國立政治大學統計學系研究所

碩 士 學 位 論 文

LASSO於羅吉斯迴歸模型之估計的應用

Application of LASSO Estimation of a Logistic Regression Model

指導教授：薛 慧 敏 博士

研究生：鍾 其 昀 撰

中 華 民 國 一 百 零 六 年 一 月

摘要

隨著資料量龐大,解釋變數過多的時代來臨,變數選取將是我們重要的議題。在線性迴歸分析中，傳統採用最小平方法(least square method)來估計模型，然而得到的迴歸係數估計值的偏差雖然比較小，但其變異程度卻較大，且預測得也不夠精準。若是考慮對迴歸係數加入限制式時,則估計量將與原本的最小平方法有何差異,偏差與標準差之間的比較。接著將此估計法應用至羅吉斯迴模型時,利用三筆實際資料，比較與最大概似估計(maximum likelihood estimate，簡稱MLE)法建立的迴歸模型及預測準確率，並於模擬實驗中，以表格及圖型呈現兩方法在估計量上的差異。

目錄

表目錄

圖目錄

1. 緒論

迴歸分析目的主要是建立反應變數(Y)與解釋變數($X_1,\ldots,X_p$)之間的系統關係式，未來當給定解釋變數後，則可對反應變數做預測。當數據大、解釋變數多時，則探索過程中最重要的就是決定模型中應納入哪些重要變數。在線性迴歸分析中，傳統的變數選取方法有普通最小平方法(ordinal least square method)、逐步迴歸(stepwise regression)、最佳子集挑選(best subset selection)。有別於傳統的變數選取方法，近來，壓縮係數(shrinkage)類型的方法則越來越受歡迎，如脊迴歸(ridge regression)(Hoerl and Kennard，1970)、LASSO(Tibshirani，1996)、Garrote(Breidman，1995)，或是 Elastic Net(Zou and Hastie，2005)等。在計算上，壓縮係數法並不會比傳統的最小平方法來得繁雜，甚至速度還更快，這些方法已經成為各領域在處理多維度資料的重要工具。

Tibshirani(1996)提出 LASSO 統計方法，它在解最小估計法的問題中加入懲罰(penalty)函數，此懲罰函數為係數絕對值和的遞增函數。所以為了降低懲罰項，部分係數將被壓縮為零，達到變數選取的目的。另有研究顯示這一方法用於高維度(變量個數遠大於樣本量)強相關、小樣本的生存資料分析非常有效(Sill，2014)。這類數據資料常發生在基因數據(genetic dataset)、臨床醫學(clinical medical science)、影像辨識(image identification)等。

跟普通最小平方法(OLS)相較，OLS 估計量偏差(bias)較小，但變異程度(variability)較大。在預測與模擬實驗中，得到 LASSO 估計量的均方誤差(mean square error)比較小。另一理由是當部分的係數縮減至零時，這些非零係數的自變數能展示與因變數更強烈的關係，解釋上也能比較清楚，當新的資料進來，也能避免過度配適，造成預測誤差太大，模型變得不穩定。

本文首先將探討在一般線型迴歸模型中，LASSO 與傳統普通最小平方法的差

異。針對估計值的計算，我們也將介紹相關的凸函數最佳化問題。另外，我們將針對二元型態的反應變數，將 LASSO 方法運用在羅吉斯迴歸模型的估計上，並且與傳統的最大概似估計法(MLE)做比較。第三章則將利用三組實際資料，比較 LASSO 與 MLE 在迴歸係數估計以及預測分類上的結果。在第四章中，我們以電腦模擬來探討兩個估計量的差異，包括其抽樣分配、不偏性以及變異程度。

2. 研究方法

2.1 線性迴歸之 LASSO

假設一組共有 n 筆資料的樣本$(X_i, Y_i), i = 1, \ldots, \mathrm{n}$，其中$y_i$是第$i$個觀察值的反應變數，$X_i = (X_{i1}, \ldots, X_{ip})$是第$i$個觀察值的解釋變數向量，假設有 p 個解釋變數，則X_{ij}為第i個觀察值之第j個解釋變數，$i = 1, \ldots, \mathrm{n}$，$j = 1, \ldots, \mathrm{p}$。當反應變數為連續型時，考慮一般的多元線性迴歸模型，則可得

$$\mathrm{Y} = X\beta + \varepsilon \tag{2.1}$$

其中 $\mathrm{Y} = (Y_1, \ldots, Y_n)^T$，$\mathrm{X} = \left[\mathbf{1}, {X_1}^T, \ldots, {X_n}^T\right]^T$為一 $\mathrm{n} \times (\mathrm{p}+1)$的實驗矩陣（design matrix），$\mathbf{1} = (1, \ldots, 1)^T$；$\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T$為迴歸係數，$\varepsilon = (\varepsilon_1, \ldots, \varepsilon_{n,})^T$為隨機誤差，假設隨機誤差滿足$\mathrm{E}(\varepsilon) = 0$，$\mathrm{Var}(\varepsilon) = \sigma^2 I$，I 為 n×n 單位矩陣。

傳統上在估計迴歸係數時採用最小平方法(least square method)，目的是想找到與樣本觀察值最接近的迴歸模型，以式子(2.1)線性迴歸為例，即求解

$$\mathrm{Min} \ \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

其中$\hat{Y}_i = \hat{\beta}_0 + \sum_{j=1}^{p} X_{ij}\hat{\beta}_j$為$Y_i$的預測值，$\hat{\beta}_0, \hat{\beta}_j$為截距$\beta_0$及$\beta_j$的最小平方法估計量。

Tibshirani(1996)提出最小絕對壓縮挑選機制（least absolute shrinkage and selection operator）或簡稱 LASSO。他們在原先的最小平方法估計法上加入對迴歸係數長度的限制式，則相對應迴歸係數估計量的定義為下列問題之解:

$$\mathrm{Min} \ \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \quad \text{subject to} \ \sum_{j=1}^{p}|\beta_j| \le \mathrm{t} \tag{2.2}$$

其中 $\mathrm{t} \in R^+$ 是迴歸係數長度總和的上界值。

2.1.1 LASSO 的估計量與其變異數矩陣

由式子(2.2)可以清楚得到，在這個最佳化問題中，目標函數是最小化殘差平方和$\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2$，限制式為$\sum_{j=1}^{p}|\beta_j|\le$ t。藉由拉格朗日函數來求解此問題，可得下列式子:

$$\mathrm{L}(\beta_0,\beta_1,\dots,\beta_p)=\sum_{i=1}^{n}(Y_i-\beta_0-\beta^T X_i^T)^2+\lambda(\sum_{j=1}^{p}|\beta_j|-\mathrm{t})$$

其$\lambda\ge 0$為拉格朗日乘數(Lagrange multiplier)。

在估計迴歸係數時，我們得先對樣本資料矩陣X的每一行解釋變數標準化，得到X_{ij}，i=1, … , n, j=1, … , p + 1。其將滿足$\bar{X}_{\cdot j}=0,\widehat{\mathrm{var}}(X_{ii})=1,\widehat{\mathrm{corr}}(X_{ij},X_{ik})=0,j,k=1,\dots,\mathrm{p},j\ne k$，其中$\bar{X}_{\cdot j}$為樣本行平均，$\widehat{\mathrm{var}}(X_{ii})$為樣本變異數，$\widehat{\mathrm{corr}}(X_{ij},X_{ik})$為樣本相關係數。此時資料矩陣若為一正交矩陣，即$X^TX=I$，其中 I 是p × p的單位矩陣。則我們可以導出$\widehat{\beta_0}=\bar{y}$，

$$\hat{\beta}_j=\mathrm{sign}(\hat{\beta}_j^0)(|\hat{\beta}_j^0|-\frac{\lambda}{2})^+=\begin{cases}\hat{\beta}_j^0-\frac{\lambda}{2} & if\ \ \hat{\beta}_j^0>\frac{\lambda}{2}\\ \hat{\beta}_j^0+\frac{\lambda}{2} & if\ \ \hat{\beta}_j^0<-\frac{\lambda}{2}\\ 0 & if\ \ -\frac{\lambda}{2}\le\hat{\beta}_j^0\le\frac{\lambda}{2}\end{cases}\qquad(2.3)$$

其中$\bar{y}$是樣本反應變數平均值，$\hat{\beta}_j^0$為最小平方法的估計量，證明請見附錄一、二。

LASSO 與脊迴歸的差異只在限制式中採用L^1或L^2長度。若將式子(2.2)定義中的限制式變更成$\sum_{j=1}^{p}{\beta_j}^2\le$ t，則形成脊迴歸(Ridge regression)最佳化問題。則其迴歸係數之對應關係如：

$$\sum_{j=1}^{p}|\beta_j|=\sum_{j=1}^{p}\beta_j^2/|\beta_j|\qquad(2.4)$$

同理，已知脊迴歸的正規方程式為$(X^T\mathrm{X}+\lambda\mathrm{I})\beta=X^T\mathrm{Y}$，其中$\beta=(\beta_1,\dots,\beta_P)^T$，則估計量為$\widehat{\beta_R}=(X^T\mathrm{X}+\lambda\mathrm{I})^{-1}X^T\mathrm{Y}$。

由於資料矩陣 X 已經過標準化，藉由上述式子(2.4)的關係，LASSO 估計量因此可寫成$\widehat{\beta_L}=(X^T\mathrm{X}+\lambda W^{-1})^{-1}X^T\mathrm{Y}$。其中 W 為一對角矩陣，其第 j 個對角元素為$|\beta_j|$, j=1, ⋯, p，則W^{-1}為其廣義反矩陣。又因誤差項ε具相等的常數變異數

σ^2，所以 Var(Y)= σ^2，迴歸係數估計量$\widehat{\beta_L}$的變異數是

$$\mathrm{Var}(\widehat{\beta_L}) = (X^T\mathrm{X} + \lambda W^{-1})^{-1} X^T X (X^T\mathrm{X} + \lambda W^{-1})^{-1}\sigma^2 。$$

2.1.2 LASSO 與 OLS 估計量的比較

當樣本資料 X 為一正交矩陣時，可得迴歸係數的 LASSO 解析解，見上一節中式子(2.3)。一般情形下，LASSO 的根通常是用數值(numeric)方法近似來求得。當我們對迴歸係數加入限制式($\sum_{j=1}^{p}|\beta_j| \leq \mathrm{t}$)時，若解釋變數個數多時，則某些迴歸係數將被截斷成零。所以在做係數估計時，同時也做了模型選取(model selection)的動作。以下我們將利用平面圖呈現，當解釋變數個數為 2 時(p=2)，LASSO 與 OLS 在幾何上的意義。

圖 2.1 引用自一篇關於 LASSO 的應用的碩士論文，圖中定義橫軸為迴歸係數β_1的方向，縱軸是迴歸係數β_2的方向。則右上橢圓為 LSE 方法的目標函數的等高線圖，中心點$\hat{\beta}$為最小平方估計量。當加入限制式$\sum_{j=1}^{p}|\beta_j| \leq \mathrm{t}$時，呈現的圖形是左下方之實心菱形區域，凡在菱形內的點都是滿足限制式的可行解。LASSO 找根的過程就是從橢圓中心點$\hat{\beta}$向外放射出去，直到碰觸菱形為止，由此可知 LASSO 解將發生在菱形邊界上。若是解恰好落到座標軸，則另一方向的迴歸係數將為零。

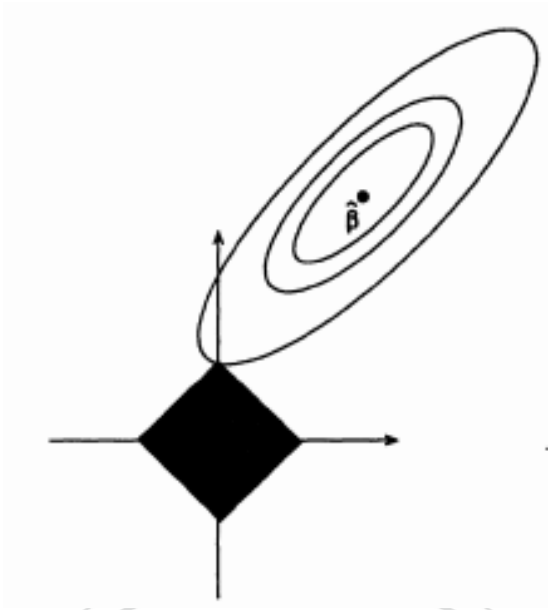


圖 2.1 當 p=2 時，LSE 之$\hat{\beta}$與 LASSO 的可行解集合。

由式子(2.3)得知，LASSO 可表示為 LSE 估計值的函數，見圖 2.2。圖中的橫軸為 LSE 估計值，紅色函數線則為 LASSO 估計值，我們可看出 LASSO 的值較為保守，其估計值較靠近零。根據(2.3)，LASSO 的估計值與門檻值 λ/2 密切相關，此處我們討論在不同的門檻值下 LASSO 估計量的偏差(bias)。已知最小平方估計量滿足不偏性(unbiased)，$E[\hat{\beta}_j^0]=\beta_j$，$\forall j=1,\cdots,p$。則我們可推得以下結果:

1. 當$\frac{\lambda}{2} \cong |\beta_j|$, $|E(\hat{\beta}_j-\beta_j)| = |E(\hat{\beta}_j^0-\frac{\lambda}{2}-\beta_j)| \leq \frac{\lambda}{2}$
2. 當$\frac{\lambda}{2} \gg |\beta_j|$, $|E(\hat{\beta}_j-\beta_j)| = |E(0-\beta_j)| = |\beta_j| < \frac{\lambda}{2}$
3. 當$\frac{\lambda}{2} \ll |\beta_j|$, $|E(\hat{\beta}_j-\beta_j)| = |E(\hat{\beta}_j^0-\frac{\lambda}{2}-\beta_j)| = \frac{\lambda}{2}$

圖 2.3 為此三種情況下兩種估計方法的結果，其中圓點代表 LSE 估計值，x 標記為真實參數值，三角點則為 LASSO 估計值，虛線為門檻值 λ/2，實線橫軸表估計值的範圍。當門檻值接近真實參數值時，LASSO 傾向為零；當門檻值明顯大於參數的絕對值時，LASSO 將多為零；當門檻值明顯低於參數絕對值時，LASSO 比 LSE 更靠近零的位置，距離為$\lambda/2$。因此在各種情況下，LASSO 都是一個有偏估計量，其偏度至多為 λ/2。

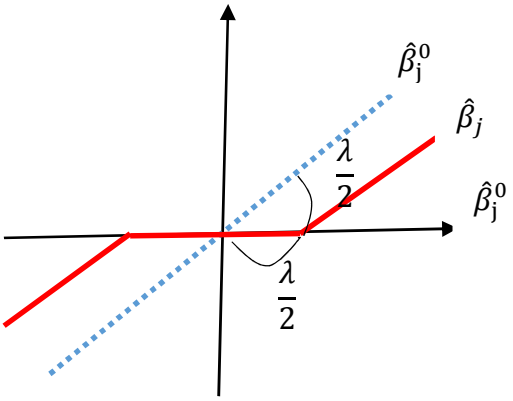


圖 2.2 LASSO 與 LSE 的關係圖。

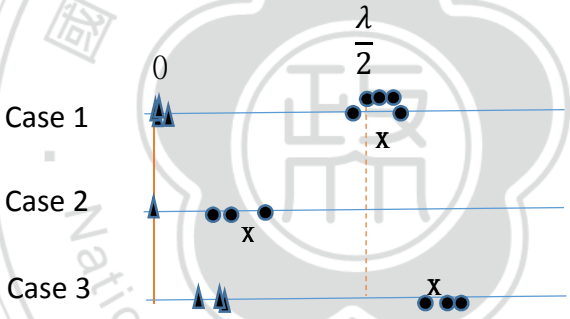


圖 2.3 LSE 與 LASSO 的比較。其中圓點代表 LSE 估計值，x 標記為真實參數值，三角點則為 LASSO 估計值，虛線為門檻值 $\lambda/2$，實線橫軸表估計值的範圍。

另外由於 LASSO 在某些樣本中的結果為 LSE 加減門檻值，另一些樣本則會得到零的結果，所以 LASSO 的分配是由 LSE 與退化至零的分配所組合的混合分配 (mixture distribution)。故可推斷 LASSO 估計量的變異程度將不超過 LSE 的變異程度。

2.1.3　凸函數最佳化與對偶理論

在線性規劃中，凸函數最佳化是實務上經常使用的方法。此節主要是針對具有門檻參數的迴歸模型的迴歸係數估計，我們將介紹運用凸函數最佳化在解根上的原理與方法。在求解最佳化問題時，通常原始問題(primal problem)與對偶問題(dual problem)是成雙成對的，有時候原始問題並不容易求解，反之從其對偶問題著手，可以簡化問題。

假設原始問題為在給定一組限制函數為$f_i(x) \leqq 0,\ i$=1,…,m,下，求解目標函數為$f_0(x)$的最小點，其中 x$\in R^n$，即:

$$min_x\ f_0(x) \tag{2.5}$$

$$\text{subject to } f_i(\text{x}) \leq 0, i = 1, \ldots, \text{m}$$

若f_0, f_i為$R^n \rightarrow$ R可微的凸函數，則此為一凸函數最佳化問題。設可行解(feasible solution)區域為D $= \{\text{x} \in R^n | f_i(\text{x}) \leq 0,\ \forall i\}$。令$\lambda_i$為相對於(2.5)中第$i$個不等限制式$f_i(\text{x}) \leq 0$之拉格朗日乘數，且$\lambda = (\lambda_1, \ldots, \lambda_m)^T$，則此最佳化問題之對偶問題為給定限制$\lambda_i \geq 0$下，求解目標函數$g(\lambda)$的最大點，其中g：$R^m \rightarrow$ R為拉格朗日對偶函數(Lagrange dual function)，其定義為

$$g(\lambda) = \inf_{\text{x} \in R^n}\{f_0(x) + \sum_{i=1}^m \lambda_i f_i(x)\}。$$

函數$g(\lambda)$亦為可微函數，定義其可行解區域為$D_\lambda = \{\lambda \in R^m | \lambda_i \geq 0,\ \forall i\}$。

原始問題(2.5)中的任一可行解$x' \in$ D，和對偶問題的任一可行解$\lambda' \in D_\lambda$將滿足不等式$\text{g}(\lambda') \leq f_0(x')$。因為

$$\text{g}(\lambda') = inf_{\text{x} \in R^n}\{f_0(x) + \sum_{i=1}^m \lambda_i' f_i(x)\} \leq f_0(x') + \sum_{i=1}^m \lambda_i' f_i(x') \leq f_0(x')。$$

此特性稱為弱對偶。

當原始問題為

$$min_x f_0(x)$$

$$\text{subject to} \quad f_i(x) \leq 0,\ i = 1, \ldots, \text{m} \tag{2.6}$$

$$h_j(x) = 0,\ j = 1, \dots, \mathrm{l}$$

其中f_0, f_i, h_j皆為可微凸函數，$\mathrm{x} \in R^n$，可行解區域$D' = \{\mathrm{x} \in R^n | f_i(x) \le 0, h_j(x) = 0, \forall \mathrm{i}, \forall \mathrm{j}\} \subset R^n$。令$\lambda_i, \mathrm{v}_j$分別為相對於(2.6)中第$i$個不等限制式$f_i(\mathrm{x}) \le 0$和第$j$個等式限制式$h_j(\mathrm{x}) = 0$之拉格朗日乘數，且$\lambda = (\lambda_1, \dots, \lambda_m)^T, v = (v_1, \dots, v_l)^T$，則在給定限制$\lambda_i \ge 0, v_j \in \mathrm{R}$下，它的對偶問題為

$$\max \quad \mathrm{g}(\lambda, \mathrm{v})$$

$$\text{subject} \quad \lambda_i \ge 0,\ v_j \in \mathrm{R}$$

其中拉格朗日對偶函數$\mathrm{g}(\lambda, \mathrm{v})$為

$$\mathrm{g}(\lambda, \mathrm{v}) = inf_{\mathrm{x} \in R^n} \{f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{l} \mathrm{v}_j h_j(x)\}。$$

g 函數的定義域為$D^* = \{\lambda \in R^m, v \in R^l | \lambda_i \ge 0, \forall i\} \subset R^m \times R^l$。

當原始問題的任一最佳解$x^* \in D'$和對偶問題的任一最佳解$(\lambda^*, v^*) \in D^*$，滿足$\mathrm{g}(\lambda^*, v^*) = f_0(x^*)$等式時，則稱此問題滿足強對偶性質。當強對偶成立時，由於$\lambda_i \ge 0, f_i(\mathrm{x}) \le 0$，則$\lambda_i^* f_i(x^*) = 0, \forall i$ 也必定成立。此定理之驗證參考自中國賈金柱教授的上課講義：

$$f_0(x^*) = \mathrm{g}(\lambda^*, \mathrm{v}^*) = inf_{\mathrm{x} \in R^n} \{f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{j=1}^{l} v_j^* h_j(x)\}$$

$$\le f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) \le f_0(x^*)$$

可得$\lambda_i^* f_i(x^*) = 0, \forall i$，此性質稱為互補差餘(complementary slackness)。

在最佳化問題(2.6)中，若函數$f_0(x), \dots, f_m(x), h_1(x), \dots, h_l(x)$為可微，定義以下 KKT 條件 (Karush-Kuhn-Tucker condition)：

1. 原始可行性(primal feasibility)：$x \in D'$
2. 對偶可行性(dual feasibility)：$(\lambda, \mathrm{v}) \in D^*$
3. 互補差餘(complementary slackness)：$\lambda_i \mathrm{f}_i(x) = 0, \forall i$
4. 平穩性(stationarity)：

$$\nabla_x f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla_x f_i(x) + \sum_{j=1}^{l} \mathrm{v}_j \nabla_x h_j(x) = 0$$

其中∇_x表示函數對點x作偏微分。若$(x^*, \lambda^*, \mathrm{v}^*)$滿足 KKT 條件，則$x^*$, $(\lambda^*, \mathrm{v}^*)$將分別是原始問題與對偶問題之最佳解，同時強對偶性質亦會成立，即$f_0(x^*) = \mathrm{g}(\lambda^*, v^*)$。

回顧第一節的資料與迴歸模型(2.1)，可以知道 LASSO 的原始問題(2.2)為凸函數最佳化問題，其目標函數為$f_0(\beta) = \sum_{i=1}^{n}(Y_i - X_i\beta)^2$，限制函數為$\mathrm{f}(\beta) = \sum_{j=1}^{p}|\beta_j| - \mathrm{t} \leq 0$。其中 t 是事先決定的正實數。則其對偶問題為$max_\beta g(\lambda)$，限制在$\lambda \geq 0$，其中g為拉格朗日對偶函數，$g(\lambda) = \inf_{\beta \in R^p}\{f_0(\beta) + \lambda \mathrm{f}(\beta)\}$；$\lambda$為相對於不等限制式$\mathrm{f}(\beta) \leq 0$之拉格朗日乘數，其可行解區域為$D_\lambda = \{\lambda \in \mathrm{R} | \lambda \geq 0\}$。

當凸函數最佳化中的函數為不可微時，可採取其次微分(sub-differential)，則 KKT 條件依然可以推導至強對偶性。所謂的次微分是當在凸集合的某一方向量測時，它呈現的是一非空封閉區間，此區間的上、下邊界值是由函數在該點右、左兩側之割線斜率的極限值所形成。例如 LASSO 中之限制函數是絕對值函數，在$\beta_j = 0$上不可微，所以我們對該點採取次微分，因此函數的微分與次微分可表示成：給定$j = 1, \ldots, p$,

$$\frac{\partial f(\beta)}{\partial \beta_j} = \begin{cases} 1 & if \quad \beta_j > 0 \\ -1 & if \quad \beta_j > 0 \end{cases}。$$

則迴歸係數的估計量必須滿足以下 KKT 條件：

1. $\beta \in D' = \{\beta \in R^p | \mathrm{f}(\beta) \leq 0\}$；
2. $\lambda \geq 0 \in D^*$；
3. $\lambda \mathrm{f}(\beta) = 0$；
4. $\nabla_{\beta_j} f_0(\beta) + \lambda \nabla_{\beta_j} f(\beta) = 0$。

2.1.4 標準差的估計方式

在上一節中，我們介紹了 LASSO 迴歸係數估計值相關的解根問題。當求解出迴歸係數的估計值後，接著便是該估計值的標準差的估計問題。Tibshirani (1996)提出一種有別於一般普通的拔靴法來估計迴歸係數的標準差，稱之為重抽殘差法。當迴歸模型為真、誤差項符合常態分配、且實驗矩陣是一個常數矩陣時，藉由重抽殘差的方式，將保留了原本解釋變數的特性。以下介紹拔靴殘差重抽法(bootstrap residual resampling)的步驟:

1. 給定一組包含 p 個解釋變數、n 個觀察值的樣本$(X_{i1},\dots,X_{ip},Y_i)$, i=1, ⋯, n。首先利用最小平方法將這組樣本配適模型並獲得預測值，$\hat{Y}_i=\sum_{j=0}^{p}X_{ij}\hat{\beta}_j^0$，其中$\hat{\beta}_j^0$為迴歸係數最小平方估計量，$X_{i0}=1$。則殘差為$e_i=Y_i-\hat{Y}_i, i=1,\dots,n$。
2. 從原始殘差$\{e_i, i=1,\dots,n\}$以抽出放回(sampling with replacement)的方式隨機抽取 n 筆觀測值，令抽出的殘差為$e_i^*, i=1,\dots,n$，則重抽反應變數$Y_i^*=\hat{Y}_i+e_i^*, i=1,\dots,n$。
3. 根據重抽樣本$(X_{i1},\dots,X_{ip},Y_i^*), i=1,\dots,n$，以特定的估計方式(本章主要探討 LSE 和 LASSO)，重新配適迴歸模型得到新的迴歸係數估計值$\hat{\beta}_j^{*(l)}$。
4. 重複以上步驟 2、3 共 B 次, 依序得到 B 個估計值$\hat{\beta}_j^{*(1)},\dots,\hat{\beta}_j^{*(B)}$。
5. 最後計算這 B 個估計值的標準差。

由 2.1.2 節估計量的比較可知，理論上 LASSO 的標準誤低於 LSE 的標準誤。藉著重抽殘差法，我們得以透過所獲得的標準誤估計值來驗證此結論。

2.1.5 t 值的決定

在估計 LASSO 迴歸係數與它的標準誤時，結果與參數 t 值有關。一般而言，t 值通常經過正規化，即計算所謂的相對界(relative bound):s=t/$\sum_{j=0}^{p}\hat{\beta}_j^0$，其

中 $\hat{\beta}_j^0$為最小平方估計量。故$0 \le s \le 1$。應用上我們可採用固定 s 值，或可在分析中再透過其他準則挑選 s 值。

考慮一般線性迴歸模型$Y = \beta^T X + \varepsilon$，假設誤差項的平均數是零，變異數為$\sigma^2$。令反應變數的預測值$\hat{Y} = \widehat{\beta^T} X$，則預測誤差(prediction error)為$E\{Y - \hat{Y}\}^2$。一般我們透過樣本資料來估計預測誤差，但為了避免過度配適(over-fitting)問題、造成過度樂觀的結論， Breidman and Spector(1992)提出了使用五折(5-fold)或十折(10-fold)的交叉驗證方法以估計預測誤差。在十折交叉驗證中，將觀察值分成 90%的訓練集(training set)和 10%檢測集(testing set)兩部分，在給定不同的 s 值下，利用訓練集配適模型、得出係數估計值，再根據檢測集來計算預測誤差。由 10 組的交叉驗證的結果取平均值當作參考。最終我們可以選取最低預測誤差相對應的 s 值，以及迴歸係數 LASSO 估計量。

另一種常用方法，是由使用者先決定選取解釋變數的個數，藉此尋求符合需求的 s 值範圍。

2.2 羅吉斯迴歸之 LASSO

生活中，二元型的羅吉斯迴歸廣泛應用在各個領域，如：信用風險之違約或未違約、金融危機的發生或不發生、生物醫學、辨識系統等等。

假設一組共有 n 筆資料的樣本(X_i, Y_i), i=1, …, n，其中Y_i是第i個二元型態(0, 1)的反應變數，$X_i = (X_{i1}, \ldots, X_{ip})$是第$i$個觀察值的解釋變數向量，假設有 p 個解釋變數，則X_{ij}為第i個觀察值之第j個解釋變數, i=1, …, n, j=1, …, p。定義給定解釋變數x_i，$Y_i = 1$的條件機率為 $p(x_i)$；則$Y_i = 0$的條件機率為$1 - p(x_i)$。因此Y_i服從於成功機率是 $p(x_i)$的百努力分配(Bernoulli distribution)。透過勝算比取自然對數(nature log odds)的轉換，把 p(x)轉換成 $\text{logit}(p(x)) = \ln(\frac{p(x)}{1-p(x)})$，則二元型羅吉斯迴歸(binary logistic regression)模型如下：

$$\text{logit}(\text{p}(x_i)) = \beta^T x_i^* \ , \ i = 1, \dots, \text{n} \qquad (2.7)$$

其中$\text{E}(Y_i) = \text{p}(x_i) = p_i$, $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$為迴歸係數, $x_i^* = (1, x_{i1}, \dots, x_{ip})^T$。由式子(2.7)可推導得到機率$\text{p}(x_i) = \frac{e^{-(\Sigma_{j=0}^{p} \beta_j x_{ij}^*)}}{1+e^{-(\Sigma_{j=0}^{p} \beta_j x_{ij}^*)}}$。

在估計羅吉斯迴歸中的迴歸係數時，一般採用最大概似函數(maximum likelihood function)法，找出使觀察值$(Y_1, \dots Y_n)$之可能性為最大的參數估計。假設樣本間獨立，且$Y_i \sim ber\big(p(x_i)\big), \text{i} = 1, \dots, \text{n}$，則概似函數為

$$\text{L}(\beta_0, \beta_1, \dots \beta_p) = \prod_{i=1}^{n} p_i{}^{y_i}(1-p_i)^{1-y_i}$$

上式取自然對數之後，則成為

$$l\big(\beta_0, \beta_1, \dots \beta_p\big) = \sum_{i=1}^{n} y_i \ln\left(\frac{p_i}{1-p_i}\right) + \sum_{i=1}^{n} \ln(1-p_i)$$

其中$p_i = \frac{e^{-(\Sigma_{j=0}^{p} \beta_j x_{ij}^*)}}{1+e^{-(\Sigma_{j=0}^{p} \beta_j x_{ij}^*)}}$，我們也能表示成

$$l\big(\beta_0, \beta_1, \dots, \beta_p\big) = \sum_{i=1}^{n} y_i \ln\left(e^{-(\Sigma_{j=0}^{p} \beta_j x_{ij}^*)}\right) + \sum_{i=1}^{n} \ln\left(\frac{1}{1+e^{-(\Sigma_{j=0}^{p} \beta_j x_{ij}^*)}}\right)$$

$$= -\sum_{i=1}^{n} y_i\big(\sum_{j=0}^{p} \beta_j x_{ij}^*\big) - \sum_{i=1}^{n} \ln\left(1 + e^{-\left(\sum_{j=0}^{p} \beta_j x_{ij}^*\right)}\right) 。$$

若是應用在凸函數最佳化問題中，則解迴歸係數$\beta_j, \ j = 0,1, \dots, \text{p}$之最大概似估計量為以下最大化問題：$max_{\beta \in R^{p+1}} L(\beta_0, \beta_1, \dots, \beta_p)$，其等價於$min_{\beta \in R^{p+1}}\{-L\big(\beta_0, \beta_1, \dots, \beta_p\big)\}$。若在最大概似估計法上加入對迴歸係數長度總和的限制式，則我們同樣可得 LASSO 迴歸係數估計量，即考慮下列問題：

$$min_{\beta \in R^{p+1}}\{-L\big(\beta_0, \beta_1, \dots, \beta_p\big)\}$$

$$\text{subject to } \sum_{j=1}^{p} |\beta_j| \leq \text{t}$$

其中 $\text{t} \in R^+$ 是迴歸係數長度總和的上界值。則其拉格朗日函數為

$$-\sum_{i=1}^{n} y_i \left(\sum_{j=0}^{p} \beta_j x_{ij}^*\right) - \sum_{i=1}^{n} \ln\left(1 + e^{-\left(\sum_{j=0}^{p} \beta_j x_{ij}^*\right)}\right) + \lambda\left(\sum_{j=1}^{p} |\beta_j| - \text{t}\right)$$

其中$\lambda \geq 0$為拉格朗日乘數。

羅吉斯迴歸模型標準差的估計

在第一節的一般線性迴歸裡，LASSO 迴歸係數估計量的標準差是用殘差拔靴的方式抽取，在羅吉斯迴歸中，Sartori(2010)提到藉由拔靴法，即可建立 LASSO 估計量的標準差，如以下步驟：

1. 假設一組共 n 筆資料的標準化隨機樣本(X_i, Y_i)，i=1, …, n，首先利用最大概似法將這組樣本配飾羅吉斯迴歸模型並得到迴歸係數估計值$\hat{\beta}$。若是具有迴歸係數總和限制式時，則透過交叉驗證的方式找出最適的懲罰(penalty)係數$\hat{\lambda}$，再根據$\hat{\lambda}$配飾羅吉斯迴歸。
2. 由這組樣本$\{(X_i, Y_i), i = 1, \dots, n\}$以抽出放回的方式抽取 n 筆觀察值，形成一組拔靴資料集$\{(X_i^*, Y_i^*), i = 1, \dots, n\}$。
3. 根據重抽樣本$(X_i^*, Y_i^*), i = 1, \dots, n$，重新配適迴歸模型得到新的迴歸係數估計值$\widehat{\beta^*}^{(l)}$(若是迴歸係數有限制式，則根據原本的$\hat{\lambda}$重新配飾羅吉斯迴歸)。
4. 重複以上步驟 2、3 共 B 次，依序得到 B 個估計值$\widehat{\beta^*}^{(1)}, \dots, \widehat{\beta^*}^{(B)}$。
5. 最後計算這 B 個估計值的標準差。

原則上以拔靴法估計出的 LASSO 標準誤會比 MLE 的標準誤小。另外，在估計迴歸係數的標準誤時，t 值會影響到標準誤的估計結果。令第 i 筆資料觀察值的預測機率為$\hat{p}_i$, $i = 1, \dots, n$，定義樣本的平均預測誤差為$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{p}_i)^2$。同 2.1.5 節敘述，為避免過度配適的情況，採用交叉驗證的方式決定預測誤差最小的，然後我們可以得到其相對應的 t 值。

於 R 之 glmnet 套件中，lambda.min 為參數值 t 選取選項，代表著選取預測誤差最小的當作標準。若是主觀認定 t 值太小，也可考慮在最小預測誤差上加一個本身標準差，即 lambda.lse 的選項。

3. 實例資料分析

本章將 LASSO 方法運用在三組實際資料集。我們運用羅吉斯迴歸模型配飾資料，並且將比較 LASSO 與最大概似估計法，比較在設限制與未設限制下的估計結果。

3.1 脊椎後凸的預測

第一個例子出自 Tibshirani(1996)的文章，是有關於脊椎後凸病人(Vertebrae kyphosis)的實證資料。此資料源自於約翰·霍普金斯大學布隆博格公共衛生學院，研究目的為研究病人之脊椎後凸現象與某些臨床測量值是否相關。資料中包括 81 位病人的三個解釋變數：X_1=年齡(age)、X_2=病人之脊椎從第幾節脊椎開始異常(number)、以及X_3=病歷中從第幾節脊椎開始手術(start)。變數皆為連續型態的資料。其反應變數為二元型，分顯性 (present≡1)和隱性(absent≡0)。

已知 LASSO 的結果與限制式$\sum_{j=1}^{3}|\beta_j| \le t$的上限 t 值相關，不同的 t 值對應的拉格朗日乘數λ也不同，且 t 與λ呈反向關係。圖 3.1 為各拉格朗日乘數λ與交互驗證的離差(deviance)的結果，其中橫軸為 log(λ)值，縱軸為離差。離差之定義為 $2(\log L_s - \log L_p)$，其中L_s是迴歸係數皆非零之概似函數，L_p則是在 p 個迴歸係數於非零之下的概似函數。圖形的上方橫軸的數字表示非零迴歸係數（不包含截距)的個數。圖中之信賴束上、下界則定義為離差之一個標準差的寬度。左邊垂直虛線表最小平均離差對應的 log(λ)；右邊垂直虛線則代表最小平均離差其對應的信賴束上界值所對應的 log(λ)值。由圖 3.1 可知，當 log(λ)愈大，t 值越低，則越多迴歸係數被設成零，且離差增加。另一方面，圖 3.2 為迴歸係數估計值之路徑圖，橫軸為 t 值，縱軸是迴歸係數的路徑。其中紅色線是X_2變數相

對應的迴歸係數估計值；黑線是X_1變數相對應的係數估計值；綠色線則是X_3變數迴歸係數估計值。當門檻值愈小，迴歸係數會逐一往零靠近，依序的變數是X_1、X_2、X_3。由全民人體力學保健教室網站中，提到多種原因會導致後凸，除了年紀以外，如幼年性基因位置異常等，皆是常見的因子。另外，當各項迴歸係數估計量超過其各自門檻時，係數絕對值會近似直線型增加。

此組資料利用最大概似法得到的羅吉斯迴歸模型如下:

$$\text{Logit}(p(x))=-1.8335+0.6351X_1+0.6649X_2-1.0086X_3。$$

另外，當運用 LASSO，且設 t=0.5457 時，則將獲得最小預測誤差，其配適的羅吉斯迴歸模型為:

$$\text{Logit}(p(x))= -1.8192+0.6154X_1+0.6507X_2-0.9961X_3。$$

比較以上結果可知，運用 LASSO 將使所有自變數的迴歸係數變小。此外在 LASSO 中，預測誤差的標準差，是根據不同的門檻值，藉由 10 次交叉驗證取平均得到的。

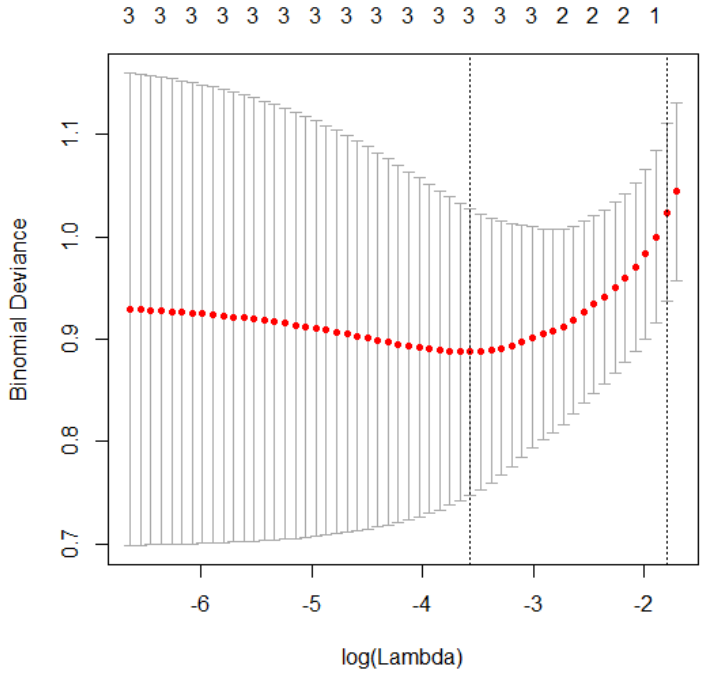


圖 3.1 給定不同門檻值 t 與離差的整體趨勢。其中橫軸為 log(λ)值，縱軸為離差。

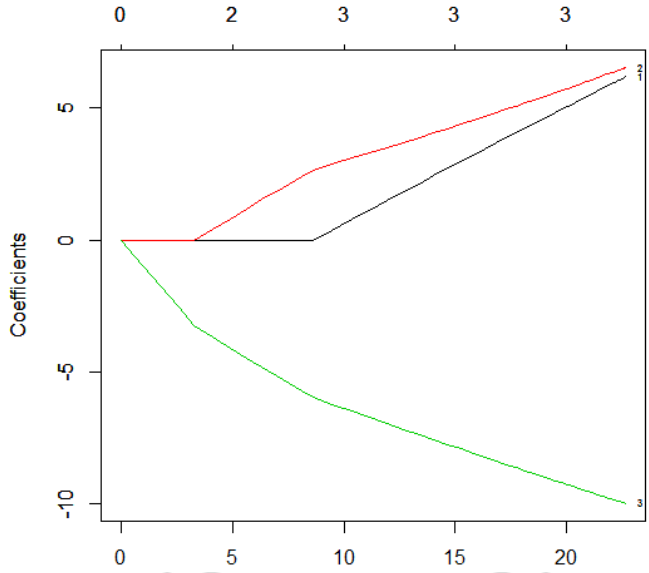


圖 3.2 門檻值與迴歸係數的路徑。其中橫軸表門檻值，縱軸是迴歸係數值。黑色線是X_1變數；紅色線是X_2變數；綠色線是X_3變數。

接著我們比較這些方法在預測上的準確性。首先我們依據全體資料所建立的迴歸模型，針對原 81 筆資料作預測。我們估計 Y=1 的機率，以 0.5 作為分界以預測真實反應變數，預測正確率見表 3.1。由表 3.1 中可以發現不論是 MLE 或 LASSO，它們的正確率皆一樣，約略 84%。接著為了避免過度配適的問題，於是我們將 81 筆資料隨機分成兩組，其中的 72 筆當作訓練資料，用以配適羅吉斯迴歸模型。其餘 9 筆則為檢測資料，我們根據所獲得的模型預測檢測資料的反應變數。我們總共作了 5 次的交叉驗證，得到預測檢測資料的平均正確率與正確率的標準差，見表 3.2。由於全部資料作訓練集時，兩種方法預測正確率並沒有達到很高，所以預測檢測資料時，最大概似法的正確率表現就很差，LASSO 預測得也不是很準。

表 3.1 當 100%的資料作訓練集時，MLE 和 LASSO 的分類正確率

真實		1 (17)	0 (64)	正確率
MLE	1 (10)	7	3	$\frac{68}{81}$
	0 (71)	10	61	
LASSO	1 (8)	6	2	$\frac{68}{81}$
	0 (73)	11	62	

表 3.2 檢測資料的平均預測正確率與正確率的標準差

正確率	MLE	LASSO
平均數	0.244	0.733
標準差	0.122	0.127

3.2 貓與狗影像的辨識

接下來我們考慮一筆貓與狗的影像資料。資料總共有貓、狗各兩百張影像。圖 3.3 為資料集中部分貓狗影像。每張影像為 64x64 個像素(pixel)，我們利用 Dalal 和 Triggsy(2005)提出方向梯度直方圖(Histogram of oriented gradients，簡稱 HOG)，由中萃取出 1,764 個解釋變數，此些解釋變數主要反映影像上各區塊位置上 9 個方向梯度的特徵量值，經過正規化後，這些特徵量值為介於 0 至 1 的連續型資料，令這些解釋變數為$X_1,\ldots,X_{1764}$。另一方面，此資料的反應變數為該影像中的動物為貓(Y=0)或狗(Y=1)。

圖 3.3 資料集中部分貓狗影像。

考慮羅吉斯迴歸模型:$\text{Logit}(p(x)) = \beta_0 + \sum_{j=1}^{1764} \beta_j X_j$，限制式$\sum_{j=1}^{1764} |\beta_j| \leq t$，其中$p(x) = P(Y = 1|X_1, \ldots, X_{1764})$，$t \epsilon R^+$。圖 3.4 為給定不同拉格朗日乘數log(λ)下，相對應配適模型之均方差(Mean Square Error)，上方橫軸表示非零迴歸係數之個數(不含截距)，已知λ與 t 呈反向的關係。左邊虛線為對應最小預測誤差的log(λ)值。在交互驗證過程中，當λ = 0.01149187，t=116.81，則均方差達到最小。此時將有 124 個變數之迴歸係數非零，見圖 3.5 左邊虛線上方之橫軸。圖 3.5 是門檻值與各項迴歸係數估計值的路徑圖。其中橫軸為門檻值 t，縱軸為 t 對應的迴歸係數估計值。當各項估計量超過各自的門檻時，係數絕對值未必呈線性遞增。

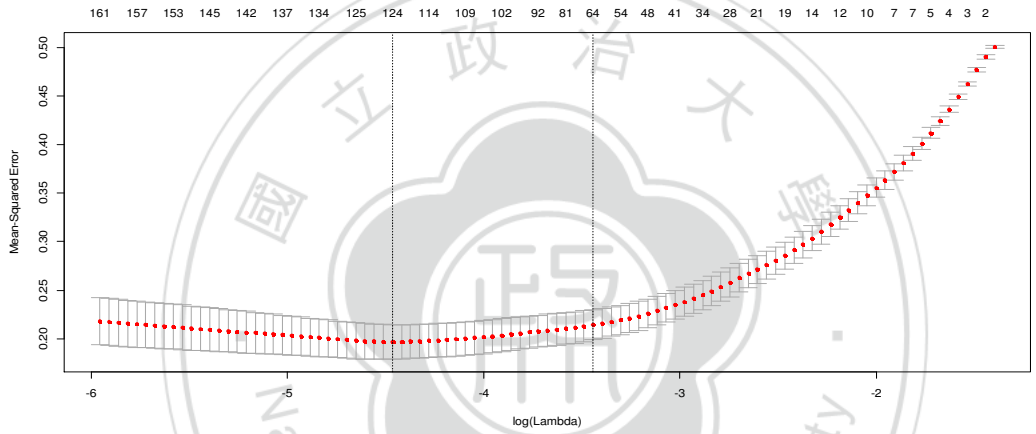


圖 3.4 給定不同門檻值 t 與預測誤差的整體趨勢。其中橫軸為 log(λ)值，縱軸為預測誤差。

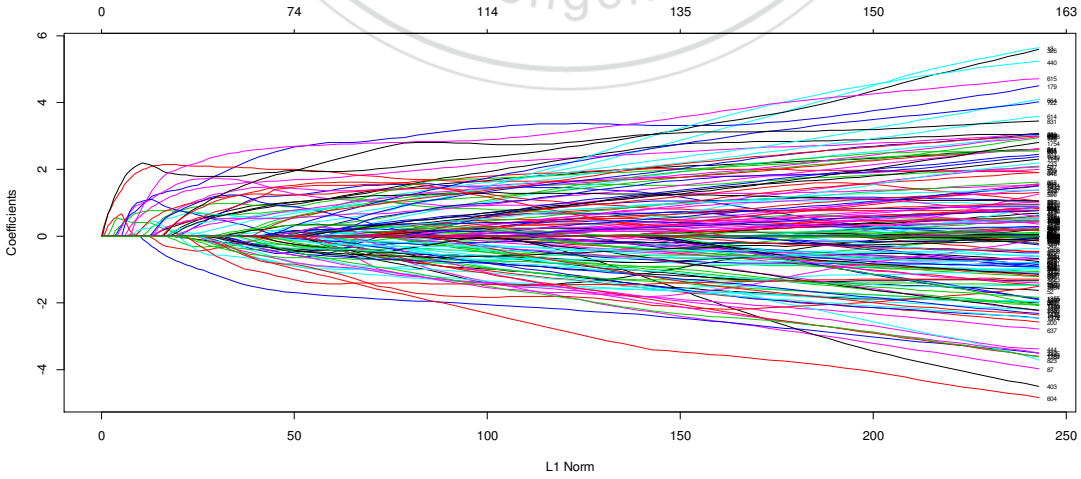


圖 3.5 門檻值與各項迴歸係數的路徑。其中橫軸表門檻值，縱軸是迴歸係數值。

接下來我們評估所建立的迴歸模型之預測能力。我們同樣以 0.5 作為切點，當$p(x) \geq 0.5$ 時，則預測該影像屬於狗影像，反之則為貓影像。若運用全部資料作為訓練集時，則對原資料之預測準確度達 100%。為了避免過度配飾，我們採用 10 次交叉驗證方式，每次取各 180 張貓、狗影像作為訓練資料，剩下各 20 張影像當作檢測資料集進行預測。表 3.3 為預測分類的平均正確率與正確率的標準差，其中又分成對訓練集資料的預測和對檢測集資料的預測結果。由表 3.3 可以發現利用訓練資料所建立的預測模型可正確預測 82%的測試資料，表現不錯。

表 3.3 分類的平均正確率與正確率的標準差(訓練集、檢測集)

正確率	訓練集	檢測集
平均	0.995	0.820
標準差	0.007	0.054

3.3 鐵達尼號倖存者的預測

在第一節脊椎後凸的例子中解釋變數皆為連續性，而上一節影像的辨識針對解釋變量超過樣本數時的特殊情況。最後本節將考慮分析鐵達尼號的資料。資料來自生物統計學院網站，樣本為 1,045 人，反應變數為倖存者(Y=1)或已故者(Y=0)。我們考慮以下七個解釋變數:艙等X_1(經濟艙 $\equiv 1$、商業倉 $\equiv 2$、頭等艙 $\equiv 3$)、性別X_2(男 $\equiv 1$、女 $\equiv 0$)、年齡X_3、朋友或配偶數X_4、家人數X_5、票價X_6、艙等與性別之交互作用項X_7，其中X_1、X_7為次序的(ordinal)；X_2為二元型態，其餘變數皆為連續型的資料型態，接著我們將利用 MLE 及 LASSO 來建立以及估計羅吉斯迴歸模型。

圖 3.6 為拉格朗日乘數與預測誤差圖。當$\lambda = 0.0002969288$，我們得到最小的預測誤差，如圖 3.6 左邊虛線，右邊虛線則代表最小預測誤差其對應的信賴束上界值相對應的 log(λ)值，此時λ約為 0.007705。圖 3.7 為 log(λ)與迴歸係

數估計值路徑圖。其中橫軸表門檻值，縱軸是迴歸係數值。利用最大概似法估計所得之羅吉斯迴歸模型如下:

$$\text{Logit}(p(x))=7.5907-2.1934X_1-6.0179X_2-0.0417248X_3-0.3601X_4$$

$$+0.1164X_5+0.0008746X_6+1.4665X_7。$$

另一方面，當利用 LASSO 方法時，t=9.8311 可得到最小預測誤差，其配適的羅吉斯迴歸模型為:

$$\text{Logit}(p(x))=7.3926-2.1252X_1-5.8079X_2-0.0411X_3-0.3544X_4$$

$$+0.1113X_5+0.0008X_6+1.3904X_7。$$

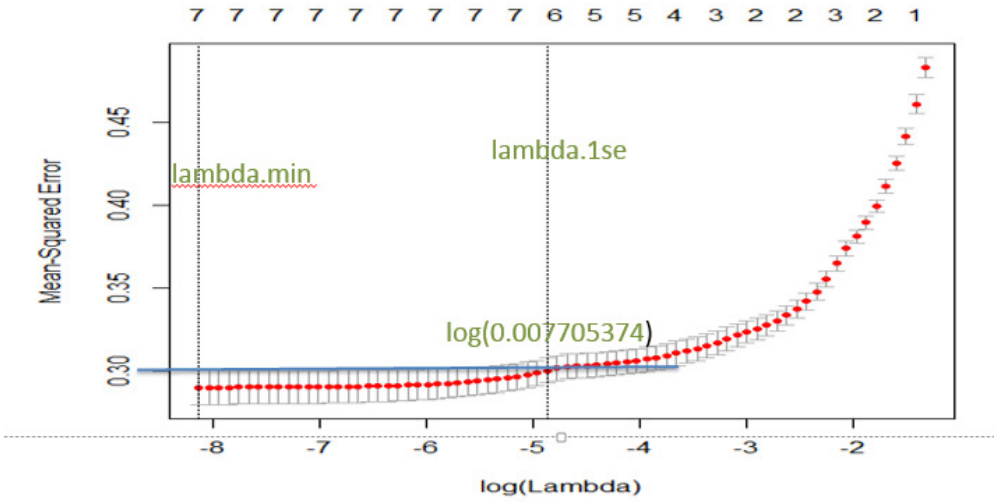


圖 3.6 給定不同門檻值 t 與預測誤差的整體趨勢。其中橫軸為 log(λ)值，縱軸為預測誤差。

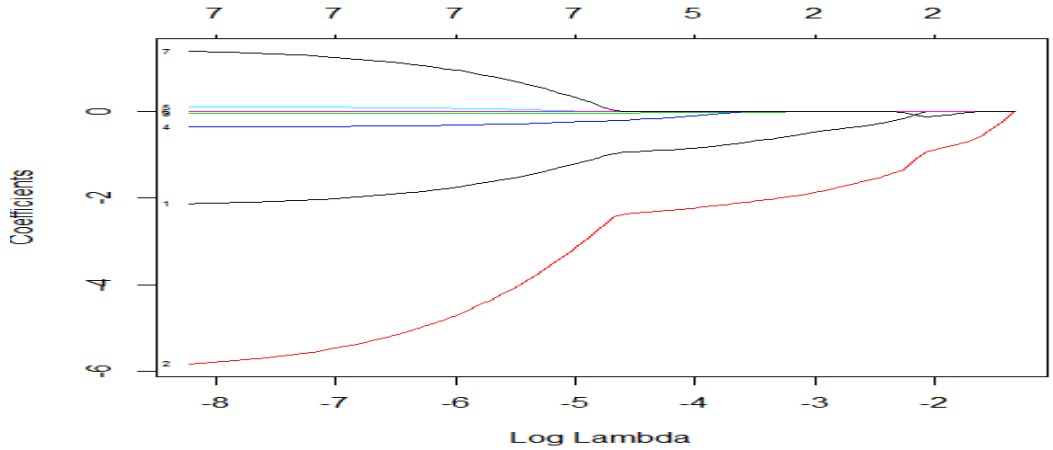


圖 3.7 門檻值與各項迴歸係數的路徑。其中橫軸表門檻值，縱軸是迴歸係數值。

最後我們將建立的迴歸模型，用來預測這1045個觀察值分類到的結果。結果顯示當全部資料都作訓練集時，最大概似估計法的分類正確率達80.77%，LASSO是80.86%，兩者差不多。接著再作10次交互驗證，我們將觀察值依三種不同艙等劃分,在各組內取90%合併為訓練集資料,預測剩下的10%測試集資料，最後得到分類正確率，如表3.4。我們觀察得知，不論使用MLE或LASSO，這組資料之預測結果差別不大。原因是因為此時在採用的門檻值下，並無解釋變數之迴歸係數縮減至零，故兩個模型差異不大。

表3.4 比較MLE和LASSO測試資料的分類正確率

正確率	MLE	LASSO
平均	0.800	0.801
標準差	0.020	0.021

4. 模擬實驗

在實務應用中母體真實迴歸方程式為未知，所以無法獲知估計結果的好壞。本章藉由電腦輔助模擬，以比較迴歸係數估計值與真實參數值之間的差異，評估估計方法的表現。本章模擬實驗考慮不同設定的模型，比較最大概似法和 LASSO 方法在羅吉斯迴歸上估計量的表現。我們將探討當解釋變數間相關係數的改變或樣本數大小，在估計結果上出現的規律或趨勢。接著由這些評估標準，比較兩種方法間在估計量上的差異，並將結果以表格和圖來呈現，詳細的流程設計請見 4.1 節。

4.1 模擬流程與參數設計

首先我們先介紹模擬流程如下:

(i) 假設有 p=5 個解釋變數，給定參數真值$\beta_0,\beta_1,\dots,\beta_5$，則羅吉斯迴歸模型為$\mathrm{Logit}\big(\mathrm{p(x)}\big)=\beta_0+\beta_1X_1+\cdots+\beta_5X_5$。

(ii)給定樣本數為 n，由電腦生成來自多變量常態分佈之自變數：

$$x_i=(x_{i1},\dots,x_{i5})\sim N_5(\mathbf{0},\Sigma),i=1,\dots,n。$$

(iii) 生成隨機反應變數$Y_i\sim \mathrm{ber}(\mathrm{p}(x_i))$，其中$\mathrm{p}(x_i)=$

$$\mathrm{p(Y=1}|x_{i1},\dots,x_{i5})=\frac{e^{(\beta_0+\sum_{j=1}^{5}\beta_jx_{ij})}}{1+e^{(\beta_0+\sum_{j=1}^{5}\beta_jx_{ij})}},\ i=1,\dots,n。$$

(iv)利用(ii)與(iii)中得到的樣本$\{(x_i,\ Y_i),\ i=1,\dots,n\}$，估計迴歸係數$(\beta_0,\beta_1,\dots,\beta_5)$，令估計值為$\widehat{\beta_0},\widehat{\beta_1},\dots,\widehat{\beta_5}$。

(v) 重複模擬 1000 次，估計$\widehat{\beta_0},\widehat{\beta_1},\dots,\widehat{\beta_5}$的偏差(bias)，標準差(standard error)，與均方差(mean squared error)。

在參數設定上，我們考慮三種迴歸係數真值$\beta_0,\beta_1 \ldots \beta_5$的情境，包括：全部皆非零(1,1,1,1,1,1)、5 個非零(1,1,1,1,1,0)、3 個非零(1,1,1,0,0,0)。自變數服從多變量常態分配$X \sim N_5(\mathbf{0},\Sigma)$，各解釋變數的平均數皆為 0，變異數為 1，令解釋變數彼此間的相關係數皆為ρ，ρ值由低至高為 0、0.2、0.5、0.9。考慮三個樣本數：n=100、200、500。在 LASSO 中門檻值 t 則是取其最低預測誤差對應之值，故每次模擬試驗的 t 皆不一樣。

4.2 估計量的比較

以下模擬主要是針對在不同情境底下，藉由 1000 次模擬，比較 MLE 與 LASSO 估計量與真實值之間的差異。在第二章中，我們已得知在線性迴歸中，理論上 LASSO 的變異數將不超過最小平方法的變異數。另一方面，最小平方估計量滿足不偏性，但 LASSO 估計量則通常低估真實參數。本節羅吉斯迴歸模型的模擬實驗中，我們除了以估計量的偏差、標準差作為評估標準，另外亦將均方差(mean square error, MSE)納入本節的評估標準。以下將分段討論在不同的迴歸模型參數下，探討相關係數與樣本數的變化對估計法之偏差、標準差及均方差產生的影響，最後再比較兩種方法在這三個標準下的差異。

壹、所有迴係數皆非零時

表 4.1、4.2、4.3 分別為當真實參數是(1,1,1,1,1,1)時，估計量的偏差、標準差、均方差。三個表格的設計，由左至右側縱列方向依序為相關係數、樣本數、估計方法、估計量。相關係數分為不相關、低度、中度及高度相關；樣本數由小至大為 100、200、500；估計法分成 MLE 和 LASSO 兩種；估計量共有六個，包括截距項與斜率項。

由表 4.1 可以得到，隨著樣本數的增加，所有估計量的偏差皆逐漸變小。當

相關係數上升，所有估計量的偏差並非隨之增加。MLE 估計量的偏差均為正值，故通常高估真實參數值。LASSO 則相反，均低估參數值。另外仔細觀察斜率的部分可得知，當樣本數 100 時，MLE 估計量的偏差絕對值全比 LASSO 大，但當樣本數為 200 以上時，MLE 估計量的偏差則未必高於 LASSO。至於截距項的估計量上，MLE 皆有較高的偏差。

由表 4.2 可以得到，隨著樣本數的增加，所有估計量的標準差皆逐漸變小。當相關係數上升，所有估計量的標準差隨之上升。LASSO 估計量的標準差均比 MLE 小，且隨相關係數越大，兩者標準差的差距也變得更大。

由表 4.3 可以得到，隨著樣本數的增加，所有估計量的均方差變得越來越小。當相關係數上升，估計量的均方差逐漸增加，且增加的幅度越來越大。不論相關係數與樣本數的大小，LASSO 的均方差皆比 MLE 的均方差來得小。

表 4.1　真實參數是(1,1,1,1,1,1)，估計量的偏差

相關係數	樣本數	方法	$\widehat{\beta_0}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_4}$	$\widehat{\beta_5}$
$\rho = 0$	n=100	MLE	0.1552	0.1448	0.1377	0.1371	0.1405	0.1392
		LASSO	-0.0091	-0.0981	-0.1042	-0.1042	-0.1010	-0.1024
	n=200	MLE	0.0483	0.0508	0.0575	0.0615	0.0516	0.0562
		LASSO	-0.0280	-0.0665	-0.0602	-0.0565	-0.0656	-0.0618
	n=500	MLE	0.0276	0.0257	0.0222	0.0257	0.0258	0.0280
		LASSO	-0.0031	-0.0214	-0.0247	-0.0211	-0.0212	-0.0191
$\rho = 0.2$	n=100	MLE	0.1507	0.1696	0.1599	0.1645	0.1615	0.1443
		LASSO	-0.0699	-0.1179	-0.1242	-0.1205	-0.1221	-0.1374
	n=200	MLE	0.0672	0.0687	0.0811	0.0692	0.0651	0.0677
		LASSO	-0.0373	-0.0683	-0.0569	-0.0674	-0.0709	-0.0688
	n=500	MLE	0.0333	0.0355	0.0231	0.0346	0.0282	0.0374
		LASSO	-0.0081	-0.0187	-0.0307	-0.0195	-0.0258	-0.0169
$\rho = 0.5$	n=100	MLE	0.1958	0.2072	0.2054	0.2367	0.1896	0.1821
		LASSO	-0.0927	-0.1373	-0.1333	-0.1109	-0.1437	-0.1535
	n=200	MLE	0.0749	0.0848	0.0903	0.0743	0.0781	0.0726
		LASSO	-0.0590	-0.0767	-0.0703	-0.0851	-0.0819	-0.0865
	n=500	MLE	0.0362	0.0472	0.0374	0.0361	0.0351	0.0396
		LASSO	-0.0178	-0.0174	-0.0269	-0.0280	-0.0289	-0.0248
$\rho = 0.9$	n=100	MLE	0.5790	0.5150	0.5700	0.4380	0.7830	1.1280
		LASSO	-0.0870	-0.1367	-0.0861	-0.0647	-0.1346	-0.1028
	n=200	MLE	0.0890	0.0940	0.0720	0.0900	0.1170	0.1000
		LASSO	-0.0693	-0.0804	-0.1012	-0.0868	-0.0708	-0.0783
	n=500	MLE	0.0420	0.0360	0.0600	0.0310	0.0510	0.0360
		LASSO	-0.0247	-0.0380	-0.0174	-0.0419	-0.0253	-0.0376

表 4.2　真實參數是(1,1,1,1,1,1)，估計量的標準差

相關係數	樣本數	方法	$\widehat{\beta_0}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_4}$	$\widehat{\beta_5}$
$\rho = 0$	n=100	MLE	0.3990	0.4112	0.3974	0.4243	0.4149	0.4272
		LASSO	0.3176	0.3257	0.3155	0.3345	0.3250	0.3391
	n=200	MLE	0.2354	0.2536	0.2467	0.2524	0.2505	0.2532
		LASSO	0.2140	0.2315	0.2241	0.2276	0.2255	0.2281
	n=500	MLE	0.1433	0.1546	0.1461	0.1505	0.1482	0.1472
		LASSO	0.1369	0.1473	0.1397	0.1452	0.1416	0.1410
$\rho = 0.2$	n=100	MLE	0.4680	0.5144	0.4876	0.4839	0.4667	0.4916
		LASSO	0.3449	0.3729	0.3580	0.3582	0.3442	0.3611
	n=200	MLE	0.2743	0.2974	0.3038	0.3010	0.2927	0.2963
		LASSO	0.2398	0.2610	0.2659	0.2624	0.2559	0.2588
	n=500	MLE	0.1573	0.1750	0.1690	0.1691	0.1708	0.1680
		LASSO	0.1487	0.1658	0.1601	0.1607	0.1621	0.1602
$\rho = 0.5$	n=100	MLE	0.5480	0.6896	0.6700	0.6754	0.6529	0.6863
		LASSO	0.3711	0.4512	0.4594	0.4513	0.4442	0.4648
	n=200	MLE	0.3006	0.3919	0.3870	0.3812	0.3987	0.3813
		LASSO	0.2533	0.3314	0.3277	0.3228	0.3366	0.3242
	n=500	MLE	0.1738	0.2324	0.2205	0.2219	0.2290	0.2254
		LASSO	0.1614	0.2184	0.2069	0.2082	0.2157	0.2117
$\rho = 0.9$	n=100	MLE	9.4410	8.8780	7.7300	2.1360	16.806	26.374
		LASSO	0.4152	0.8444	0.8559	0.8950	0.8424	0.9410
	n=200	MLE	0.3399	0.9019	0.8313	0.8230	0.8553	0.8659
		LASSO	0.2794	0.6903	0.6322	0.6242	0.6613	0.6587
	n=500	MLE	0.1939	0.5193	0.4816	0.5062	0.5147	0.4914
		LASSO	0.1775	0.4751	0.4439	0.4610	0.4739	0.4513

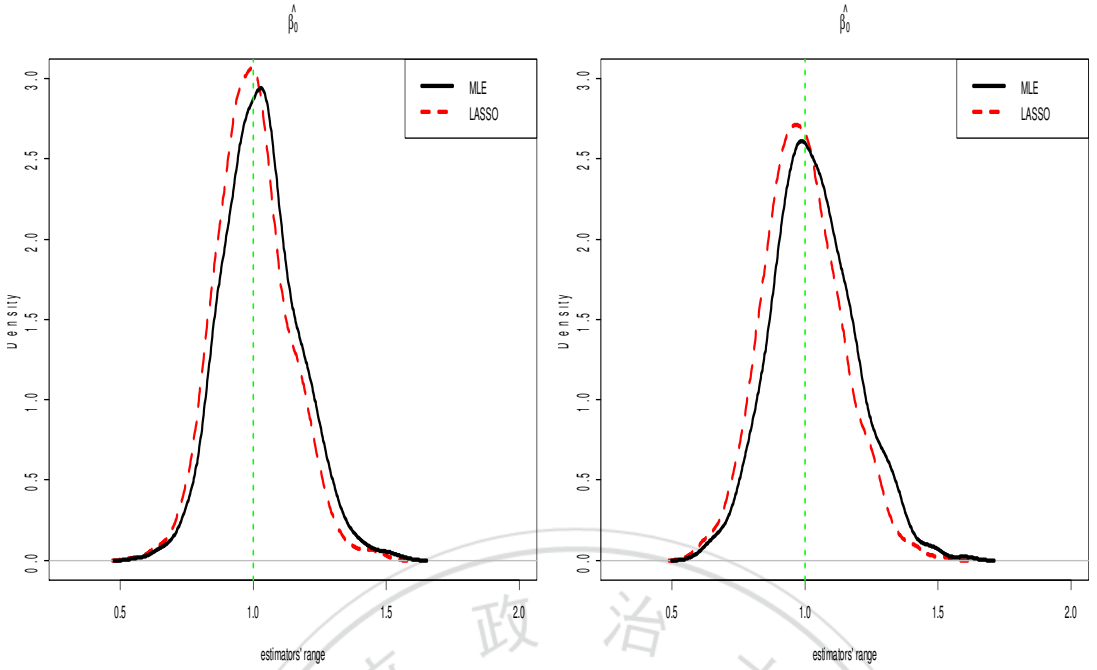
表 4.3　真實參數是(1,1,1,1,1,1)，估計量的均方差

相關係數	樣本數	方法	$\widehat{\beta_0}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_4}$	$\widehat{\beta_5}$
$\rho = 0$	n=100	MLE	0.1833	0.1901	0.1769	0.1989	0.1919	0.2019
		LASSO	0.1010	0.1157	0.1104	0.1227	0.1158	0.1255
	n=200	MLE	0.0578	0.0669	0.0642	0.0675	0.0654	0.0673
		LASSO	0.0466	0.0580	0.0539	0.0550	0.0552	0.0559
	n=500	MLE	0.0213	0.0246	0.0218	0.0233	0.0226	0.0225
		LASSO	0.0188	0.0221	0.0201	0.0215	0.0205	0.0203
$\rho = 0.2$	n=100	MLE	0.2418	0.2934	0.2633	0.2613	0.2439	0.2625
		LASSO	0.1238	0.1530	0.1436	0.1428	0.1334	0.1492
	n=200	MLE	0.0798	0.0932	0.0989	0.0954	0.0899	0.0924
		LASSO	0.0589	0.0728	0.0739	0.0734	0.0705	0.0717
	n=500	MLE	0.0259	0.0319	0.0291	0.0298	0.0300	0.0296
		LASSO	0.0222	0.0278	0.0266	0.0262	0.0269	0.0259
$\rho = 0.5$	n=100	MLE	0.3387	0.5185	0.4910	0.5122	0.4622	0.5042
		LASSO	0.1463	0.2224	0.2288	0.2160	0.2180	0.2396
	n=200	MLE	0.0959	0.1607	0.1579	0.1509	0.1650	0.1506
		LASSO	0.0676	0.1157	0.1123	0.1114	0.1200	0.1126
	n=500	MLE	0.0315	0.0563	0.0500	0.0505	0.0537	0.0524
		LASSO	0.0264	0.0480	0.0436	0.0441	0.0474	0.0454
$\rho = 0.9$	n=100	MLE	89.472	79.082	60.073	4.7530	283.06	696.89
		LASSO	0.1800	0.7317	0.7400	0.8052	0.7277	0.8960
	n=200	MLE	0.1234	0.8223	0.6961	0.6854	0.7451	0.7598
		LASSO	0.0829	0.4830	0.4100	0.3972	0.4423	0.4400
	n=500	MLE	0.0393	0.2710	0.2355	0.2572	0.2676	0.2428
		LASSO	0.0321	0.2271	0.1974	0.2143	0.2253	0.2051

除了比較估計量在各種標準下的差異外，我們也觀察截距項$\widehat{\beta_0}$與非零項迴歸係數估計量$\widehat{\beta_3}$的分佈情況。圖 4.1、4.2 為當樣本數固定為 500 時，β_0,β_3的 MLE 與 LASSO 的抽樣分配圖。

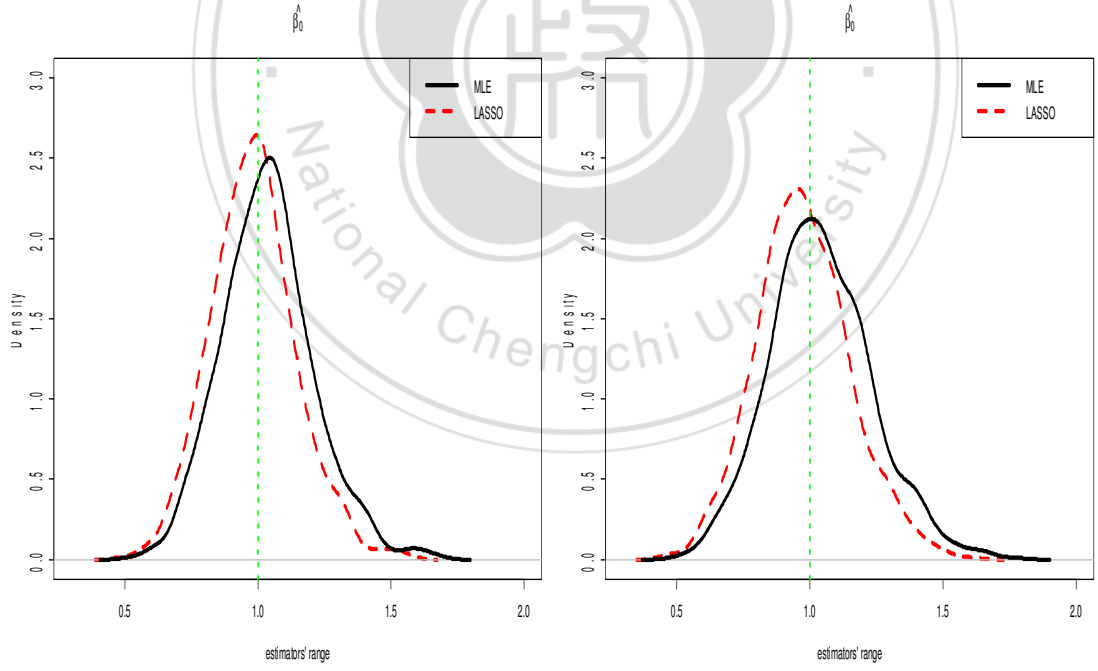
由圖 4.1 可知，兩個方法的截距項估計量抽樣分佈圖形大致滿足對稱性，且不受相關係數大小的影響。兩估計量與參數真值的偏差很小，唯當高度相關時，LASSO 低估現象稍加明顯。由於 LASSO 的機率函數圖較 MLE 集中一點，反映了其變異程度略小一點。此外，當解釋變數間的相關係數愈大，兩估計量的集中程度下降，反映了其變異程度隨之增加，而且 MLE 估計量的全域(range)還逐漸上升，其變異程度若與 LASSO 相比，差距擴大至更明顯。

由圖 4.2 可知，兩方法之非零項估計量抽樣分佈並不對稱，呈右偏，尤其 MLE 比 LASSO 明顯。當解釋變數間的相關係數愈大時，右偏情況更加顯著。與參數真值的偏差，LASSO 傾向低估，但是偏度並不受相關係數的影響，而 MLE 則不偏。LASSO 的機率函數圖較 MLE 集中，反映了其變異程度較小。當解釋變數間的相關係數愈大，兩估計量的變異程度將隨之增加。MLE 估計量的變異程度若與 LASSO 變異程度之間的差距將更明顯。



(a)

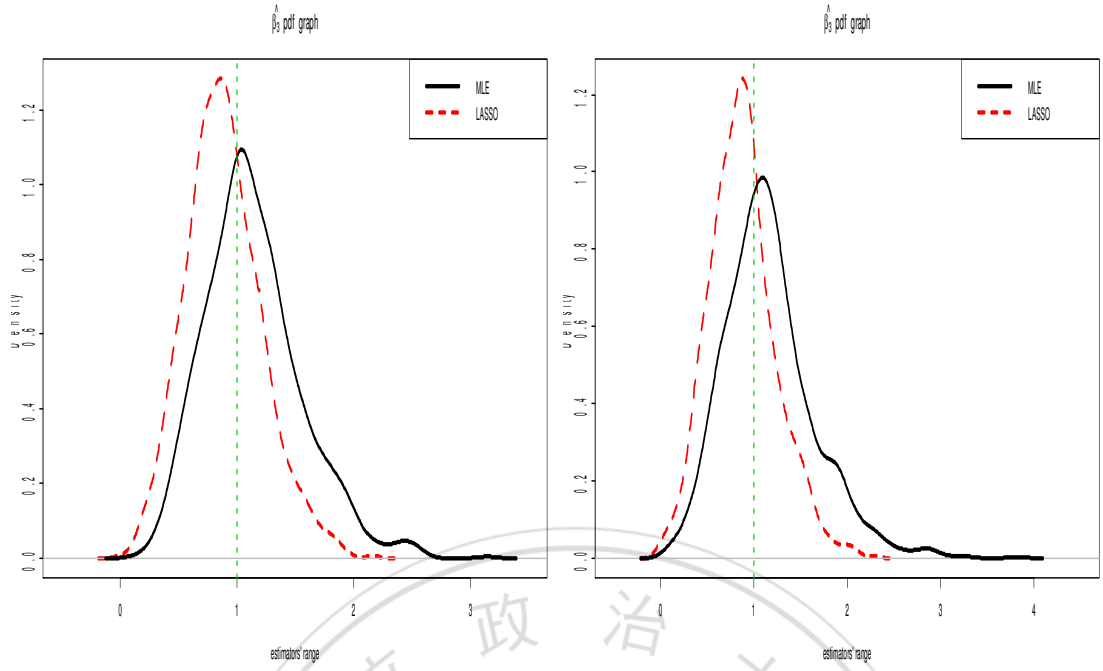
(b)



(c)

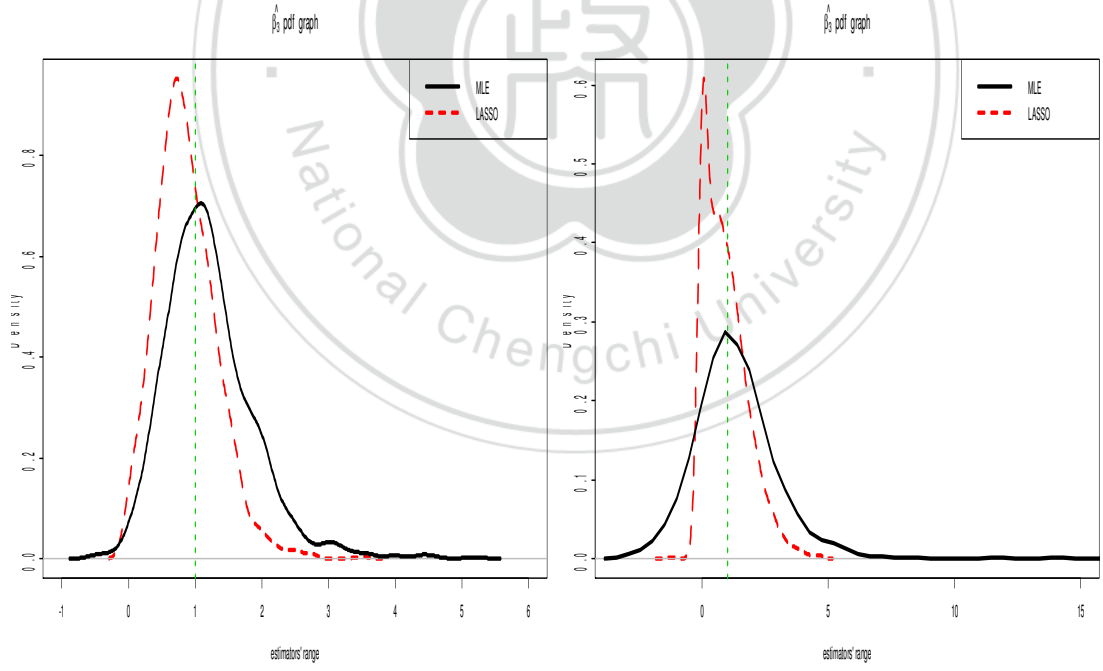
(d)

圖 4.1　給定樣本數為 500 時，真實迴歸係數是$(1,1,1,1,1,1)$，$\widehat{\beta_0}$的抽樣分佈
（a）ρ=0　（b）ρ=0.2　（c）ρ=0.5　（d）ρ=0.9



(a)

(b)



(c)

(d)

圖 4.2　給定樣本數為 500 時，真實迴歸係數是$(1,1,1,1,1,1)$，$\widehat{\beta_3}$的抽樣分佈
（a）ρ=0　（b）ρ=0.2　（c）ρ=0.5　（d）ρ=0.9

貳、五個迴歸係數非零時：

表 4.4、4.5、4.6 分別為當真實參數是(1,1,1,1,1,0)時，估計量的偏差、標準差、均方差。三個表格的設計，由左至右側縱列方向依序為相關係數、樣本數、估計方法、估計量。相關係數分為不相關、低度、中度及高度相關；樣本數由小至大為 100、200、500；估計法分成 MLE 和 LASSO 兩種；估計量共有六個，包括截距項與斜率項。

由表 4.4 可以得到，隨著樣本數的增加，所有估計量的偏差變得愈來愈小。當解釋變數間不相關時，兩方法的零斜率估計量$\widehat{\beta_5}$，偏差並非隨樣本增加而變小。當相關係數上升，所有估計量的偏差並非隨之增加。MLE 在非零係數估計量中，偏差均為正值，故通常高估真實參數值。LASSO 則相反，均低估參數值。另外仔細觀察斜率的部分可得知，當樣本數 100 時，MLE 估計量的偏差絕對值超過一半比LASSO大，但當樣本數為200以上時，MLE估計量的偏差則不到一半高於LASSO。至於截距項的估計量上，MLE 皆有較高的偏差。當變數間相關程度為零時，MLE 零斜率估計量的偏差均比 LASSO 大，當相關程度為非零時，LASSO 零斜率估計量的偏差則較大。

由表 4.5 可以得到，隨著樣本數的增加，所有估計量的標準差皆逐漸變小。當相關係數上升，所有估計量的標準差隨之上升。LASSO 估計量的標準差均比 MLE 小，且隨相關係數越大，兩者標準差的差距也變得更大。LASSO 零斜率估計量的標準差，明顯比非零項估計量小很多。

由表4.6可以得到，隨著樣本數的增加，所有估計量的均方差變得越來越小。當相關係數上升，估計量的均方差逐漸增加，且增加的幅度越來越大。不論相關係數與樣本數的大小，LASSO 的均方差幾乎比 MLE 的均方差來得小。

表 4.4　真實參數是(1,1,1,1,1,0)，估計量的偏差

相關係數	樣本數	方法	$\widehat{\beta_0}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_4}$	$\widehat{\beta_5}$
ρ = 0	n=100	MLE	0.1142	0.1353	0.1200	0.1412	0.1201	0.0137
		LASSO	-0.0441	-0.1236	-0.1350	-0.1171	-0.1358	0.0078
	n=200	MLE	0.0526	0.0542	0.0590	0.0561	0.0515	0.0007
		LASSO	-0.0296	-0.0805	-0.0757	-0.0783	-0.0830	0.0005
	n=500	MLE	0.0296	0.0257	0.0273	0.0339	0.0170	-0.0088
		LASSO	-0.0078	-0.0360	-0.0344	-0.0279	-0.0443	-0.0072
ρ = 0.2	n=100	MLE	0.1272	0.1412	0.1319	0.1636	0.1347	-0.0097
		LASSO	-0.0699	-0.1406	-0.1463	-0.1210	-0.1446	0.0333
	n=200	MLE	0.0616	0.0661	0.0730	0.0777	0.0623	-0.0062
		LASSO	-0.0412	-0.0825	-0.0762	-0.0713	-0.0853	0.0202
	n=500	MLE	0.0350	0.0255	0.0277	0.0316	0.0302	-0.0048
		LASSO	-0.0105	-0.0400	-0.0380	-0.0340	-0.0355	0.0090
ρ = 0.5	n=100	MLE	0.1547	0.1621	0.1430	0.1914	0.1949	-0.0169
		LASSO	-0.0832	-0.1559	-0.1695	-0.1334	-0.1304	0.0792
	n=200	MLE	0.0655	0.0734	0.0774	0.0778	0.0601	-0.0134
		LASSO	-0.0526	-0.0891	-0.0862	-0.0845	-0.1009	0.0500
	n=500	MLE	0.0320	0.0275	0.0207	0.0331	0.0344	0.0064
		LASSO	-0.0202	-0.0442	-0.0512	-0.0391	-0.0382	0.0348
ρ = 0.9	n=100	MLE	0.1941	0.1277	0.2099	0.2900	0.1939	-0.0074
		LASSO	-0.0698	-0.2171	-0.1538	-0.1259	-0.1605	0.2522
	n=200	MLE	0.0794	0.0541	0.1176	0.0816	0.0683	0.0012
		LASSO	-0.0468	-0.1337	-0.0805	-0.1183	-0.1189	0.1861
	n=500	MLE	0.0345	0.0478	0.0245	0.0230	0.0357	0.0237
		LASSO	-0.0179	-0.0420	-0.0619	-0.0621	-0.0516	0.1274

表 4.5　真實參數是(1,1,1,1,1,0)，估計量的標準差

相關係數	樣本數	方法	$\widehat{\beta_0}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_4}$	$\widehat{\beta_5}$
$\rho = 0$	n=100	MLE	0.3687	0.3804	0.3951	0.3881	0.3866	0.3252
		LASSO	0.2999	0.3110	0.3205	0.3139	0.3148	0.2205
	n=200	MLE	0.2253	0.2504	0.2439	0.2427	0.2530	0.2037
		LASSO	0.2030	0.2270	0.2208	0.2188	0.2268	0.1583
	n=500	MLE	0.1422	0.1454	0.1420	0.1455	0.1448	0.1223
		LASSO	0.1358	0.1394	0.1382	0.1405	0.1399	0.1055
$\rho = 0.2$	n=100	MLE	0.4026	0.4252	0.4529	0.4382	0.4432	0.3626
		LASSO	0.3114	0.3256	0.3482	0.3391	0.3447	0.2208
	n=200	MLE	0.2481	0.2917	0.2805	0.2758	0.2848	0.2352
		LASSO	0.2182	0.2584	0.2471	0.2423	0.2506	0.1724
	n=500	MLE	0.1548	0.1576	0.1587	0.1621	0.1641	0.1396
		LASSO	0.1459	0.1514	0.1521	0.1573	0.1579	0.1163
$\rho = 0.5$	n=100	MLE	0.4721	0.5863	0.5831	0.5916	0.5899	0.4900
		LASSO	0.3452	0.4187	0.4138	0.4139	0.4178	0.2654
	n=200	MLE	0.2693	0.3632	0.3533	0.3533	0.3514	0.3141
		LASSO	0.2314	0.3135	0.3068	0.3102	0.3039	0.2173
	n=500	MLE	0.1622	0.2039	0.2076	0.2089	0.2079	0.1869
		LASSO	0.1526	0.1928	0.1965	0.1973	0.1934	0.1497
$\rho = 0.9$	n=100	MLE	0.5322	1.3131	1.3069	1.2584	1.2513	1.2868
		LASSO	0.3669	0.7496	0.7777	0.7809	0.7517	0.6025
	n=200	MLE	0.2978	0.8005	0.7984	0.7694	0.8050	0.7496
		LASSO	0.2538	0.6350	0.6314	0.6160	0.6283	0.4592
	n=500	MLE	0.1797	0.4634	0.4510	0.4611	0.4610	0.4443
		LASSO	0.1678	0.4301	0.4170	0.4245	0.4254	0.3285

表 4.6　真實參數是(1,1,1,1,1,0)，估計量的均方差

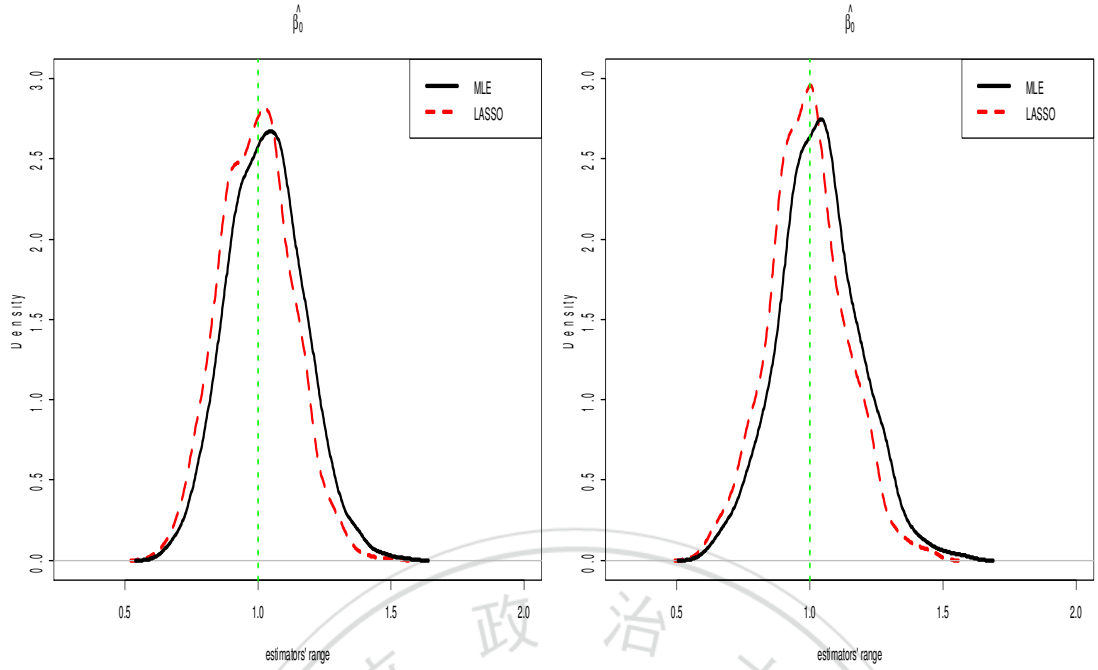
相關係數	樣本數	方法	$\widehat{\beta_0}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_4}$	$\widehat{\beta_5}$
$\rho = 0$	n=100	MLE	0.1490	0.1630	0.1705	0.1706	0.1638	0.1059
		LASSO	0.0919	0.1120	0.1209	0.1123	0.1175	0.0487
	n=200	MLE	0.0535	0.0657	0.0630	0.0621	0.0667	0.0415
		LASSO	0.0421	0.0580	0.0545	0.0540	0.0583	0.0251
	n=500	MLE	0.0211	0.0218	0.0209	0.0223	0.0213	0.0150
		LASSO	0.0185	0.0207	0.0203	0.0205	0.0215	0.0112
$\rho = 0.2$	n=100	MLE	0.1782	0.2007	0.2225	0.2188	0.2146	0.1316
		LASSO	0.1019	0.1258	0.1426	0.1296	0.1397	0.0488
	n=200	MLE	0.0653	0.0895	0.0840	0.0821	0.0850	0.0553
		LASSO	0.0493	0.0736	0.0668	0.0638	0.0701	0.0308
	n=500	MLE	0.0252	0.0255	0.0259	0.0273	0.0278	0.0199
		LASSO	0.0214	0.0245	0.0246	0.0259	0.0262	0.0136
$\rho = 0.5$	n=100	MLE	0.2468	0.3700	0.3604	0.3866	0.3860	0.2404
		LASSO	0.1261	0.1996	0.2000	0.1891	0.1916	0.0767
	n=200	MLE	0.0768	0.1373	0.1308	0.1309	0.1271	0.0988
		LASSO	0.0563	0.1062	0.1016	0.1034	0.1026	0.0497
	n=500	MLE	0.0273	0.0423	0.0435	0.0447	0.0444	0.0350
		LASSO	0.0237	0.0391	0.0412	0.0404	0.0389	0.0236
$\rho = 0.9$	n=100	MLE	0.3209	1.7406	1.7520	1.6676	1.6033	1.6559
		LASSO	0.1395	0.6090	0.6284	0.6257	0.5909	0.4266
	n=200	MLE	0.0950	0.6437	0.6513	0.5986	0.6527	0.5619
		LASSO	0.0666	0.4210	0.4052	0.3935	0.4088	0.2455
	n=500	MLE	0.0335	0.2170	0.2040	0.2132	0.2138	0.1980
		LASSO	0.0285	0.1868	0.1778	0.1840	0.1837	0.1241

除了比較估計量在各種標準下的差異外，我們也觀察截距項$\widehat{\beta_0}$、非零項迴歸係數估計量$\widehat{\beta_1}$與零項迴歸係數估計量$\widehat{\beta_5}$的分佈情況。圖 4.3、4.4、4.5 為當樣本數固定為 500 時，β_0,β_1,β_5的 MLE 與 LASSO 的抽樣分配圖。

由圖 4.3 可知，兩個方法的截距項估計量抽樣分佈圖形大致滿足對稱性，且不受相關係數大小的影響。當解釋變數間不相關時，兩估計量與參數真值的偏差很小。但隨著相關性的上升，MLE 估計量仍滿足不偏性，而 LASSO 低估的現象變得越來越明顯。由於 LASSO 的機率函數圖較 MLE 集中一點，反映了其變異程度略小一點。此外，當解釋變數間的相關係數愈大，兩估計量的集中程度下降並不明顯，但是 MLE 估計量的全距略為上升，反映了其變異程度稍微增加。

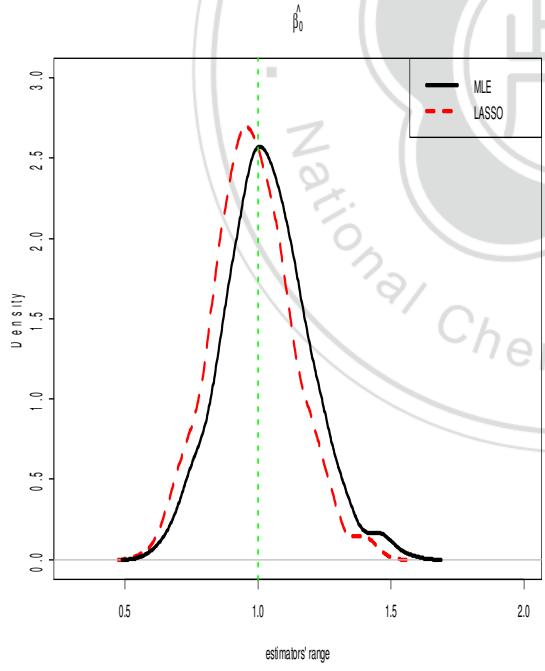
由圖 4.4 可知，兩方法之非零項斜率估計量抽樣分佈圖形大致滿足對稱性，且不受相關係數大小的影響。與參數真值的偏差，兩估計量與參數真值的偏差很小，隨著相關性的上升，MLE 估計量仍滿足不偏性，而 LASSO 低估的現象稍微增加。LASSO 的機率函數圖較 MLE 集中一點，反映了其變異程度較小一點，但並不是很明顯。當解釋變數間的相關係數愈大，兩估計量的變異程度將隨之增加。

由圖 4.5 可知，MLE 之零項斜率估計量抽樣分佈圖形大致滿足對稱性，且不受相關係數大小的影響；反之，LASSO 估計量於相關係數為零時，圖形還稱得上對稱，但隨著相關係數上升，就變得很不對稱。與參數真值的偏差，由於 LASSO 圖形不對稱，所以偏度不易辨別，不過原則上兩估計量大致上是不偏的，且偏度亦不受解釋變數間相關性高低的影響。LASSO 的機率函數圖較 MLE 集中許多，反映了其變異程度較小很多。當解釋變數間的相關係數愈大，兩估計量的變異程度將隨之增加。

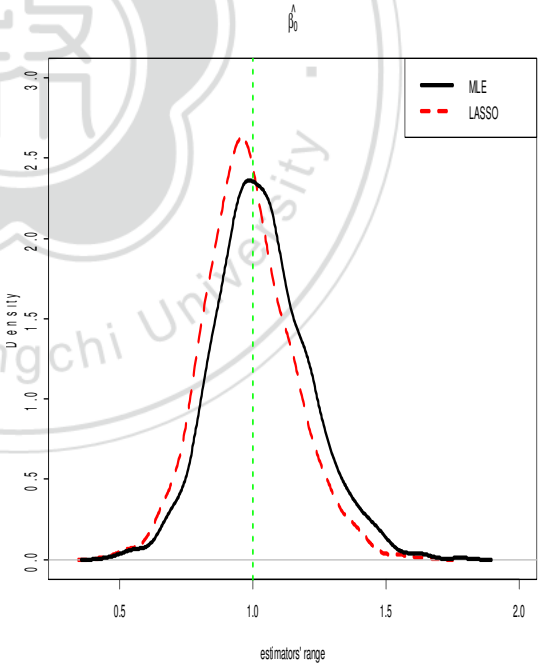


(a)

(b)



(c)



(d)

圖 4.3　給定樣本數為 500 時，真實迴歸係數是$(1,1,1,1,1,0)$，$\widehat{\beta_0}$的抽樣分佈
（a）ρ=0　（b）ρ=0.2　（c）ρ=0.5　（d）ρ=0.9

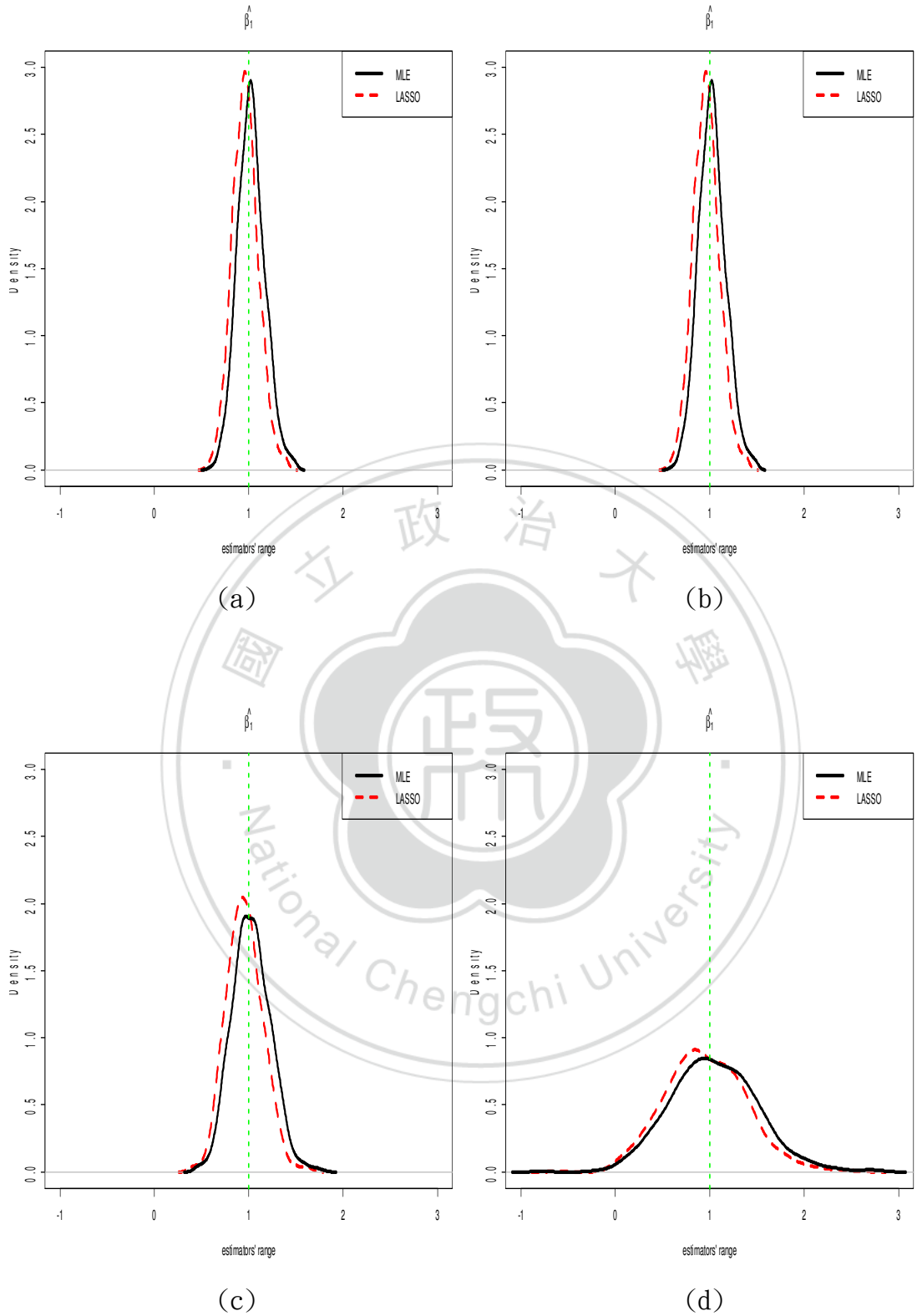
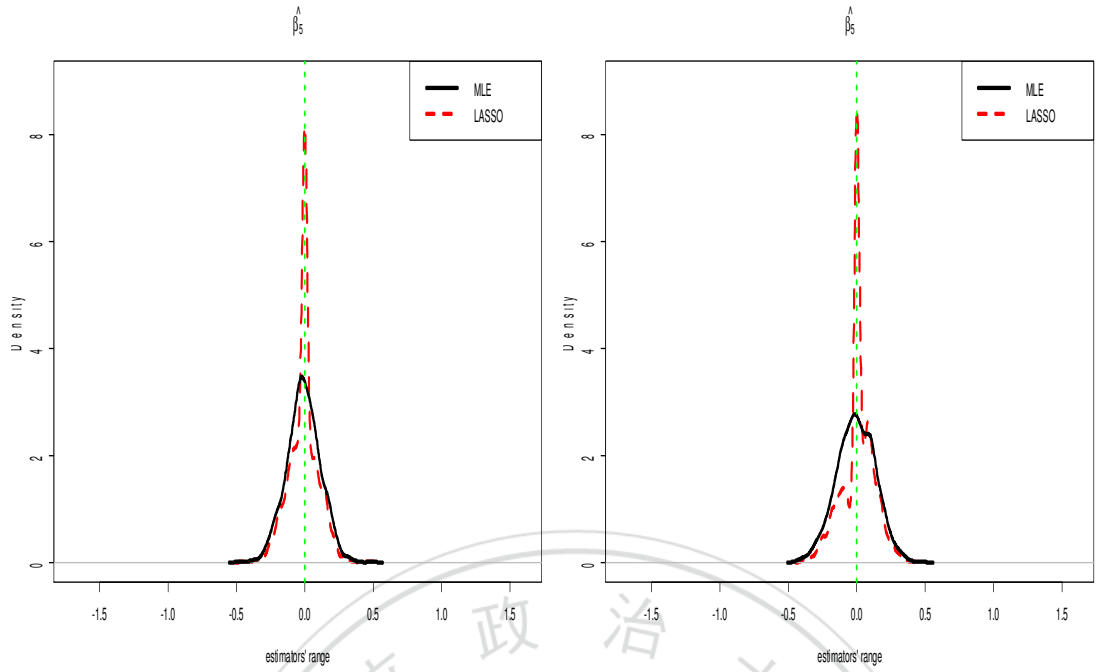
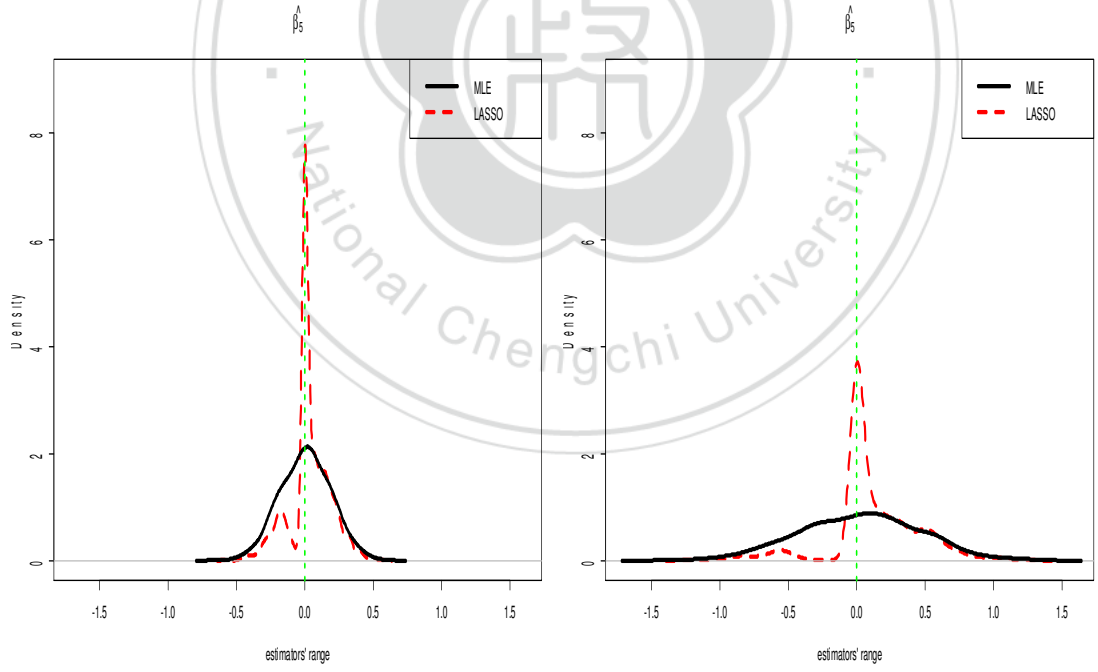


圖 4.4　給定樣本數為 500 時，真實迴歸係數是$(1,1,1,1,1,0)$，$\widehat{\beta_1}$的抽樣分佈
（a）ρ=0　（b）ρ=0.2　（c）ρ=0.5　（d）ρ=0.9



(a)　　(b)



(c)　　(d)

圖 4.5　給定樣本數為 500 時，真實迴歸係數是(1, 1, 1, 1, 1, 0)，$\widehat{\beta_5}$的抽樣分佈
（a）ρ=0　（b）ρ=0.2　（c）ρ=0.5　（d）ρ=0.9

叁、三個迴歸係數非零時：

表 4.7、4.8、4.9 分別為當真實參數是(1,1,1,0,0,0)時，估計量的偏差、標準差、均方差。三個表格的設計，由左至右側縱列方向依序為相關係數、樣本數、估計方法、估計量。相關係數分為不相關、低度、中度及高度相關；樣本數由小至大為 100、200、500；估計法分成 MLE 和 LASSO 兩種；估計量共有六個，包括截距項與斜率項。

由表4.7可以得到,隨著樣本數的增加,非零項估計量的偏差變得越來越小。當相關係數上升，所有估計量的偏差並非隨之增加。MLE 非零估計量的偏差均為正值，故通常高估真實參數值。LASSO 則相反，均低估參數值。另外仔細觀察斜率的部分可得知， LASSO 非零估計量的偏差絕對值皆比 MLE 大。至於截距項的估計量上，MLE 估計量的偏差幾乎比 LASSO 大。

由表4.8可以得到,隨著樣本數的增加,所有估計量的標準差皆一致地變小。當相關係數上升，所有估計量的標準差有上升的趨勢。LASSO 零斜率估計量的標準差，明顯比非零項估計量小很多。LASSO 估計量的標準差幾乎比 MLE 小。除了截距項，當相關係數越大，兩者標準差的差距也變得更大。

由表4.9可以得到,隨著樣本數的增加,所有估計量的均方差皆一致地變小。當相關係數上升，除了截距項，斜率估計量的均方差逐漸增加，且增加的幅度越來越大。LASSO 的零斜率估計量皆一致比 MLE 來得小，且隨著相關係數上升，兩方法之零斜率估計量的均方差差距隨之增加。

表 4.7　真實參數是(1,1,1,0,0,0)，估計量的偏差

相關係數	樣本數	方法	$\widehat{\beta_0}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_4}$	$\widehat{\beta_5}$
$\rho = 0$	n=100	MLE	0.0996	0.1097	0.1059	-0.0055	-0.0049	-0.0016
		LASSO	-0.0513	-0.1994	-0.2000	-0.0013	-0.0046	-0.0002
	n=200	MLE	0.0415	0.0407	0.0448	-0.0015	-0.0070	-0.0064
		LASSO	-0.0434	-0.1434	-0.1389	-0.0030	-0.0034	-0.0055
	n=500	MLE	0.0235	0.0262	0.0238	0.0067	-0.0117	-0.0013
		LASSO	-0.0221	-0.0736	-0.0757	0.0042	-0.0081	-0.0023
$\rho = 0.2$	n=100	MLE	0.1044	0.1061	0.1014	-0.0053	-0.0025	0.0024
		LASSO	-0.0488	-0.1974	-0.1971	0.0152	0.0224	0.0235
	n=200	MLE	0.0480	0.0437	0.0431	-0.0021	-0.0090	-0.0089
		LASSO	-0.0406	-0.1385	-0.1388	0.0135	0.0083	0.0058
	n=500	MLE	0.0243	0.0252	0.0255	0.0065	-0.0074	-0.0011
		LASSO	-0.0246	-0.0765	-0.0760	0.0119	0.0020	0.0088
$\rho = 0.5$	n=100	MLE	0.1076	0.1234	0.1085	-0.0010	-0.0037	0.0090
		LASSO	-0.0461	-0.1902	-0.1979	0.0518	0.0453	0.0512
	n=200	MLE	0.0419	0.0515	0.0401	0.0080	-0.0058	-0.0107
		LASSO	-0.0450	-0.1372	-0.1487	0.0368	0.0315	0.0269
	n=500	MLE	0.0248	0.0248	0.0219	0.0075	-0.0085	-0.0020
		LASSO	-0.0203	-0.0763	-0.0788	0.0232	0.0165	0.0159
$\rho = 0.9$	n=100	MLE	0.1197	0.1572	0.0931	0.0317	-0.0382	0.0015
		LASSO	-0.0243	-0.2609	-0.3137	0.1543	0.1275	0.1382
	n=200	MLE	0.0499	0.0429	0.0439	0.0202	-0.0109	0.0080
		LASSO	-0.0187	-0.2001	-0.2042	0.1151	0.0945	0.1028
	n=500	MLE	0.0243	0.0190	0.0177	0.0027	0.0017	0.0063
		LASSO	-0.0066	-0.1232	-0.1221	0.0703	0.0686	0.0652

表 4.8　真實參數是(1,1,1,0,0,0)，估計量的標準差

相關係數	樣本數	方法	$\widehat{\beta_0}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_4}$	$\widehat{\beta_5}$
$\rho = 0$	n=100	MLE	0.3245	0.3488	0.3467	0.2879	0.3007	0.2980
		LASSO	0.2667	0.2903	0.2868	0.1670	0.1769	0.1700
	n=200	MLE	0.2094	0.2228	0.2291	0.1921	0.1800	0.1884
		LASSO	0.1921	0.2111	0.2199	0.1257	0.1159	0.1212
	n=500	MLE	0.1258	0.1362	0.1327	0.1147	0.1162	0.1148
		LASSO	0.1213	0.1359	0.1349	0.0829	0.0845	0.0813
$\rho = 0.2$	n=100	MLE	0.3290	0.3613	0.3661	0.3037	0.3194	0.3183
		LASSO	0.2725	0.2951	0.2972	0.1657	0.1775	0.1751
	n=200	MLE	0.2089	0.2472	0.2372	0.2038	0.1921	0.2058
		LASSO	0.1903	0.2262	0.2218	0.1300	0.1187	0.1301
	n=500	MLE	0.1315	0.1451	0.1370	0.1231	0.1245	0.1177
		LASSO	0.1249	0.1421	0.1347	0.0856	0.0873	0.0789
$\rho = 0.5$	n=100	MLE	0.3432	0.4441	0.4415	0.3951	0.4237	0.4022
		LASSO	0.2777	0.3504	0.3421	0.2166	0.2402	0.2147
	n=200	MLE	0.2118	0.2942	0.2815	0.2494	0.2480	0.2555
		LASSO	0.1903	0.2632	0.2513	0.1503	0.1514	0.1547
	n=500	MLE	0.1344	0.1767	0.1624	0.1509	0.1546	0.1540
		LASSO	0.1272	0.1642	0.1540	0.1017	0.1048	0.1048
$\rho = 0.9$	n=100	MLE	0.3541	0.9431	0.9382	0.9148	0.9200	0.9507
		LASSO	0.2907	0.6161	0.6261	0.4445	0.4359	0.4779
	n=200	MLE	0.2193	0.5849	0.5711	0.5685	0.5654	0.5817
		LASSO	0.2012	0.4814	0.4690	0.3293	0.3383	0.3498
	n=500	MLE	0.1387	0.3571	0.3531	0.3548	0.3495	0.3413
		LASSO	0.1329	0.3206	0.3229	0.2349	0.2274	0.2222

表 4.9　真實參數是(1,1,1,0,0,0)，估計量的均方差

相關係數	樣本數	方法	$\widehat{\beta_0}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_4}$	$\widehat{\beta_5}$
$\rho = 0$	n=100	MLE	0.1152	0.1337	0.1314	0.0829	0.0904	0.0888
		LASSO	0.0738	0.1241	0.1223	0.0279	0.0313	0.0289
	n=200	MLE	0.0456	0.0513	0.0545	0.0369	0.0324	0.0355
		LASSO	0.0388	0.0651	0.0676	0.0158	0.0134	0.0147
	n=500	MLE	0.0164	0.0192	0.0182	0.0132	0.0136	0.0132
		LASSO	0.0152	0.0239	0.0239	0.0069	0.0072	0.0066
$\rho = 0.2$	n=100	MLE	0.1192	0.1418	0.1443	0.0923	0.1020	0.1013
		LASSO	0.0766	0.1260	0.1272	0.0277	0.0320	0.0312
	n=200	MLE	0.0459	0.0630	0.0581	0.0415	0.0370	0.0424
		LASSO	0.0379	0.0704	0.0685	0.0171	0.0142	0.0170
	n=500	MLE	0.0179	0.0217	0.0194	0.0152	0.0156	0.0139
		LASSO	0.0162	0.0260	0.0239	0.0075	0.0076	0.0063
$\rho = 0.5$	n=100	MLE	0.1294	0.2124	0.2067	0.1561	0.1796	0.1618
		LASSO	0.0793	0.1589	0.1562	0.0496	0.0598	0.0487
	n=200	MLE	0.0466	0.0892	0.0809	0.0623	0.0615	0.0654
		LASSO	0.0382	0.0881	0.0853	0.0239	0.0239	0.0247
	n=500	MLE	0.0187	0.0318	0.0269	0.0228	0.0240	0.0237
		LASSO	0.0166	0.0328	0.0299	0.0109	0.0113	0.0112
$\rho = 0.9$	n=100	MLE	0.1397	0.9141	0.8888	0.8379	0.8479	0.9038
		LASSO	0.0851	0.4476	0.4904	0.2214	0.2063	0.2475
	n=200	MLE	0.0506	0.3439	0.3281	0.3236	0.3198	0.3384
		LASSO	0.0408	0.2718	0.2616	0.1217	0.1234	0.1330
	n=500	MLE	0.0198	0.1279	0.1250	0.1259	0.1222	0.1166
		LASSO	0.0177	0.1180	0.1192	0.0601	0.0564	0.0536

除了比較估計量在各種標準下的差異外，當樣本數固定為 500 時，我們也觀察截距項$\widehat{\beta_0}$、非零項迴歸係數估計量$\widehat{\beta_2}$與零項迴歸係數估計量$\widehat{\beta_4}$的抽樣分配情況。

由圖 4.6 可知，兩個方法的截距項估計量抽樣分佈圖形大致滿足對稱性，尤以 MLE 較為對稱，但隨相關係數的上升，兩方法之估計量抽樣分佈稍微右偏。與參數真值的偏差，MLE 傾向高估，LASSO 傾向低估，且兩方法不受相關係數大小的影響。由於LASSO的機率函數圖較MLE集中一點，反映了其變異程度略小一點。此外，當解釋變數間的相關係數愈大，兩估計量的集中程度下降，而且估計量的全距還逐漸上升，反映了其變異程度隨之增加。

由圖 4.7 可知，MLE 非零項估計量抽樣分佈圖形大致滿足對稱，且不受相關係數影響；LASSO 的抽樣分佈圖則呈右偏，且隨相關係數上升，右偏情況越明顯。與參數真值的偏差，LASSO 傾向低估，但是偏度並不受相關係數的影響，而 MLE 則不偏。除了變數間相關係數為零時，LASSO 的機率函數圖皆較 MLE 集中一點，反映了其變異程度小一點。當解釋變數間的相關係數愈大，兩估計量的集中程度下降且全距也逐漸上升，反映了其變異程度隨之增加。

由圖 4.8 可知，MLE 之零項估計量抽樣分佈圖形大致滿足對稱性，且不受相關係數大小的影響；反之，LASSO 估計量抽樣分佈圖，不論相關係數大小，圖形左右尾呈波浪狀放射出去，故不對稱。與參數真值的偏差，兩估計量大致上是不偏的，且偏度亦不受解釋變數間相關性高低的影響。LASSO 的機率函數圖較 MLE 集中許多，反映了其變異程度較小很多。當解釋變數間的相關係數愈大，兩估計量的集中程度有下降的趨勢，唯當LASSO在低度相關時例外，且全距也逐漸上升，代表兩估計量的變異程度逐漸增加。

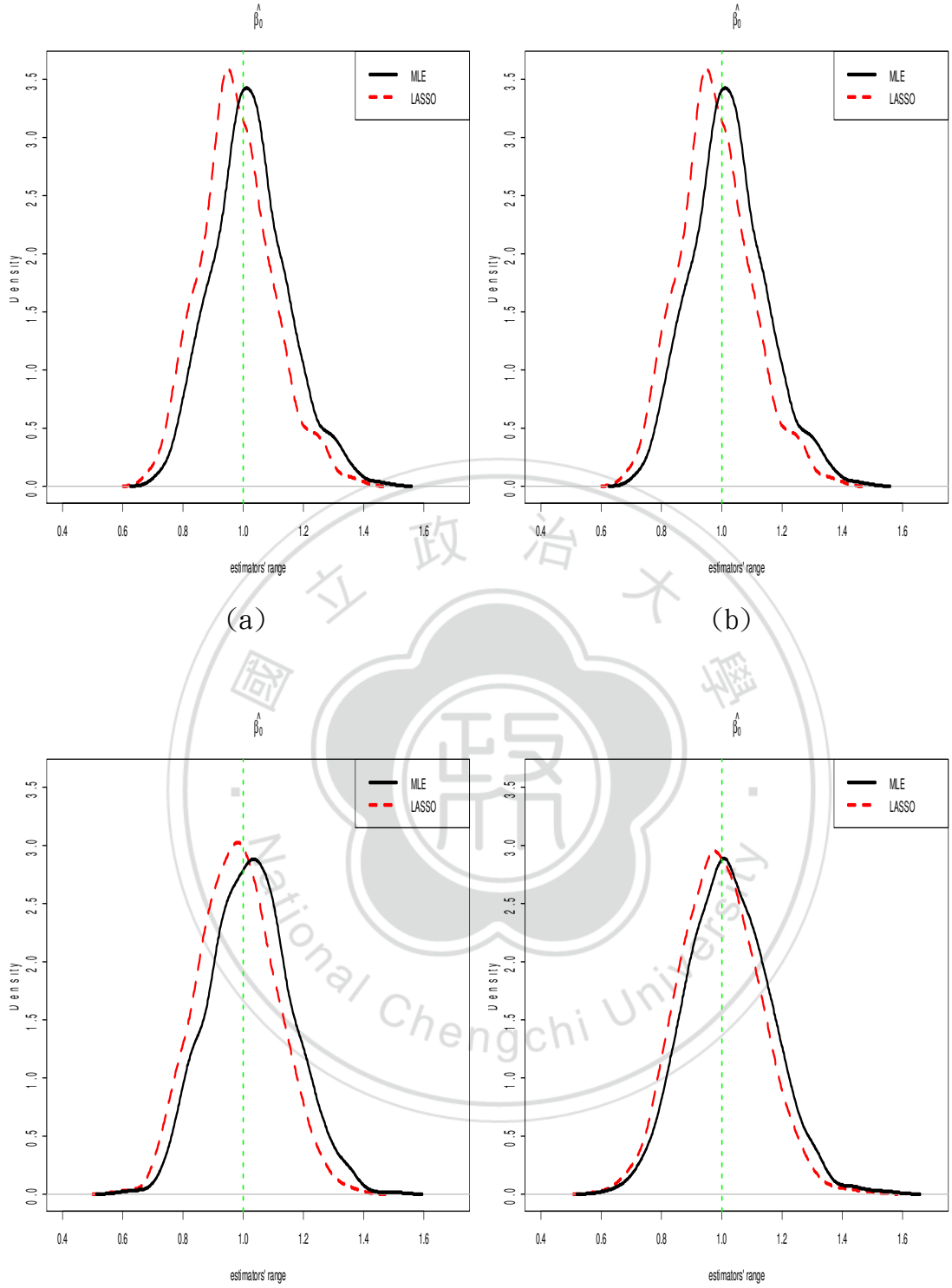


圖 4.6　給定樣本數為 500 時，真實迴歸係數是$(1,1,1,0,0,0)$，$\widehat{\beta_0}$的抽樣分佈
（a）ρ=0　（b）ρ=0.2　（c）ρ=0.5　（d）ρ=0.9

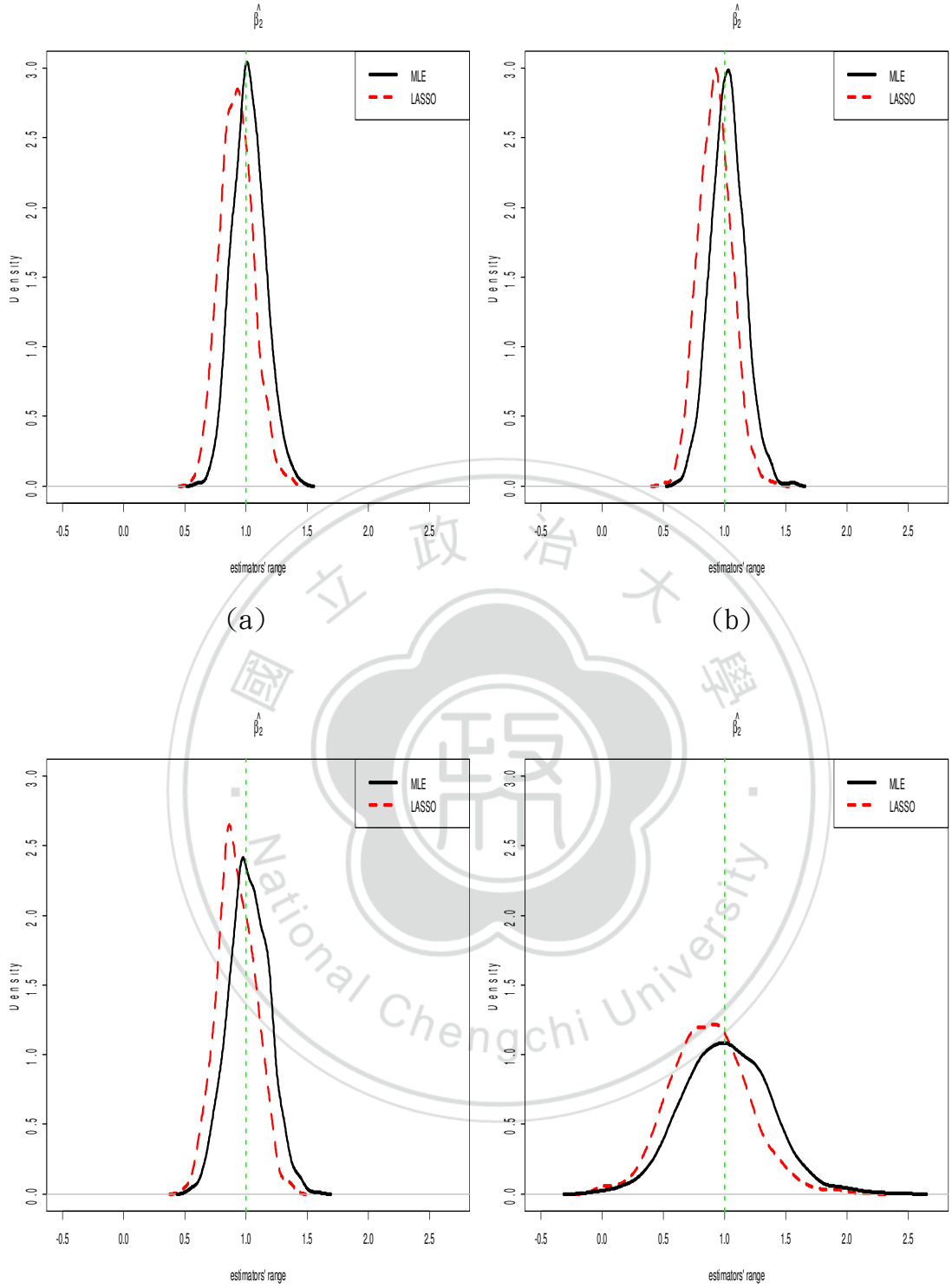


圖 4.7　給定樣本數為 500 時，真實迴歸係數是$(1,1,1,0,0,0)$，$\widehat{\beta_2}$的抽樣分佈
（a）ρ=0　（b）ρ=0.2　（c）ρ=0.5　（d）ρ=0.9

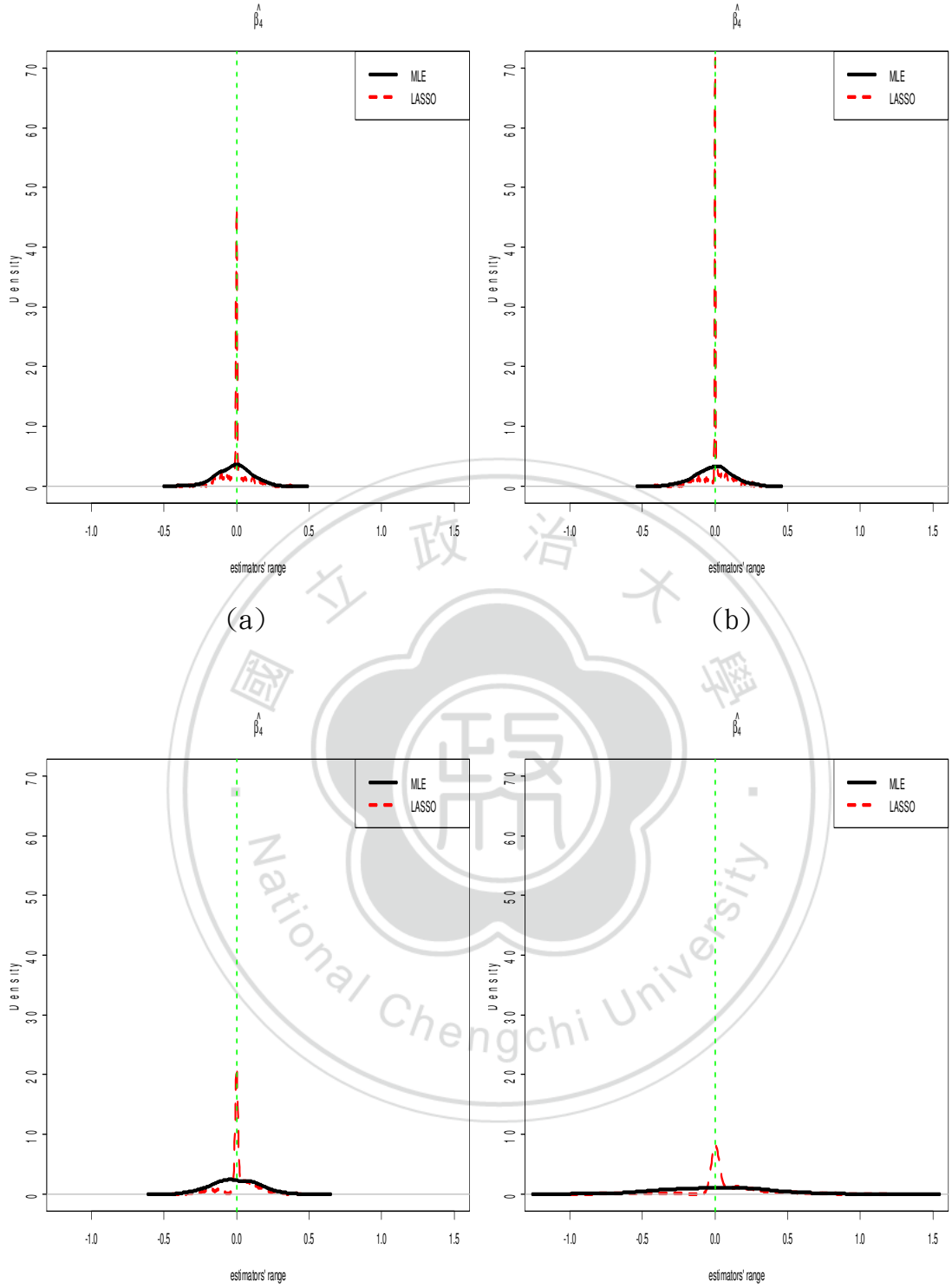


圖 4.8　給定樣本數為 500 時，真實迴歸係數是$(1,1,1,0,0,0)$，$\widehat{\beta_4}$的抽樣分佈
（a）ρ=0　（b）ρ=0.2　（c）ρ=0.5　（d）ρ=0.9

5. 結論

本文主要探討在迴歸分析中，LASSO 與傳統估計方法上的差異。在線性迴歸中，我們比較了最小平方法與 LASSO 之偏差、變異數。已知在計算上，最小平方法有解析解，而 LASSO 則是牽涉限制式下最小化工作，並無解析解。為探討其背後解根原理，本文介紹凸函數最佳化問題以及藉由 Karush-Kuhn-Tucker 條件推導至強對偶性以求得最佳解。我們也介紹如何利用拔靴法估計估計量的標準差，以及利用交叉驗證方式以選擇門檻值。之後我們將 LASSO 應用在羅吉斯迴歸的迴歸係數估計上，包括估計量標準差的估計方式和門檻值的決定。

我們將 LASSO 法應用在三組實際資料之羅吉斯迴歸分析中。在脊椎後凸的例子裡，LASSO 較 MLE 有更準確的預測結果。在貓狗影像資料裡，解釋變數個數多於觀測值個數，LASSO 也能獲得不錯的表現。我們也透過模擬實驗，比較兩種方法的優劣性。當模型中存在與反應變數獨立之解釋變數，則在 LASSO 中，此變數相對應的零斜率估計量的標準差明顯比 MLE 小更多。另外，有關截距與非零斜率的估計上，LASSO 則略為低估參數真值，但有較低的變異。當解釋變數間的相關係數增加時，兩方法的變異數間的差距也隨之增加。

參考文獻

一、英文文獻

1. Boyd, S. and Vandenberghe, L. (2004), *Convex Optimization*, Cambridge University Press, 215-244.

2. Breiman, L. (1995) Better Subset Regression Using the Nonnegative Garrotte, *American Statistical Association*, **37**, 373-384.

3. Breiman, L. and Spector P. (1992) Submodel Selection and Evaluation in Regression. The X-Random Case, *International Statistical Review*, **60**, 291-319.

4. Dalal, N. Triggs, B. (2005) Histograms of Oriented Gradients for Human Detection, http://lear.inrialpes.fr/.

5. Friedl, I., Tilg, N. (1995) Variance estimates in logistic regression using the bootstrap, *Communications in Statistic-Theory and Methods*, **24**(2) 473-486.

6. Hoerl, E. and Kennard, R. (1970) Ridge Regression: Biased Estimation for Nonorthogonal Problems, *American Statistical Association*, **12**, 55-67.

7. Osborne, M., Presnell, B. and Turlach, B. (2000) On the LASSO and its dual, *Journal of Computational and Graphical Statistics*, **9**, 319–337.

8. Sill, M.,Hielscher, T. A and Becker M. (2014) Extended Inference with Lasso and Elastic-Net Regularized Cox and Generalized Linear Models, *Journal of Statistical Software*, **62**, 1-22.

9. Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society*, **58**, 267-288.

10. Zhao, X. (2008) *Lasso and Its Applications*, University of Minnesota Duluth, 4-17.

11. Zou, H. and Hastie, T. (2005) Regularization and Variable Selection via the Elastic Net, *Journal of the Royal Statistical Society*, **67**, 301-320.

二、中文文獻

1. 全民人體力學健康教室，淺談三種脊椎歪斜。

2. 賈金柱，高等統計選講，高等統計入門分析，2.2 節 Duality。

附錄一(2.3)之推導證明

假設 t 是事先給定的實數值，且X_{ij}已經經過標準化，藉由拉格朗日函數可得:

$$\mathrm{L}(\beta_0,\beta,\lambda)=\sum_{i=1}^{n}(y_i-\beta_0-\beta^T X_i^T)^2+\lambda(\sum_{j=1}^{p}|\beta_j|-\mathrm{t})。$$

欲求估計量β_0，假設

$$\frac{\partial L(\beta_0,\beta,\lambda)}{\partial\beta_0}=-2\sum_{i=1}^{n}(y_i-\beta_0-\beta^T X_i^T)=0。$$

因為

$$\sum_{i=1}^{n}(y_i-\beta_0-\beta^T X_i^T)=n\bar{y}-n\,\beta_0-\sum_{i=1}^{n}\beta^T X_i^T\ ,$$

$$\sum_{i=1}^{n}\beta^T X_i^T=\sum_{i=1}^{n}(\beta_1,\ldots,\beta_p)({X_1}^T,\ldots,{X_n}^T)=\sum_{i=1}^{n}(\sum_{j=1}^{p}\beta_j X_{ij})$$

$$=(\beta_1 X_{11}+,\ldots,\beta_p X_{1p})+,\ldots,+(\beta_1 X_{n1}+,\ldots,\beta_p X_{np})$$

$$=\beta_1(X_{11}+,\ldots,+X_{n1})+\beta_p(X_{1p}+,\ldots,+X_{np})=\sum_{j=1}^{n}\beta_j(\sum_{i=1}^{n}X_{ij})\ 。$$

因為 $\sum_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{X_{ij}}=0$，所以

$$\sum_{\mathrm{i}=1}^{\mathrm{n}}\beta^{\mathrm{T}}\mathrm{X_i^T}=\sum_{\mathrm{i}=1}^{\mathrm{n}}(\sum_{\mathrm{j}=1}^{\mathrm{p}}\beta_{\mathrm{j}}\mathrm{X_{ij}})=\sum_{\mathrm{j}=1}^{\mathrm{n}}\beta_{\mathrm{j}}(\sum_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{X_{ij}})=0。$$

且 $\frac{\partial^2 L(\beta_0,\beta,\lambda)}{\partial{\beta_0}^2}=2>0$，故當 $n\bar{y}-n\beta_0=0$ 時，可以最小化$L_{\beta_0}(\beta_0,\beta,\lambda)$

得證 $\widehat{\beta_0}=\bar{y}$。

附錄二(2.3)之推導證明

同(2.3)假設條件，假設 X 是一個$\mathrm{n}\times\mathrm{p}$的設計矩陣，當 X 滿足$\mathrm{X^TX}=\mathrm{I}$，其中$I$是$\mathrm{p}\times\mathrm{p}$的單位矩陣。欲求出估計量β，即 $min_{\beta}\mathrm{L}(\beta_0,\beta,\lambda)$。

以下皆用矩陣的方式呈現，首先定義 $\mathrm{g}(\beta)=\sum_{j=1}^{p}|\beta_j|-\mathrm{t}$是一個$R^p\to R$的函數，$\beta^0=(\beta_1^0,\ldots,\beta_p^0)^T$是普通最小平方估計量。

我們知道拉格朗日函數

$$\mathrm{L}(\beta_0,\beta,\lambda)=(\mathrm{y}-\beta_0-\beta^TX)^T(\mathrm{y}-\beta_0-\beta^TX)+\lambda\mathrm{g}(\beta)$$
$$=y^Ty-2y^T\beta_0-2\beta_0^T\beta^TX-2y^TX\beta+\beta^TX^TX\beta+\lambda\mathrm{g}(\beta)。$$

假設

$$\nabla_{\beta}\mathrm{L}(\beta_0,\beta,\lambda)=-2\beta_0^TX-2y^TX+2X^TX\beta+\lambda\,sign(\beta)=0 \qquad (*)$$

已知普通最小平方法的正規方程式(normal equation)

$$X'^TX'\beta^0=X'^Ty\Longrightarrow\beta^0=(X'^TX')^{-1}X'^Ty，$$

其中$X'=(\mathbf{1},X)$為$\mathrm{n}\times(\mathrm{p}+1)$的矩陣，$\mathbf{1}=(1,\ldots,1)^T$為$\mathrm{n}\times1$的向量。由式子(*)可得 LASSO 正規方程式:

$$X'^TX'\beta'+\frac{\lambda}{2}\,sign(\beta')=X'^Ty$$

其中$\beta'=(\beta_0,\beta_1,\ldots,\beta_p)^T$為$(\mathrm{p}+1)\times1$的向量。

由於$X^TX=I$，$\lambda\geq0$，所以估計量

$$\hat{\beta}=-\frac{\lambda}{2}\,sign(\hat{\beta})+(X^TX)^{-1}X^Ty=\widehat{\beta^0}-\frac{\lambda}{2}\,sign(\hat{\beta})=$$

$$\begin{cases}\widehat{\beta^0}-\frac{\lambda}{2} & ;\,if\ sign(\hat{\beta})=1\\ \widehat{\beta^0}+\frac{\lambda}{2} & ;\,if\ sign(\hat{\beta})=-1\end{cases}。$$

於是我們發現 $sign(\hat{\beta})=sign(\widehat{\beta^0})$，故 $\hat{\beta}=sign(\widehat{\beta^0})(|\widehat{\beta^0}|-\frac{\lambda}{2})^{+}$。