

統計計算與模擬

政治大學統計系余清祥

2026年5月27日～6月3日

第八單元：貝氏計算

<http://csyue.nccu.edu.tw>



Bayesian vs. Frequentist

■ 參數(Parameter)是已知或未知？

→ 參數的詮釋可分為頻率學派、貝氏學派。

→ 兩者在觀念、分析、解釋上的特色？

問題：初等統計學多半採納那種方式？

■ 多數統計教學採取頻率學派，但有些概念可能引起誤解。

→ 95%信賴區間的詮釋？

%

Probability of the
events observed
given a theory

FREQUENTIST
STATISTICS



%

Probability of the
multiple theories
given the observed events

BAYESIAN
STATISTICS



貝氏計算(Bayesian Computing)

- 什麼是貝氏計算(Bayesian Computing)
→ 根據新蒐集的資訊，修正更新對某個未知事物發生的機率預測。
- 貝氏方法將過去經驗作為先驗資訊(Prior Information)，結合本次蒐集的資料(Data)，綜合可得新的經驗(驗後結果, Posterior)：

先驗資訊 + 實驗結果 → 驗後結果

(Prior + Data → Posterior)

貝氏(Bayesian)定理

$$f(\theta | D) = \frac{f(D | \theta)f(\theta)}{f(D)} = \frac{f(D | \theta)f(\theta)}{\int f(D | \theta')f(\theta')d\theta'}$$

- 貝氏觀點：根據新蒐集資料更新認知。
→ 有人認為貝氏分析偏向「主觀機率」，
但先驗分配不應主宰驗後結果。

貝氏計算的種類

- 目前主流的計算技術主要分為抽樣模擬與近似優化兩大類：
 - 本課程介紹MCMC、EM兩種。

方法	EM	MCMC
類型	Optimization	Sampling
目標	MLE/MAP	Posterior
輸出	點估計	分配
速度	快	慢
不確定性	較少	完整
local optimum	可能	較少

關於先驗分配

- 貝氏方法於實際應用時，可將過去經驗或是其他相關研究的結果視為先驗資訊。
 - 然而，使用時需注意假設條件及參數是否改變，如果情況不盡相同，可透過調整先驗分配的權數因應。
- 如果缺乏相關資訊，通常採用：
 - Non-informative Prior
 - Weakly Informative Prior
 - Empirical Bayes

貝氏(Bayesian)分析



■ 貝氏理論(Bayesian Theory)

→ 由Thomas Bayes於1763年發表的貝氏定理而得名。貝氏分析可合併經驗與資料，可隨著資料累積更新想法，較為符合知識發展的路徑。(註：我站在巨人的肩膀上...)

→ 以往受限於計算較為複雜，因為電腦科技發展使得貝氏計算更為可行，自1990年代起愈發受到重視。

■ 範例一、某君投擲一硬幣100次，出現42次正面；第二天再投擲同一硬幣100次，出現52次正面。檢驗此硬幣是否為公平硬幣。

→ 投擲硬幣100次，出現正面的次數可視為二項分配的變數，與觀察某地區同年齡居民在一年內的死亡人數類似，投擲兩天就像是觀察連續兩年的死亡率。

→ 以估算死亡率的角度來看，我們應該使用第一年、第二年、或是兩年合併的資料估計出現正面的機率，又該如何合併兩年的資料？（例如：加權平均、直接相加？）

- 範例一(續)、你/妳覺得硬幣出現正面的機會(p)等於多少？ $X \sim B(100, p)$

→ 只使用第一天資料， $\hat{p} = \frac{42}{100} = 0.42$ 。

→ 只使用第二天資料， $\hat{p} = \frac{52}{100} = 0.52$ 。

→ 兩天資料相加， $\hat{p} = \frac{42 + 52}{100 + 100} = \frac{94}{200} = 0.47$ 。

- 問題：第二天的權重是否應該較重？

貝氏分析

- 若無特別指定，貝氏分析原則上將所有資料直接合併，也就是平均加權。
- 先驗與驗後屬於同一類型的分配，稱為共軛先驗分配(Conjugate Prior Distribution)，常見的例子有：

Beta + Binomial → Beta

Gamma + Poisson → Gamma

Normal + Normal → Normal

Dirichlet + Multinomial → Dirichlet

- 以範例一為例，硬幣出現正面的次數大多視為二項分配，如果假設正面出現的機會服從Beta分配，則在加入資料的分析後，出現正面的機會仍舊服從驗後Beta分配。
→ 正面、反面個數分別相加。

$$p \sim \text{Beta}(\alpha, \beta) \Rightarrow \pi(p) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1}$$

$$X \sim B(n, p) \Rightarrow f(x | p) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$p | x \sim \text{Beta}(\alpha + x, \beta + n - x) \Rightarrow \pi(p | x) \propto p^{\alpha+x-1} (1-p)^{\beta+n-x-1}$$

- 上述的先驗Beta分配中，參數 α 及 β 可視為過去的經驗中，正面及反面出現的次數，因此驗後Beta分配中的參數，可解釋為累積的正面及反面出現次數。

→ Beta分配的期望值滿足 $E(p) = \alpha/(\alpha+\beta)$ ，因此

$$\begin{aligned} E(p | x) &= \frac{\alpha + x}{\alpha + \beta + n} \\ &= \frac{\alpha + \beta}{\alpha + \beta + n} \cdot \frac{\alpha}{\alpha + \beta} + \frac{n}{\alpha + \beta + n} \cdot \frac{x}{n} \end{aligned}$$

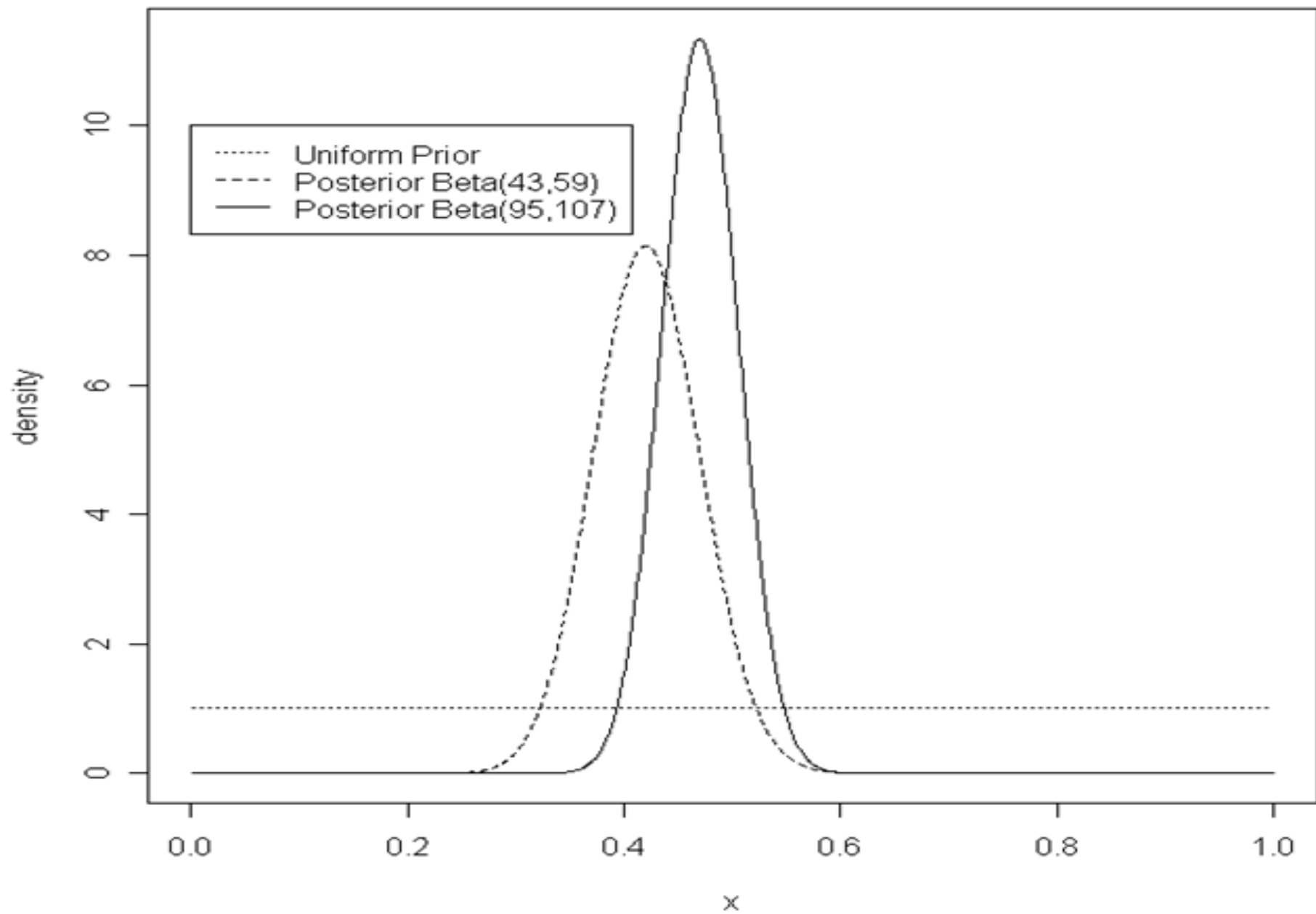
為先驗分配及實驗結果的加權平均，權數為個別的樣本數。

- 如果過去經驗記為 s_x (標準生命表的數值)，死亡率的觀察值為 u_x ，範例一可視為：

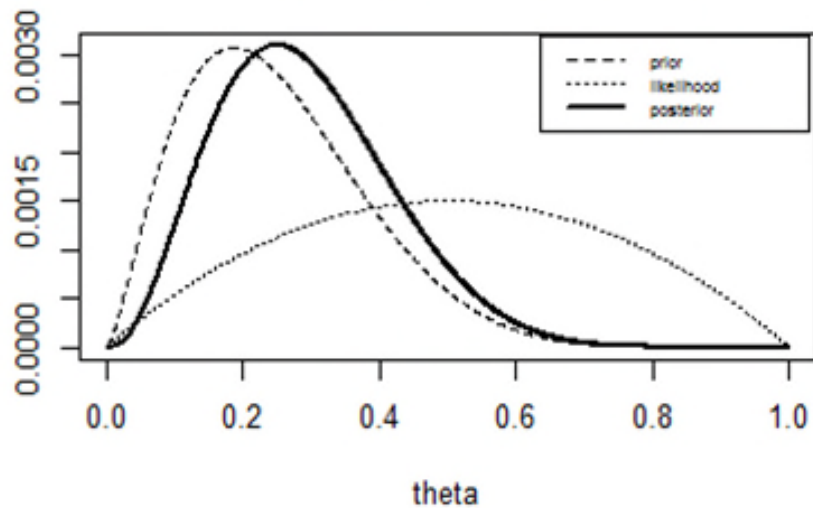
$$v_x = \frac{\alpha + \beta}{\alpha + \beta + n} \cdot s_x + \frac{n}{\alpha + \beta + n} \cdot u_x$$

修勻值為經驗值與死亡觀察值的加權平均。過去經驗的樣本數如果較多，則修勻值較接近經驗值；反之，如果本次蒐集的資料比數較多 (n 較大)，最近的經驗佔較大的比重，甚至可推翻過去的經驗。（「三人成虎」！）

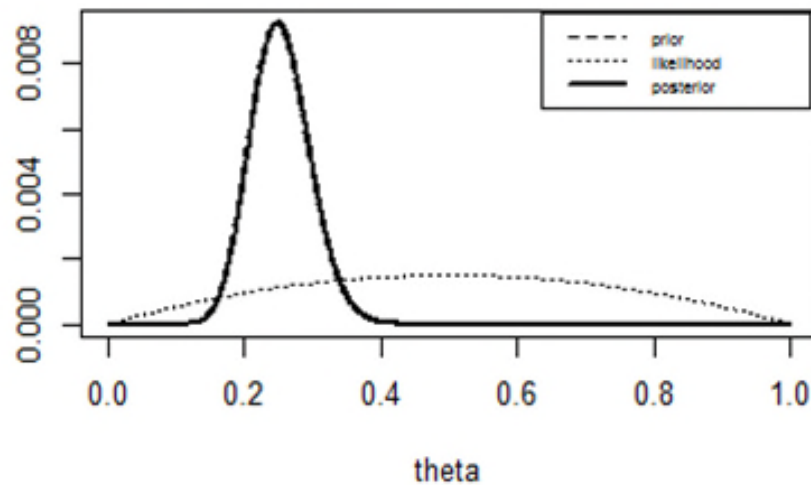
Posterior Density (Coin Tossing)



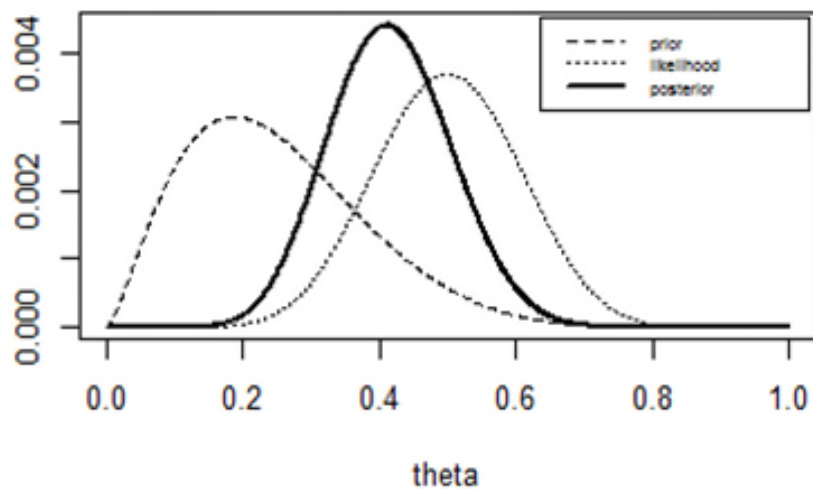
Beta(2.5,7.5) & 1 Heads, 1 Tail



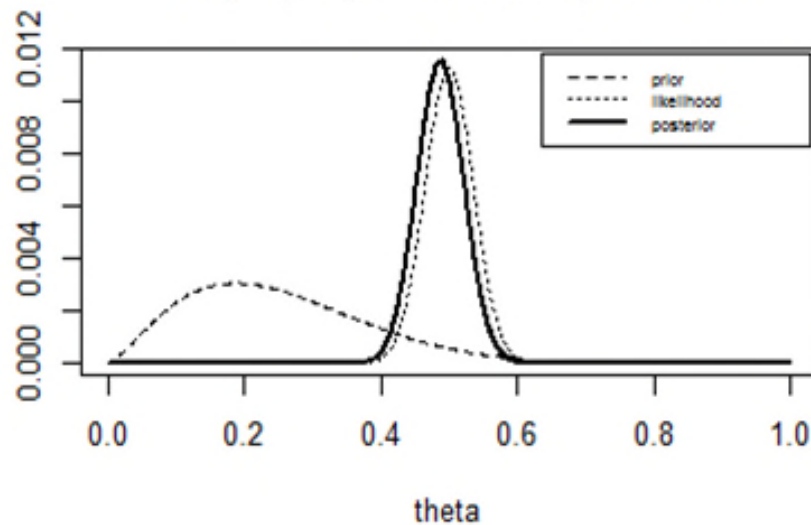
Beta(25,75) & 1 Heads, 1 Tail



Beta(2.5,7.5) & 10 Heads, 10 Tails

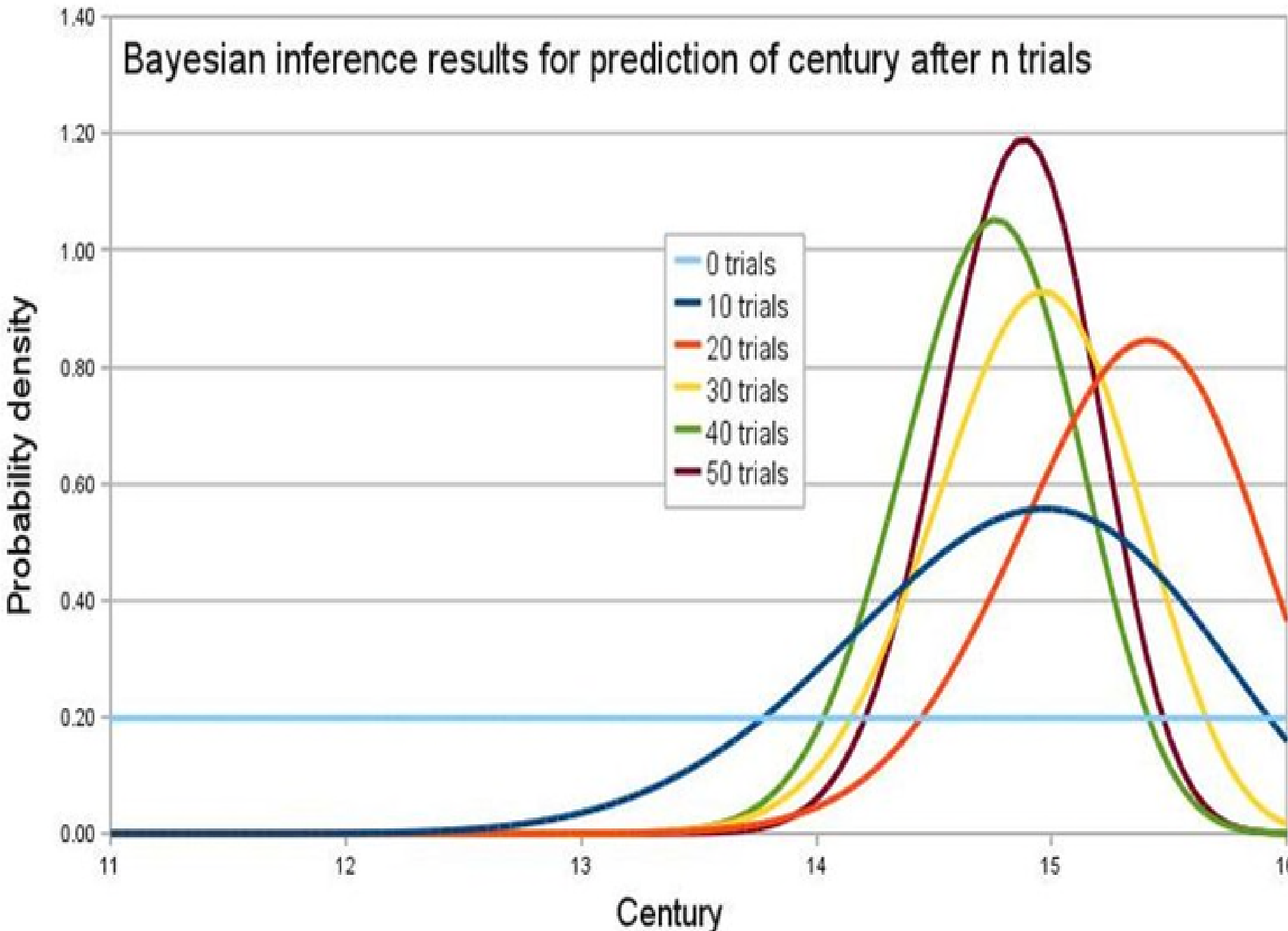


Beta(2.5,7.5) & 100 Heads, 100 Tails

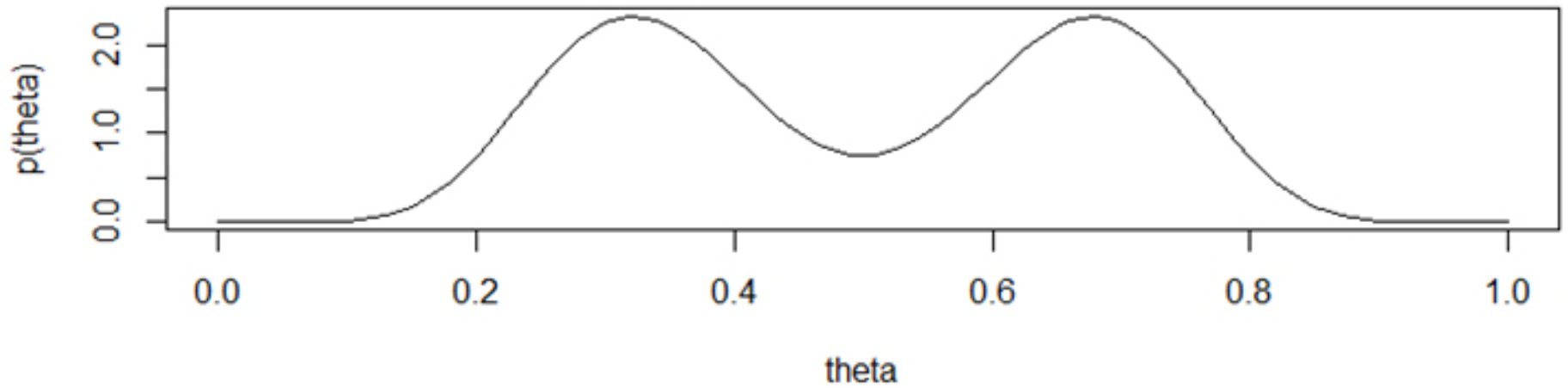


Prior與Data的樣本數對Posterior的影響

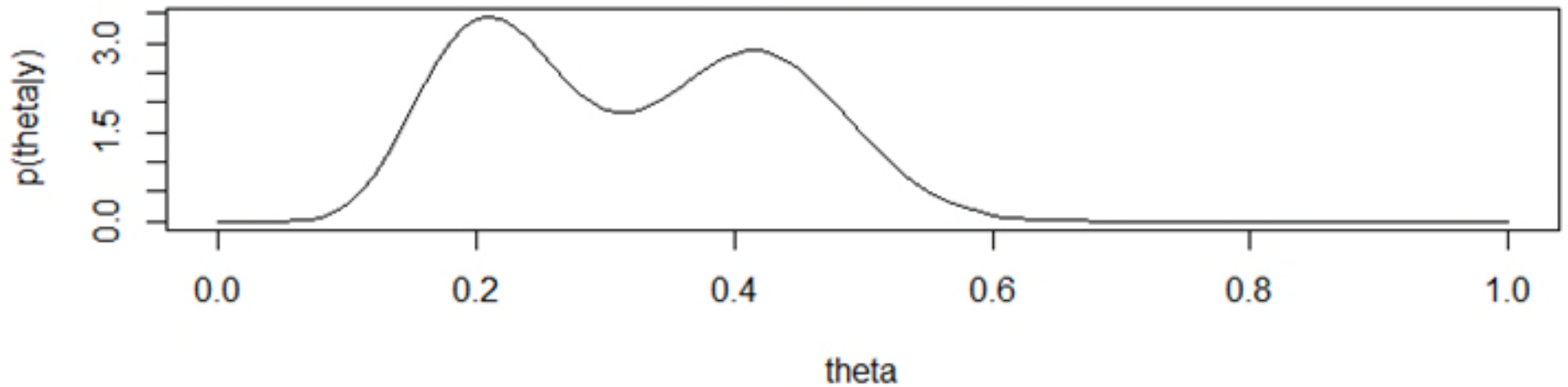
Bayesian inference results for prediction of century after n trials



Prior



Posterior



Data樣本數夠大就能扭轉Prior的影響！

貝氏分析的理念

- 一般對參數的假設是「未知但固定」，貝氏則認為參數「具有某種分配」，對參數的瞭解通常隨著樣本數增加而更確定。

→ 以上述投擲硬幣為例，正面出現機會的驗後分配，滿足變異數

$$\text{Var}(p | x) = \frac{(\alpha + x)(\beta + n - x)}{(\alpha + \beta + n + 1)(\alpha + \beta + n)^2}$$

隨著樣本數增加趨近於0，即收斂至一點。

貝氏分析的理念(續)

- 如果缺乏過去經驗，先驗分配可假設無資訊的先驗分配(Non-informative Prior)，意謂先驗分配對整體的影響力極小。

→ 出現正面機會中的先驗Beta分配，一般會給定均勻分配 $U(0,1) = \text{Beta}(1,1)$ 為無資訊先驗分配，也就是過去只有一次正面、一次反面的經驗。因此驗後期望值為 $\frac{x+1}{n+2}$ ，與

最大概似估計量MLE $\frac{x}{n}$ 非常接近。

貝氏分析的理念(續)

- 常態分配是另一個常見的共軛先驗分配，因為各年齡的人數動輒以萬計算，以常態分配近似還算適當。(註：死亡率大小?)

→ 假設某年齡的死亡率滿足 $\theta \sim N(\mu, \tau^2)$ ，

而某年度蒐集的該年齡死亡資料也滿

足 $X | \theta \sim N(\theta, \sigma^2)$ ，則死亡率的驗後分配

服從 $\theta | X = x \sim N\left(\frac{\sigma^2 \mu + \tau^2 x}{\tau^2 + \sigma^2}, \frac{\tau^2 \sigma^2}{\tau^2 + \sigma^2}\right)$

或是 $E(\theta | X = x) = \frac{1/\tau^2}{1/\sigma^2 + 1/\tau^2} \mu + \frac{1/\sigma^2}{1/\sigma^2 + 1/\tau^2} x$ 。

Normal + Normal \rightarrow Normal

- 常態分配的共軛先驗分配，其型態和計算方法與二項分配類似。

\rightarrow 若真實死亡率的先驗分配滿足 $\underline{t} \sim N(\underline{m}, A)$

觀察的死亡率服從常態分配 $\underline{u} | \underline{t} \sim N(\underline{t}, B)$

$$\pi(\underline{t}) = k_1 \cdot \exp\left[-\frac{1}{2}(\underline{t} - \underline{m})' A^{-1}(\underline{t} - \underline{m})\right]$$

$$f(\underline{u} | \underline{t}) = k_1 \cdot \exp\left[-\frac{1}{2}(\underline{u} - \underline{t})' B^{-1}(\underline{u} - \underline{t})\right]$$

- 驗後分配需要先計算出 \underline{u} 的邊際分配 (Marginal distribution) $f(\underline{u})$ ，藉由

$$f(\underline{t}, \underline{u}) = \pi(\underline{t}) \cdot f(\underline{u} | \underline{t}) = f(\underline{u}) \cdot \pi(\underline{t} | \underline{u})$$

算出真實死亡率的驗後分配。

- 因為計算不易，一般只考慮與 \underline{t} 有關的項次，或是令 $\underline{t} | \underline{u} \sim N(\underline{v}, C)$ ，其中

$$(\underline{t} - \underline{v})' C^{-1} (\underline{t} - \underline{v}) \propto$$

$$(\underline{t} - \underline{m})' A^{-1} (\underline{t} - \underline{m}) + (\underline{u} - \underline{t})' B^{-1} (\underline{u} - \underline{t})$$

$$C^{-1} = A^{-1} + B^{-1}$$

- 上式對 \underline{t} 微分可得

$$2C^{-1}(\underline{t} - \underline{v}) = 2A^{-1}(\underline{t} - \underline{m}) + 2B^{-1}(\underline{u} - \underline{t})$$

或是修勻值 $\underline{v} = C(A^{-1}\underline{m} + B^{-1}\underline{u})$ 。

- 重新排列，令 I 為單位矩陣，修勻值可表達為 $\underline{v} = \underline{u} + (I + AB^{-1})^{-1}(\underline{m} - \underline{u})$ ，也就是根據觀察死亡率 \underline{u} ，再加入經驗值 \underline{m} 的效果來修勻。同理，也可以經驗值再加入觀察死亡率來調整 $\underline{v} = \underline{m} + (I + BA^{-1})^{-1}(\underline{u} - \underline{m})$ 。

Choices of \underline{m} , A, & B

- 先驗分配中的參數 \underline{m} 及 A 對結果有相當的影響，以下列出參數選擇的建議：
 - \underline{m} 的選擇大多是經驗中最有可能的數值。以修勻今年生命表為例，通常會以上一次、或是去年死亡經驗當作經驗值。
 - 因為A代表各年齡死亡率間的關係，一般會先決定每個年齡變異數，再藉由年齡間的相關程度決定共變異數。

Choices of \underline{m} , A, & B (cont.)

→ 共變數矩陣A需要滿足

(i) 正定、對稱矩陣 (ii) 可逆矩陣

建議可選取 $A_{xy} = \sqrt{A_{xx}} \sqrt{A_{yy}} r^{|x-y|}$, $0 \leq r < 1$

其中 A_{xx} 及 A_{yy} 為 x 歲及 y 歲死亡率的變異數。

→ 因為原始死亡率服從二項分配，變異數矩陣 B 通常假設為對角矩陣，各年齡死亡率的變異數為

$$B_{xx} = \text{Var}(u_x) = \frac{t_x(1-t_x)}{n_x} \cong \frac{m_x(1-m_x)}{n_x}$$



Correlation Matrix (Positive Definite)

```
PDM=function(n,rho) {  
  A=NULL  
  j=c(1:n)  
  for (i in 1:n) {  
    A=rbind(A,rho^abs(i-j))  
  }  
  return(A)  
}
```

```
#
```

```
# Use “eigen” to check the result.
```

```
eigen(PDM(5,0.9))
```

```
$values
```

```
[1] 4.26213409 0.45446614 0.14592141 0.07943386 0.05804450
```

參數A與B的特例

→ 共變數矩陣A與B均為對角矩陣

$$v_x = \frac{b_{xx}}{a_{xx} + b_{xx}} m_x + \frac{a_{xx}}{a_{xx} + b_{xx}} u_x$$

→ 如果

$$A = \underline{\underline{1}} \underline{\underline{1}}' \sigma^2 = \begin{bmatrix} 1 & 1 & \cdot & \cdot & 1 \\ 1 & 1 & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot & 1 \end{bmatrix} \sigma^2, \quad B = I_{n \times n}$$

$$\Rightarrow E(\underline{\underline{t}} | \underline{\underline{u}}) = \underline{\underline{m}} + \mu \underline{\underline{1}}, \quad \mu = \frac{1}{2} \left(\underline{\underline{1}} B^{-1} \underline{\underline{1}} + \frac{n^2}{\sigma^2} \right) \underline{\underline{1}}' B^{-1} (\underline{\underline{u}} - \underline{\underline{m}})$$

其它共轭贝氏分配

- 死亡率除了假设常态分配，常见的是令其服从多项分配，

$$f(d_1, d_2, \dots, d_k | \underline{t}) = \frac{d!}{k^d} \prod_{i=1}^k t_i^{d_i}$$

其中 d 为所有死亡人数， d_i 为第 i 年龄组死亡人数。

→ 如果死亡率服从Dirichlet假设，记为

$$\underline{t} = (t_1, t_2, \dots, t_k) \sim \text{Dirichlet}(a_1, a_2, \dots, a_k)$$

其它共軛貝氏分配(續)

→ Dirichlet分配的密度函數為

$$f(t_1, t_2, \dots, t_k) = \frac{\Gamma(\sum_{i=1}^k a_i)}{\prod_{i=1}^k \Gamma(a_i)} \prod_{i=1}^k t_i^{a_i-1}$$

驗後的分配變成

$$\underline{t} | \underline{d} \sim \text{Dirichlet}(a_1 + d_1, a_2 + d_2, \dots, a_k + d_k)$$

→ 其中第*i*組死亡率的修勻值滿足

$$v_i = \frac{a}{a+d} \cdot \frac{a_i}{a} + \frac{d}{a+d} \cdot \frac{d_i}{d}$$

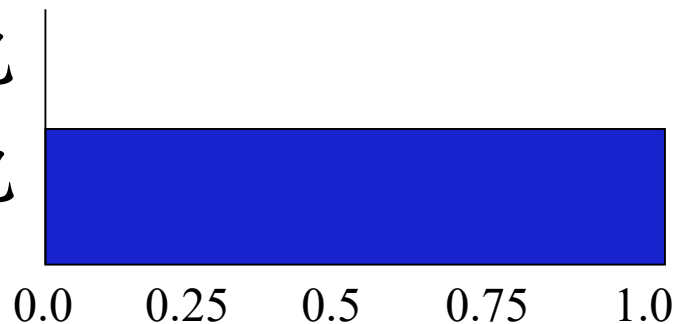
非共軛分配的貝氏計算

- 共軛分配可簡化計算，但也限制貝氏分析的應用範圍，只要先驗分配與假設略有差異，計算複雜度將大為提高。
- 近年電腦科技進步快速，貝氏分析因為蒙地卡羅馬可夫鏈(Monte Carlo Markov Chain ; MCMC)而更為可行，不需要共軛分配假設、或是指定先驗分配。

關於貝氏計算

■ 透過貝氏計算可估計出參數的驗後分配。

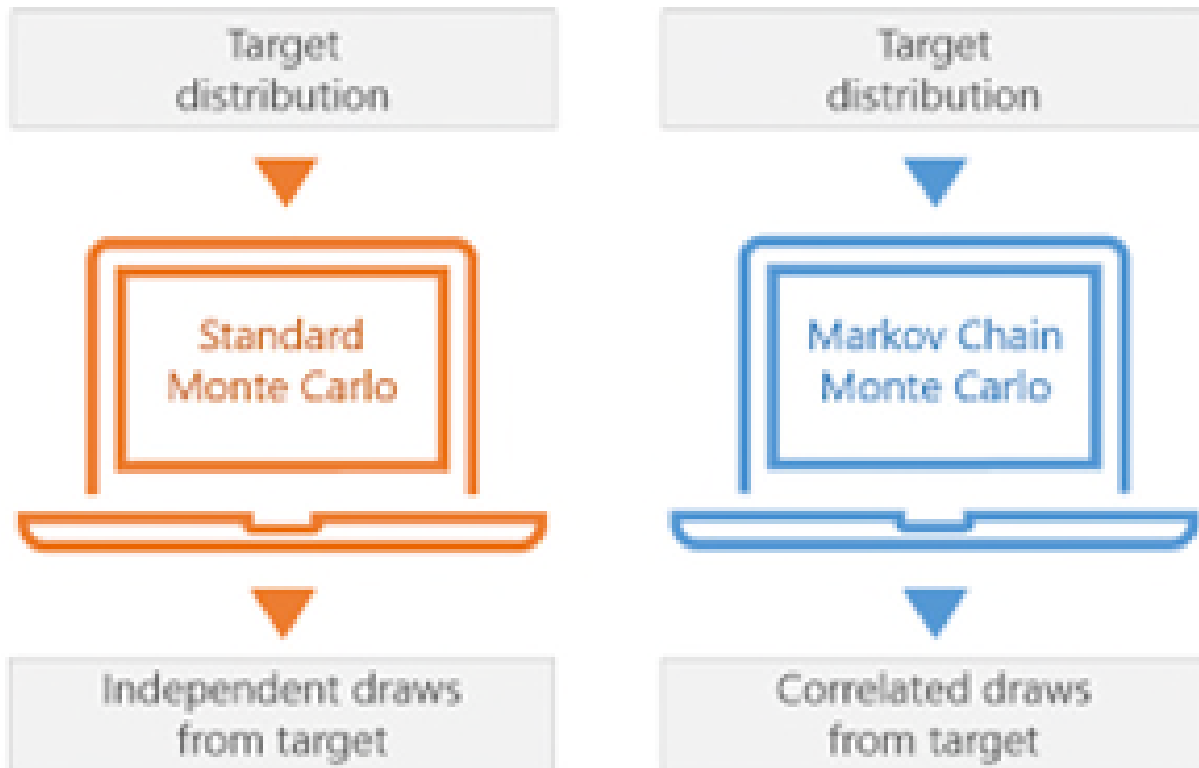
→ 根據經驗值決定先驗分配，或是使用無資訊先驗分配 (Non-informative Prior)；



→ 經驗貝氏 (Empirical Bayes)：可將資料拆成兩部分，一部份用於估計先驗分配。

■ MCMC 是近年常見的貝氏計算方法，透過多次遞迴的電腦模擬，估計出參數的驗後分配。

什麼是MCMC？



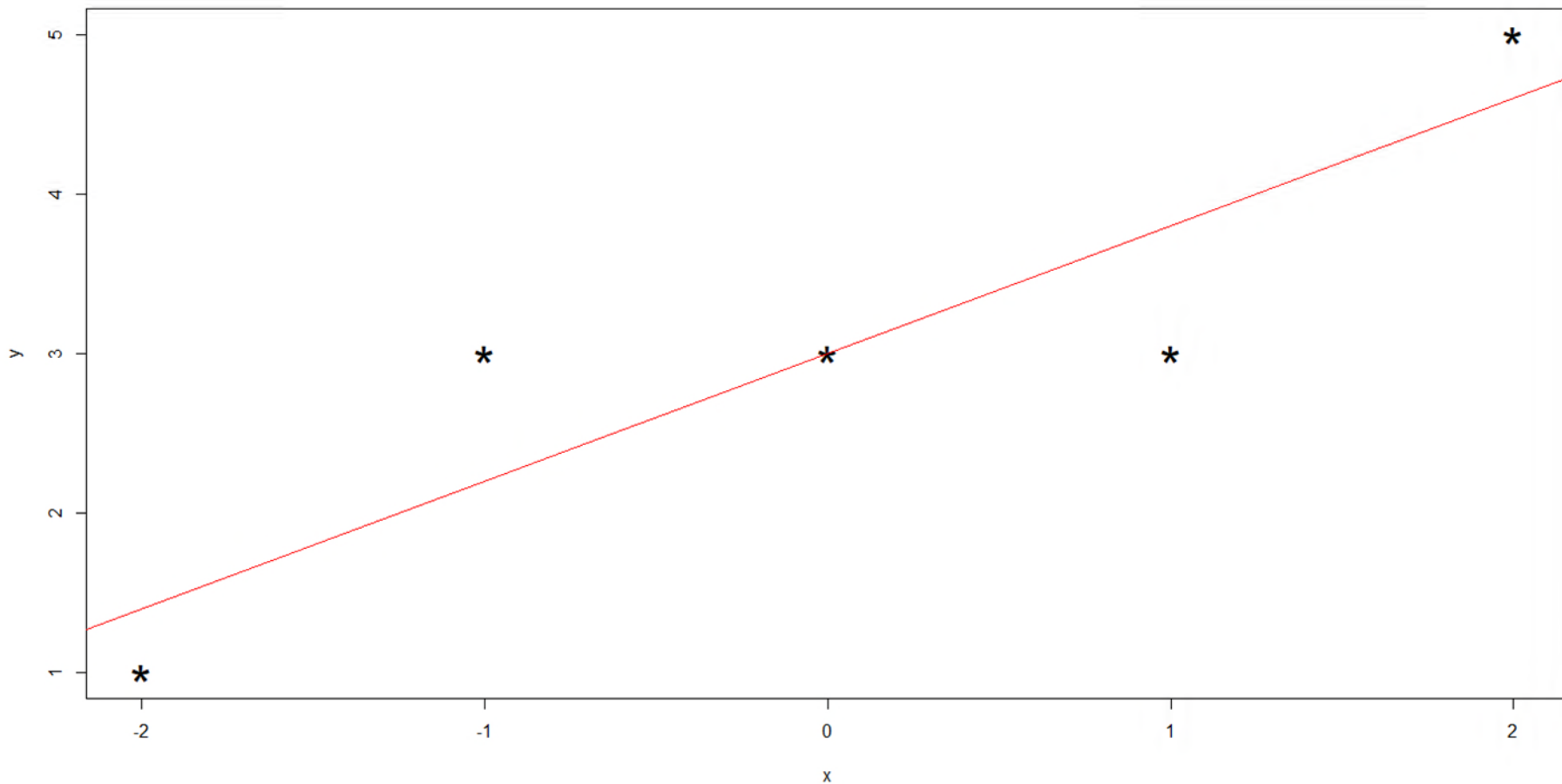
MCMC方法簡介

- 若先驗分配與驗後分配屬於不同形態的函數，必須仰賴數值積分之類的方法求出驗後機率，而MCMC以另一種方式簡化問題。
- MCMC結合 Markov Chain 與 Monte Carlo simulation，主要概念為建構一個以目標分配為平穩分配的 Markov Chain，透過反覆遞迴抽樣，使分布逐漸收斂至目標。收斂後的樣本即可視為來自該目標分配，進而用以估計期望值、變異數及其他統計量。

範例一：簡單線性迴歸

■ X與Y似乎存有線性關係

→ 由散佈圖判斷X與Y呈現線性關係。



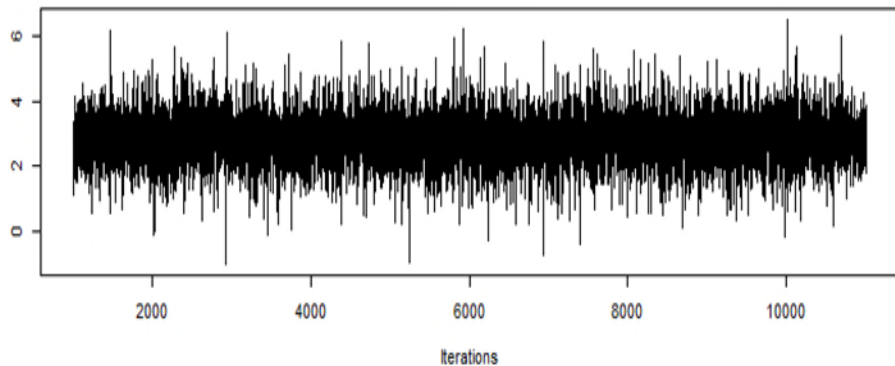
```
library(devtools)
library(ggplot2)
library(HDIInterval)
library(MCMCpack)
library(mcmc)
library(mcmcse)
library(Rcpp)
library(RcppArmadillo)
library(stableGR)
```

```
#
```

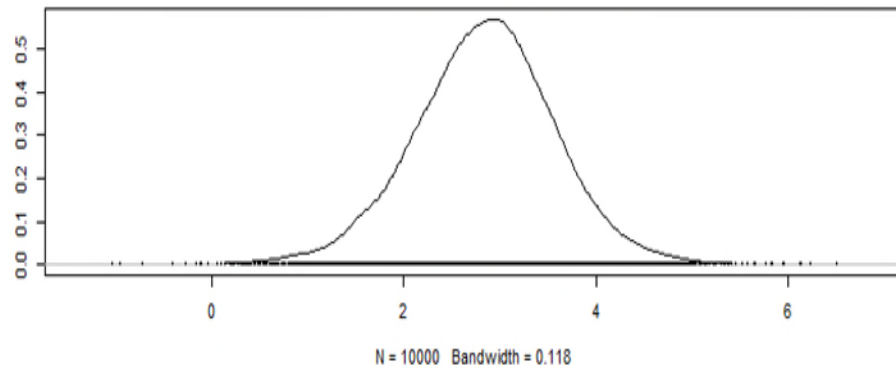
```
line = list(X = c(-2,-1,0,1,2), Y = c(1,3,3,3,5))
posterior = MCMCregress(Y~X, b0=0, B0 = 0.1,
                        sigma.mu = 5, sigma.var = 25, data=line, verbose=1000)
plot(posterior)
raftery.diag(posterior)
summary(posterior)
```

```
#           2.5%    25%    50%    75%    97.5%
# Intercept = 3    1.3085  2.3765  2.8594  3.3180  3.3290
# Slope = 0.8     -0.3465  0.4353  0.7788  1.1100  1.8460
# Sigma = 0.7303   1.1552  1,8859  2.5745  3.6240  7.8270
```

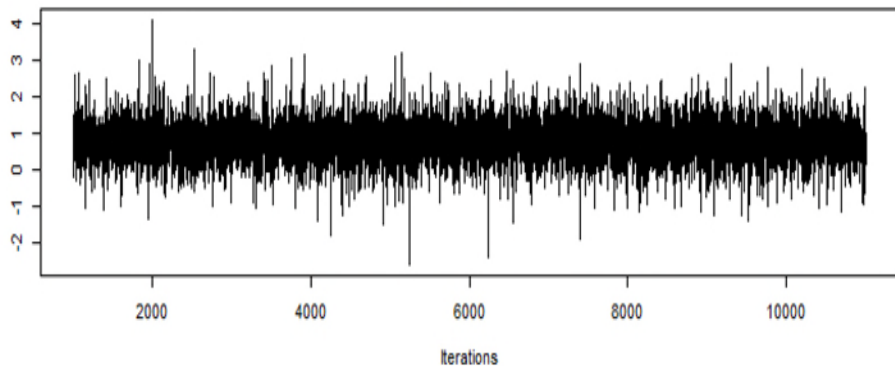
Trace of (Intercept)



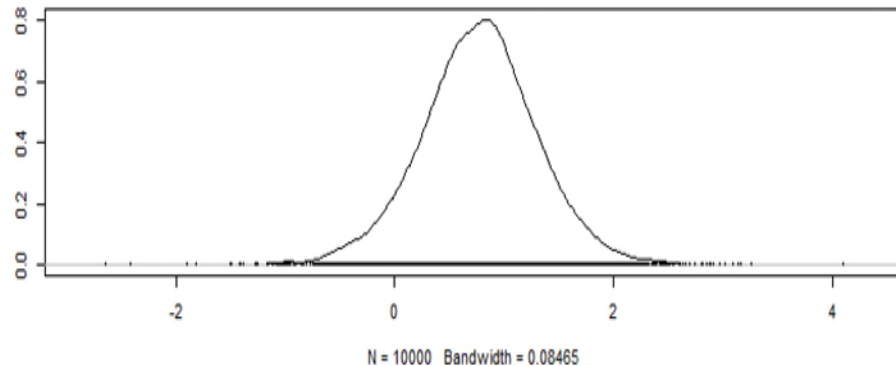
Density of (Intercept)



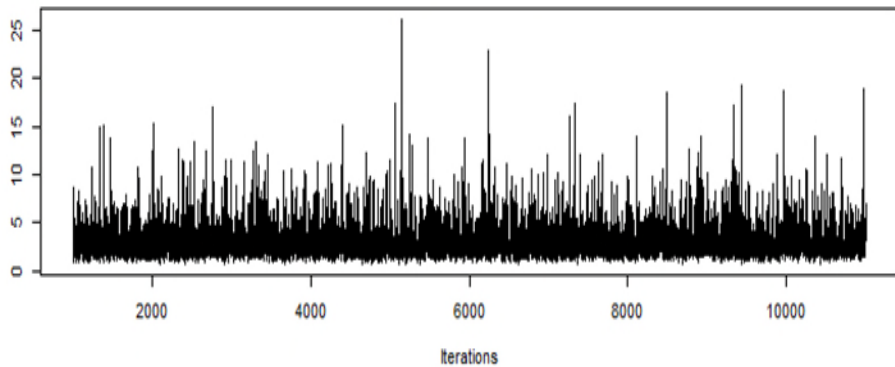
Trace of X



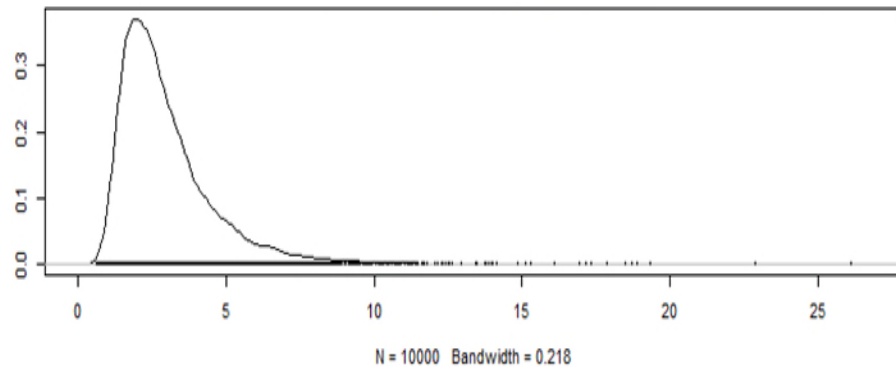
Density of X



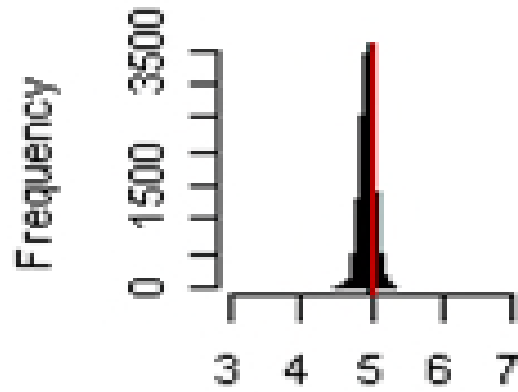
Trace of sigma2



Density of sigma2

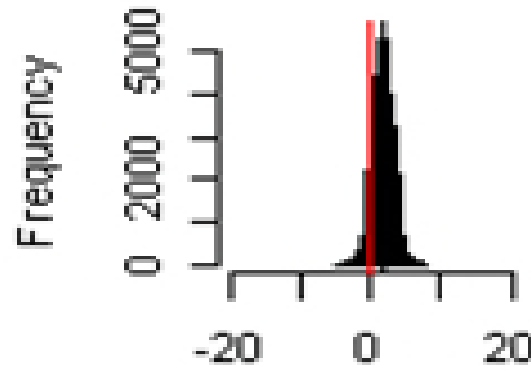


Posterior of a



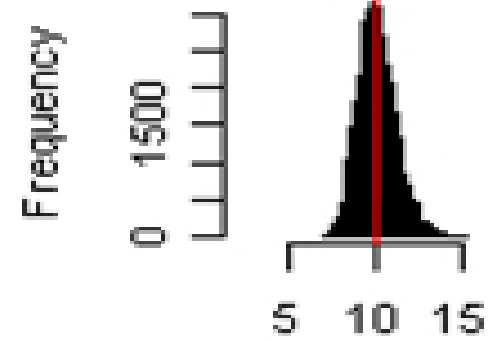
True value = red line

Posterior of b



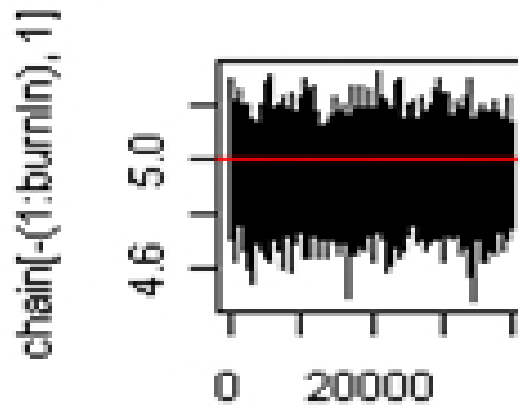
True value = red line

Posterior of sd



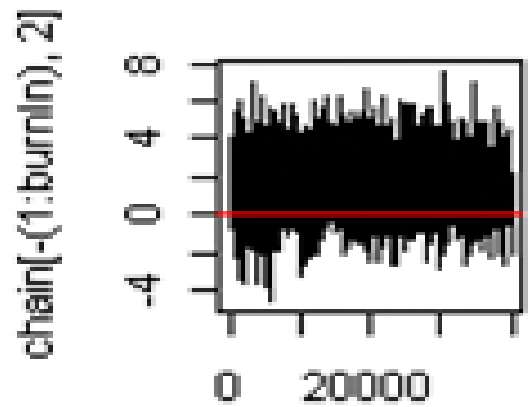
True value = red line

Chain values of a



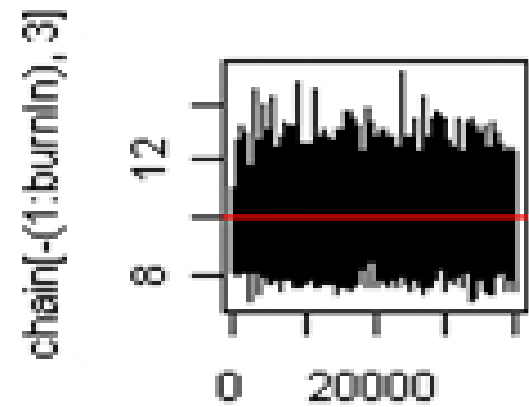
True value = red line

Chain values of b



True value = red line

Chain values of sd



True value = red line

An Example MCMC – Simple Linear Regression

範例二：簡單線性迴歸「bikes.csv」

■ 報名腳踏車比賽人數(Y)、體感溫度(X)

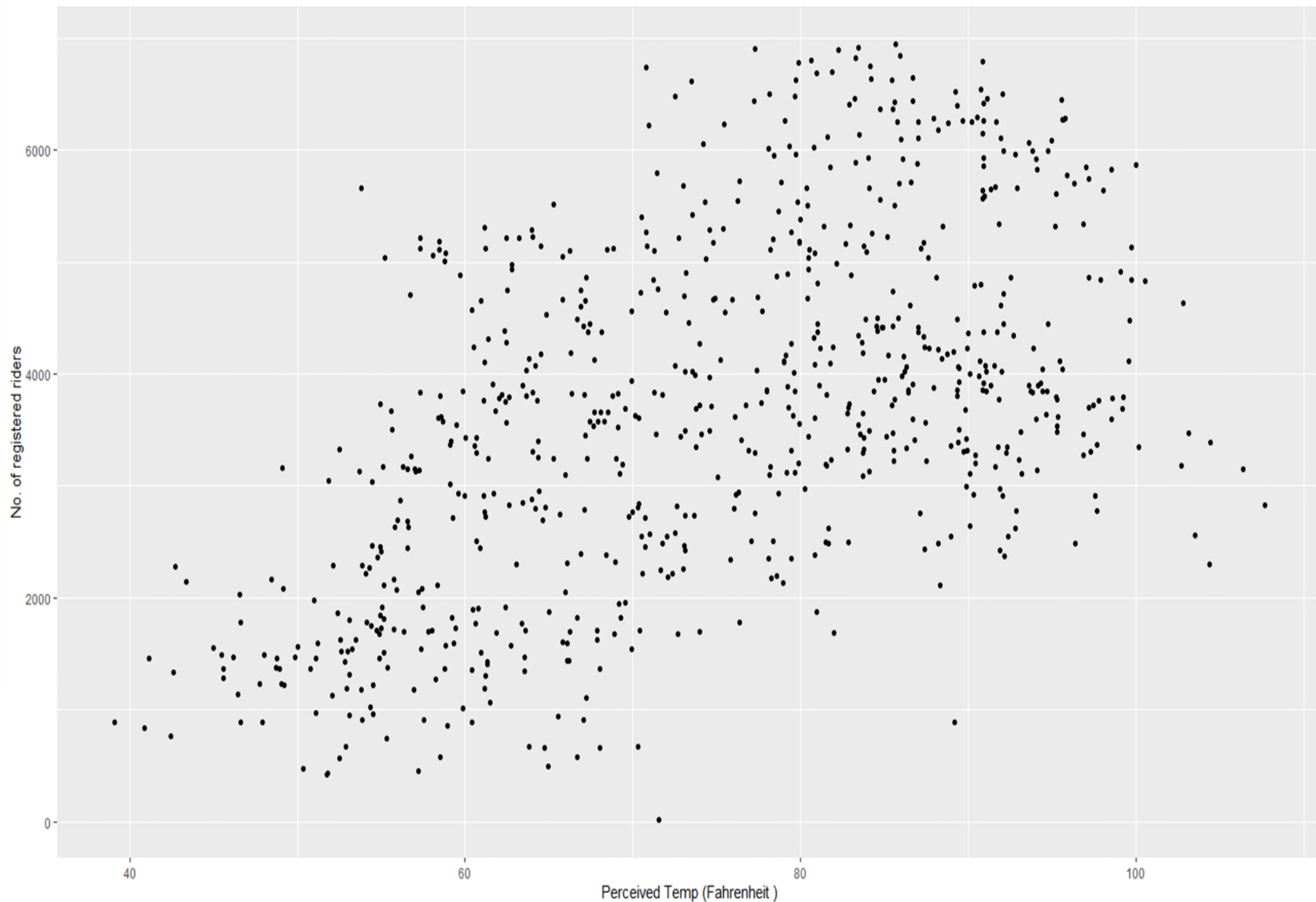
→ 由散佈圖判斷，兩者呈現線性關係，可透過一般簡單線型迴歸、或是貝氏估計求取Y與X的關係。

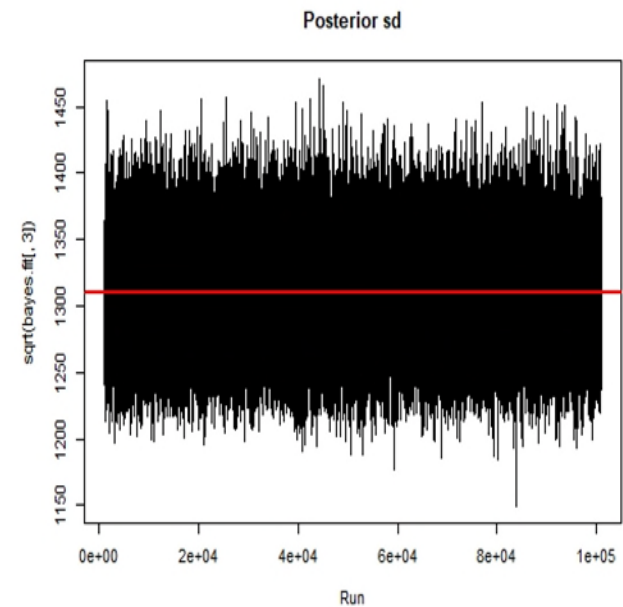
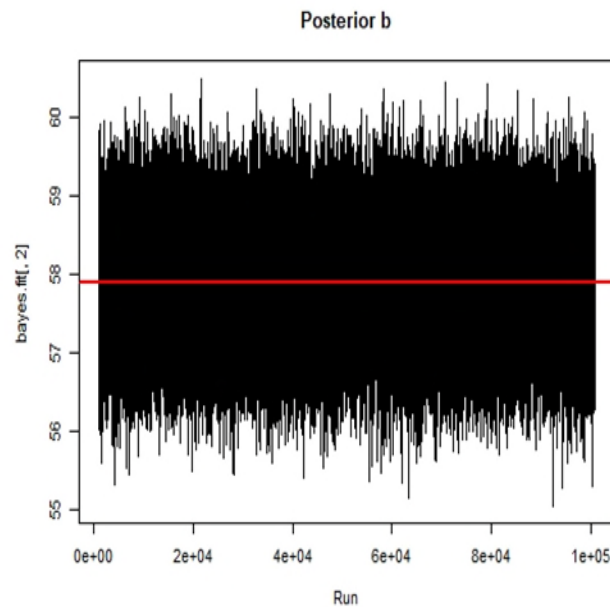
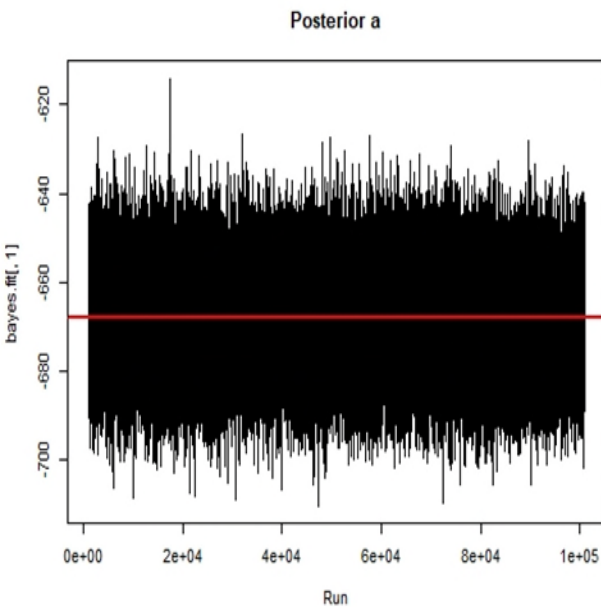
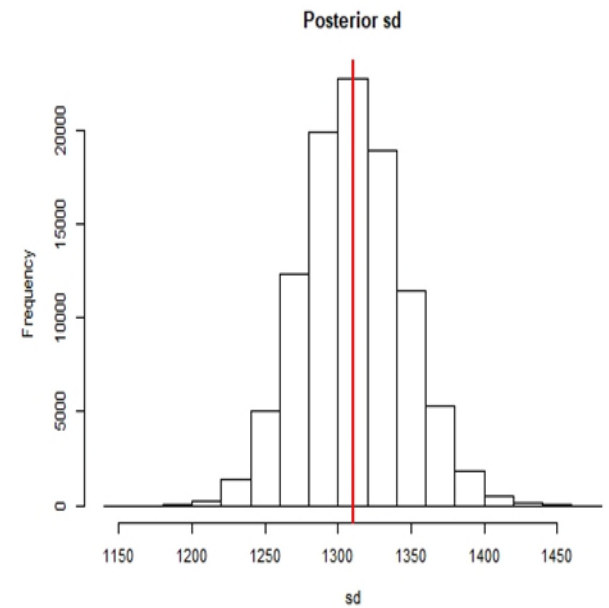
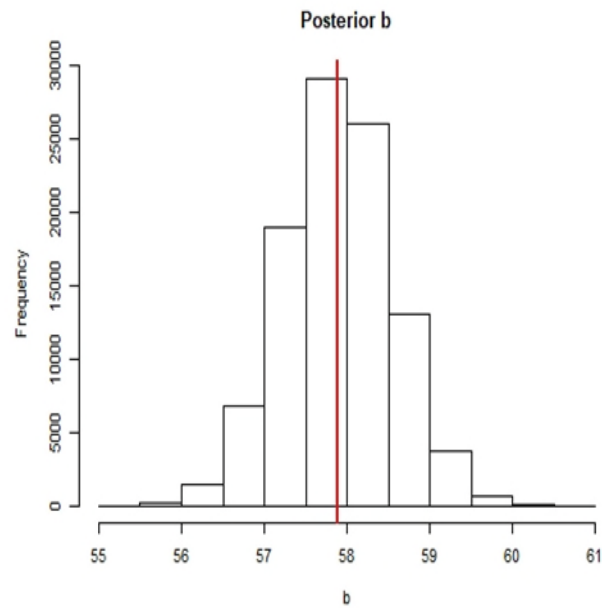
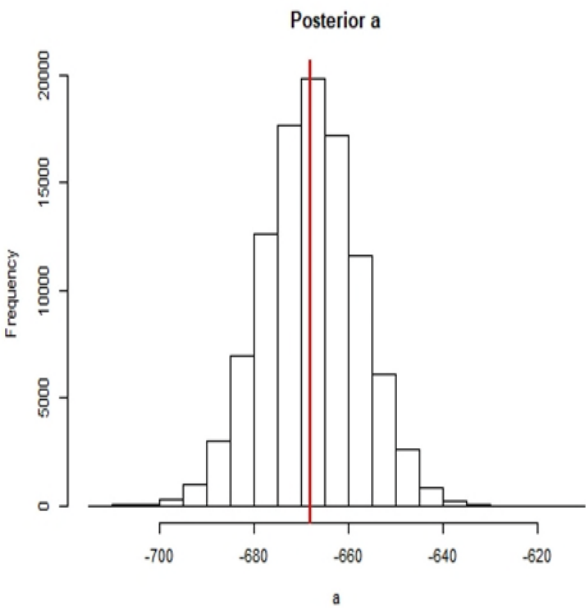
→ 一般線性迴歸結果：

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-667.916	251.608	-2.655	0.00811	**
temp_feel	57.892	3.306	17.514	< 2e-16	***

Residual standard error: 1310 on 729 degrees of freedom
Multiple R-squared: 0.2961, Adjusted R-squared: 0.2952
F-statistic: 306.7 on 1 and 729 DF, p-value: < 2.2e-16

報名腳踏車比賽人數(Y)、體感溫度(X)

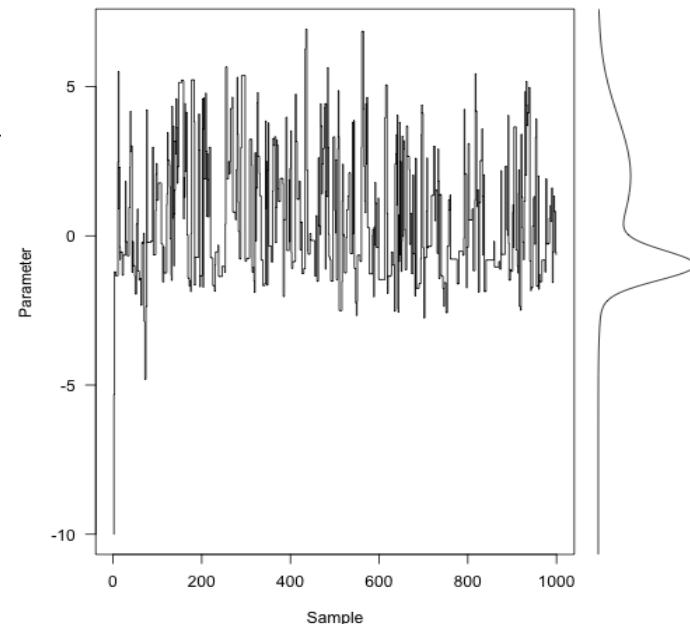




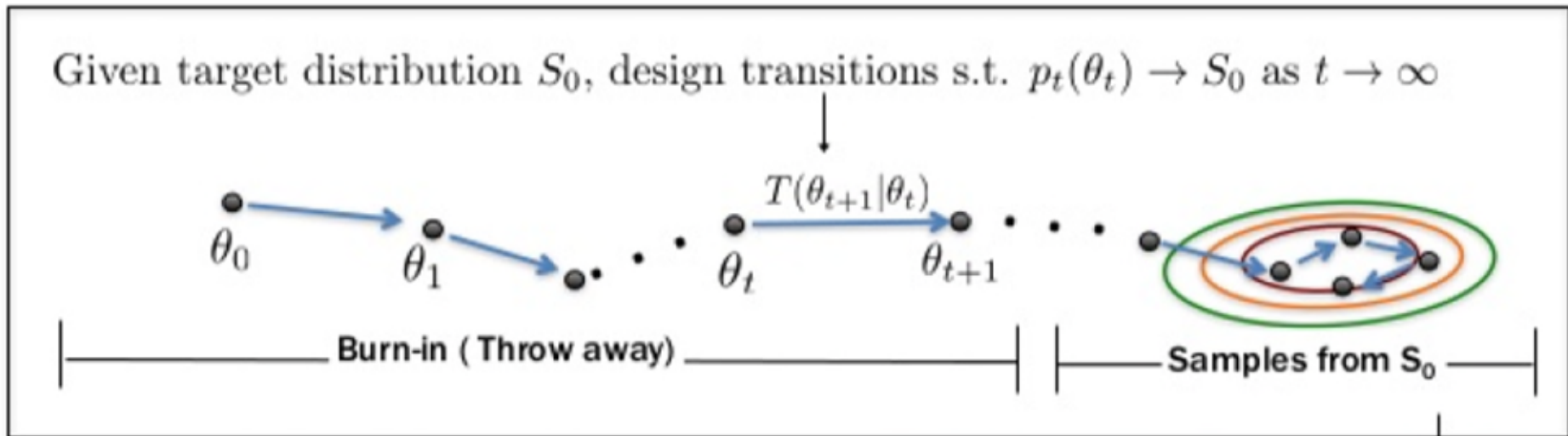
MCMC Estimates of Bikes Data (Simple Linear Regression)

MCMC方法簡介(續)

- MCMC的關鍵之一在於如何透過馬可夫鏈，遞迴產生近似驗後分配的觀察值。
- 為了確保MCMC產生的亂數是否平穩，通常捨棄較前面的觀察值(Burn-in sample)。
- 例如：從5萬個MCMC樣本刪除前面2萬個，以之後的3萬個觀察值估計參數，也可建立Credible Set、類似信賴區間的區間估計值。



MCMC—以馬可夫鏈遞迴式的產生亂數



Q：如何檢查MCMC
樣本的平均性？

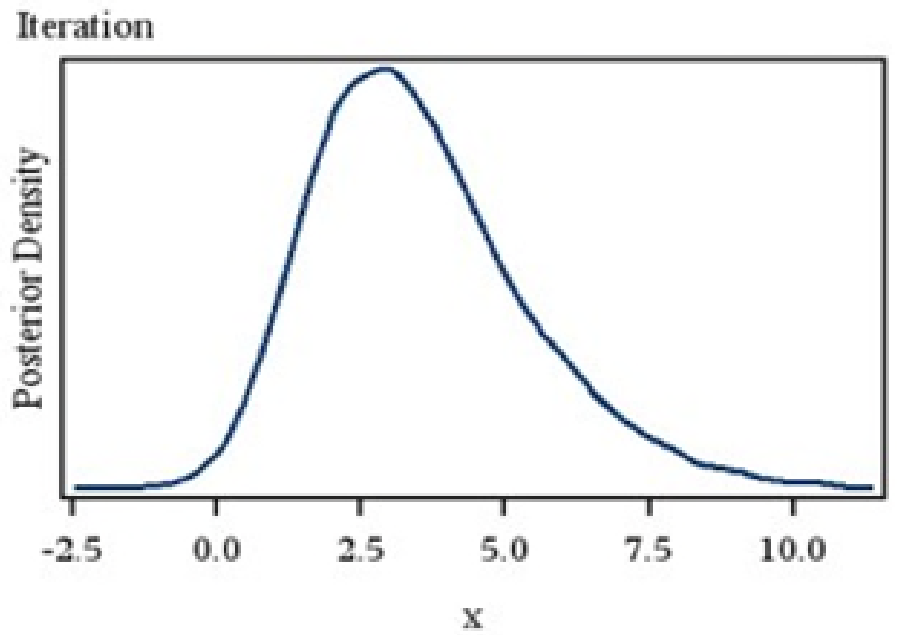
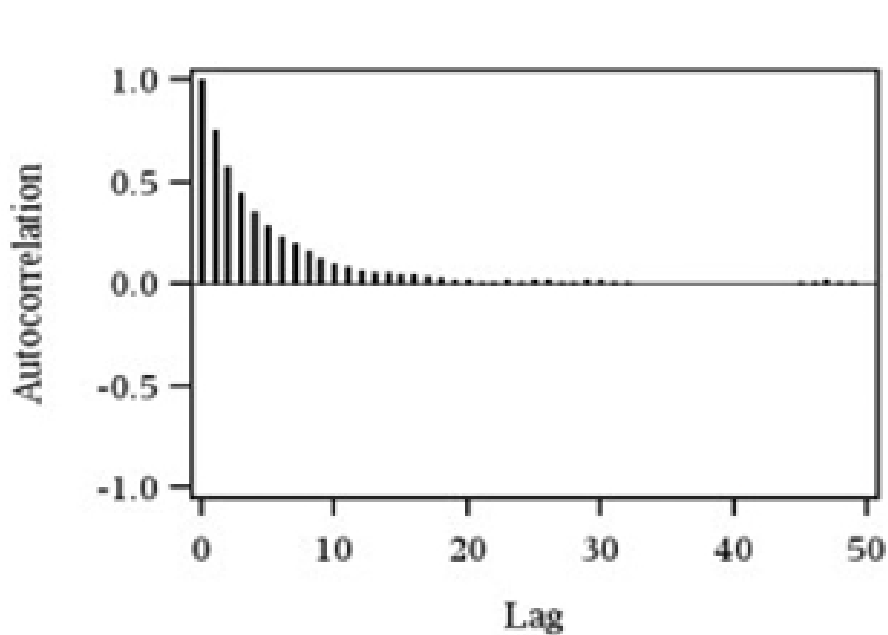
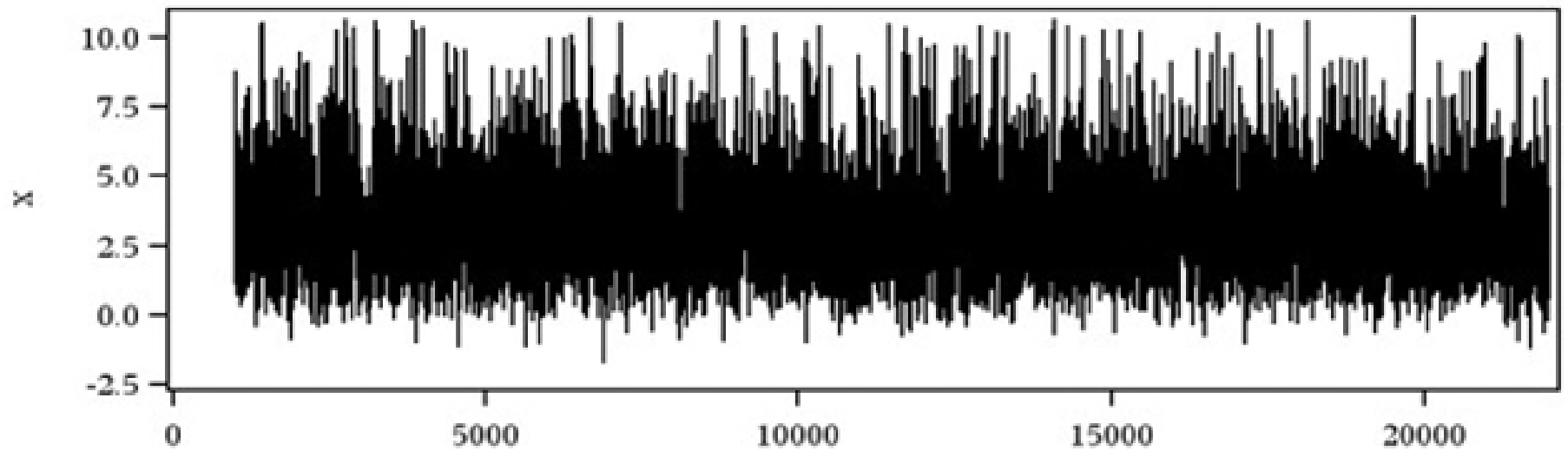
$$I = \langle f \rangle_{S_0} \approx \hat{I} = \frac{1}{T} \sum_{t=1}^T f(\theta_t)$$

$$\text{Bias}(\hat{I}) = \mathbb{E}[\hat{I} - I] = 0$$

$$\text{Var}(\hat{I}) = \tau \frac{\text{Var}(f)}{T}$$

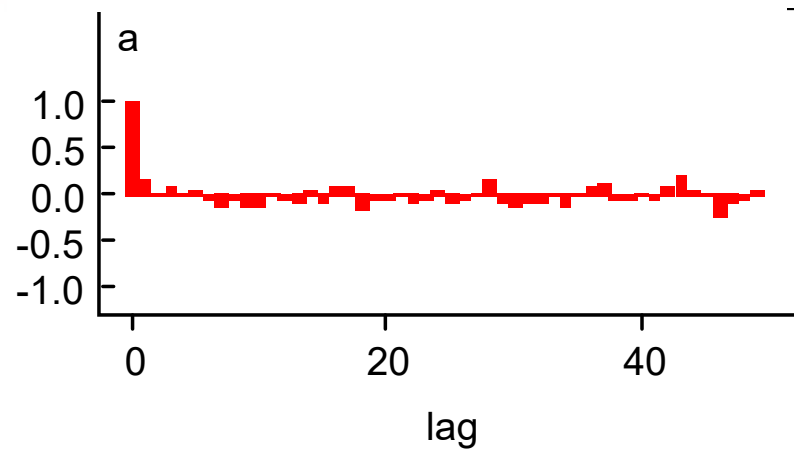
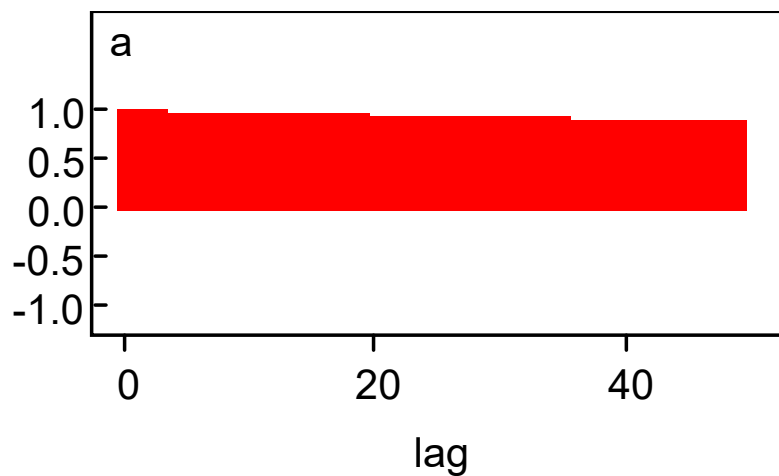
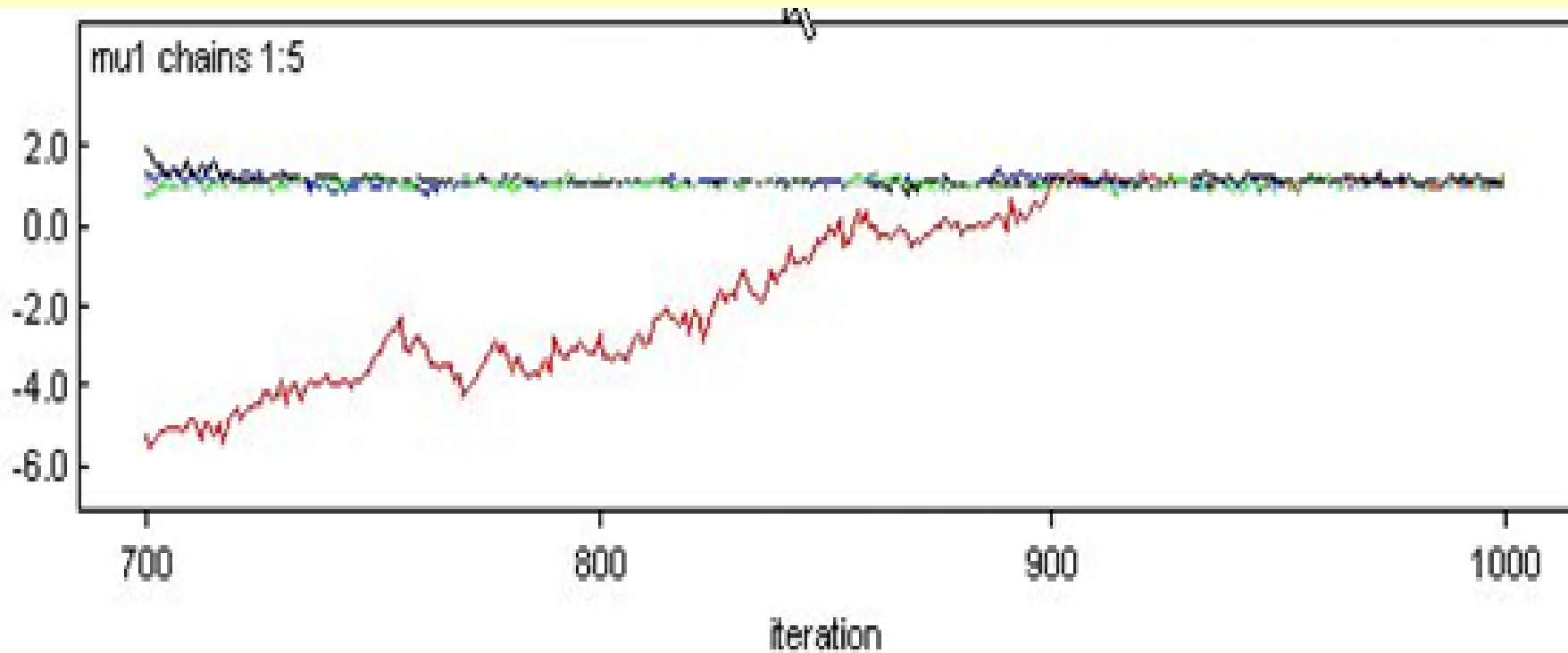
Auto correlation time

Diagnostics for x



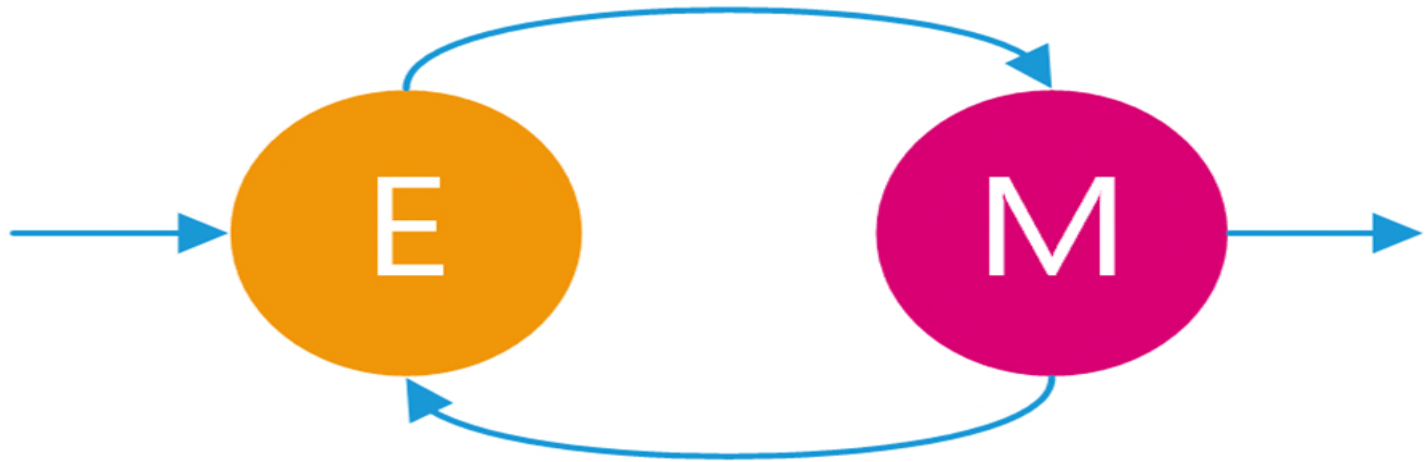
MCMC樣本平穩性-需符合亂數(Randomness)特性

關於 Burn-in Sample ○ ○ ○

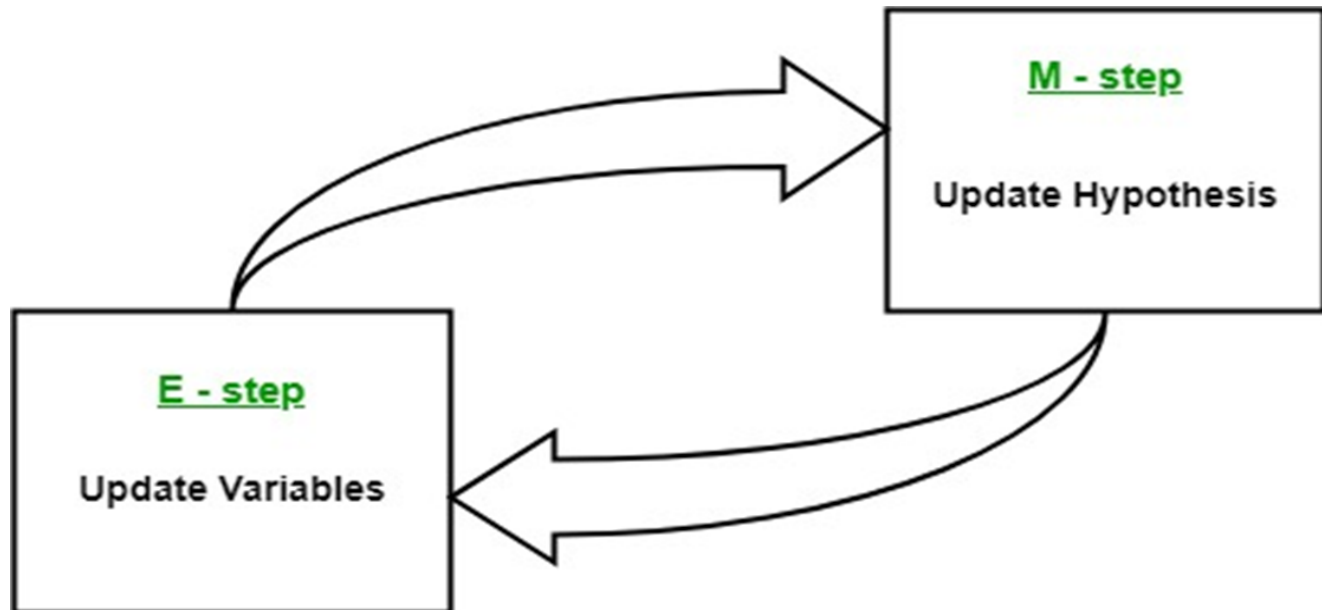


MCMC的使用建議

- MCMC可以結合過去經驗與觀察值，同時也可加入限制式。
 - 根據經驗值、觀察值的樣本人數決定共變異數矩陣，進一步結合MCMC進行估計。
 - 如果缺乏過去經驗，可透過無資訊、弱資訊先驗分配，或是敏感度分析檢驗結果。
 - MCMC模擬長度更為關鍵，建議至少去掉一萬個Burn-in Sample。



ALGORITHM



EM Algorithm (期望最大化演算法)

■ EM (Expectation-Maximization) 是一種常用於含有隱藏變數(latent variables)或資料不完整時的參數估計方法。

→ 在無法直接最大化 likelihood function，透過反覆迭代逼近最大概似估計(MLE)。

■ 完整資料的Likelihood $\rightarrow L(\theta | Y, Z)$

但只能觀察到Y $\rightarrow L(\theta | Y) = \sum_Z L(\theta | Y, Z)$

EM Algorithm的想法

■ EM 的進行步驟：

- 先「猜測」隱藏變數的分布(E-step)；
- 再將其視為已知資料更新參數(M-step)；
- 重複直到收斂。

註：E-step (Expectation Step)

M-step (Maximization Step)

- ## ■ EM能處理隱藏變數問題，likelihood在每次迭代不會下降，而且實作相對穩定，也可用於 mixture model 中。

EM Algorithm的範例

- 假設資料由兩個常態分配構成，但不清楚各組的參數、發生機率。

→ R可執行EM的常見模組之一「mixtools」：

```
set.seed(456)
```

```
x=c(rnorm(100),rnorm(100,mean=3,sd=1))
```

```
fit1=normalmixEM(x[c(1:30,101:130)], k=2)
```

```
plot(fit1,which=2)
```

```
summary(fit1)
```

註：以kernel為EDA檢視工具，在樣本數較少時不易區分群集邊界。

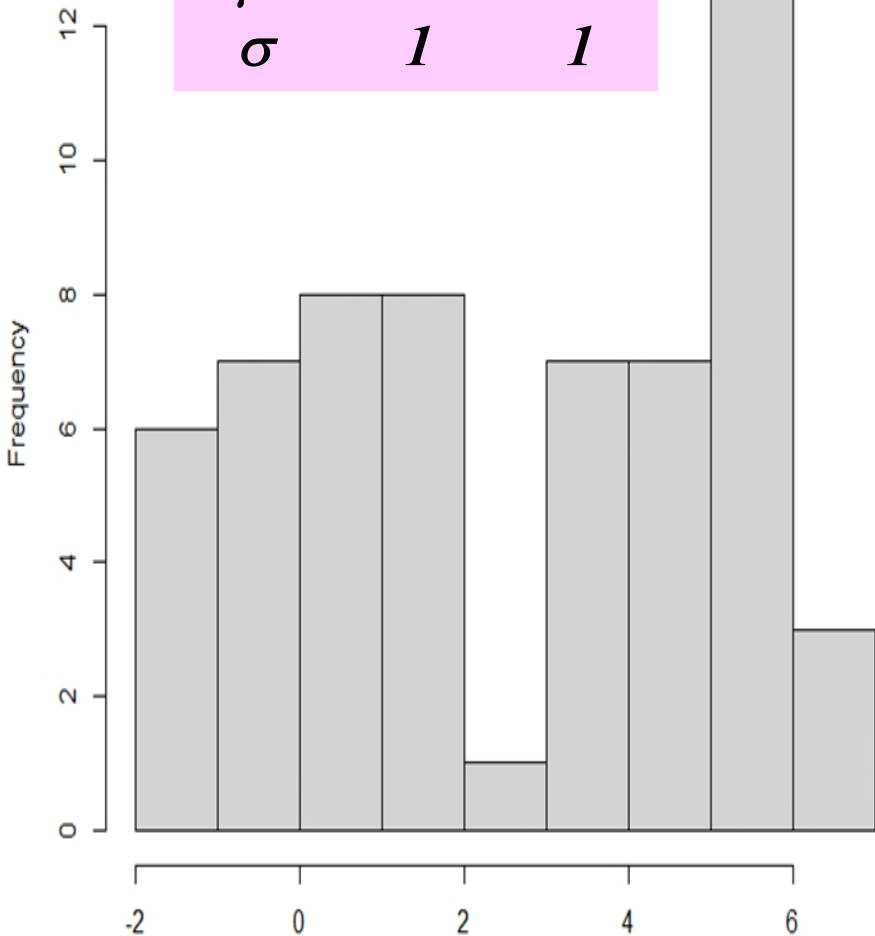
EM範例:兩個常態混合分配 ($n=30,100$)

真實 分量1 分量2

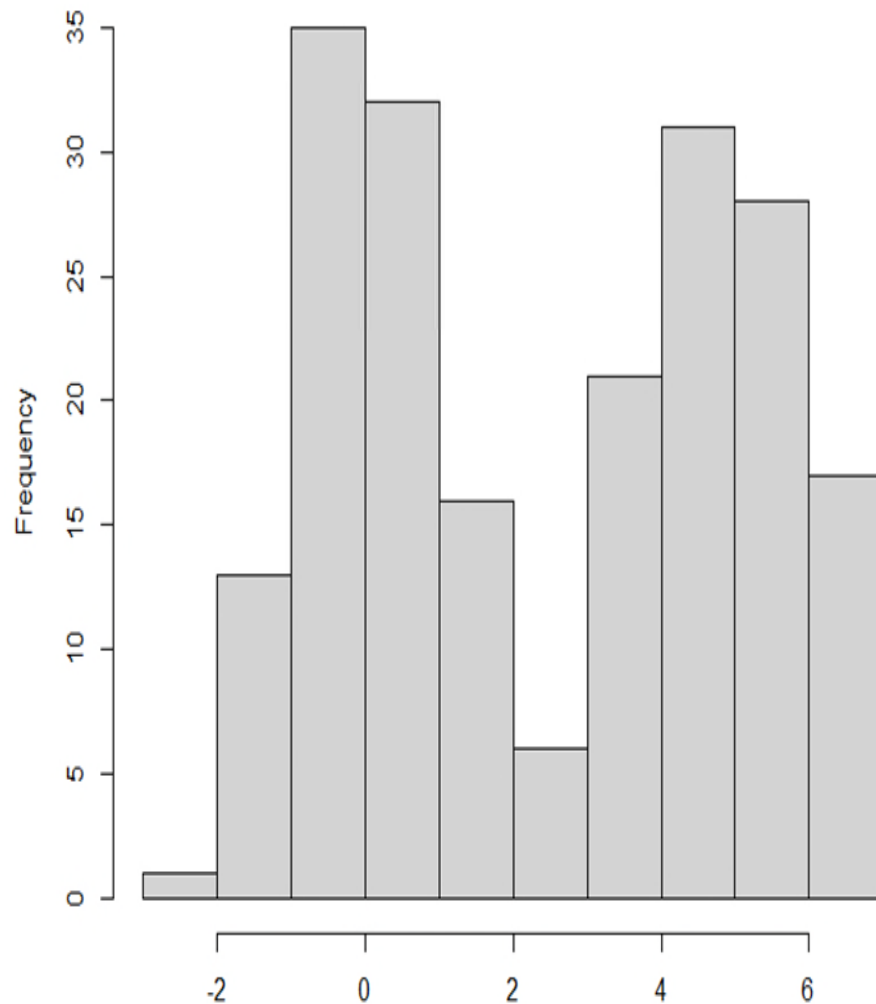
λ .5 .5

μ 0 5

σ 1 1

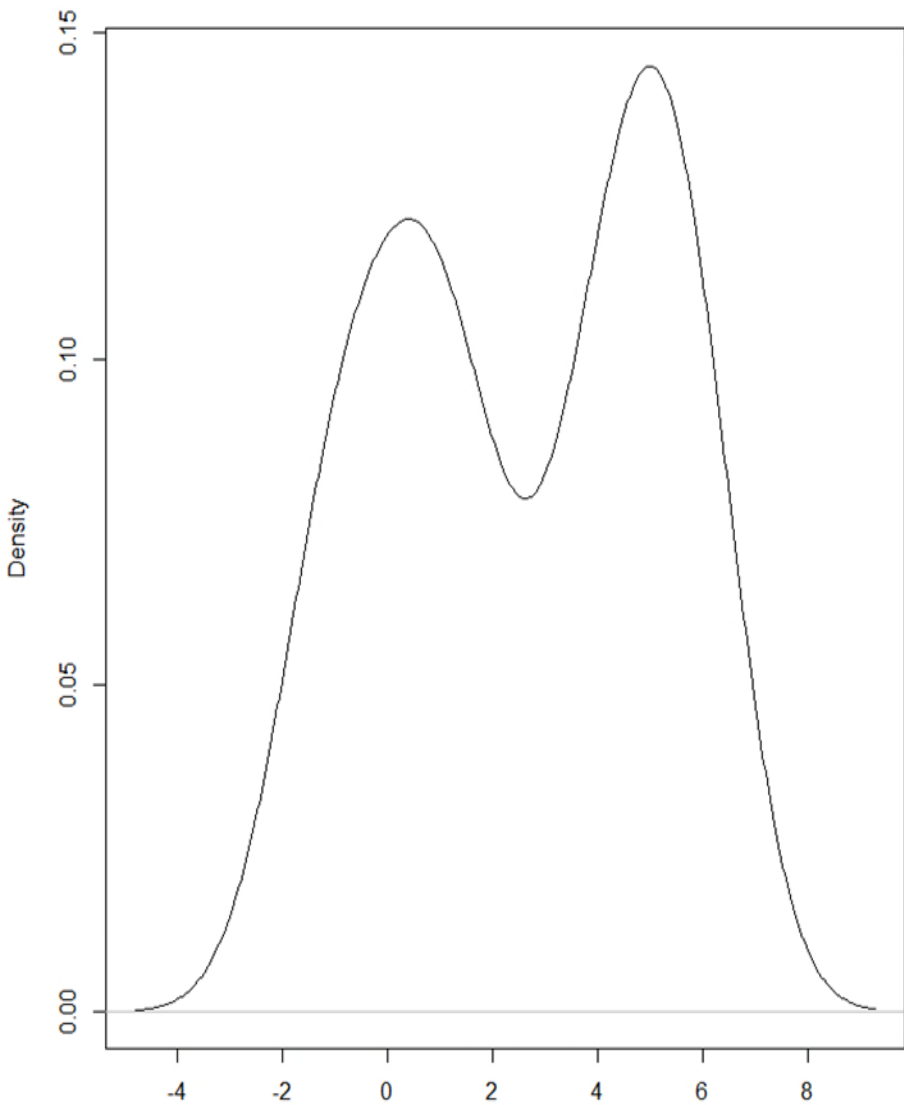


60 observation

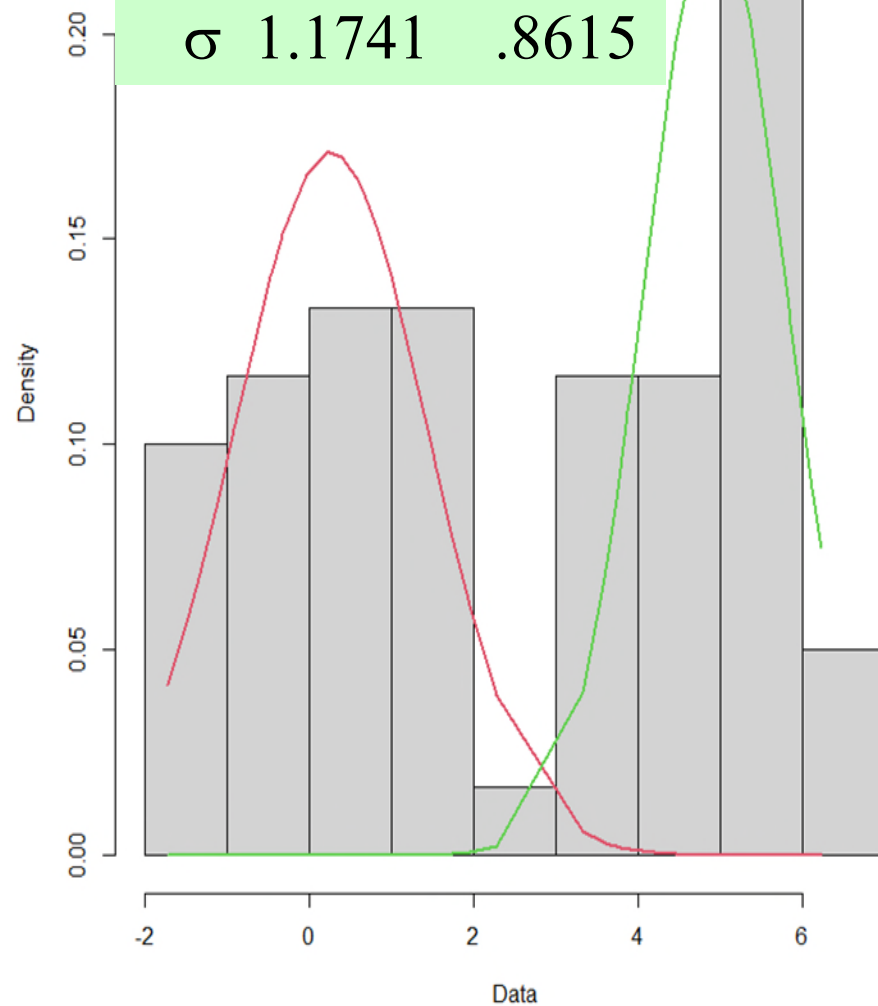


200 observation

Kernel與EM分析比較 ($n=30$)

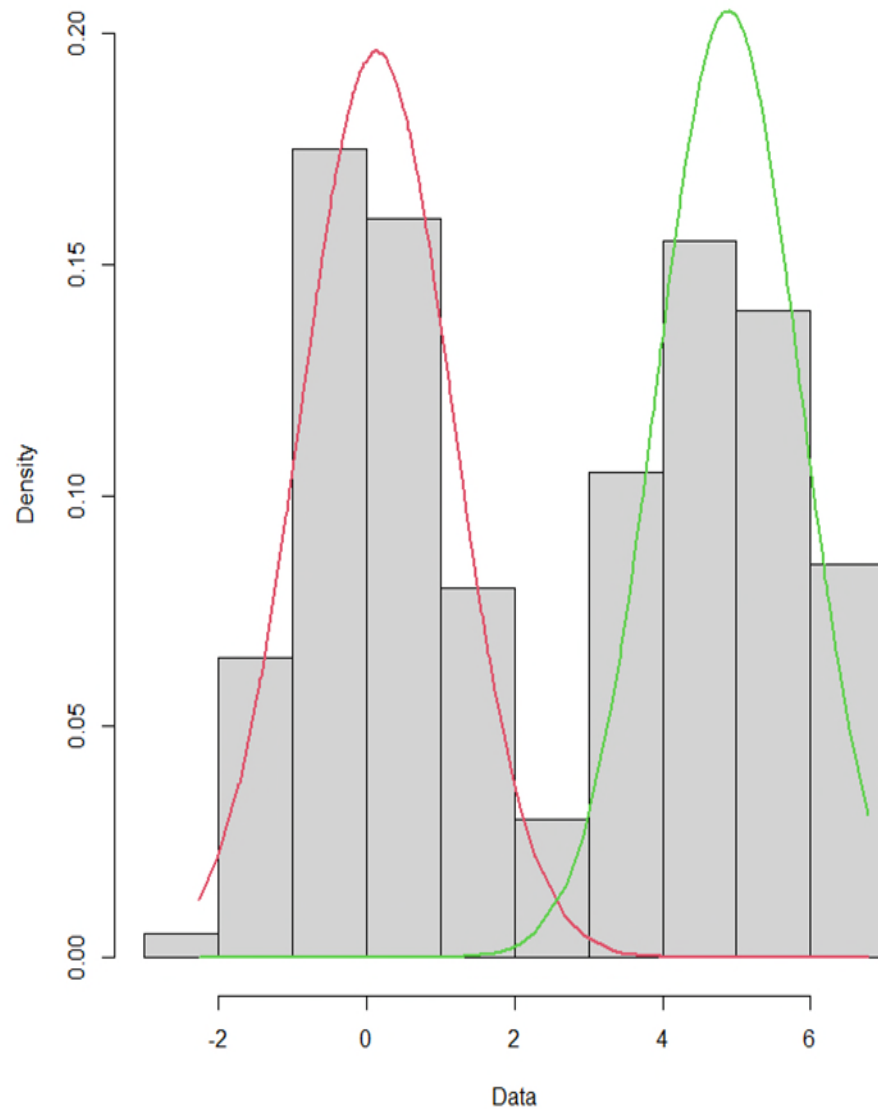
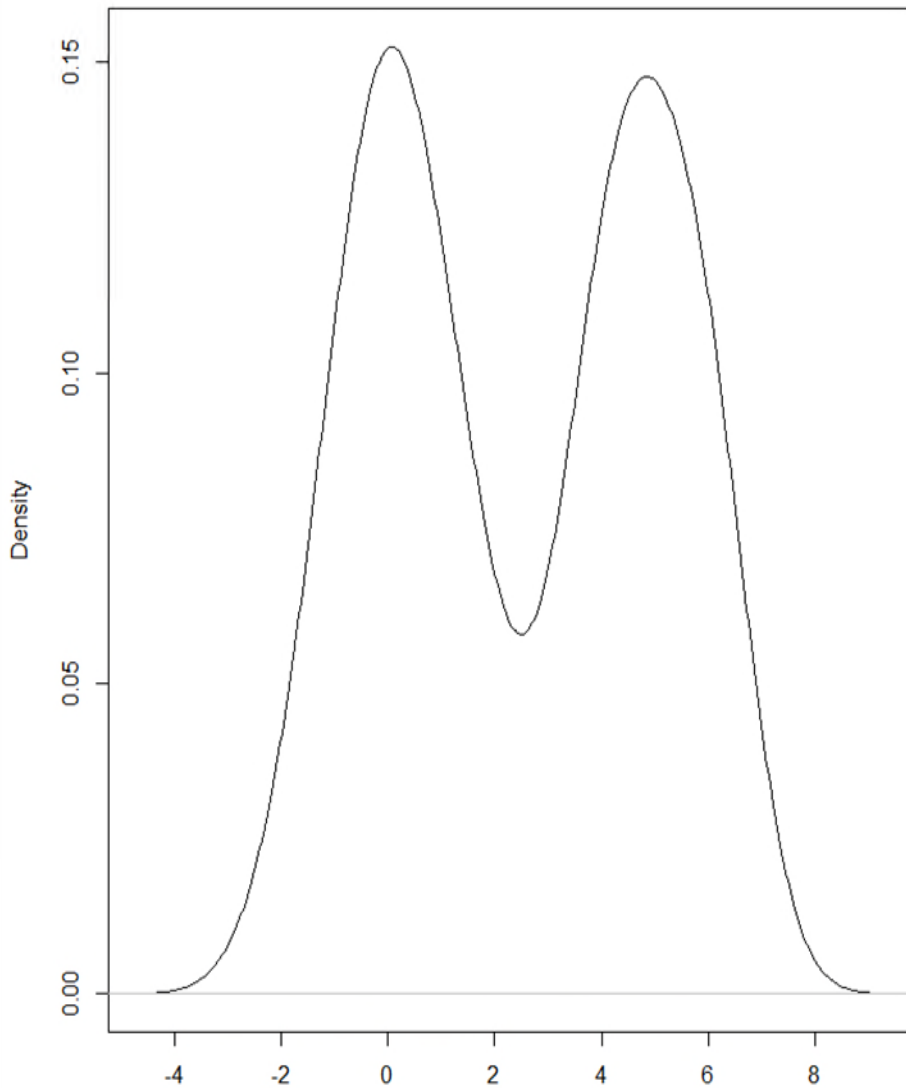


估計	分量1	分量2
λ	.5036	.4964
μ	.2587	4.9322
σ	1.1741	.8615



Kernel與EM分析比較 ($n=100$)

估計	分量1	分量2	μ	.1385	4.8897
λ	.5026	.4974	σ	1.0222	.9680



另一個EM Algorithm範例

- 多項分配中如果沒有遺漏值：

$$\pi_{\theta} = \left(\frac{1}{2} + \frac{1}{4}\theta, \frac{1}{4}(1 - \theta), \frac{1}{4}(1 - \theta), \frac{1}{4}\theta \right).$$

則對數概似函數為

$$l(\theta; \mathbf{x}) = C + x_1 \ln \left(\frac{1}{2} + \frac{1}{4}\theta \right) + (x_2 + x_3) \ln(1 - \theta) + x_4 \ln(\theta)$$

可套用優化方法求取參數值，但屬於非線性的估計。

EM Algorithm 範例二(續)

- 可將第一項拆成兩項：

$$\pi_{\theta}^* = \left(\frac{1}{2}, \frac{1}{4}\theta, \frac{1}{4}(1-\theta), \frac{1}{4}(1-\theta), \frac{1}{4}\theta \right)$$

或 $(y_1 + y_2, y_3, y_4, y_5)$ ，原先的對數概似函數變成

$$l(\theta; \mathcal{X}, \mathcal{Y}) = C + y_2 \ln(\theta) + (y_3 + y_4) \ln(1 - \theta) + y_5 \ln(\theta)$$

其中 $\mathcal{X} = (y_1 + y_2, y_3, y_4, y_5)$ 、 $\mathcal{Y} = y_2$ ，或是

$$f(\mathcal{Y} | \mathcal{X}, \theta) = \text{Bin} \left(y_1 + y_2, \frac{\theta/4}{1/2 + \theta/4} \right)$$

EM Algorithm 範例二(程式碼)

- 以下為R程式碼： $(x_1, x_2, x_3, x_4) = c(125, 18, 20, 34)$

```
theta_old = 0.5
# EM iteration
tol = 1e-8
max_iter = 1000
theta_trace = c(theta_old)
for (iter in 1:max_iter){ # ----- E-step -----
  p_old = (theta_old / 4) / (0.5 + theta_old / 4) # expected y2
  Ey2 = (x1) * p_old # ----- M-step -----
  theta_new = (x4 + Ey2) / (x2 + x3 + x4 + Ey2)
  theta_trace <- c(theta_trace, theta_new) # convergence check
  if(abs(theta_new - theta_old) < tol){
    break
  }
  theta_old <- theta_new
}
cat("Estimated theta =", theta_new, "\n")
cat("Iterations =", iter, "\n")
```