

國立政治大學統計學系研究所

碩士學位論文

一種基於BIC的B-Spline節點估計方式



指導教授:黃子銘 博士

研究生:何昕燁

中華民國一百零二年六月

摘要

在迴歸分析中，若變數間具有非線性的關係時，B-Spline線性迴歸是以無母數的方式建立模型。B-Spline函數為具有節點(knots)的分段多項式，選取合適節點的位置對B-Spline的估計有重要的影響，在近年來許多的文獻中已提出一些尋找節點位置的估計方法，而本文中我們提出了一種基於Bayesian information criterion(BIC)的節點估計方式。

我們想要深入了解在不同類型的迴歸函數間，各種選取節點方法的配適效果與模擬時間，並且加以比較，在使用B-Spline函數估計時，能夠使用合適的方法尋找節點。



Abstract

In regression analysis, when the relation between the response variable and the explanatory variable is nonlinear, one can use nonparametric methods to estimate the regression function.

B-Spline regression is one of the popular nonparametric regression methods. B-Splines are piecewise polynomial joint at knots, and the choice of knot locations is crucial.

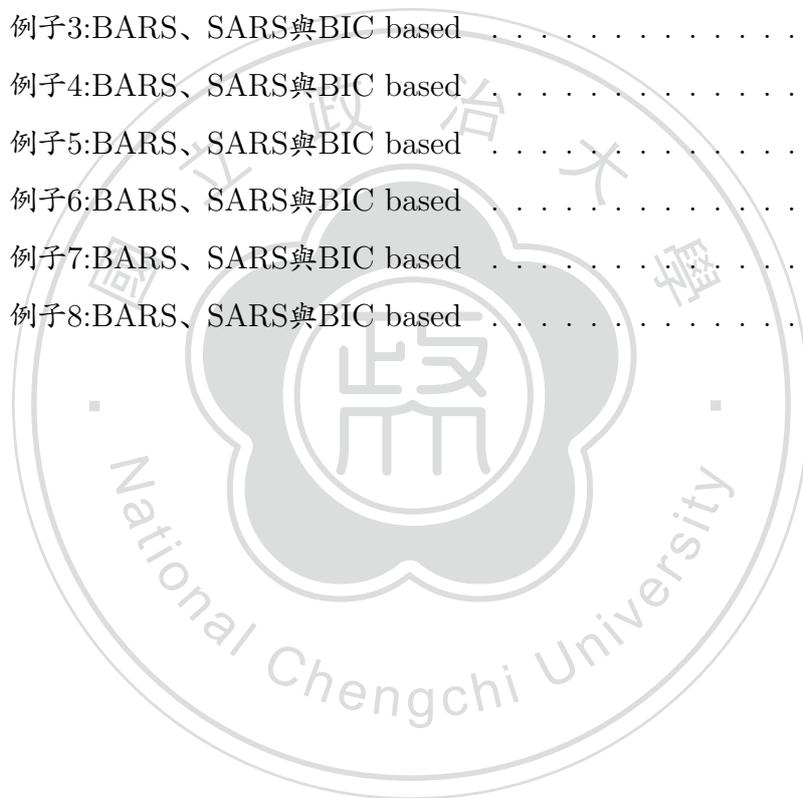
Zhou and Shen (2001) proposed to use spatially adaptive regression splines (SARS), where the knots are estimated using a selection scheme. Dimatteo, Genovese, and Kass (2001) proposed to use Bayesian adaptive regression splines (BARS), where certain priors for knot locations are considered. In this thesis, a knot estimation method based on the Bayesian information criterion (BIC) is proposed, and simulation studies are carried out to compare BARS, SARS and the proposed BIC-based method.

目錄

| | |
|---|----|
| 1 緒論 | 1 |
| 2 文獻回顧 | 2 |
| 3 研究方法 | 3 |
| 3.1 建立B-Spline迴歸模型 | 3 |
| 3.2 Spatially Adaptive Regression Splines | 4 |
| 3.2.1 尋找節點的起始值 | 4 |
| 3.2.2 對起始節點做增加、刪除或平移 | 6 |
| 3.3 Bayesian Adaptive Regression Splines | 6 |
| 3.4 基於BIC的節點估計方式 | 8 |
| 4 模擬與比較 | 9 |
| 4.1 模擬 | 9 |
| 4.2 比較 | 11 |
| 5 結論與建議 | 13 |
| 5.1 結論 | 13 |
| 5.2 建議 | 13 |

圖目錄

| | | |
|----|--|----|
| 1 | 選取節點位置的重要性 | 1 |
| 2 | BARS在間斷型迴歸函數之圖形 | 12 |
| 3 | 例子2中使用SARS、BARS與BIC-based之圖形 | 12 |
| 4 | 例子1:BARS、SARS與BIC based | 17 |
| 5 | 例子2:BARS、SARS與BIC based | 17 |
| 6 | 例子3:BARS、SARS與BIC based | 18 |
| 7 | 例子4:BARS、SARS與BIC based | 18 |
| 8 | 例子5:BARS、SARS與BIC based | 19 |
| 9 | 例子6:BARS、SARS與BIC based | 19 |
| 10 | 例子7:BARS、SARS與BIC based | 20 |
| 11 | 例子8:BARS、SARS與BIC based | 20 |



表目錄

| | | |
|---|---------------------------|----|
| 1 | 平滑型函數之估計效果 | 9 |
| 2 | 間斷型函數之估計效果 | 10 |
| 3 | 尋找節點之模擬時間(單位:分) | 10 |
| 4 | 平均選取節點個數 | 11 |



1 緒論

在非線性(nonlinear)迴歸當中，當迴歸函數類型已知，例如倒數、二次、三次多項式、指數、對數函數、羅吉斯(logistic)函數等等，可利用最小平方方法(least square method)估計迴歸函數。但是如果迴歸函數類型未知，則需要使用無母數(nonparametric)的方法找出迴歸函數估計。

B-Spline線性迴歸是以無母數的方式估計迴歸函數，使用B-Spline函數估計時，需要設定函數最高次數與節點的個數與位置，其中節點位置的設定對B-Spline的估計有重要的影響，以下以圖1中的資料舉例說明：

圖1中兩圖均來自相同資料，圖1(a)將節點等距排放於函數區間上，對節點的個數做估計，而圖1(b)則是同時估計了節點的個數與位置。由圖形可知，若將節點等距排放則無法達到良好的表現，在近年來許多文獻中提出一些估計B-Spline節點的方法，因此我們想要深入了解不同選取節點方法的配適效果，並且加以比較。

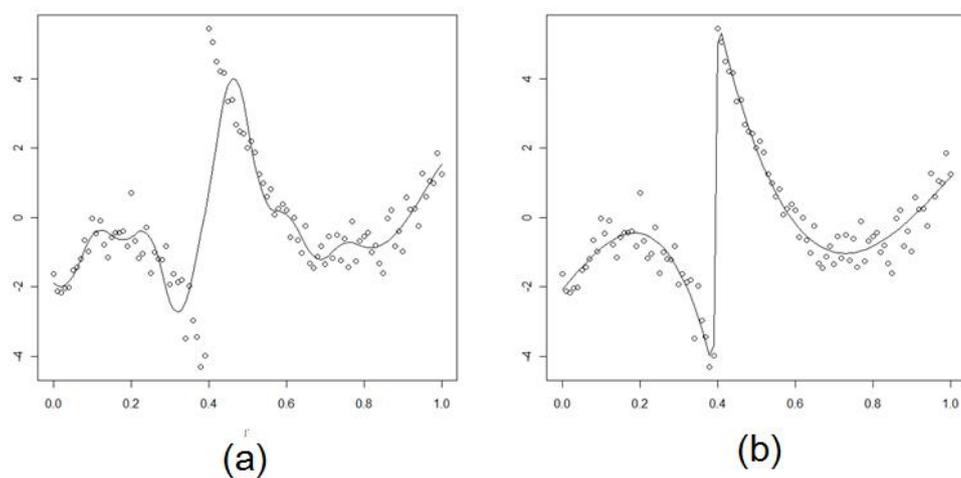


圖 1: 選取節點位置的重要性

2 文獻回顧

B-Spline曲線在1946年由Schoenberg所提出，此曲線由B-Spline的基底 (basis)函數組成，可以使用在平滑函數的近似。B-Spline函數為分段多項式 (piecewise polynomial)，分段點稱為節點(knots)，在統計方面的迴歸分析中，迴歸函數可以透過B-Spline曲線近似，再以最小平方法估計。使用B-Spline估計時需要設定節點，其中一種設定方式為平均的放入節點到函數區間上，但是需要選取節點個數，例如 Halpern[10]、Lindstorm[12]與 Ruppert[13]均使用這種方式並提出了不同選取節點個數的方法。另外一種方法需要估計節點的位置，例如 Friedman[7]與 Zhou與 Shen[16]的論文使用了Stein的標準[15]對節點的位置做估計，其中 Zhou與 Shen的文章提出了新的演算法尋找到節點估計。此外，Denison, Mallick與 Smith[4]、Biller[1]與 Dimatteo, Genovese與 Kass[5]的文章中使用貝氏的方式設定節點的事前分配，再由reversible-jump Markov chain Monte Carlo[9]的方法對節點的事後分配做抽樣，再進行迴歸函數的估計。

3 研究方法

假設一筆資料中，解釋變數 x_i 與被解釋變數 y_i 間具有非線性的關係如模型(1):

$$y_i = f(x_i) + \varepsilon_i, i = 1, \dots, n. \quad (1)$$

其中

$$\varepsilon_i \sim N(0, \sigma^2)$$

當 f 為一平滑函數且型式不確定時，則我們可以使用B-Spline函數近似 f ，在3.1節我們將進一步說明如何估計 f 。

3.1 建立B-Spline迴歸模型

假設 f 為一定義於 $[a, b]$ 區間上的B-Spline函數，則 f 為一分段多項式，且可以表示為B-Spline基底函數的線性組合，分段多項式的最高次數為 m 之下，B-Spline的基底函數由一組節點向量 $t = (t_1, \dots, t_k)$ 所決定，基底函數共有 $k + m + 1$ 個，記為 e_1, \dots, e_{k+m+1} ，可利用de Boor遞迴公式[3]算出。此時模型(1)中的 f 可改寫成:

$$f(x; \beta, t) = \sum_{j=1}^{k+m+1} \beta_j e_j(x; t), a \leq x \leq b. \quad (2)$$

其中 $\beta = (\beta_1, \dots, \beta_{k+m+1})$ 為基底的係數。以下我們均將 m 設定為3。基底的係數可以由最小平方方法估計，令 $y = (y_1, \dots, y_n)^T$ 且

$$B_{k,t} = \begin{pmatrix} e_1(x_1; t) & \cdots & e_{k+m+1}(x_1; t) \\ \vdots & \vdots & \vdots \\ e_1(x_n; t) & \cdots & e_{k+m+1}(x_n; t) \end{pmatrix}$$

則 β 的最小平方方法估計為:

$$\hat{\beta} = (B_{k,t}^T B_{k,t})^{-1} B_{k,t}^T y$$

再由 $\beta = \hat{\beta}$ 代回式(2)可得到 f 的最小平方估計，記為 $\hat{f}(\cdot; t)$ 。

以下將介紹幾種節點位置的估計方法，在3.2節介紹 Zhou與 Shen提出的 spatially adaptive regression splines[16]、3.3節介紹 Dimatteo,Genovese與 Kass提出的 Bayesian adaptive regression splines[5]，而3.4節為本論文中提出的新方法，是基於 Bayesian information criterion(以下均簡稱 BIC)的一種節點估計方式。

3.2 Spatially Adaptive Regression Splines

本節介紹spatially adaptive regression splines(以下均簡稱SARS)的方法。在SARS的方法中使用Stein[15]所提出的標準選取節點，其標準說明如下。在給定節點向量 t 與多項式最高次數 m 時，定義 $R(\hat{f}(t))$ ：

$$R(\hat{f}(t)) = \sum_{i=1}^n (y_i - \hat{f}(x_i; t))^2/n + 2 \times (k + m + 1)\sigma^2/n, \quad (3)$$

其中 \hat{f} 為式(2)的 f 最小平方估計，而 σ^2 為迴歸模型(1)中 ε_i 的變異數，因為 σ 未知，所以我們使用 $\hat{\sigma}$ 估計：

$$\hat{\sigma} = \text{median}_{1 \leq i \leq 2/n} (|y_{2i} - y_{2i-1}| / (0.6745\sqrt{2})). \quad (4)$$

Stein的選取標準是選取節點使得式(3)的 $R(\hat{f}(t))$ 為最小，而 $R(\hat{f}(t))$ 中 $2 \times (k + m + 1)\sigma^2/n$ 為一懲罰項(penalty)，加上懲罰項的原因在於避免選取過多的節點時，產生過度配適之現象。

在SARS的方法中使用了一個特定的演算法將一組起始節點做增加、刪除或平移使得 $R(\hat{f}(t))$ 為最小尋找到新節點。以下在3.2.1節說明如何設定起始節點，而3.2.2節介紹如何由起始節點做增加、刪除或平移尋找到新節點。

3.2.1 尋找節點的起始值

Zhou與Shen[16]提出了一種選取起始節點的方式，起始節點從 $x_1 \cdots, x_n$ 中挑選，在決定 x_i 是否作為起始節點時，會依照以下標準判斷。

考慮以 x_i 為中心的一區間 $[x_i - d, x_i + d]$, $i = 1, \dots, n$, (d 的大小在本小節之後介紹), 令 $J = \{j \mid 1 \leq j \leq n, x_j \in [x_i - d, x_i + d]\}$, 為資料 $(x_j, y_j), j \in J$ 配適迴歸模型:

$$y_j = \alpha_{i,d}(x_j - x_i)_+^m + \sum_{v=0}^{m-1} \beta_{v,d}(x_j - x_i)^v + \epsilon_{i,j} \quad (5)$$

其中

$$(x_j - x_i)_+ = \max\{0, (x_j - x_i)\}$$

令 $\hat{\alpha}_{i,d}$ 為 $\alpha_{i,d}$ 的最小平方估計, 若 $|\hat{\alpha}_{i,d}|$ 夠大時, 則 x_i 會被選為起始的節點, 以下詳細說明之。令 X 為模型(5)中的解釋變數矩陣, 即

$$X = ((x_j - x_i)_+^m, (x_j - x_i)_+^0, \dots, (x_j - x_i)^m), j \in J,$$

令 $\widehat{Var}(\hat{\alpha}_{i,d}) = \hat{\sigma}^2((X^T X)^{-1})_{11}$, 其中 $((X^T X)^{-1})_{11}$ 為 $((X^T X)^{-1})$ 第一行第一列的元素, 而 $\hat{\sigma}$ 定義於式(4), 令 $r(x_i) = |\hat{\alpha}_{i,d} / \sqrt{\widehat{Var}(\hat{\alpha}_{i,d})}|$, 若 $r(x_i) > C$, 則認為 $|\hat{\alpha}_{i,d}|$ 夠大, 並將 x_i 加到起始節點, 此時 C 取5做為臨界值。

在選取節點的過程使用演算法呈現, 假設一開始的起始值集合 Δ 為一空集合, 再根據上述的規則, 逐漸增加合適的 x_i 到集合 Δ 內:

階段1 給定一 p 值, 使得 J 內的元素個數大於5。

階段2 $d = ((\max(x_i) - \min(x_i)) / n) \times 1.2^p$, $x_j \in [x_i - d, x_i + d]$, 求出 $r(x_i)$, $i = 1, \dots, n$, 若所有的 $r(x_i)$ 均小於 C 則跳到**階段5**, 否則繼續**階段3**。

階段3 存在 x_i 使得 $r(x_i) > C$, 則增加 x_i 到 Δ 中。

階段4 若已搜尋到的起始值為 x_i , 則在 $[x_i - d, x_i + d]$ 的區間內不再檢定是否需加入到起始節點。

階段5 增加 p 為 $p + 1$, 回到**階段2**。如果 $d < (\max(x_i) - \min(x_i)) / 5$, 則停止。

由以上演算法可求得更新後的起始值為 Δ , 但如果 Δ 依然為一空集合, 則可以考慮取 $C = 4$, 重新回演算法求得起始值。

3.2.2 對起始節點做增加、刪除或平移

由3.2.1節尋找到的起始節點中，我們對起始節點做增加、刪除或平移尋找新節點：

增加節點 假設起始節點向量為 $\mathbf{t} = (t_1, \dots, t_k)$ ，決定是否要在區間 $[t_i, t_{i+1}]$ 中增加潛在的節點 t_i^a ，其中 $i = 1 \dots, k$ ，而 t_i^a 則是由最佳化的方式尋找在區間 $[t_i, t_{i+1}]$ 中式(3)的最小值發生位置，令 \mathbf{t}_a 為 t_i^a 增加到 $[t_i, t_{i+1}]$ 後的節點向量，若 $R(\hat{f}(\mathbf{t}_a)) < R(\hat{f}(\mathbf{t}))$ ，則增加 t_i^a 到 $[t_i, t_{i+1}]$ 。

刪除平移 在**增加節點**階段後尋找出的節點為 $\mathbf{t}_a = (t_1, \dots, t_{k_a})$ ，決定是否將 t_i 平移到區間 $[t_i - 1, t_i + 1]$ 為 t_i^r 或是刪除 t_i ， $i = 1 \dots, k_a$ ，而 t_i^r 則是由最佳化的方式尋找在區間 $[t_i - 1, t_i + 1]$ 中式(3)的最小值發生位置，其中 \mathbf{t}_r 為 t_i 平移到 t_i^r 的節點向量，刪除 t_i 後的節點向量為 \mathbf{t}_d ，若 $R(\hat{f}(\mathbf{t}_r)) < R(\hat{f}(\mathbf{t}_d))$ ，則平移 t_i 到區間 $[t_i - 1, t_i + 1]$ 為 t_i^r ，若 $R(\hat{f}(\mathbf{t}_r)) > R(\hat{f}(\mathbf{t}_d))$ ，則刪除節點 t_i 。

在**增加節點**階段中若沒有任何的起始節點被增加，則先到**刪除平移**階段做平移或刪除為起始節點，重新回到**增加節點**階段，若依然無節點被增加，則代表起始節點不需要再做更改之程序。

因此以上階段可將起始節點直接尋找到合適的新節點配適模型(1)。

3.3 Bayesian Adaptive Regression Splines

在Bayesian adaptive regression splines(以下均簡稱BARS)估計B-Spline迴歸模型時，利用貝式的方法，將節點與其他的未知參數來自某一事前分配(prior distribution)，然後再使用Green[9]中reversible-jump Markov chain Monte Carlo(以下均簡稱reversible-jump MCMC)的方法對節點的事後分配(posterior distribution)抽樣，估計迴歸模型(1)。以下我們簡略的說明BARS事前分配的設定以及在抽樣時，BARS與BIC的關係：

$$\pi(k) = 1/\lambda \quad , \quad k = 1, \dots, \lambda$$

其中 λ 為常數，本文中設定為15. 而節點向量 $t = (t_1, \dots, t_k)$ 生成方式為隨機由解釋變數 x_1, \dots, x_n 抽取 k 個再排序得到 t_1, \dots, t_k ,

$$\pi_\sigma(\sigma) = 1/\sigma,$$

$$\beta \mid k, t, \sigma \sim N_{k+m+1}\{0, \sigma^2 n(B_{k,t}^T B_{k,t})^{-1}\},$$

$$\pi(\beta, k, t, \sigma) = \pi_\beta(\beta \mid t, k, \sigma) \pi_t(t \mid k) \pi_k(k) \pi_\sigma(\sigma),$$

由上述事前分配的設定之下，可求得 $p(y \mid \beta, k, t, \sigma)$ ，而邊際分配 $p(y \mid k, t)$ 可透過下述積分為：

$$p(y \mid k, t) = \int p(y \mid \beta, k, t, \sigma) \pi(\beta, \sigma \mid k, t) d\beta d\sigma.$$

我們的目的在於由節點的事後分配尋找到節點的個數與位置，因此使用reversible-jump MCMC的方式抽樣。在抽樣過程中，只需對節點向量做抽樣，而抽樣方式為將目前的節點向量 t 做更新得到新節點向量 t^c 作為下一次抽樣的樣本，或者是不更新節點向量而以 t 作為下一次抽樣的樣本。由 t 更新到 t^c 的機率受到 likelihood ratio = $\frac{p(y \mid k^c, t^c)}{p(y \mid k, t)}$ 的影響，likelihood ratio較大時，比較容易由 t 更新到 t^c 。

以下由likelihood ratio可以看出BARS與BIC的關係：

$$\begin{aligned} \text{likelihood ratio} &= \frac{p(y \mid k^c, t^c)}{p(y \mid k, t)} \\ &= \left(\frac{1}{\sqrt{n+1}} \right)^{k^c - k} \left(\frac{y^T \{I_n - n(n+1)^{-1} B_{k,t} (B_{k,t}^T B_{k,t})^{-1} B_{k,t}^T\} y}{y^T \{I_n - n(n+1)^{-1} B_{k^c, t^c} (B_{k^c, t^c}^T B_{k^c, t^c})^{-1} B_{k^c, t^c}^T\} y} \right)^{n/2} \\ &\simeq \left(\frac{1}{\sqrt{n}} \right)^{k^c - k} \left(\frac{(y - B_{k,t} \hat{\beta})^T (y - B_{k^c, t} \hat{\beta})}{(y - B_{k^c, t^c} \hat{\beta}^c)^T (y - B_{k, t^c} \hat{\beta}^c)} \right)^{n/2} = \exp \left(- \frac{(BIC_{k^c} - BIC_k)}{2} \right), \end{aligned}$$

由上式可知，當 BIC_{k^c} 較小時，likelihood ratio較大，因此節點向量 t 較容易移動到新節點向量 t^c 。

使用reversible-jump MCMC抽樣時，我們總共抽出 $N + N_0$ 組節點，令 t_l 為第 l 次抽出的節點向量，則可得到式(2)中 f 的最小平方估計 $\hat{f}(\cdot; t_l)$ ，最後 f 由 \bar{f} 所估計：

$$\bar{f}(x) = \sum_{l=N_0+1}^{N+N_0} \hat{f}(x; t_l) / N.$$

因為一開始抽樣所得到的樣本分佈尚未接近事後分配，所以我們將前 N_0 組所抽出的節點捨去，使用剩下的 N 組節點向量估計 f 。

3.4 基於BIC的節點估計方式

我們提出以BIC的標準來選取節點，其標準說明如下。給定正整數 k 時，令 S_k 為 (a, b) 區間上長度為 k 的起始節點向量所成的集合：

$$\begin{aligned} \hat{t}(k) &= \operatorname{argmin}_{t \in S_k} \sum_{i=1}^n (y_i - \hat{f}(x_i; t))^2, \\ RSS(\hat{t}(k)) &= \sum_{i=1}^n (y_i - \hat{f}(x_i; \hat{t}(k)))^2, \\ BIC(k) &= n \ln \left(\frac{RSS(\hat{t}(k))}{n} \right) + k \ln(n). \end{aligned}$$

其中 $k = 1, \dots, k^*$ ，取 $\hat{k} = \operatorname{argmin}_{1 \leq k \leq k^*} BIC(k)$ ，則 $\hat{t}(\hat{k})$ 為基於BIC標準選取的節點。

我們使用軟體R的optim函數提供的一種最佳化的方式解上述 $\hat{t}(k)$ ，而使用此種方式需要設定起始節點向量。以下將說明如何處理起始節點，而最佳化的方式將於附錄1說明。給定節點向量長度為 k 時，我們每次隨機在區間 (a, b) 上抽取 k 個節點作為一起始節點向量，並用軟體所解出 $\hat{t}(k)$ ，共進行 j^* 次。若第 \hat{j} 次解出的 $\hat{t}(k)$ 所對應的 $RSS(\hat{t}(k))$ 最小，則以第 \hat{j} 次解出的 $\hat{t}(k)$ 作為 $\hat{t}(k)$ 的解。¹

¹在第四章節模擬時，我們取 $k^* = 10, j^* = 5$ 。

4 模擬與比較

4.1 模擬

在第三章節使用了BARS、SARS與基於BIC的節點估計方式(以下我們簡稱BIC based)三種方法尋找B-Spline函數的節點。

本章節的模擬資料是由模型(1)所生成，其中 x_1, \dots, x_n 為區間 $[0,1]$ 中等距排放的 n 個點， n 設定為101，我們考慮8種不同 f 與 ε_i 的設定，詳見在附錄2，每一種設定下生出100組的模擬資料。令 MSE_j 為第 j 組資料透過BARS、SARS與BIC based的方法估計 f 所得到的均方差：

$$MSE_j = \sum_{i=1}^n (f(x_i) - \hat{f}_j(x_i; t))^2 / (n - 1), j = 1, \dots, 100$$

以下估計效果則是將此100筆資料由BARS、SARS與BIC based的方法所估計出的均方差取平均值：²

$$\sum_{j=1}^{100} MSE_j / 100$$

表 1: 平滑型函數之估計效果

| 選取節點方法 | BARS | BIC based | SARS |
|--------|--|--|--|
| 1 | 0.0092 (0.0005) | 0.0107(0.0007) | 0.0112(0.0007) |
| 2 | 0.6047(0.0838) | 0.6840(0.0945) | 0.3792 (0.0802) |
| 3 | 0.0221(0.0020) | 0.0225(0.0035) | 0.0131 (0.0015) |
| 4 | 1.30×10^{-6} (9.48×10^{-8}) | 1.49×10^{-6} (1.59×10^{-7}) | 1.13×10^{-6} (1.44×10^{-7}) |

²我們將例子分為兩類，一類為平滑型迴歸函數，另一類為間斷型迴歸函數

表 2: 間斷型函數之估計效果

| 選取節點方法 | BARS | BIC based | |
|--------|----------------|------------------------|----------------|
| 5 | 0.0490(0.0030) | 0.0398 (0.0036) | 0.1252(0.0102) |
| 6 | 0.1921(0.0215) | 0.1854 (0.0139) | 0.2705(0.0443) |
| 7 | 1.8330(0.1198) | 1.1571 (0.0983) | 1.6073(0.2812) |
| 8 | 0.9018(0.0567) | 0.8710 (0.0984) | 1.1217(0.1237) |

表 3: 尋找節點之模擬時間(單位:分)

| 選取節點方法 | BARS | SARS | BIC based |
|--------|------|------|-----------|
| 1 | 8 | 3 | 66 |
| 2 | 8 | 3 | 60 |
| 3 | 9 | 3 | 60 |
| 4 | 8 | 3 | 59 |
| 5 | 8 | 4 | 53 |
| 6 | 8 | 3 | 54 |
| 7 | 8 | 4 | 58 |
| 8 | 9 | 4 | 66 |

表 4: 平均選取節點個數

| 選取節點方法 | 真實迴歸函數 | BARS | SARS | BIC based |
|--------|--------|------|------|-----------|
| 1 | 無 | 6 | 7 | 6 |
| 2 | 3 | 2 | 1 | 2 |
| 3 | 2 | 2 | 1 | 2 |
| 4 | 無 | 3 | 1 | 2 |
| 5 | 5 | 4 | 6 | 5 |
| 6 | 6 | 5 | 7 | 5 |
| 7 | 6 | 4 | 7 | 5 |
| 8 | 6 | 3 | 7 | 4 |

4.2 比較

我們將模擬結果做出比較:

1. BARS的方法中，若為間斷型迴歸函數時，估計效果較無平滑型迴歸函數準確，原因在於BARS的方法在使用reversible-jump MCMC的抽樣過程中，將 N 組的估計函數取平均，在尖點或間斷點通常無法達到真實迴歸函數的圖形。如圖2中我們分別使用例子5與例子8為代表，實線為真實迴歸函數，虛線使用BARS的方法估計，發現在尖點或間斷點無法達到良好的表現。
2. SARS的方法中，若使用B-Spline函數所估計出的節點較少時，則SARS由式(3)所挑選出節點的估計效果較其他方法好。如圖3我們以例子2為代表，實線為真實迴歸函數，虛線分別使用 SARS、BARS與 BIC based所估計。
3. BIC based的方法中，若為間斷型迴歸函數時，配適出的效果較其他方法好。
4. 若考慮到模擬時間的因素時，因為BIC based將 k^* 組起始節點作更新尋

找到新節點，再使用BIC的標準挑選合適的函數估計，因此模擬時間比BARS與SARS長。

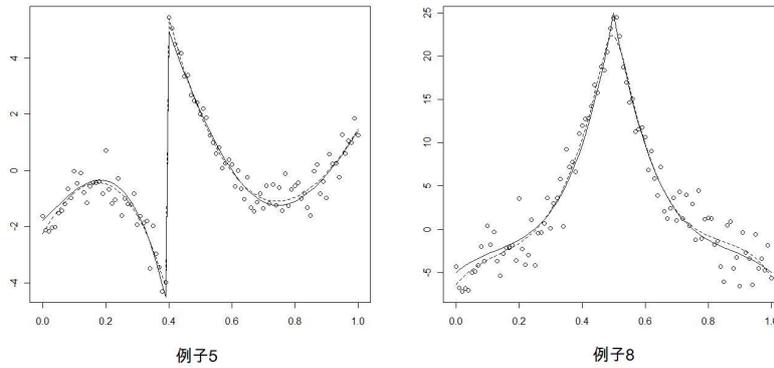


圖 2: BARS在間斷型迴歸函數之圖形

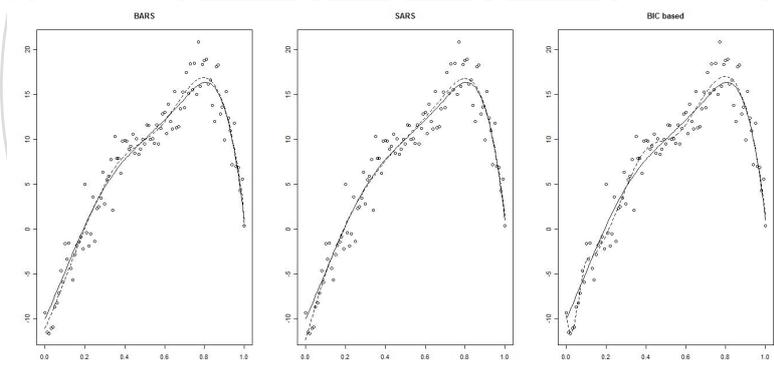


圖 3: 例子2中使用SARS、BARS與BIC-based之圖形

5 結論與建議

5.1 結論

在第四章節中我們將模擬出的結果做結論，若資料呈現為間斷型資料時，使用BIC based的方法估計較合適，若考量到時間因素之下，可考慮選擇BARS的方法，原因在於在尖點或間斷點需要選取較多的節點估計，雖然BARS的方法在間斷型資料較無平滑型資料準確，但BARS的方法也是使用BIC的標準選取節點，與SARS所使用式(3)的標準相比較不會有過度配適之現象。

若資料呈現為平滑型資料且所需的節點較少時，使用SARS估計較精準，若資料呈現為平滑型資料且所需的節點較多時，則BARS的估計效果較精準。

5.2 建議

在本文中我們提出了一種基於BIC的B-Spline節點估計方式，並且與 BARS跟 SARS的方法做比較，但這些方法中均存在優缺點，因此提出幾點建議作為未來研究方向：

1. BIC based的方法雖然在間斷型資料具有良好的表現，但同時也加長了模擬時間，因此若能提早確定節點個數，則可以縮短估計時間。
2. 在平滑型資料當中，可以考慮將節點等距排放，估計節點的個數，例如 Halpern[10]、Lindstorm[12]與 Ruppert[13]的文獻提出不同選取節點個數的方法，並且與 SARS、BARS與基於BIC的節點估計方式做比較做為未來研究方向。

附錄1

在這裡我們介紹起始節點如何由最佳化的方式尋找到 $\hat{t}(k)$ 。固定 k 值之下隨機抽取 j 組節點後，定義 $\mathbf{t}_l = (t_{1,j,k}, \dots, t_{k,j,k})_l$ 為第 l 次疊代之節點向量，若 $g(\mathbf{t})$ 定義為 $RSS(t(k))$ 的函數：

$$g(\mathbf{t}) = \sum_{i=1}^n (y_i - \hat{f}(x_i; \mathbf{t}))^2$$

利用泰勒展開式 $g(\mathbf{t}) \simeq g(\mathbf{t}_l) + \nabla g(\mathbf{t}_l)(\mathbf{t} - \mathbf{t}_l) + \frac{1}{2}(\mathbf{t} - \mathbf{t}_l)^T H_l(\mathbf{t} - \mathbf{t}_l)$ ，其中 $\nabla g(\mathbf{t})$ 為一階導數， $H = \nabla^2 g(\mathbf{t})$ 為函數的二階導數，稱之為海森矩陣：

$$H = \begin{pmatrix} \frac{\partial^2 g}{\partial t_{1jk}^2} & \cdots & \frac{\partial^2 g}{\partial t_{1jk} \partial t_{kjk}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 g}{\partial t_{kjk} \partial t_{1jk}} & \cdots & \frac{\partial^2 g}{\partial t_{kjk}^2} \end{pmatrix}$$

將以上的泰勒展開式做一階導數為0，目的在於一階導數能求得函數之局部極值， $\nabla g(\mathbf{t}) \simeq \nabla g(\mathbf{t}_l) + H_l(\mathbf{t} - \mathbf{t}_l) = 0$ ，且假設 $\mathbf{t} = \mathbf{t}_{l+1}$ ，可得：³

$$\mathbf{t}_{l+1} = \mathbf{t}_l - H_l^{-1} \nabla g(\mathbf{t}_l) \quad (6)$$

若 $A_l = H_l^{-1}$ ，在計算式(6)的過程中，使用Broyden[2]、Goldfarb[8]、Fletcher[6]與Shanno[14]四人所提出的BGSF法更新 A_{l+1} ：

$$A_{l+1} = A_l + \frac{s_l s_l^T}{s_l^T q_l} - \frac{A_l q_l q_l^T A_l}{q_l^T A_l q_l} + q_l^T A_l q_l w w^T \quad (7)$$

其中

$$w = \frac{s_l}{s_l^T q_l} - \frac{A_l q_l}{q_l^T A_l q_l}$$

$$q_l = \nabla g(\mathbf{t}_{l+1}) - \nabla g(\mathbf{t}_l)$$

$$s_l = \mathbf{t}_{l+1} - \mathbf{t}_l$$

³在式(6)中我們取 $l=100$ 為疊代總次數

而我們可將式(7)化簡為:

$$A_{l+1} = A_l + \frac{q_l q_l^T}{q_l^T s_l} - \frac{A_l s_l s_l^T A_l}{s_l^T A_l s_l}$$

值得注意的是, 在疊代過程中, 泰勒展開式的一階導數為0時:

$$-H^{-1} \nabla g(\mathbf{t}_l) = \mathbf{t} - \mathbf{t}_l$$

$$\Rightarrow \nabla g(\mathbf{t}_l) = -(\mathbf{t} - \mathbf{t}_l) H_l$$

$$\Rightarrow \nabla g(\mathbf{t}_l)(\mathbf{t} - \mathbf{t}_l) = -(\mathbf{t} - \mathbf{t}_l)^T H_l (\mathbf{t} - \mathbf{t}_l)$$

$$\Rightarrow g(\mathbf{t}) - g(\mathbf{t}_l) \simeq -\frac{1}{2}(\mathbf{t} - \mathbf{t}_l)^T H_l (\mathbf{t} - \mathbf{t}_l)$$

由以上的最後一個式子中, \mathbf{t} 為 \mathbf{t}_{l+1} 取代時, 若函數有局部極小值, $g(\mathbf{t}_{l+1})$ 必須小於 $g(\mathbf{t}_l)$, 則 $(\mathbf{t} - \mathbf{t}_l)^T H_l (\mathbf{t} - \mathbf{t}_l) > 0$, 因此海森矩陣必須為正定矩陣。

附錄2

第四章節中，使用8個不同例子比較各種選取節點的方法，而資料則是由真實迴歸函數加上一干擾變數 ε_i 產生：

$$y_i = f(x_i) + \varepsilon_i$$

例子1: 真實迴歸函數為 $f(x_i) = \sin(20 \times x_i)$ ，其中 $\varepsilon_i \sim N(0, 0.23^2)$

例子2: 真實迴歸函數為一B-Spline函數，節點為 $t = (0.3, 0.5, 0.7)$ ，係數為 $\beta = (-10, -5, 5, 10, 15, 20, 1)$ ，其中 $\varepsilon_i \sim N(0, 2.49^2)$

例子3: 真實迴歸函數為一B-Spline函數，節點為 $t = (0.5, 0.7)$ ，係數為 $\beta = (10, 8, 8, 10, 12, 2)$ ，其中 $\varepsilon_i \sim N(0, 0.48^2)$

例子4: 真實迴歸函數為一四次多項式 $f(x_i) = x_i^4 - 2 \times x_i^3 + 1.4 \times x_i^2 - 0.4 \times x_i + 0.0384$ ，其中 $\varepsilon_i \sim N(0, 0.003^2)$

例子5: 真實迴歸函數為一B-Spline函數，節點為 $t = (0.4, 0.4, 0.4, 0.4, 0.7)$ ，係數為 $\beta = (-1.78, -0.33, 1.62, -5, 5, 1.85, -3, -0.35, 1.39)$ ，其中 $\varepsilon_i \sim N(0, 0.57^2)$

例子6: 真實迴歸函數為一B-Spline函數，節點為 $t = (0.3, 0.3, 0.3, 0.3, 0.7, 0.8)$ ，係數為 $\beta = (2, 5, 4, 20, 4, 16, 10, 8, 5, 4)$ ，其中 $\varepsilon_i \sim N(0, 1.1^2)$

例子7: 真實迴歸函數為一B-Spline函數，節點為 $t = (0.1, 0.5, 0.5, 0.5, 0.5, 0.7)$ ，係數為 $\beta = (-10, -5, -2, 20, 1, 20, 25, 10, 1, -1)$ ，其中 $\varepsilon_i \sim N(0, 2.7^2)$

例子8: 真實迴歸函數為一B-Spline函數，節點為 $t = (0.2, 0.5, 0.5, 0.5, 0.5, 0.8)$ ，係數為 $\beta = (-5, -3, -2, 5, 25, 25, 5, -2, -3, -5)$ ，其中 $\varepsilon_i \sim N(0, 2.59^2)$

圖4為例子1中所產生的資料，實線為真實迴歸函數，虛線為B-Spline函數分別使用BARS、SARS與BIC based的方法選取節點估計。

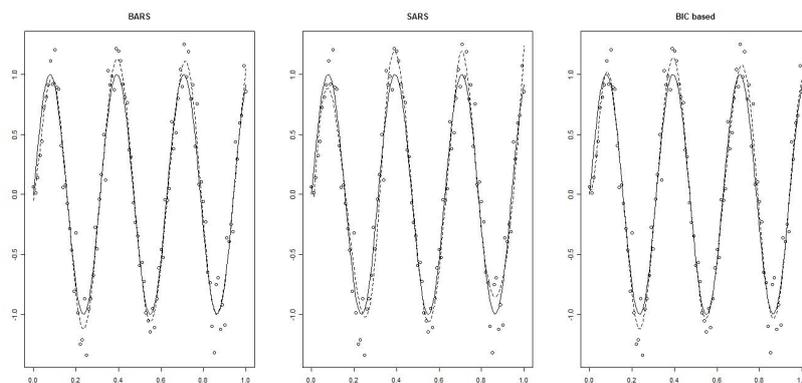


圖 4: 例子1:BARS、SARS與BIC based

圖5為例子2中所產生的資料，實線為真實迴歸函數，虛線為B-Spline函數分別使用BARS、SARS與BIC based的方法選取節點估計。

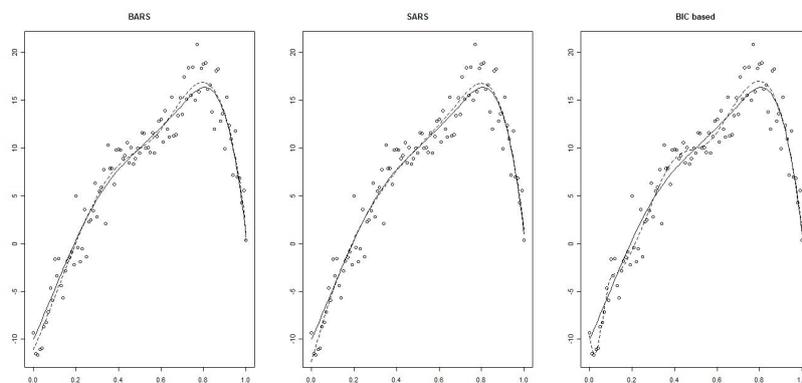


圖 5: 例子2:BARS、SARS與BIC based

圖6為例子3中所產生的資料，實線為真實迴歸函數，虛線為B-Spline函數分別使用BARS、SARS與BIC based的方法選取節點估計。

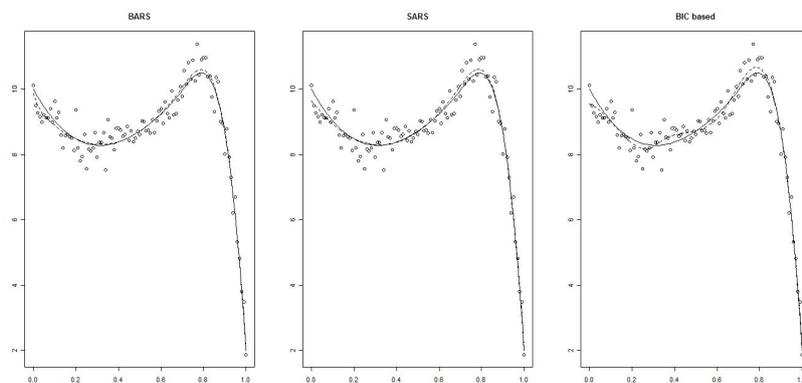


圖 6: 例子3:BARS、SARS與BIC based

圖7為例子4中所產生的資料，實線為真實迴歸函數，虛線為B-Spline函數分別使用BARS、SARS與BIC based的方法選取節點估計。

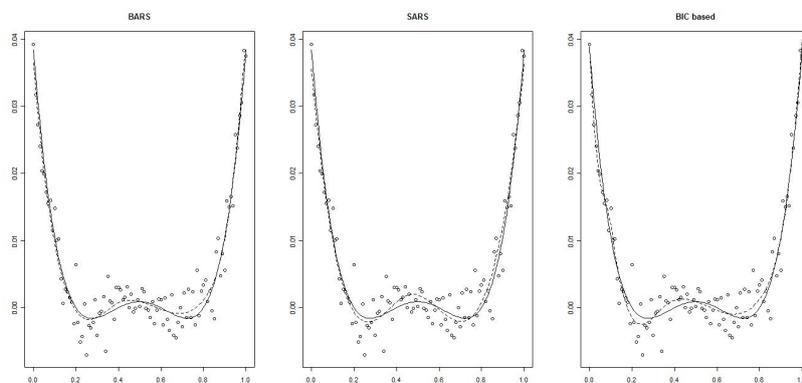


圖 7: 例子4:BARS、SARS與BIC based

圖8為例子5中所產生的資料，實線為真實迴歸函數，虛線為B-Spline函數分別使用BARS、SARS與BIC based的方法選取節點估計。

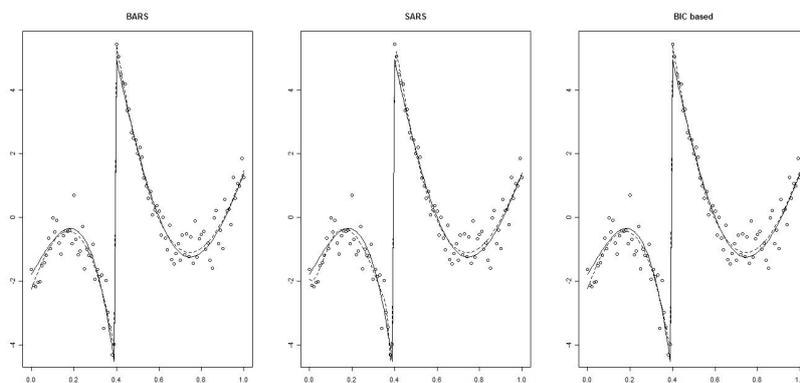


圖 8: 例子5:BARS、SARS與BIC based

圖9為例子6中所產生的資料，實線為真實迴歸函數，虛線為B-Spline函數分別使用BARS、SARS與BIC based的方法選取節點估計。

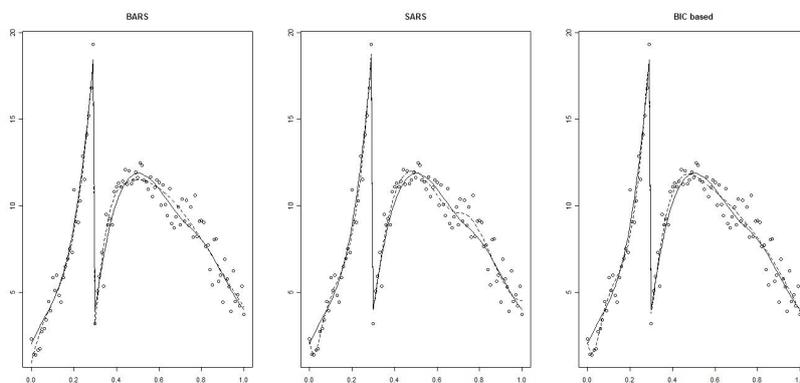


圖 9: 例子6:BARS、SARS與BIC based

圖10為例子7中所產生的資料，實線為真實迴歸函數，虛線為B-Spline函數分別使用BARS、SARS與BIC based的方法選取節點估計。

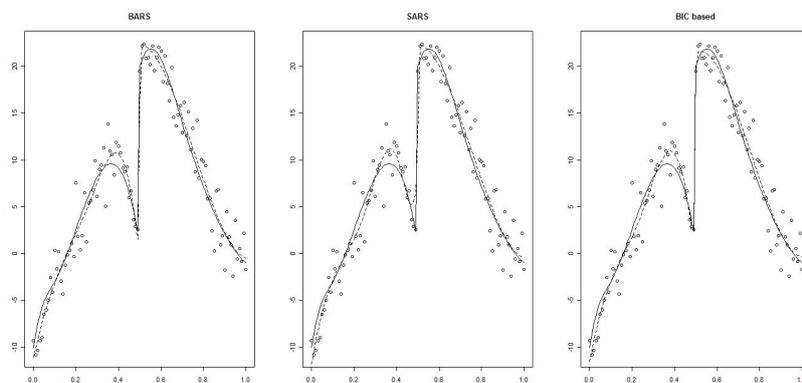


圖 10: 例子7:BARS、SARS與BIC based

圖11為例子8中所產生的資料，實線為真實迴歸函數，虛線為B-Spline函數分別使用BARS、SARS與BIC based的方法選取節點估計。

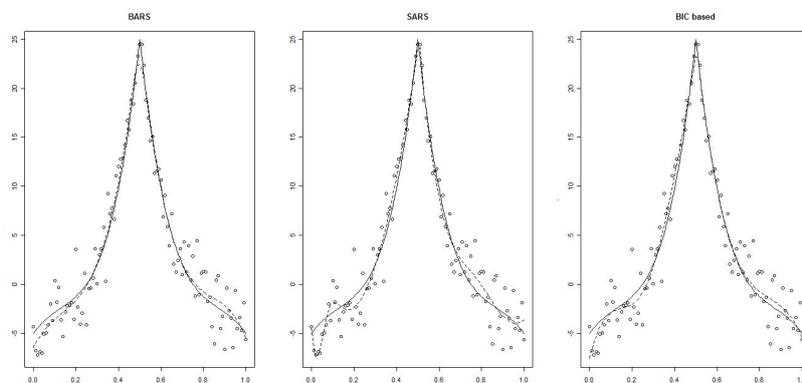


圖 11: 例子8:BARS、SARS與BIC based

References

- [1] C. Biller. Adaptive Bayesian regression splines in semiparametric generalized linear models. *Journal of Computational and Graphical Statistics*, 9:122–40, 2000.
- [2] C.G. Broyden. The convergence of a class of double-rank minimization algorithms. *Journal of the Institute of Mathematics and Its Applications*, (6):222–231, 1970.
- [3] C. de Boor. On calculating with B-Splines. *Journal of approximation theory*, (6):50–62, 1972.
- [4] D.G.T. Denison, B.K. Mallick, and A.F.M. Smith. Automatic Bayesian curve fitting. *Journal of Royal Statistical Society: Series B*, (60):333–350, 1998.
- [5] I. Dimatteo, C.R. Genovese, and R.E. Kass. Bayesian curve-fitting with free-knot splines. *Biometrika*, 88(4):1055–1071, 2001.
- [6] R. Fletcher. A new approach to variable metric algorithms. *Computer Journal*, 13(3):317–322, 1970.
- [7] J.H. Friedman. Multivariate adaptive regression splines. *The Annals of Statistics*, 19:1–141, 1991.
- [8] D. Goldfarb. A family of variable metric updates derived by variational means. *Mathematics of Computation*, 24(109):23–26, 1970.
- [9] P.J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–32, 1995.

- [10] E.F. Halpern. Bayesian spline regression when the number of knots is unknown. *Journal of Royal Statistical Society: Series B*, 35:347–60, 1973.
- [11] R.E. Kass and L. Wasserman. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of American Statistical Association*, 90:928–34, 1995.
- [12] M.J. Lindstrom. Penalized estimation of free-knot splines. *Journal of Computational and Graphical Statistics*, 8:333–52, 1999.
- [13] David Ruppert. Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, 11(4):735–757, 2002.
- [14] D.F. Shanno. Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computation*, 24(111):647–656, 1970.
- [15] C.M. Stein. Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, 9(6):1135–1151, 1981.
- [16] S.ZHOU and X.SHEN. Spatially adaptive regression splines and accurate knot selection schemes. *Journal of American Statistical Association*, 96(453):247–259, 2001.