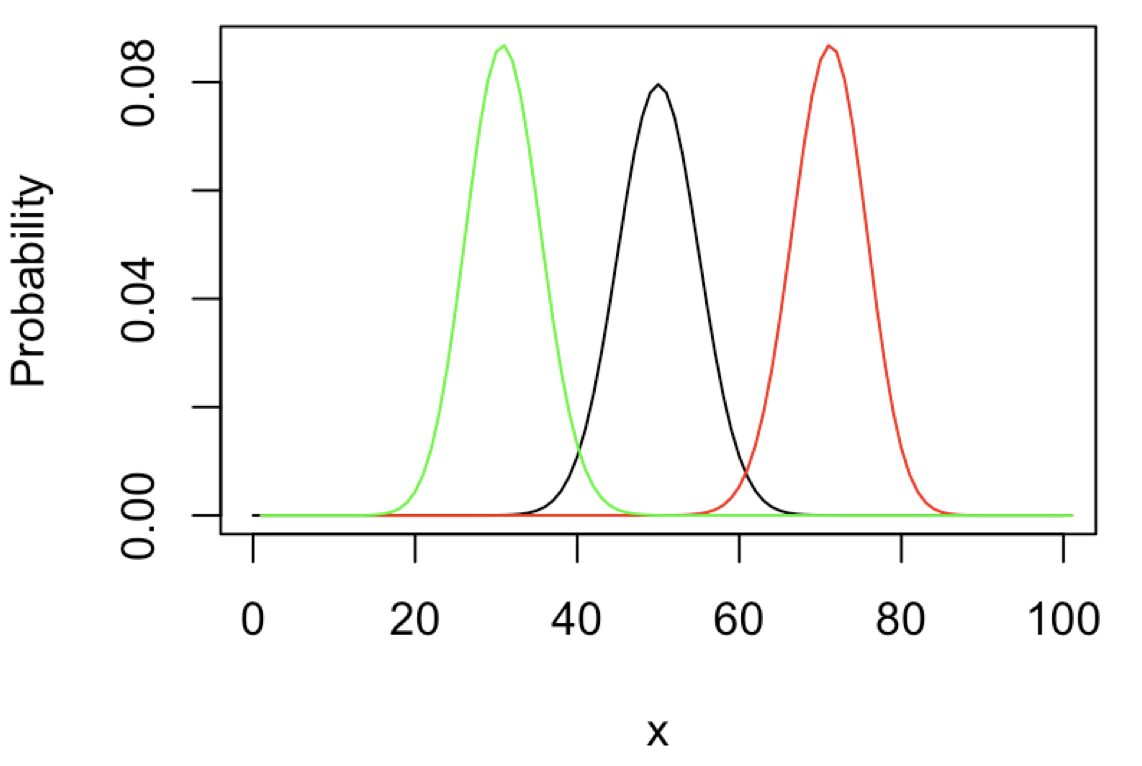


圖 a2-6 二項式分布的機率分布（2）

### pbinom() 函數

回傳對應輸入的二項式分布累積機率值：

> pbinom(50, size = 100, prob = 0.5)  
[1] 0.5397946

### qbinom() 函數

回傳對應累積機率值的二項式分布輸入：

> qbinom(0.53, size = 100, prob = 0.5)  
[1] 50

### rbinom() 函數

回傳 n 個符合二項式分布的隨機值：

> x <- rbinom(1000, size = 100, prob = 0.5)  
> hist(x, ylab = "Frequency")

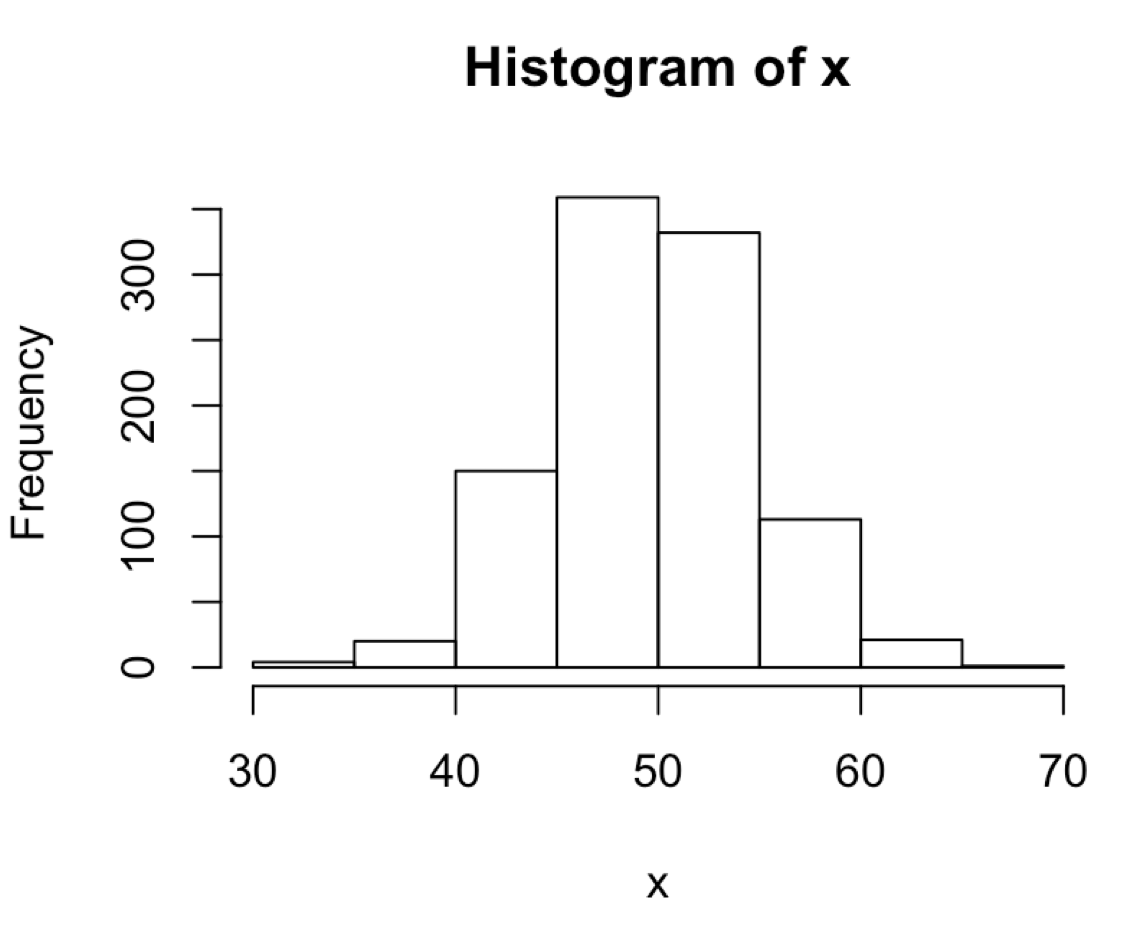


圖 a2-7 二項式分布的隨機值

## Poisson 分布

Poisson 分布是單位時間內某特定事件發生次數的機率分布，因此在呼叫函數時必須要指定單位時間參數 lambda。

### dpois() 函數

回傳單位時間 4 以內某特定事件沒有發生到發生 20 次的機率分布：

> x <- 0:20  
> y <- dpois(x, lambda = 4)  
> plot(x, y, type = "l", ylab = "Probability")

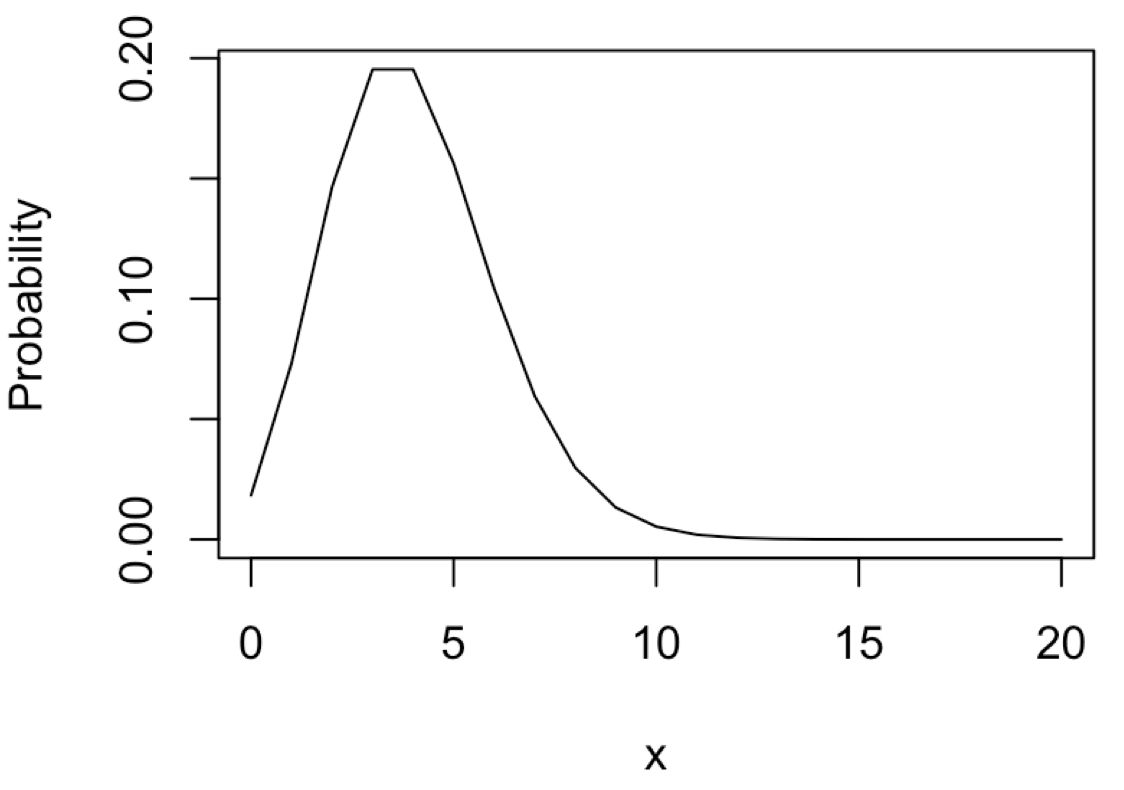


圖 a2-8 Poisson 分布的機率分布

### ppois() 函數

回傳對應輸入的 Poisson 分布累積機率值：

> ppois(4, lambda = 4)  
[1] 0.6288369

### qpois() 函數

回傳對應累積機率值的 Poisson 分布輸入：

> qpois(0.62, lambda = 4)  
[1] 4

### rpois() 函數

回傳 n 個符合 Poisson 分布的隨機值：

> x <- rpois(1000, lambda = 4)  
> hist(x, ylab = "Frequency")

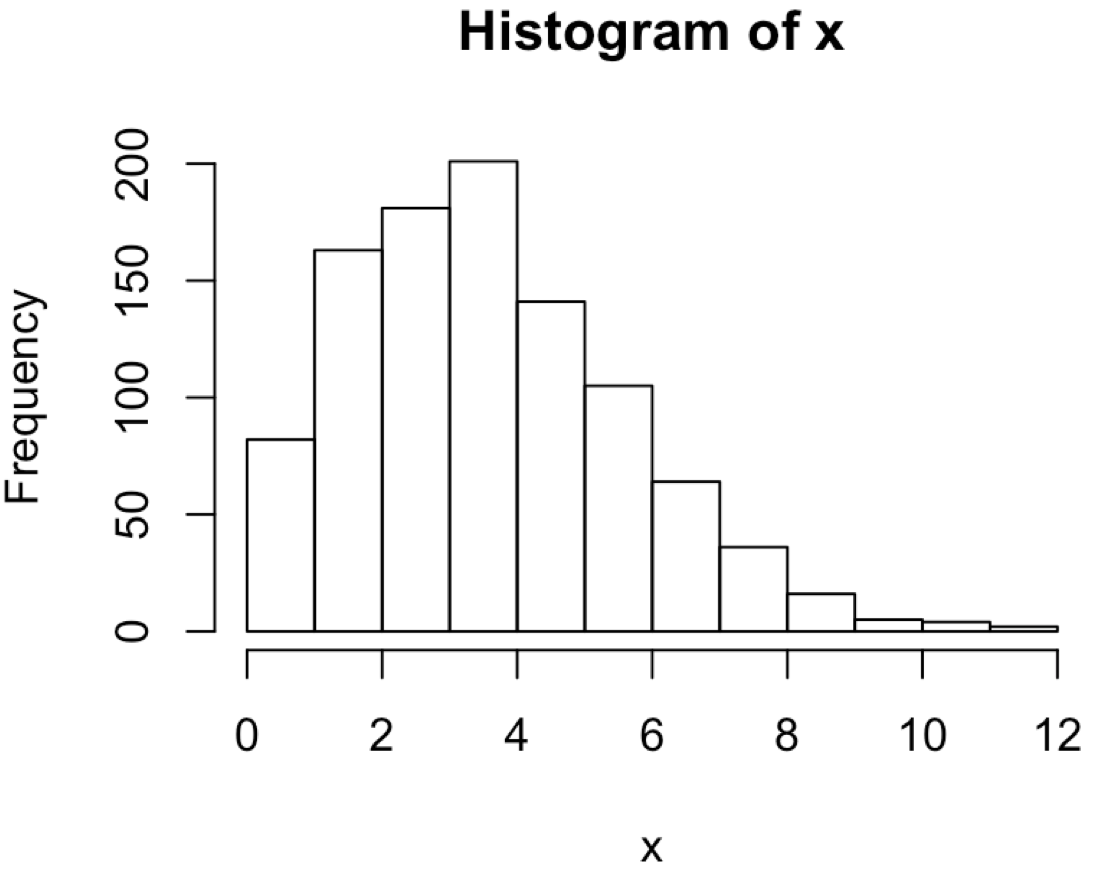


圖 a2-9 Poisson 分布的隨機值

## 卡方分布

在呼叫函數時必須要指定自由度參數 df。

### dchisq() 函數

回傳卡方分布的機率密度值：

> x <- 1:50  
> y <- dchisq(x, df = 5)  
> plot(x, y, type = "l", ylab = "Probability")

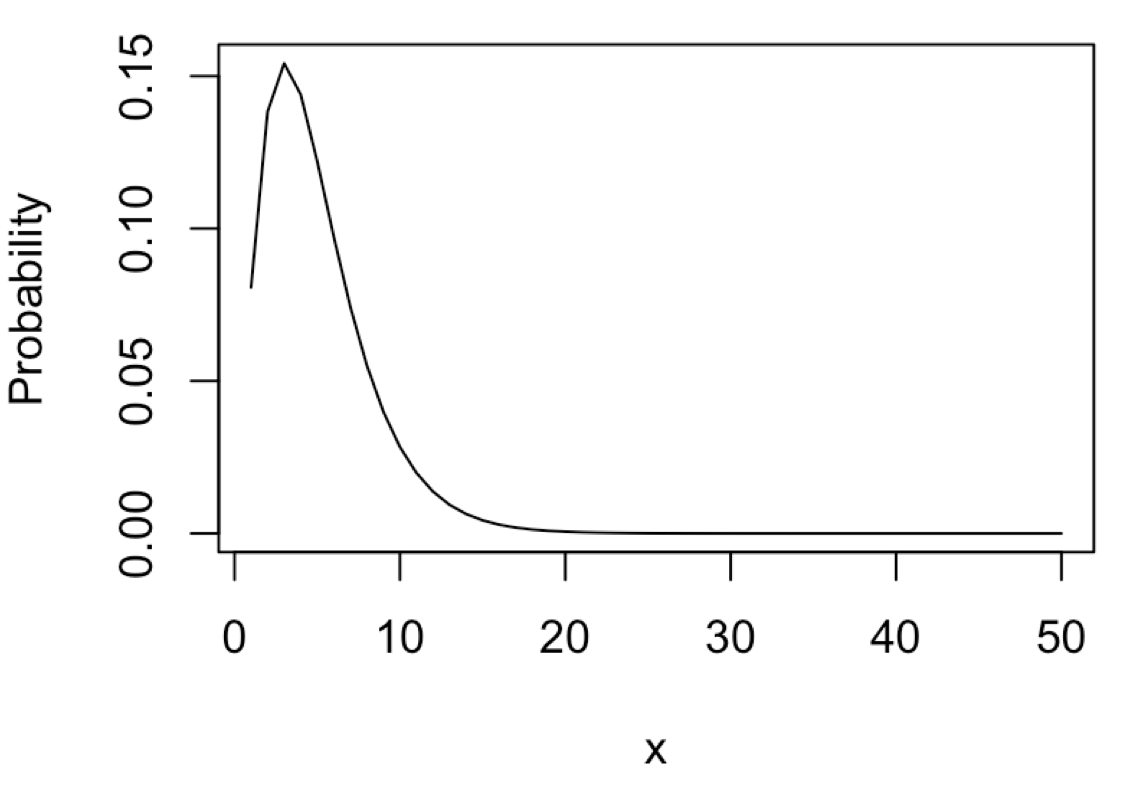


圖 a2-10 卡方分布的機率密度值

### pchisq() 函數

回傳對應輸入的卡方分布累積機率值：

> pchisq(5, df = 5)  
[1] 0.5841198

### qchisq() 函數

回傳對應累積機率值的卡方分布輸入：

> qchisq(0.58, df = 5)  
[1] 4.966356

### rchisq() 函數

回傳 n 個符合卡方分布的隨機值：

> x <- rchisq(1000, df = 5)  
> hist(x, ylab = "Frequency")

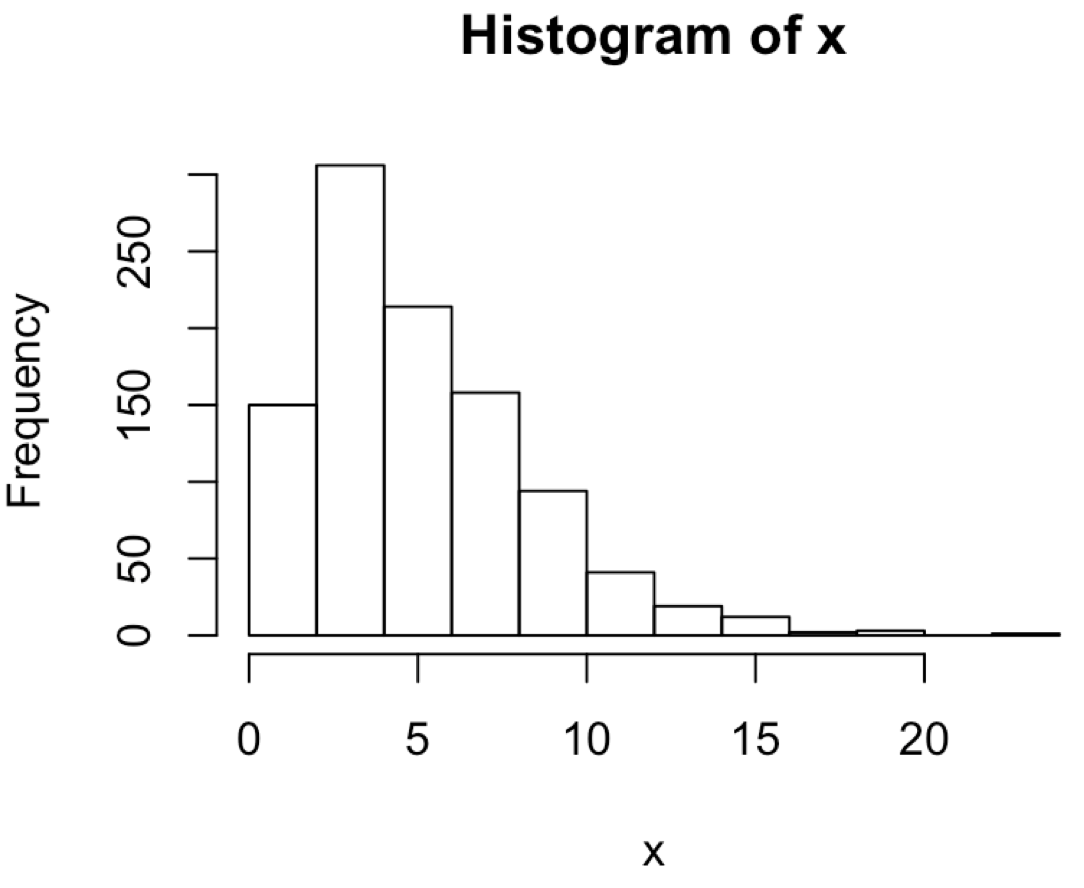


圖 a2-11 卡方分布的機率密度值