

109-2 電腦概論與程式設計

作業 (6)

學號: 410978040 姓名: 黃冠翔

21 五月 2021

Contents

注意事項	1
寫作業要點	1
上傳答題檔案	2
其它事項	2
base 繪圖: 機率分配圖形	2
ggplot2: 機率分配圖形	3
二項式分布算機率	5
小例子	5
寫成函數	6
以常態分佈逼近布瓦松 (Poisson) 分佈	6
格式 (額外加分)	8

注意事項

寫作業要點

- 繳交期限: 2021/05/21(五), 24:00 前上傳完畢。
- 於課程網站 (<http://www.hmwu.idv.tw>) 下載題目卷。
- 可參考課本、上課講義 (包含電子檔) 及其它資料。
- 自己親手做, 儘量不要與別人 (或同學) 討論, 不可參考同學的答案, 不可抄襲。
- 程式設計題, 若程式碼直接複製 (或照抄) 講義上的以不給分為原則。
- 有問題者, 請 FB 群組或私訊問助教或老師。
- 程式直接寫在本 Rmd 檔。經 knitr 編譯, 產生.html、.pdf、.doc 檔, 需印出 R 程式碼及執行結果。
- 不按照規定作答者, 酌量扣分。

上傳答題檔案

- 於教師網站首頁登入 [作業考試上傳區]，帳號: **r1092**。密碼: **xxx**。
- 選取「正確的」資料夾上傳，若傳錯，請最終要上傳一份正確的答題檔案。
- 請上傳「學號-姓名-R-HW6.Rmd」、「學號-姓名-R-HW6.html」、「學號-姓名-R-HW6.pdf」及「學號-姓名-R-HW6.doc」。(學號及姓名，改成自己)
- 若上傳檔案格式錯誤，內容亂碼，空檔等等問題。請自行負責。
- 若要重覆上傳 (第 2 次以上)，請在檔名最後加「-2」、「-3」，例如: 「學號-姓名-R-HW6-2.Rmd」等等。
- 上傳兩次 (含) 以上、格式不合等等酌量扣分。
- 如果上傳網站出現「You can modify the html file, but please keep the link www.wftpservers.com at least.」，請將滑鼠移至「網址列」後，按「Enter」即可。若再不行，請換其它瀏覽器 (IE/Edge/Firefox/Chrome)。
- 有問題者，請 FB 私訊老師。

其它事項

- 若有題目不會寫、或只會寫一半、或結果是有錯的，導致 **knit** 無法編譯產生文件，則可以「不執行有錯的程式碼」，但必需列印此段程式碼。助教會依照狀況部份給分。
- 此份作業，助教以 **pdf** 檔批改為準。**Rmd** 則是做為比對之用 (比對同學們的 **Rmd** 是否相似)。
- 總分 100 分，由助教決定每題配分。全部改完會上傳，答案卷同學們可自由下載。

base 繪圖：機率分配圖形

t 分佈在 wikipedia 中的介紹如下列網頁: https://en.wikipedia.org/wiki/Student%27s_t-distribution。以 R 基礎套件 (base graphics)，畫出 t 分佈在自由度為 1 及自由度為 5 的 (a) 機率密度函數圖，(b) 累積機率分佈函數圖，(c) 分位數函數圖及 (d) 隨機抽樣 ($n = 100$) 直方圖。(要求: (1) 前三個圖上各有兩條不同自由度之函數曲線 (以不同顏色表示)，直方圖則為重疊 (以不同顏色表示)。(2) 需加註: 標題、 x 及 y 標號及 legend (以不同顏色表示相對應的自由度)。(3) 4 張圖一頁: 2 by 2)。

```
# your source code here
par(mfcol=c(2,2))
library(showtext)
```

```
## Warning: package 'showtext' was built under R version 4.0.5
```

```
## Loading required package: sysfonts
```

```
## Warning: package 'sysfonts' was built under R version 4.0.5
```

```
## Loading required package: showtextdb
```

```
## Warning: package 'showtextdb' was built under R version 4.0.5
```

```
showtext.auto()
```

```
## 'showtext.auto()' is now renamed to 'showtext_auto()'
```

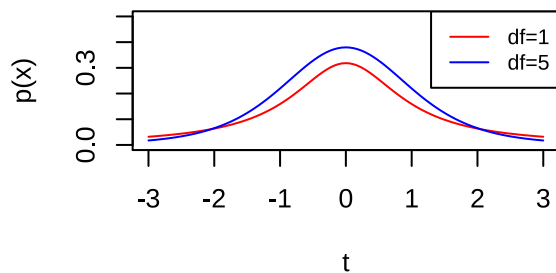
```
## The old version still works, but consider using the new function in future code
```

```

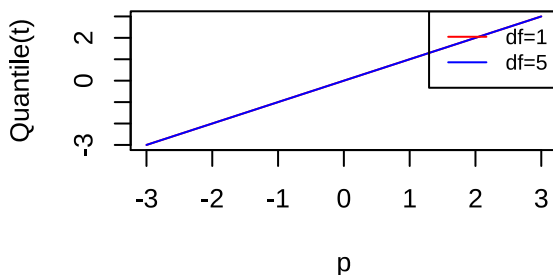
curve(dt(x, 1), -3, 3, ylim=c(0, 0.5), col="red", main=" 機率密度函數圖", xlab="t", ylab="p(x)")
curve(dt(x, 5), -3, 3, add = T, col="blue")
legend("topright", c("df=1", "df=5"), col = c("red", "blue"), lty = 1, cex = 0.8)
curve(pt(x, 1), -3, 3, col="red", main=" 累積機率分佈函數圖", xlab="t", ylab="p(X<=x)")
curve(pt(x, 5), add = T, col="blue")
legend("bottomright", c("df=1", "df=5"), col = c("red", "blue"), lty = 1, cex = 0.8)
curve(qt(pt(x, 1), 1), -3, 3, col="red", main=" 分位數函數圖", xlab="p", ylab="Quantile(t)")
curve(qt(pt(x, 5), 5), -3, 3, add = T, col="blue")
legend("topright", c("df=1", "df=5"), col = c("red", "blue"), lty = 1, cex = 0.8)
hist(rt(100, 1), xlim=c(-50, 50), col="red", main=" 隨機抽樣 (n=100) 直方圖", xlab="t")
hist(rt(100, 5), xlim=c(-50, 50), add=T, col="blue")
legend("topright", c("df=1", "df=5"), col = c("red", "blue"), lty = 1, lwd=5, cex = 0.8)

```

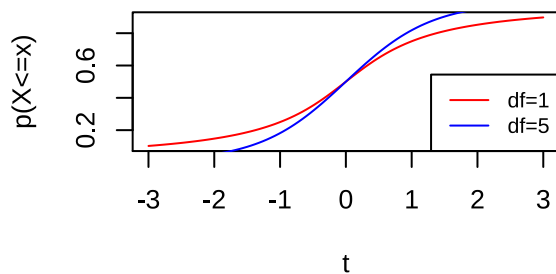
機率密度函數圖



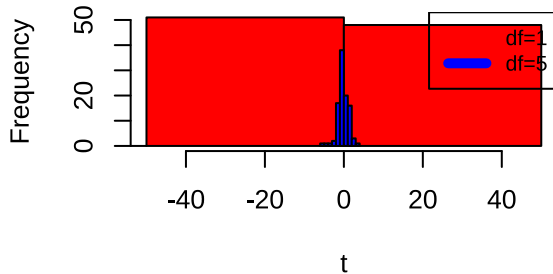
分位數函數圖



累積機率分佈函數圖



隨機抽樣(n=100)直方圖



ggplot2: 機率分配圖形

同上題，以 ggplot2 套件，畫出上述之圖形，要求同上。

```

# your source code here
library(ggplot2)

```

```

## Warning: package 'ggplot2' was built under R version 4.0.5

```

```

library(Rmisc)

## Warning: package 'Rmisc' was built under R version 4.0.5

## Loading required package: lattice

## Loading required package: plyr

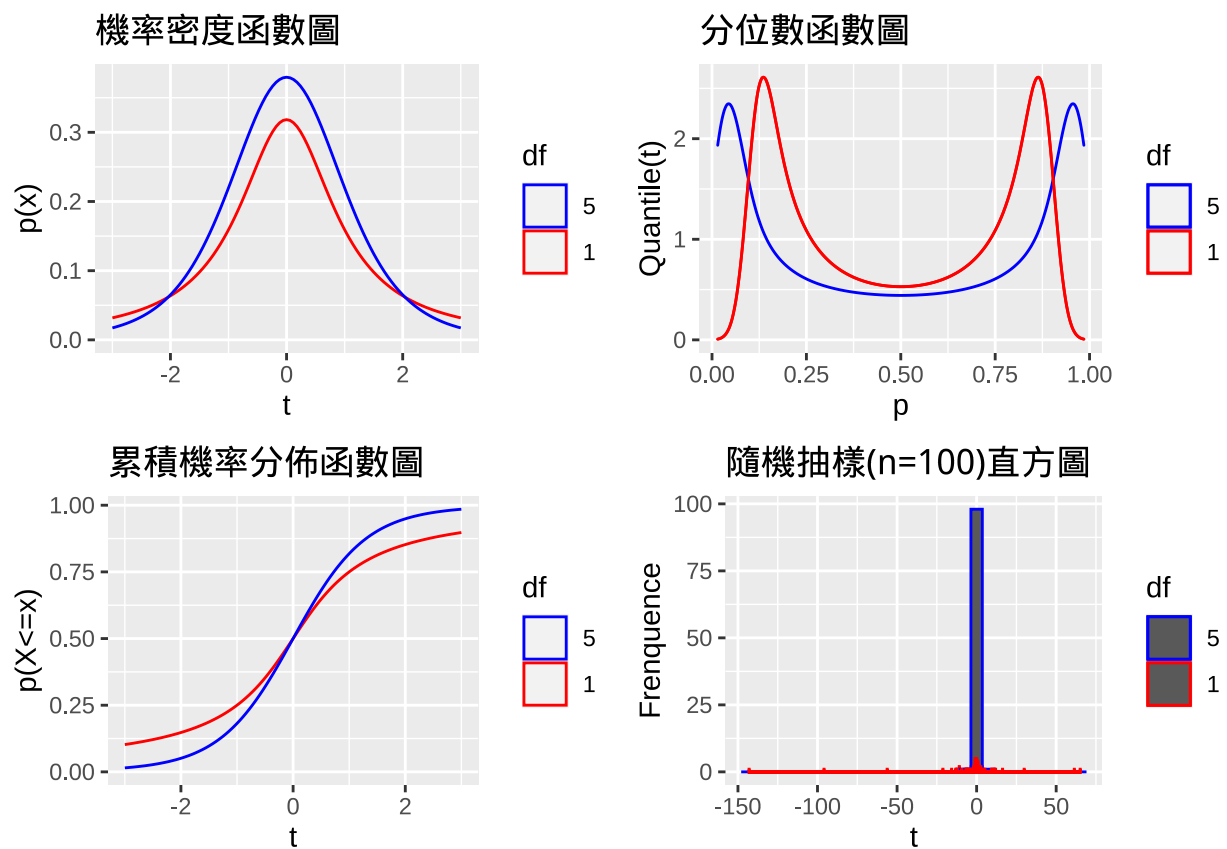
## Warning: package 'plyr' was built under R version 4.0.5

library(showtext)
showtext.auto()
p1 <- ggplot(data.frame(x = c(-3, 3)), aes(x = x, color=c("red", "blue"))) +
  stat_function(fun = dt, args = list(df = 1), color="red")
p5 <- p1 + stat_function(fun = dt, args = list(df = 5), color="blue") +
  ggtitle(" 機率密度函數圖") +
  scale_colour_manual(name = "df", values = c("blue", "red"), labels=c(5, 1)) +
  theme(legend.position = "right") +
  xlab("t") +
  ylab("p(x)") +
  geom_density()
p2 <- ggplot(data.frame(x = c(-3, 3)), aes(x = x, color=c("red", "blue"))) +
  stat_function(fun = pt, args = list(df = 1), color="red")
p6 <- p2 + stat_function(fun = pt, args = list(df = 5), color="blue") +
  ggtitle(" 累積機率分佈函數圖") +
  scale_colour_manual(name = "df", values = c("blue", "red"), labels=c(5, 1)) +
  theme(legend.position = "right") +
  xlab("t") +
  ylab("p(X<=x)") +
  geom_density()
x <- seq(-3, 3, 0.0001)
p3 <- ggplot(data.frame(qt(pt(x, 1), 1)), aes(x = pt(x, 1), color="red")) +
  geom_density(col="red")
p3 <- p3 + geom_density(data = data.frame(qt(pt(x, 5), 5)), aes(x = pt(x, 5), color="blue")) +
  ggtitle(" 分位數函數圖") +
  xlab("p") +
  ylab("Quantile(t)") +
  scale_colour_manual(name = "df", values = c("blue", "red"), labels=c(5, 1)) +
  theme(legend.position = "right") +
  geom_density()
p <- ggplot(data.frame(rt(100, 1)), aes(x=rt(100, 1), col="red")) +
  geom_histogram(fill="red", binwidth = 0.1)

p <- p + geom_histogram(data=data.frame(rt(100, 5)), aes(x=rt(100, 5), col="blue")) +
  ggtitle(" 隨機抽樣 (n=100) 直方圖") +
  xlab("t") +
  ylab("Frenquence") +
  scale_colour_manual(name = "df", values = c("blue", "red"), labels=c(5, 1)) +
  theme(legend.position = "right") +
  geom_histogram(binwidth = 0.1)
multiplot(p5, p6, p3, p, cols=2)

```

```
## Warning: Groups with fewer than two data points have been dropped.
## Warning: Groups with fewer than two data points have been dropped.
## Warning in max(ids, na.rm = TRUE): max 中沒有無漏失的引數；回傳 -Inf
## Warning in max(ids, na.rm = TRUE): max 中沒有無漏失的引數；回傳 -Inf
## Warning: Groups with fewer than two data points have been dropped.
## Warning: Groups with fewer than two data points have been dropped.
## Warning in max(ids, na.rm = TRUE): max 中沒有無漏失的引數；回傳 -Inf
## Warning in max(ids, na.rm = TRUE): max 中沒有無漏失的引數；回傳 -Inf
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



二項式分布算機率

小例子

袋中有 3 紅球、4 白球，從中每次任取一球，取後放回，共取 6 次，則 6 次取球中恰取得 2 次紅球的機率為 $C_2^6 \left(\frac{3}{7}\right)^2 \left(\frac{4}{7}\right)^4$ 。請用 R 計算出此機率。

```
# your source code here
choose(6, 2)*(3/7)^2*(4/7)^4
```

```
## [1] 0.2937552
```

寫成函數

袋中有 r 顆紅球、 w 顆白球，從中每次任取一球，取後放回，共取 k 次，則 k 次取球中恰取得 p 次紅球的機率為何？試寫一 R 函數（命名為 `draw_ball_prob`），計算此抽球機率。執行程式時，以 $r = 3, w = 4, k = 6, p = 2$ 為例。

```
# your source code here
draw_ball_prob <- function(r, w, k, p){
  x <- choose(k, p)*(r/(r+w))^p*(w/(r+w))^(k-p)
  x
}
draw_ball_prob(3, 4, 6, 2)
```

```
## [1] 0.2937552
```

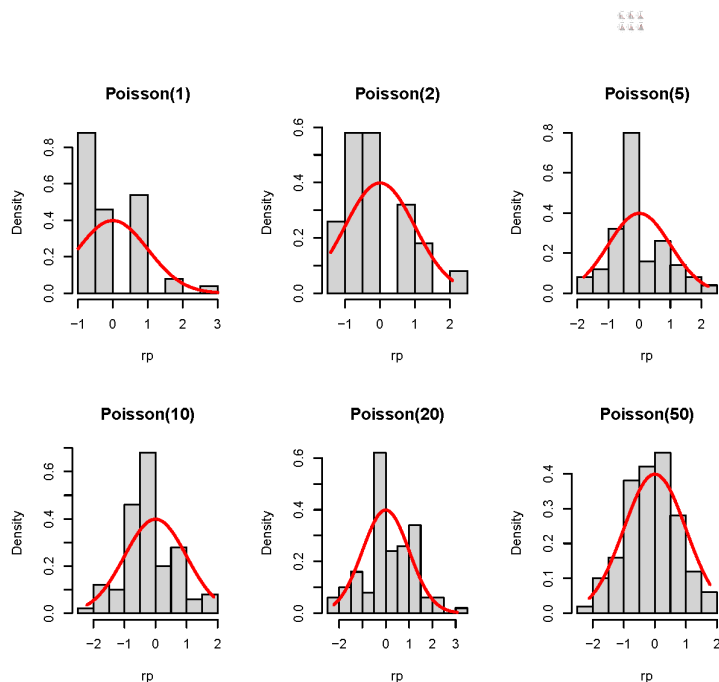
以常態分佈逼近布瓦松 (Poisson) 分佈

Normal Approximation to Poisson Distribution 的定理如下：

If $X \sim \text{Poisson}(\lambda)$ then $\frac{X - \lambda}{\sqrt{\lambda}} \xrightarrow{d} \text{Normal}(0, 1)$ for a sufficient large λ .

使用 $\lambda = 1, 2, 5, 10, 20, 50$ 重覆下列步驟來驗證。（共 6 個圖，請畫成一頁 2×3 ：

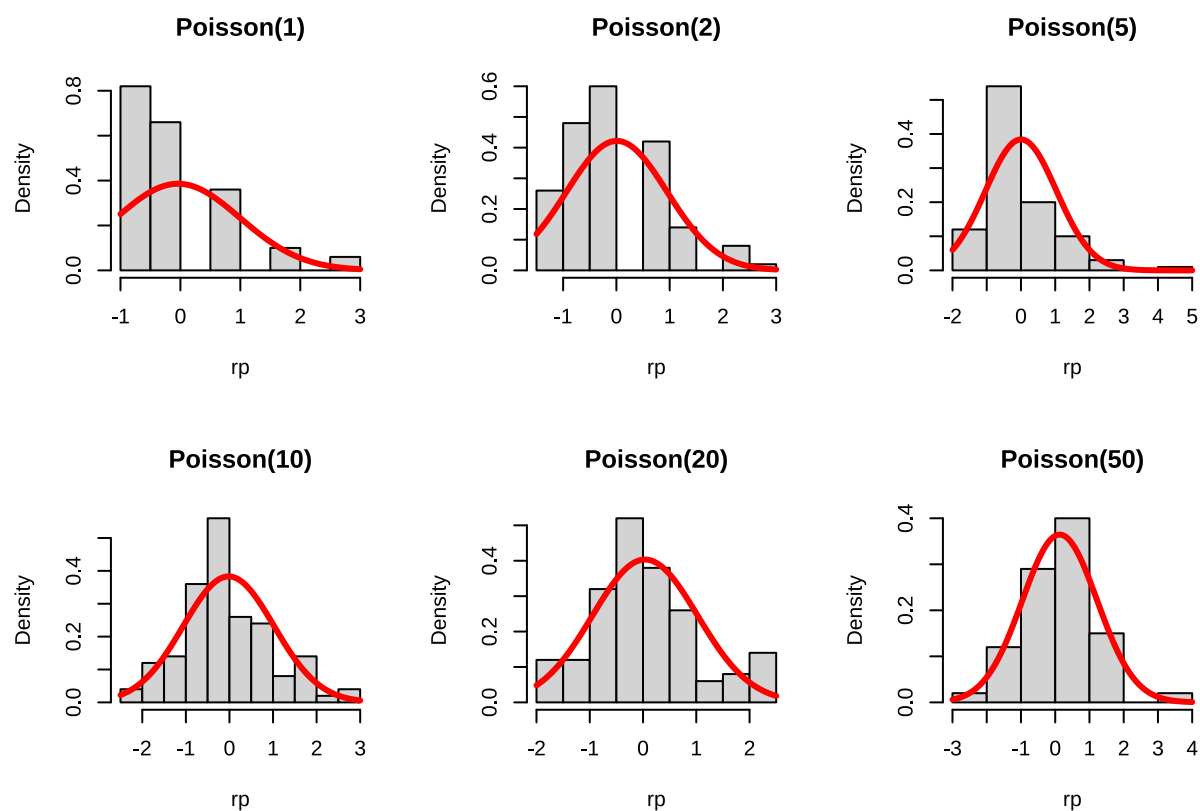
- 隨機產生 100 個 $\text{Poisson}(\lambda)$ 隨機數，將資料利用 $\frac{x-\lambda}{\sqrt{\lambda}}$ 轉換後，畫出其直方圖（圖標題是 $\text{Poisson}(\lambda)$ ， λ 需換成數字）。
- 在直方圖上加上（紅色）標準常態分佈曲線。



```

# your source code here
a <- rpois(100, 1)
b <- rpois(100, 2)
c <- rpois(100, 5)
d <- rpois(100, 10)
e <- rpois(100, 20)
f <- rpois(100, 50)
x1 <- (a-1)/sqrt(1)
x2 <- (b-2)/sqrt(2)
x5 <- (c-5)/sqrt(5)
x10 <- (d-10)/sqrt(10)
x20 <- (e-20)/sqrt(20)
x50 <- (f-50)/sqrt(50)
par(mfcol=c(2,3))
hist(x1, freq=F, main="Poisson(1)", xlab="rp")
curve(dnorm(x, mean=mean(x1), sd=sd(x1)), add=TRUE, col="red", lwd=3)
hist(x10, freq=F, main="Poisson(10)", xlab="rp")
curve(dnorm(x, mean=mean(x10), sd=sd(x10)), add=TRUE, col="red", lwd=3)
hist(x2, freq=F, main="Poisson(2)", xlab="rp")
curve(dnorm(x, mean=mean(x2), sd=sd(x2)), add=TRUE, col="red", lwd=3)
hist(x20, freq=F, main="Poisson(20)", xlab="rp")
curve(dnorm(x, mean=mean(x20), sd=sd(x20)), add=TRUE, col="red", lwd=3)
hist(x5, freq=F, main="Poisson(5)", xlab="rp")
curve(dnorm(x, mean=mean(x5), sd=sd(x5)), add=TRUE, col="red", lwd=3)
hist(x50, freq=F, main="Poisson(50)", xlab="rp")
curve(dnorm(x, mean=mean(x50), sd=sd(x50)), add=TRUE, col="red", lwd=3)

```



格式 (額外加分)

有成功將「學號-姓名-R-HW5.Rmd」編譯出正確的「學號-姓名-R-HW5.html」、「學號-姓名-R-HW5.pdf」及「學號-姓名-R-HW5.doc」，並上傳。以下數學式是測試 MikTeX/LaTeX，請勿刪。這是常態分佈的機率密度函數：

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$