

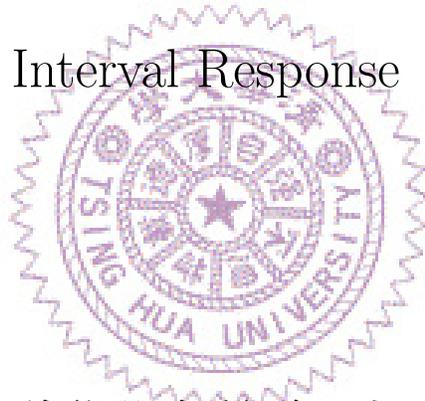
國立清華大學

碩士論文

題目：區間反應變數之預測與模型選取

Prediction and Model Selection

for Interval Response



所 別 : 統計學研究所 組別:工業統計組

指導教授 : 鄭少為(Shao-Wei Cheng) 博士

姓 名 : 張明中(Ming-Chung Chang)

學 號 : 9524503

中華民國九十七年六月

摘要

在迴歸模型中, 反應變數形式除了常見的一個值之外, 還有可能會是一個區間。我們稱前者為單一反應變數, 後者為區間反應變數。在單一反應變數的情形下, 不論在線性模型亦或非線性模型, 都已經有了相當多的研究結果, 像是參數估計、反應值的預測、模型選取等等。而當反應值為區間反應變數時, 處理的方式不外乎是將之視為兩個單一反應變數, 然後各自處理。但如此並未考慮到區間的特性, 有可能喪失區間特有的資訊。以模型選取為例, 在單一反應變數中有不少準則供我們使用, 像是 C_p 、 $PRESS$ 等等。對區間反應變數做模型選取時, 如果將區間的上下界, 或是區間中心與區間長度這些一一對應的量各自建立迴歸模型後, 再各自利用 C_p 或 $PRESS$ 等準則來做模型選取, 則會忽略了區間特有的資訊, 而可能產生不恰當的結論。本文在決策理論的架構下, 研究當反應變數為區間時, 該如何做模型選取以及區間預測才是最適當的。

關鍵字: 區間反應變數, 模型選取, 決策理論, 貝氏預測函數, 區間預測。

致謝辭

研究所求學的兩年時光一下子就過去了，在清華待的日子雖然不長，但在這段時間留下的回憶卻比之前任何時期要深刻。因為在這裡，我有一群很棒的同學，也有很棒的老師。

首先要感謝我的指導教授鄭少為老師。老師是『教』我們怎麼做論文，而不是『幫』我們做論文。每次討論研究進度的時候，他都會適當的給一些建議。如果研究遇到瓶頸，老師也會點出一些可行的方向供你參考，但下一步要怎麼做完全靠自己的思考與努力。這樣的方式讓我真的有在做研究的感覺，而不只是像個機器人一般在接收指令。因此當我論文的雛形出來後，心中的歡喜真是不可言喻。在我跟著老師的這一年，可以說是來到清華後收穫最多的一年。做研究跟平常上課考試真的差很多，一個有制式的答案在那，一個卻似一片完全未探索過的迷霧，需要費盡心思找出一條路。而我的老闆充分的讓我體會到這兩者的差異，如果說我來到清華有學到什麼，無疑的就是這個求學態度與精神。

接著我要謝謝我在這裡認識的好朋友們，也可以說是球友。原本以為研究所兩年應該不太會打球了，但沒想到打球的頻率反而比大學多出很多。感謝小民、問賢、神龍、Ginobili、BigWu、大鵬、安國，陪我在球場上飆汗，讓我知道跟校隊一起打全場是什麼感覺。而我們去年拿下了統研盃籃球亞軍，沒有奪冠雖然可惜，但能跟你們一起享受比賽的感覺，真的很棒！

接著我要謝謝我的家人，他們在背後辛苦賺錢讓我念書，卻從來不在我面前流露過任何疲累的表情。每當我要去火車站搭車時，他們也不厭其煩的接送。爸、媽，真的很謝謝你們，沒有你們的辛苦我就不會有現在的成就。

最後要謝謝我的寶貝雅馨，我們從大學至今一起度過了5個多年頭。每當我心情不好、研究遇到挫折，都會得到你滿滿的關懷。你的溫柔與貼心真的是我上輩子修到的福氣，有你的陪伴，真的很幸福！

目錄

1	緒論與文獻回顧	1
2	統計建模	5
2.1	模型一	6
2.2	模型二	7
2.3	模型三	8
2.4	貝氏觀點	11
3	模型選取與區間預測	12
3.1	區間預測	12
3.2	模型選取準則	17
3.3	矩陣 t 分配的應用	18
4	套用區間距離函數	22
4.1	第一種區間距離函數	22
4.1.1	區間預測	24
4.1.2	模型選取準則	26
4.2	第二種區間距離函數	29
4.2.1	區間預測與模型選取準則	33
5	實例分析	35
5.1	選模準則: <i>PRESS</i>	37
5.2	選模準則: 最小貝氏預測風險函數	38
5.2.1	第一種區間距離函數	39

5.2.2	第二種區間距離函數	45
5.2.3	兩種區間距離函數的比較	50
6	結論	52
	參考文獻	53



圖目錄

1.1 電流強度區間	2
4.1 第二種區間距離函數	30
4.2 第二種區間距離函數將上半平面的切割	32
5.1 PRESS與觀測區間之比較	38
5.2 距離一, $k=0$, 與觀測區間之比較	40
5.3 距離一, $k=0$, 與PRESS之比較	40
5.4 距離一, $k=1$, 與觀測區間之比較	41
5.5 距離一, $k=1$, 與PRESS之比較	42
5.6 距離一, $k=100$, 與觀測區間之比較	43
5.7 距離一, $k=100$, 與PRESS之比較	44
5.8 距離二, $k=0$, 與觀測區間之比較	46
5.9 距離二, $k=0$, 與PRESS之比較	46
5.10 距離二, $k=1$, 與觀測區間之比較	47
5.11 距離二, $k=1$, 與PRESS之比較	48
5.12 距離二, $k=100$, 與觀測區間之比較	49
5.13 距離二, $k=100$, 與PRESS之比較	49
5.14 $k=0$, 兩種距離的比較	50
5.15 $k=1$, 兩種距離的比較	51
5.16 $k=100$, 兩種距離的比較	51

表目錄

5.1	實驗因子描述	35
5.2	設計矩陣與觀測值	36
5.3	<i>PRESS</i> 選出的模型	37
5.4	距離一, $k=0$ 選出的模型 M_0^*	39
5.5	距離一, $k=1$ 選出的模型 M_1^*	41
5.6	距離一, $k=100$ 選出的模型 M_{100}^*	42
5.7	距離二, $k=0$ 選出的模型 M_0^*	45
5.8	距離二, $k=1$ 選出的模型 M_1^*	47
5.9	距離二, $k=100$ 選出的模型 M_{100}^*	48



第 1 章

緒論與文獻回顧

在迴歸模型中，反應變數形式除了常見的一個值之外，還有可能會是一個區間，我們稱前者為單一反應變數 (single response)，後者為區間反應變數 (interval response)。在單一反應變數的情形下，不論是在線性模型亦或非線性模型，都已經有了相當多的研究結果，像是參數估計、反應值的預測、模型選取等等。而反應值為區間的實驗較不常見，在Li et al. (2001)的論文中談到了一個焊接實驗，此實驗的目的在找出合適的電流強度以產生良好的焊接點，此時的電流強度範圍就是一個區間。然而電流強度太低會使得焊接點無法產生，而電流強度太高會使焊接點產生噴濺 (expulsion)，此現象可由圖 1.1 來展示。圖 1.1 展示了電流強度變化的情形，由該圖可看到當其他因子設定固定時，隨著焊接時間長短的不同，合適的電流強度範圍也不一樣，因而可以找到兩條曲線，將焊接時間與電流強度形成的二維平面分成『不足以產生焊接點』、『良好的焊接點』與『焊接點噴濺』這三個區域。由於生產良好的焊接點是我們所希望的，因此找出合適的電流強度區間就是一個很重要的步驟。然而受限於經費、資源等限制，不可能把每種因子設定下合適的電流區間都找出來，故對電流區間與因子效應來建立模型，以及對未實驗過的電流區間做預測是必要的。在這種情形下，很自然的我們便需處理區間反應變數的數據。

在Li et al. (2001)中的作法是對區間中點與 $\ln(\text{區間半長})$ 分別建立線性迴歸模型，再將之視為兩個單一反應變數來作分析。但如此並未考慮到區間的特性，有可能喪失區間特有的資訊。以模型選取為例，在單一反應變數中有不少準則供我們使用，像是 C_p 、PRESS(predicted sum of square) 等等。對區間反應變數做模型選取時往往是將區間的上下界，或是區間中心與區間長度這些一一對應的量來各自建立迴歸模

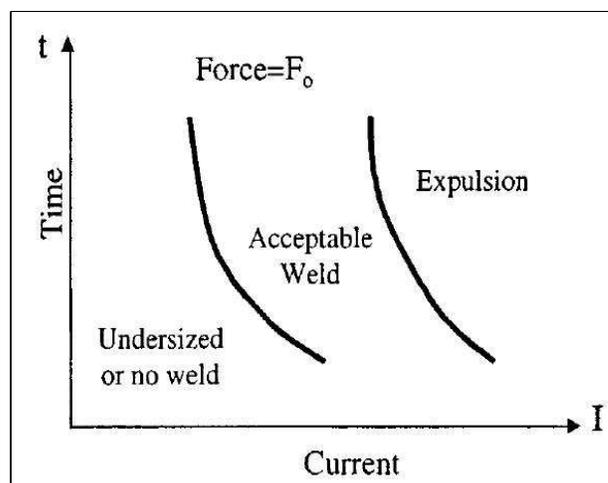


圖 1.1: 電流強度區間

型後,再各自利用 C_p 或 $PRESS$ 等準則來做模型選取。但如此會忽略了區間特有的資訊,而可能產生不恰當的結論。

本文在決策理論的架構下,研究當反應變數為區間時,該如何做模型選取以及區間預測才是最適當的。使用決策理論是因我們將區間之間的差異視為一種損失的概念,這恰好與該理論中的損失函數 (loss function) 相符。且若我們能對一般的損失函數發展出模型選取與區間預測的方法的話,則在不同的實驗要求下,只需套用符合該目的的損失函數即可。下段文獻回顧的部分,介紹了針對單一反應變數的模型選取準則,以及矩陣 t 分配 (matrix t distribution) 的定義與一些性質。接著在第2章提出了三種用來描述區間反應變數的模型。第3章探討在第2章的第一種模型假設之下,該如何做區間預測以及模型選取。第4章將我們定義的兩種區間距離函數套用在第3章的結果上。第5章用了Li et al. (2001)中的一個實驗數據來比較本文提出的方法與之前的作法。第6章是結論。

在Gelfand and Ghosh (1998)的文章中提到單一反應變數值之下,如何利用決策理論的概念做貝氏模型選取。若該實驗在因子水準組合(level combination) 為 x_1, x_2, \dots, x_n (x_j 為一向量) 下各獲得實驗值 $y_{x_1}, y_{x_2}, \dots, y_{x_n}$, 令 $\psi_j(x_k)$ 表示在因子設定 x_k 下第 j 個效應,

如主效應、二階交互作用等等。則此數據的模型可表示為：

$$\begin{pmatrix} y_{x_1} \\ y_{x_2} \\ \vdots \\ y_{x_n} \end{pmatrix} = \begin{pmatrix} \beta_0 + \sum_{j=1}^p \beta_{1j} \psi_j(x_1) \\ \beta_0 + \sum_{j=1}^p \beta_{2j} \psi_j(x_2) \\ \vdots \\ \beta_0 + \sum_{j=1}^p \beta_{nj} \psi_j(x_n) \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix},$$

其中

$$\begin{pmatrix} \epsilon_1 & \dots & \epsilon_n \end{pmatrix}' | \beta, \sigma^2 \sim N_n(0, \sigma^2 I_n), \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \sim N_{p+1}(\mu_\beta, \Sigma_\beta),$$

I_n 為 $n \times n$ 的單位矩陣。此處 Σ_β 只須為正半定矩陣 (positive semidefinite matrix) 即可。該篇文章定義了如下的損失函數

$$L(y_{rep,j}, a_j; y_{x_j}) = (a_j - y_{rep,j})^2 + k(a_j - y_{x_j})^2, k \geq 0.$$

y_{x_j} 為 x_j 這個因子設定下得到的觀測值, $y_{rep,j}$ 為未來再做一次相同的實驗得到的觀測值, a_j 是 x_j 下反應變數的預測值。比較特別的是該篇作者將 $L(y_{rep,j}, a_j; y_{x_j})$ 定義成

$$(a_j - y_{rep,j})^2 + k(a_j - y_{x_j})^2,$$

由於 $y_{rep,j}$ 是未來做實驗才會得到的觀測值, 故此時尚未有數據, 因此 $(a_j - y_{rep,j})^2$ 可看成用 a_j 預測 $y_{rep,j}$ 的差異。而 y_{x_j} 是已經觀測到的數據, 因此 $(a_j - y_{x_j})^2$ 可看成用 a_j 擬和 y_{x_j} 的程度。也就是將損失函數定為『預測』和『擬和』這兩個部分的和, 並用 k 來呈現兩著的重要性。

Gelfand and Ghosh (1998) 也提到當 $\Sigma_\beta^{-1} \approx 0$,

$$y_{rep} | y_{obs} \approx N_n(X(X'X)^{-1}X'y_{obs}, (I_n + X(X'X)^{-1}X')\sigma^2),$$

其中

$$y_{rep} = \begin{pmatrix} y_{rep,1} & y_{rep,2} & \dots & y_{rep,n} \end{pmatrix}'$$

$$y_{obs} = \begin{pmatrix} y_{x_1} & y_{x_2} & \dots & y_{x_n} \end{pmatrix}'$$

在如此的假設下就可以推導出模型選取的準則為：

找一子模型使得 $SSE + \sigma^2(n + p + 1)$ 達到最小。

接著我們將一個在後續推導上會用到的分配，矩陣 t 分配，的定義與兩個重要的定理，陳述如下：

定義 1. 一隨機矩陣 $T(p \times m)$ 服從矩陣 t 分配 $t_{p \times m}(M, \Sigma, \Omega, \nu)$ ，若且為若該機率密度函數為

$$\frac{\Gamma_p[\frac{1}{2}(n + m + p - 1)]}{\pi^{\frac{1}{2}mp} \Gamma_p[\frac{1}{2}(n + p - 1)]} \det(\Sigma)^{-\frac{1}{2}m} \det(\Omega)^{-\frac{1}{2}p} \det(I_p + \Sigma^{-1}(T - M)\Omega^{-1}(T - M)')^{-\frac{1}{2}(n+m+p-1)}$$

其中 $T \in R^{p \times m}$, $M \in R^{p \times m}$, $\Omega(m \times m)$ 和 $\Sigma(p \times p)$ 為正定矩陣, I_p 為 $p \times p$ 的單位矩陣。

定理 1. 若 $T \sim t_{p \times m}(M, \Sigma, \Omega, \nu)$ ，則

$$E(T) = M$$

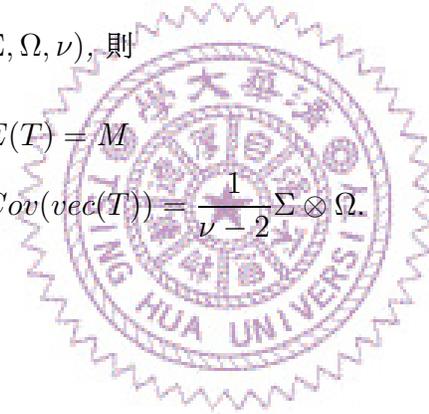
$$Cov(\text{vec}(T)) = \frac{1}{\nu - 2} \Sigma \otimes \Omega.$$

其中 \otimes 是 *Kronecker* 乘積。

定理 2. 當 $\nu \rightarrow \infty$ 時，

$$t_{p \times m}(M, \Sigma, \Omega, \nu) \rightarrow N_{pm}(\text{vec}(M), \Sigma \otimes \Omega).$$

更詳細的介紹可以參考 Gupta and Nagar (2000)。



第 2 章

統計建模

實驗會得到數據，而數據可視為由其背後某個隨機變數所產生的。是隨機變數就可以用機率模型來描述它。本章將討論針對區間反應變數的幾種可行的統計模型。

在一區間反應變數的實驗中，得到的數據會是區間的端點。對於在因子的水準組合為 x (x 為一向量) 下作實驗獲得的區間，我們可用

$$\mathcal{I}_x = [u_x, v_x]$$

來代表它，其中 u_x 是該區間之下界，而 v_x 為其上界，故 $u_x < v_x$ 。這時建模常用的做法是分別對 u_x 、 v_x 各自建立線性迴歸模型，即

$$\begin{aligned} u_x &= \Psi(x)\beta + \epsilon \\ v_x &= \Psi(x)\gamma + \epsilon \end{aligned} \tag{2.1}$$

$\Psi(x)$ 表示因子設定在 x 下各種因子效應(effect)所形成的向量。若總共有 p 個效應，並以 $\psi_j(x)$ 表示第 j 個效應，如主效應、二階交互作用等等，則

$$\Psi(x) = \begin{pmatrix} 1 & \psi_1(x) & \psi_2(x) & \dots & \psi_p(x) \end{pmatrix}$$

不過這種做法可能會產生作區間預測時下界比上界大，即 $\hat{u}_x > \hat{v}_x$ 的不合理情形。

為了避免這種情況，我們有幾種做法可以考慮。第一種是在 (2.1) 式的模型中加入 $u_x < v_x$ 的限制式，這樣所得到的 \hat{u}_x 就不會超過 \hat{v}_x 。但如此會增加參數估計的複雜度，造成計算上的困難。第二種做法則是注意到一個區間除了可用上下端點來描述之

外, 也可以用區間中心點和區間長度來描述, 並且它們是一對一的函數變換。故可令

$$c_x = \frac{u_x + v_x}{2}, c_x \in R$$

$$l_x = \frac{v_x - u_x}{2}, l_x \in R^+.$$

轉換完後, 再針對 c_x 和 l_x 各別建立線性迴歸模型。這時我們會發現 $\hat{u}_x > \hat{v}_x$ 取決於 \hat{l}_x 是否大於0。因此還需要對 l_x 的模型作一點調整。一種可行的做法是將 l_x 模型中 $E(l_x)$ 參數的部分改為非線性的指數函數, 另一種則是對 l_x 取自然對數, 即 $z_x = \ln(l_x)$, 得到

$$c_x = \frac{u_x + v_x}{2}, c_x \in R$$

$$z_x = \ln\left(\frac{v_x - u_x}{2}\right), z_x \in R. \quad (2.2)$$

再針對 c_x 和 z_x 各別建立線性迴歸模型。這些做法各有優缺點, 會在下節詳細說明。

2.1 模型一

在穩健參數設計 (robust parameter design) 中, 有一種建模方式是分別對相同的因子設定下獲得之反應值的平均數和 $\ln(\text{變異數})$ 來建立模型。將變異數取對數的原因有很多, 其中最主要的就是對數函數可將一個恆正的數映至整個實數軸上, 這時若對 $\ln(\text{變異數})$ 建立線性迴歸模型, 則變異數的預測值必定恆正。關於其他的原因可參見Wu and Hamada (2000)。

由於區間長度必定大於0, 其角色與穩健參數設計中的變異數有類似之處, 因此這裡可用類似的方式, 模仿穩健參數設計中的建模方式, 以區間中心 c_x 、區間半長 l_x 來取代平均數、變異數的角色來建立模型。

令區間反應變數 $\mathcal{I}_x = \begin{pmatrix} c_x & z_x \end{pmatrix}$, c_x 表示區間中心值, z_x 表示 $\ln(l_x)$, 並假設兩者獨立地服從二元常態, 則我們可以建立以下的迴歸模型:

$$\begin{pmatrix} c_x \\ z_x \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \mu_{c,x} \\ \mu_{z,x} \end{pmatrix}, \begin{pmatrix} \sigma_c^2 & 0 \\ 0 & \sigma_z^2 \end{pmatrix} \right)$$

其中

$$\mu_{c,x} = E(c_x) = \beta_0 + \sum_{j=1}^p \beta_j \psi_j(x), \text{Var}(c_x) = \sigma_c^2$$

$$\mu_{z,x} = E(z_x) = \gamma_0 + \sum_{j=1}^p \gamma_j \psi_j(x), \text{Var}(z_x) = \sigma_z^2$$

而 σ_c^2 和 σ_z^2 為固定常數, 不隨著 x 變動, 且在 $\mu_{c,x}$ 和 $\mu_{z,x}$ 中各有 $p+1$ 個參數。若該實驗在因子設定為 x_1, x_2, \dots, x_n 下各獲得區間實驗值 $(c_{x_1}, z_{x_1}), (c_{x_2}, z_{x_2}), \dots, (c_{x_n}, z_{x_n})$, 則此數據的模型可表示為:

$$\begin{pmatrix} c_{x_1} & z_{x_1} \\ c_{x_2} & z_{x_2} \\ \vdots & \vdots \\ c_{x_n} & z_{x_n} \end{pmatrix} = \begin{pmatrix} \beta_0 + \sum_{j=1}^p \beta_{1j} \psi_j(x_1) & \gamma_0 + \sum_{j=1}^p \gamma_{1j} \psi_j(x_1) \\ \beta_0 + \sum_{j=1}^p \beta_{2j} \psi_j(x_2) & \gamma_0 + \sum_{j=1}^p \gamma_{2j} \psi_j(x_2) \\ \vdots & \vdots \\ \beta_0 + \sum_{j=1}^p \beta_{nj} \psi_j(x_n) & \gamma_0 + \sum_{j=1}^p \gamma_{nj} \psi_j(x_n) \end{pmatrix} + \begin{pmatrix} \epsilon_{c,1} & \epsilon_{z,1} \\ \epsilon_{c,2} & \epsilon_{z,2} \\ \vdots & \vdots \\ \epsilon_{c,n} & \epsilon_{z,n} \end{pmatrix},$$

其中

$$\begin{pmatrix} \epsilon_{c,1} & \dots & \epsilon_{c,n} & \epsilon_{z,1} & \dots & \epsilon_{z,n} \end{pmatrix}' \sim N_{2n} \left(0, \begin{pmatrix} \sigma_c^2 I_n & 0 \\ 0 & \sigma_z^2 I_n \end{pmatrix} \right).$$

I_n 為 $n \times n$ 的單位矩陣。這樣的模型由於背後的機率分配為常態分佈, 在推導的過程上有很多好的性質可以運用, 而且不必擔心預測出來的區間長度會有小於0的情形。缺點則是在兩區間損失函數的計算上比較麻煩, 關於此點第5章會有介紹。

2.2 模型二

使用第一個模型在計算兩區間的損失函數上會比較複雜, 就是因為區間反應變數長度的部分為對數的形式。因此在第二個模型中, 我們讓中心值的部分仍維持模型一的假設, 但長度的部分直接採用區間半長, 而非其自然對數之轉換。但為了讓預測的區間恆正, 我們把區間半長中期望值的結構轉為指數函數的形式。

令區間反應變數 $\mathcal{I}_x = \begin{pmatrix} c_x & l_x \end{pmatrix}$, c_x 表示區間中心值, l_x 表示區間半長值, 並假設兩者獨立地服從二元常態, 則我們可以建立以下的迴歸模型:

$$\begin{pmatrix} c_x \\ l_x \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \mu_{c,x} \\ \mu_{l,x} \end{pmatrix}, \begin{pmatrix} \sigma_c^2 & 0 \\ 0 & \sigma_l^2 \end{pmatrix} \right),$$

其中

$$\mu_{c_x} = E(c_x) = \beta_0 + \sum_{j=1}^p \beta_j \psi_j(x), \text{Var}(c_x) = \sigma_c^2$$

$$\mu_{l_x} = E(l_x) = \gamma_0 + \exp\left(\sum_{j=1}^p \gamma_j \psi_j(x)\right), \text{Var}(z_x) = \sigma_l^2$$

而 σ_c^2 和 σ_l^2 為固定常數，不隨著 x 變動，且在 $\mu_{c,x}$ 和 $\mu_{l,x}$ 中各有 $p+1$ 個參數。若該實驗在因子設定為 x_1, x_2, \dots, x_n 下各獲得區間實驗值 $(c_{x_1}, l_{x_1}), (c_{x_2}, l_{x_2}), \dots, (c_{x_n}, l_{x_n})$ ，則此數據的模型可表示為：

$$\begin{pmatrix} c_{x_1} & l_{x_1} \\ c_{x_2} & l_{x_2} \\ \vdots & \vdots \\ c_{x_n} & l_{x_n} \end{pmatrix} = \begin{pmatrix} \beta_0 + \sum_{j=1}^p \beta_{1j} \psi_j(x_1) & \exp(\gamma_0 + \sum_{j=1}^p \gamma_{1j} \psi_j(x_1)) \\ \beta_0 + \sum_{j=1}^p \beta_{2j} \psi_j(x_2) & \exp(\gamma_0 + \sum_{j=1}^p \gamma_{2j} \psi_j(x_2)) \\ \vdots & \vdots \\ \beta_0 + \sum_{j=1}^p \beta_{nj} \psi_j(x_n) & \exp(\gamma_0 + \sum_{j=1}^p \gamma_{nj} \psi_j(x_n)) \end{pmatrix} + \begin{pmatrix} \epsilon_{c,1} & \epsilon_{l,1} \\ \epsilon_{c,2} & \epsilon_{l,2} \\ \vdots & \vdots \\ \epsilon_{c,n} & \epsilon_{l,n} \end{pmatrix},$$

其中

$$\begin{pmatrix} \epsilon_{c,1} & \dots & \epsilon_{c,n} & \epsilon_{l,1} & \dots & \epsilon_{l,n} \end{pmatrix}' \sim N_{2n} \left(0, \begin{pmatrix} \sigma_c^2 I_n & 0 \\ 0 & \sigma_l^2 I_n \end{pmatrix} \right).$$

D_n 為 $n \times n$ 的單位矩陣。這樣的模型由於背後的機率分配為常態分佈，在推導的過程上有很多好的性質可以運用，且在損失函數上推導出來的形式較簡單。缺點則是此模型為了讓預測出來的區間長度恆正， $\mu_{l,x}$ 的部分為一非線性函數，因此在計算上增加不少困難度。

2.3 模型三

模型三對區間反應變數的假設更一般化，其作了以下兩假設：

1. 令 R_x 與 S_x 為在因子設定 x 之下，兩獨立服從 $N(\mu_x, \sigma_x^2)$ 的隨機變數。
2. 由於 $\max(R_x, S_x) > \min(R_x, S_x)$ ，故取區間的上界為 $v_x = \max(R_x, S_x)$ ，下界為 $u_x = \min(R_x, S_x)$ 。

在此假設下, u_x 和 v_x 的聯合機率分配為:

$$f_{u_x, v_x}(u, v) = \frac{1}{\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2}((u - \mu_x)^2 + (v - \mu_x)^2)\right), -\infty < u < v < \infty.$$

此時若透過 (2.2) 式之轉換可得

$$\begin{aligned} u_x &= c_x - l_x \\ v_x &= c_x + l_x \end{aligned} \Rightarrow |J| = \left| \det \begin{pmatrix} \frac{\partial u_x}{\partial c_x} & \frac{\partial u_x}{\partial l_x} \\ \frac{\partial v_x}{\partial c_x} & \frac{\partial v_x}{\partial l_x} \end{pmatrix} \right| = \left| \det \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \right| = 2.$$

因此 c_x 與 l_x 的聯合機率分配為:

$$f_{c_x, l_x}(s, t) = \frac{2}{\pi\sigma_x^2} \exp\left(-\frac{1}{2\sigma_x^2}((s - t - \mu_x)^2 + (s + t - \mu_x)^2)\right), s \in R, t \in R^+.$$

由於

$$(s - t - \mu_x)^2 + (s + t - \mu_x)^2 = 2(s - \mu_x)^2 + 2t^2,$$

故可化簡為:

$$f_{c_x, l_x}(s, t) = \frac{2}{\pi\sigma_x^2} \exp\left(-\frac{1}{\sigma_x^2}((s - \mu_x)^2 + t^2)\right), s \in R, t \in R^+.$$

由上可看出 c_x 和 l_x 為互相獨立的隨機變數, 故 c_x 的密度函數為:

$$f_{c_x}(s) = \frac{1}{\sigma_x\sqrt{\pi}} \exp\left(-\frac{1}{\sigma_x^2}(s - \mu_x)^2\right), s \in R.$$

因此區間中心 c_x 服從常態分配 $N(\mu_x, \frac{\sigma_x^2}{2})$ 。若我們將 c_x 的期望值與因子效應以線性結構連結起來, 則可將其表示為:

$$E(c_x) = \mu_x = \beta_0 + \sum_{j=1}^p \beta_j \psi_j(x).$$

而 l_x 的密度函數為:

$$f_{l_x}(t) = \frac{2}{\sigma_x\sqrt{\pi}} \exp\left(-\frac{t^2}{\sigma_x^2}\right), t \in R^+,$$

且模仿模型二的方式, 將 l_x 的期望值以指數函數的結構連結起來, 則可將其表示為:

$$E(l_x) = \frac{\sigma_x}{\sqrt{\pi}} = \frac{1}{\sqrt{\pi}} \exp\left(\gamma_0 + \sum_{j=1}^p \gamma_j \psi_j(x)\right),$$

如此可使 \hat{l}_x 恆正。由上可知 μ_x 和 σ_x 中各有 $p+1$ 個參數。若該實驗在因子設定為 x_1, x_2, \dots, x_n 下各獲得區間實驗值 $(c_{x_1}, l_{x_1}), (c_{x_2}, l_{x_2}), \dots, (c_{x_n}, l_{x_n})$ ，則此數據的模型可表示為：

$$\begin{pmatrix} c_{x_1} & l_{x_1} \\ c_{x_2} & l_{x_2} \\ \vdots & \vdots \\ c_{x_n} & l_{x_n} \end{pmatrix} = \begin{pmatrix} \beta_0 + \sum_{j=1}^p \beta_{1j} \psi_j(x_1) & \frac{1}{\sqrt{\pi}} \exp(\gamma_0 + \sum_{j=1}^p \gamma_{1j} \psi_j(x_1)) \\ \beta_0 + \sum_{j=1}^p \beta_{2j} \psi_j(x_2) & \frac{1}{\sqrt{\pi}} \exp(\gamma_0 + \sum_{j=1}^p \gamma_{2j} \psi_j(x_2)) \\ \vdots & \vdots \\ \beta_0 + \sum_{j=1}^p \beta_{nj} \psi_j(x_n) & \frac{1}{\sqrt{\pi}} \exp(\gamma_0 + \sum_{j=1}^p \gamma_{nj} \psi_j(x_n)) \end{pmatrix} + \begin{pmatrix} \epsilon_{c,1} & \epsilon_{l,1} \\ \epsilon_{c,2} & \epsilon_{l,2} \\ \vdots & \vdots \\ \epsilon_{c,n} & \epsilon_{l,n} \end{pmatrix},$$

其中

$$\begin{pmatrix} \epsilon_{c,1} \\ \epsilon_{c,2} \\ \vdots \\ \epsilon_{c,n} \end{pmatrix} \sim N_n \left(0, \begin{pmatrix} \frac{\sigma_{x_1}^2}{2} & 0 & \dots & 0 \\ 0 & \frac{\sigma_{x_2}^2}{2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\sigma_{x_n}^2}{2} \end{pmatrix} \right),$$

$\epsilon_{l,1}, \epsilon_{l,2}, \dots, \epsilon_{l,n}$ 為獨立且具有相同機率密度函數

$$g(t) = \frac{2}{\sigma_x \sqrt{\pi}} \exp\left(-\frac{1}{\sigma_x^2} \left(t + \frac{\sigma_x}{\sqrt{\pi}}\right)^2\right), t > -\frac{\sigma_x}{\sqrt{\pi}}$$

的隨機變數，且 $\left(\epsilon_{c,1} \ \epsilon_{c,2} \ \dots \ \epsilon_{c,n} \right)'$ 和 $\left(\epsilon_{l,1} \ \epsilon_{l,2} \ \dots \ \epsilon_{l,n} \right)'$ 互相獨立。

另外一種可行的建模方式是不使用 l_x ，而對 l_x 做進一步的轉換 $z_x = \ln(l_x)$ 。在假設1與2下得 z_x 的密度函數為：

$$f_{z_x}(w) = \frac{2}{\sigma_x \sqrt{\pi}} \exp\left(w - \frac{e^{2w}}{\sigma_x^2}\right), w \in R$$

因為 l_x 和 c_x 互相獨立，故 z_x 和 c_x 也互相獨立。由於

$$\int w \exp\left(w - \frac{e^{2w}}{\sigma_x^2}\right) dw$$

沒有封閉型 (closed form)，因此 $E(z_x)$ 沒有一確切的型式，但它必為 σ_x 的函數。故我

們仍可對 (c_x, z_x) 建立線性迴歸模型如下：

$$\begin{pmatrix} c_{x_1} & z_{x_1} \\ c_{x_2} & z_{x_2} \\ \vdots & \vdots \\ c_{x_n} & z_{x_n} \end{pmatrix} = \begin{pmatrix} \beta_0 + \sum_{j=1}^p \beta_{1j} \psi_j(x_1) & \gamma_0 + \sum_{j=1}^p \gamma_{1j} \psi_j(x_1) \\ \beta_0 + \sum_{j=1}^p \beta_{2j} \psi_j(x_2) & \gamma_0 + \sum_{j=1}^p \gamma_{2j} \psi_j(x_2) \\ \vdots & \vdots \\ \beta_0 + \sum_{j=1}^p \beta_{nj} \psi_j(x_n) & \gamma_0 + \sum_{j=1}^p \gamma_{nj} \psi_j(x_n) \end{pmatrix} + \begin{pmatrix} \epsilon_{c,1} & \epsilon_{z,1} \\ \epsilon_{c,2} & \epsilon_{z,2} \\ \vdots & \vdots \\ \epsilon_{c,n} & \epsilon_{z,n} \end{pmatrix},$$

其中

$$\begin{pmatrix} \epsilon_{c,1} \\ \epsilon_{c,2} \\ \vdots \\ \epsilon_{c,n} \end{pmatrix} \sim N_n(0, \begin{pmatrix} \frac{\sigma_{x_1}^2}{2} & 0 & \dots & 0 \\ 0 & \frac{\sigma_{x_2}^2}{2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\sigma_{x_n}^2}{2} \end{pmatrix}).$$

$\epsilon_{z,1}, \epsilon_{z,2}, \dots, \epsilon_{z,n}$ 為獨立且具有相同機率密度函數

$$h(w) = \frac{2}{\sigma_x \sqrt{\pi}} \exp(-(w + E(z_x)) - \frac{1}{\sigma_x^2} e^{2(w + E(z_x))}), w \in R$$

的隨機變數, 且 $(\epsilon_{c,1} \ \epsilon_{c,2} \ \dots \ \epsilon_{c,n})'$ 和 $(\epsilon_{z,1} \ \epsilon_{z,2} \ \dots \ \epsilon_{z,n})'$ 互相獨立。這個模型的優點是對於區間反應變數的假設比之前兩個模型要弱, 或許更符合現實狀況。缺點則是所推導出的分佈只有區間中心值服從常態分配, 因此在後續的發展難度增加。

2.4 貝氏觀點

模型參數的假設一向分為頻率學派和貝氏學派, 其中頻率學派認為參數是固定未知的常數, 而貝氏學派認為參數也是隨機變數, 其背後有個機率分佈在影響它。由於在下一章發展選模準則時需要將參數視為隨機變數以推導出一些有用的性質, 而本文以模型一為主, 故對模型一中的參數 $\beta, \gamma, \sigma_c, \sigma_z$, 我們採取貝氏的觀點, 將其視為隨機變數, 而為了後續推導出矩陣 t 分配, 我們假設該先驗分配為無訊息的 (noninformative):

$$\pi(\theta, \Sigma^{-1}) d\theta d\Sigma^{-1} \propto \det(\Sigma^{-1})^{-\frac{m+1}{2}} d\theta d\Sigma^{-1}.$$

其中

$$\theta = (\beta, \gamma)$$

$$\Sigma = \begin{pmatrix} \sigma_c^2 I_n & 0 \\ 0 & \sigma_z^2 I_n \end{pmatrix}.$$

這裡無訊息的先驗分配是指 $\beta, \gamma, \sigma_c, \sigma_z$ 的聯合密度函數與 β, γ 無關, 只與 σ_c, σ_z 的乘積成比例。

第 3 章

模型選取與區間預測

Gelfand and Ghosh (1998)在決策理論的架構下針對適用迴歸模型的數據發展了一個預測實驗觀察值重複點(replication)的貝氏選模準則,此處重複點是指使用相同的因子設定再做一次實驗所得到的反應值。但該篇論文的反應值僅考慮單一反應變數的狀況,並且也沒有提到未知實驗點的預測式為何。本文套用該篇架構,以模型一為主,發展一針對區間反應變數的區間預測式與模型選取準則。以下3.1節介紹了區間距離函數與區間損失函數,並推導出區間預測式,3.2節利用了3.1節的結果推導出模型選取的準則。而我們發現前兩節的結果與矩陣 t 分配密不可分,因此在3.3節介紹該分配在此處的應用。本章是針對一般形式的損失函數來探討,而在第5章我們將套用兩種區間距離函數至前兩節的提出的結果上。

3.1 區間預測

由於本文以決策理論為架構,因此在推導區間反應變數的預測式之前,我們需要先定義何謂區間的損失函數。現假設有一反應值為區間的實驗,我們在因子設定 x_1, x_2, \dots, x_n 下做一次實驗後,會得到 n 個區間反應變數,將它們記為 $\mathcal{I}_{obs,1}, \mathcal{I}_{obs,2}, \dots, \mathcal{I}_{obs,n}$, 我們稱之為在 x_1, x_2, \dots, x_n 下的觀測區間。如果未來在相同的因子設定下再做一次實驗,得到的 n 個區間反應變數記為 $\mathcal{I}_{rep,1}, \mathcal{I}_{rep,2}, \dots, \mathcal{I}_{rep,n}$, 我們稱之為在 x_1, x_2, \dots, x_n 下的重複區間。注意到此時 $\mathcal{I}_{obs,1}, \mathcal{I}_{obs,2}, \dots, \mathcal{I}_{obs,n}$ 是已觀測到的數據,而 $\mathcal{I}_{rep,1}, \mathcal{I}_{rep,2}, \dots, \mathcal{I}_{rep,n}$ 還尚未得到。每一個區間反應變數都可視為隨機變數,其機率分配與期望值的結構我們

採用第三章的模型一來描述, 亦即:

$$\begin{pmatrix} \mathcal{I}_{obs,1} \\ \mathcal{I}_{obs,2} \\ \vdots \\ \mathcal{I}_{obs,n} \end{pmatrix} = \begin{pmatrix} c_{x_1} & z_{x_1} \\ c_{x_2} & z_{x_2} \\ \vdots & \vdots \\ c_{x_n} & z_{x_n} \end{pmatrix} = \begin{pmatrix} \beta_0 + \sum_{j=1}^p \beta_{1j} \psi_j(x_1) & \gamma_0 + \sum_{j=1}^p \gamma_{1j} \psi_j(x_1) \\ \beta_0 + \sum_{j=1}^p \beta_{2j} \psi_j(x_2) & \gamma_0 + \sum_{j=1}^p \gamma_{2j} \psi_j(x_2) \\ \vdots & \vdots \\ \beta_0 + \sum_{j=1}^p \beta_{nj} \psi_j(x_n) & \gamma_0 + \sum_{j=1}^p \gamma_{nj} \psi_j(x_n) \end{pmatrix} + \begin{pmatrix} \epsilon_c & \epsilon_z \end{pmatrix},$$

與

$$\begin{pmatrix} \mathcal{I}_{rep,1} \\ \mathcal{I}_{rep,2} \\ \vdots \\ \mathcal{I}_{rep,n} \end{pmatrix} = \begin{pmatrix} c_{rep,1} & z_{rep,1} \\ c_{rep,2} & z_{rep,2} \\ \vdots & \vdots \\ c_{rep,n} & z_{rep,n} \end{pmatrix} = \begin{pmatrix} \beta_0 + \sum_{j=1}^p \beta_{1j} \psi_j(x_1) & \gamma_0 + \sum_{j=1}^p \gamma_{1j} \psi_j(x_1) \\ \beta_0 + \sum_{j=1}^p \beta_{2j} \psi_j(x_2) & \gamma_0 + \sum_{j=1}^p \gamma_{2j} \psi_j(x_2) \\ \vdots & \vdots \\ \beta_0 + \sum_{j=1}^p \beta_{nj} \psi_j(x_n) & \gamma_0 + \sum_{j=1}^p \gamma_{nj} \psi_j(x_n) \end{pmatrix} + \begin{pmatrix} \epsilon_c & \epsilon_z \end{pmatrix},$$

其中

$$\begin{aligned} \epsilon_c &= \begin{pmatrix} \epsilon_{c,1} & \dots & \epsilon_{c,n} \end{pmatrix}', \\ \epsilon_z &= \begin{pmatrix} \epsilon_{z,1} & \dots & \epsilon_{z,n} \end{pmatrix}', \left(\begin{array}{c} \epsilon_c \\ \epsilon_z \end{array} \right) | \beta, \gamma, \Sigma \sim N_{2n} \left(\mathbf{0}, \begin{pmatrix} \sigma_c^2 I_n & 0 \\ 0 & \sigma_z^2 I_n \end{pmatrix} \right), \\ \beta &= \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \sim N_{p+1}(\mu_\beta, \Sigma_\beta), \gamma = \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_p \end{pmatrix} \sim N_{p+1}(\mu_\gamma, \Sigma_\gamma), \Sigma = \begin{pmatrix} \sigma_c^2 I_n & 0 \\ 0 & \sigma_z^2 I_n \end{pmatrix}. \end{aligned}$$

此處我們多加一項假設, 就是對所有的 i, j , $\mathcal{I}_{obs,i} | \beta, \gamma, \Sigma$ 與 $\mathcal{I}_{rep,j} | \beta, \gamma, \Sigma$ 互相獨立。

接著為了方便我們令 M 代表模型中所有效應形成的集合, 即

$$M = \{\psi_j(x) : j = 1, 2, \dots, p\},$$

稱之為母模型 (full model)。取 $m_c \subset M$ 表示 c_x 的子模型 (sub-model), $m_z \subset M$ 表示 z_x 的子模型。因此 $m = (m_c, m_z)$ 表示一組 c_x 和 z_x 的模型組合。令 $\hat{\mathcal{I}}_{m,k} = \begin{pmatrix} \hat{c}_{x_k} & \hat{z}_{x_k} \end{pmatrix}$ 為在模型 m 下, 對因子設定 x_k 的區間預測。

提出區間損失函數前，我們必須要先能衡量兩區間的距離。若定義函數

$$d: R^2 \times R^2 \rightarrow R^+ \cup \{0\}$$

為兩區間的距離函數。其中 d 必須滿足：

- 1：對任意區間 $\mathcal{I}, d(\mathcal{I}, \mathcal{I}) = 0$,
- 2：對任意兩區間 $\mathcal{I}_1, \mathcal{I}_2, d(\mathcal{I}_1, \mathcal{I}_2) \geq 0$ 。

則在模型 m 下，我們可以衡量因子設定為 x_j 時，預測區間 $\hat{\mathcal{I}}_{m,j}$ 對重複區間與觀測區間的距離，也就是 $d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j})$ 和 $d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{obs,j})$ 。由於 $\mathcal{I}_{rep,j}$ 是未來做實驗才會得到的區間，故 $d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j})$ 可視為用 $\hat{\mathcal{I}}_{m,j}$ 預測 x_j 因子設定下之區間的好壞。而 $\mathcal{I}_{obs,j}$ 是已經觀測到的數據，故 $d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{obs,j})$ 可視為用 $\hat{\mathcal{I}}_{m,j}$ 擬合 x_j 因子設定下之區間的程度。因此有了 x_j 下的觀測區間 $\mathcal{I}_{obs,j}$ 、重複區間 $\mathcal{I}_{rep,j}$ ，以及預測區間 $\hat{\mathcal{I}}_{m,j}$ 之後，我們定的區間損失函數應能同時考慮到模型預測與模型擬合的程度，故將之定義如下：

$$L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j}) = d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}) + kd(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{obs,j}), \quad (3.1)$$

其中 k 值作為 $d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j})$ 和 $d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{obs,j})$ 的一個權重比例。 $k = 0$ 代表完全不關心模型的擬合程度， $k = 1$ 代表認為模型擬合與模型預測同等重要， $k > 1$ 代表認為模型擬合較重要。本章先不對區間距離函數的形式做任何限定，而直接以一般的形式來發展模型選取準則與區間預測式。而關於我們使用的區間距離函數的介紹請見第5章。

有了區間損失函數後，我們就可以推導出在決策理論的觀點下，最佳的區間預測式。在區間損失函數 $L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})$ 中， $\mathcal{I}_{rep,j}$ 和 $\mathcal{I}_{obs,j}$ 都是隨機變數，因此當區間預測值為 $\hat{\mathcal{I}}_{m,j}$ 時， $L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})$ 也是隨機變數。故我們想看平均來說 $L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})$ 會是多少，也就是 $L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})$ 的期望值。在貝氏觀點下，由於 $\mathcal{I}_{rep,j}$ 尚未得到，而 $\mathcal{I}_{obs,j}$ 已有實驗數據。故我們將 $L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})$ 的期望值對 $\mathcal{I}_{rep,j} | \mathcal{I}_{obs,j}$ 的機率分配來求取。但其實已有的觀測區間不只第 j 筆，因此我們納入所有觀測區間的

資訊，進而對 $\mathcal{I}_{rep,j}|\mathcal{I}_{obs}$ 的分配來求期望值，其中

$$\mathcal{I}_{obs} = \begin{pmatrix} \mathcal{I}_{obs,1} \\ \mathcal{I}_{obs,2} \\ \vdots \\ \mathcal{I}_{obs,n} \end{pmatrix}.$$

$\mathcal{I}_{rep,j}|\mathcal{I}_{obs}$ 的條件密度函數可稱為給定 \mathcal{I}_{obs} 下， $\mathcal{I}_{rep,j}$ 的貝氏預測密度函數。若令 $p(\mathcal{I}_{rep,j}|\mathcal{I}_{obs})$ 為該密度函數，參數為 β, γ, Σ ，其先驗分佈為 $\pi(\beta, \gamma, \Sigma)$ 。由於對所有的 i, j ， $\mathcal{I}_{obs,i}|\beta, \gamma, \Sigma$ 與 $\mathcal{I}_{rep,j}|\beta, \gamma, \Sigma$ 互相獨立，故

$$p(\mathcal{I}_{rep,j}|\beta, \gamma, \Sigma) = p(\mathcal{I}_{rep,j}|\beta, \gamma, \Sigma, \mathcal{I}_{obs}).$$

因此 $p(\mathcal{I}_{rep,j}|\mathcal{I}_{obs})$ 可由下式得到：

$$p(\mathcal{I}_{rep,j}|\mathcal{I}_{obs}) = \int \int \int p(\mathcal{I}_{rep,j}|\beta, \gamma, \Sigma)\pi(\beta, \gamma, \Sigma|\mathcal{I}_{obs})d\beta d\gamma d\Sigma,$$

其中

$$\pi(\beta, \gamma, \Sigma|\mathcal{I}_{obs}) = \frac{\pi(\beta, \gamma, \Sigma)p(\mathcal{I}_{obs}|\beta, \gamma, \Sigma)}{\int \int \int \pi(\beta, \gamma, \Sigma)p(\mathcal{I}_{obs}|\beta, \gamma, \Sigma)d\beta d\gamma d\Sigma}.$$

有了 $p(\mathcal{I}_{rep,j}|\mathcal{I}_{obs})$ 就可以計算出貝氏預測風險函數：

$$E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})] = \int L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})p(\mathcal{I}_{rep,j}|\mathcal{I}_{obs})d(\mathcal{I}_{rep,j}).$$

由期望值的性質可將 $E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})]$ 展開得到：

$$\begin{aligned} E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})] &= E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}) + kd(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{obs,j})] \\ &= E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j})] + kd(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{obs,j}). \end{aligned}$$

至此我們就導出了 x_j 下的貝氏預測風險函數，而因子設定共有 n 組，將它們加起來得到整個實驗下的貝氏預測風險函數 $\sum_{j=1}^n E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})]$ 。很明顯的它是 $\hat{\mathcal{I}}_{m,1}, \dots, \hat{\mathcal{I}}_{m,n}$ 的函數，故可令

$$f(\hat{\mathcal{I}}_{m,1}, \dots, \hat{\mathcal{I}}_{m,n}) = \sum_{j=1}^n E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})].$$

因此 $f(\hat{\mathcal{I}}_{m,1}, \dots, \hat{\mathcal{I}}_{m,n})$ 就代表當模型為 m ，區間預測值為 $\hat{\mathcal{I}}_{m,1}, \dots, \hat{\mathcal{I}}_{m,n}$ 下，所有因子設定的貝氏預測風險函數和。假設 f 存在最小值，則該值就意謂著模型為 m 時，去預測因子設定 x_1, x_2, \dots, x_n 下區間的最小平均損失。既然我們用貝氏預測風險函數來衡量預測區間造成的平均損失，那麼最佳的區間預測值理應要能使得貝氏預測風險函數最小。故模型為 m 時， x_1, x_2, \dots, x_n 的區間預測式為

$$\begin{aligned} (\hat{\mathcal{I}}_{m,1}^*, \dots, \hat{\mathcal{I}}_{m,n}^*) &= \arg\{\min_{\hat{\mathcal{I}}_{m,1}, \dots, \hat{\mathcal{I}}_{m,n}} f(\hat{\mathcal{I}}_{m,1}, \dots, \hat{\mathcal{I}}_{m,n})\} \\ &= \arg\{\min_{\hat{\mathcal{I}}_{m,1}, \dots, \hat{\mathcal{I}}_{m,n}} \sum_{j=1}^n E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}} [L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})]\}. \end{aligned} \quad (3.2)$$

以上是針對在有實驗點的區間做預測，而還有一種情況是對一個尚未做實驗的因子設定 x_{new} 的區間做預測。記此區間為 \mathcal{I}_{new} ，並假設對所有的 j ， $\mathcal{I}_{new}|\beta, \gamma, \Sigma$ 和 $\mathcal{I}_{obs,j}|\beta, \gamma, \Sigma$ 互相獨立。由於此時沒有了觀測區間，故模型為 m ，預測區間為 $\hat{\mathcal{I}}_{m,new}$ 時，區間損失函數為

$$L(\hat{\mathcal{I}}_{m,new}, \mathcal{I}_{new}; \mathcal{I}_{obs,j}) = d(\hat{\mathcal{I}}_{m,new}, \mathcal{I}_{new}).$$

\mathcal{I}_{new} 是隨機變數，因此 $L(\hat{\mathcal{I}}_{m,new}, \mathcal{I}_{new}; \mathcal{I}_{obs,j})$ 也是隨機變數，我們想看平均來說 $L(\hat{\mathcal{I}}_{m,new}, \mathcal{I}_{new}; \mathcal{I}_{obs,j})$ 會是多少，也就是 $L(\hat{\mathcal{I}}_{m,new}, \mathcal{I}_{new}; \mathcal{I}_{obs,j})$ 的期望值。在貝氏觀點下，由於 \mathcal{I}_{new} 尚未得到，但之前實驗的數據 \mathcal{I}_{obs} 仍能提供一些資訊。故我們將 $L(\hat{\mathcal{I}}_{m,new}, \mathcal{I}_{new}; \mathcal{I}_{obs,j})$ 的期望值對 $\mathcal{I}_{new}|\mathcal{I}_{obs}$ 的機率分配來求取。若記 $\mathcal{I}_{new}|\mathcal{I}_{obs}$ 的條件密度函數為 $p(\mathcal{I}_{new}|\mathcal{I}_{obs})$ ，由於對所有的 j ， $\mathcal{I}_{new}|\beta, \gamma, \Sigma$ 與 $\mathcal{I}_{obs,j}|\beta, \gamma, \Sigma$ 互相獨立，故 $p(\mathcal{I}_{new}|\beta, \gamma, \Sigma) = p(\mathcal{I}_{new}|\beta, \gamma, \Sigma, \mathcal{I}_{obs})$ 。因此 $p(\mathcal{I}_{new}|\mathcal{I}_{obs})$ 可由下式得到：

$$p(\mathcal{I}_{new}|\mathcal{I}_{obs}) = \int \int \int p(\mathcal{I}_{new}|\beta, \gamma, \Sigma) \pi(\beta, \gamma, \Sigma|\mathcal{I}_{obs}) d\beta d\gamma d\Sigma,$$

其中

$$\pi(\beta, \gamma, \Sigma|\mathcal{I}_{obs}) = \frac{\pi(\beta, \gamma, \Sigma) p(\mathcal{I}_{obs}|\beta, \gamma, \Sigma)}{\int \int \int \pi(\beta, \gamma, \Sigma) p(\mathcal{I}_{obs}|\beta, \gamma, \Sigma) d\beta d\gamma d\Sigma}.$$

有了 $p(\mathcal{I}_{new}|\mathcal{I}_{obs})$ 就可以對 $L(\hat{\mathcal{I}}_{m,new}, \mathcal{I}_{new}; \mathcal{I}_{obs,j})$ 求期望值，令

$$f(\hat{\mathcal{I}}_{m,new}) = E_{\mathcal{I}_{new}|\mathcal{I}_{obs}} [L(\hat{\mathcal{I}}_{m,new}, \mathcal{I}_{new}; \mathcal{I}_{obs,j})].$$

因此 $f(\hat{\mathcal{I}}_{m,new})$ 就代表當模型為 m ，區間預測為 $\hat{\mathcal{I}}_{m,new}$ 下的貝氏預測風險函數值。既然我們用貝氏預測風險函數來衡量預測區間造成的平均損失，那麼最佳的區間預測值理應要能使得貝氏預測風險函數最小。故區間預測式為：

$$\begin{aligned}\hat{\mathcal{I}}_{m,new}^* &= \arg\{\min_{\hat{\mathcal{I}}_{m,new}} f(\hat{\mathcal{I}}_{m,new})\} \\ &= \arg\{\min_{\hat{\mathcal{I}}_{m,new}} E_{\mathcal{I}_{new}|\mathcal{I}_{obs}}[L(\hat{\mathcal{I}}_{m,new}, \mathcal{I}_{new}; \mathcal{I}_{obs,j})]\}.\end{aligned}\quad (3.3)$$

此處和有實驗點的情形類似，隨著區間損失函數的不同， $\hat{\mathcal{I}}_{m,new}^*$ 也會不一樣，所以實驗者一開始就要先決定哪種區間損失函數是適合的。

3.2 模型選取準則

3.1節討論了給定子模型 m 下的區間預測式，本節由上述的結果進而發展一個模型選取的準則。在決策理論的觀點下，一個最佳的子模型 m 應該要能讓貝氏預測風險函數值最小，也就是最小化

$$f(\hat{\mathcal{I}}_{m,1}, \dots, \hat{\mathcal{I}}_{m,n}) = \sum_{j=1}^n E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})].$$

但對每一個 m ， $f(\hat{\mathcal{I}}_{m,1}, \dots, \hat{\mathcal{I}}_{m,n})$ 的值並不唯一，而是隨著區間預測值的變動而改變。然而區間預測的值是我們可以決定的，故很自然的我們會將區間預測值選在 $f(\hat{\mathcal{I}}_{m,1}, \dots, \hat{\mathcal{I}}_{m,n})$ 的極小點上。這個極小點就是將 (3.1) 式代入，可得模型 m 下最小的貝氏預測風險函數值，該值亦可看成模型 m 下，預測區間造成的最小平均損失。因此我們希望所選出的模型能夠擁有最小的『最小平均損失』。故若記

$$\begin{aligned}R_k(m) &= f(\hat{\mathcal{I}}_{m,1}^*, \dots, \hat{\mathcal{I}}_{m,n}^*) \\ &= \sum_{j=1}^n E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[L(\hat{\mathcal{I}}_{m,j}^*, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})].\end{aligned}$$

則選模準則為：

$$\text{找一 } m^* \text{ 使得 } m^* = \arg\{\min_m R_k(m)\}.$$

3.3 矩陣 t 分配的應用

上兩節需要求 $E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})]$ 與 $E_{\mathcal{I}_{new}|\mathcal{I}_{obs}}[L(\hat{\mathcal{I}}_{m,new}, \mathcal{I}_{new}; \mathcal{I}_{obs,j})]$ ，因此 $\mathcal{I}_{rep}|\mathcal{I}_{obs}$ 與 $\mathcal{I}_{new}|\mathcal{I}_{obs}$ 的機率分配是我們所關心的。其中

$$\mathcal{I}_{rep} = \begin{pmatrix} \mathcal{I}_{rep,1} \\ \mathcal{I}_{rep,2} \\ \vdots \\ \mathcal{I}_{rep,n} \end{pmatrix}.$$

Ng (2000)有以下定理:

定理 3. 若

1. Y 為 $n \times m$, X 為 $n \times p$, θ 為 $p \times m$ 的矩陣, 且有 $Y = X\theta + \varepsilon$ 的線性結構, ε 為橢圓輪廓分佈 (*elliptically contoured distribution*).
2. Y_f 為 $n_f \times m$ 的矩陣, 而 X_f 為其模型矩陣, 且有 $Y_f = X_f\theta + \varepsilon_f$ 的線性結構, ε_f 的機率分配和 ε 一樣。
3. θ 和 $Cov(\varepsilon)$ 參數的先驗分配為無訊息 (*noninformative*).

則 $Y_f|Y$ 的條件分配為矩陣 t 分配 (*matrix t distribution*)

$$t_{n \times m}(X_f \hat{\theta}, (I_n - X_f(X'X + X_f'X_f)^{-1}X_f')^{-1}, (Y - X_f \hat{\theta})'(Y - X_f \hat{\theta}), n - p - m + 1),$$

其中 $\hat{\theta} = (X'X)^{-1}X'Y$, I_n 是 $n \times n$ 的單位矩陣。

利用定理 3, 我們在子模型 m 下, 可得 $\mathcal{I}_{rep}|\mathcal{I}_{obs}$ 服從

$$t_{n \times 2}(\Psi_m(X) \hat{\theta}, (I_n - 0.5\Psi_m(X)(\Psi_m(X)'\Psi_m(X))^{-1}\Psi_m(X)')^{-1}, (\mathcal{I}_{obs} - \Psi_m(X) \hat{\theta})'(\mathcal{I}_{obs} - \Psi_m(X) \hat{\theta}), n - p - 1)$$

其中 $\Psi_m(X)$ 為模型 m 下的模型矩陣, 且

$$\begin{aligned} \hat{\theta} &= (\Psi_m(X)'\Psi_m(X))^{-1}\Psi_m(X)'\mathcal{I}_{obs} \\ &= \begin{pmatrix} \hat{\beta} & \hat{\gamma} \end{pmatrix}, \end{aligned}$$

而

$$\begin{aligned}\hat{\beta} &= (\Psi_m(X)' \Psi_m(X))^{-1} \Psi_m(X)' c_{obs}, \\ \hat{\gamma} &= (\Psi_m(X)' \Psi_m(X))^{-1} \Psi_m(X)' z_{obs}. \\ c_{obs} &= \begin{pmatrix} c_{x_1} & c_{x_2} & \cdots & c_{x_n} \end{pmatrix}', \\ z_{obs} &= \begin{pmatrix} z_{x_1} & z_{x_2} & \cdots & z_{x_n} \end{pmatrix}',\end{aligned}$$

由定理 2 可知

$$E(\mathcal{I}_{rep} | \mathcal{I}_{obs}) = \Psi_m(X) \hat{\theta}, \quad (3.4)$$

$$\begin{aligned}Cov(\mathcal{I}_{rep} | \mathcal{I}_{obs}) \\ = \frac{(I_n - 0.5 \Psi_m(X) (\Psi_m(X)' \Psi_m(X))^{-1} \Psi_m(X)')^{-1} \otimes (\mathcal{I}_{obs} - \Psi_m(X) \hat{\theta})' (\mathcal{I}_{obs} - \Psi_m(X) \hat{\theta})}{n - p - 3},\end{aligned}$$

其中 \otimes 表示 Kronecker 乘積。很容易可以驗證

$$(I_n - 0.5 \Psi_m(X) (\Psi_m(X)' \Psi_m(X))^{-1} \Psi_m(X)')^{-1} = (I_n + \Psi_m(X) (\Psi_m(X)' \Psi_m(X))^{-1} \Psi_m(X)'),$$

故上式可改寫成

$$\begin{aligned}Cov(\mathcal{I}_{rep} | \mathcal{I}_{obs}) \\ = \frac{(I_n + \Psi_m(X) (\Psi_m(X)' \Psi_m(X))^{-1} \Psi_m(X)') \otimes (\mathcal{I}_{obs} - \Psi_m(X) \hat{\theta})' (\mathcal{I}_{obs} - \Psi_m(X) \hat{\theta})}{n - p - 3}.\end{aligned} \quad (3.5)$$

至於矩陣 t 分配的大樣本漸進分配可由定理 3 得到，當 $n \rightarrow \infty$ 時， $\mathcal{I}_{rep} | \mathcal{I}_{obs}$ 的大樣本分配為

$$N_{2n}(vec(\Psi_m(X) \hat{\theta}), (I_n + \Psi_m(X) (\Psi_m(X)' \Psi_m(X))^{-1} \Psi_m(X)') \otimes (\mathcal{I}_{obs} - \Psi_m(X) \hat{\theta})' (\mathcal{I}_{obs} - \Psi_m(X) \hat{\theta})).$$

由此也可推得當 $n \rightarrow \infty$ 時， $e^{z_{rep,l}} | \mathcal{I}_{obs}$ 的大樣本分配為

$$LogNormal((\Psi_m(X) \hat{\gamma})_l, Q_{l+n, l+n})$$

$$Q = (I_n + \Psi_m(X) (\Psi_m(X)' \Psi_m(X))^{-1} \Psi_m(X)') \otimes (\mathcal{I}_{obs} - \Psi_m(X) \hat{\theta})' (\mathcal{I}_{obs} - \Psi_m(X) \hat{\theta}).$$

其中 $(\Psi_m(X)\hat{\gamma})_l$ 表示 $\Psi_m(X)\hat{\gamma}$ 第 l 個元素。因此

$$E(e^{z_{rep,l}}|\mathcal{I}_{obs}) \approx e^{(\Psi_m(X)\hat{\gamma})_l + \frac{1}{2}Q_{l+n,l+n}}, \quad (3.6)$$

$$Var(e^{z_{rep,l}}|\mathcal{I}_{obs}) \approx e^{2(\Psi_m(X)\hat{\gamma})_l + Q_{l+n,l+n}}(e^{Q_{l+n,l+n}} - 1). \quad (3.7)$$

找出 $\mathcal{I}_{new}|\mathcal{I}_{obs}$ 的機率分配過程類似 $\mathcal{I}_{rep}|\mathcal{I}_{obs}$, 利用定理 1, 我們在子模型 m 下, 可得 $\mathcal{I}_{new}|\mathcal{I}_{obs}$ 服從

$$t_{1 \times 2}(\Psi_m(x_{new})\hat{\theta}, (1 - \Psi_m(x_{new})(\Psi_m(X)' \Psi_m(X) + \Psi_m(x_{new})' \Psi_m(x_{new}))^{-1} \Psi_m(x_{new})')^{-1}, \\ (\mathcal{I}_{obs} - \Psi_m(X)\hat{\theta})'(\mathcal{I}_{obs} - \Psi_m(X)\hat{\theta}), n - p - 1).$$

其中

$$\hat{\theta} = (\Psi_m(X)' \Psi_m(X))^{-1} \Psi_m(X)' \mathcal{I}_{obs} \\ = \begin{pmatrix} \hat{\beta} & \hat{\gamma} \end{pmatrix},$$

而

$$\hat{\beta} = (\Psi_m(X)' \Psi_m(X))^{-1} \Psi_m(X)' c_{obs}, \\ \hat{\gamma} = (\Psi_m(X)' \Psi_m(X))^{-1} \Psi_m(X)' z_{obs}. \\ c_{obs} = \begin{pmatrix} c_{x_1} & c_{x_2} & \cdots & c_{x_n} \end{pmatrix}', \\ z_{obs} = \begin{pmatrix} z_{x_1} & z_{x_2} & \cdots & z_{x_n} \end{pmatrix}'.$$

由定理 2 可得

$$E_{\mathcal{I}_{new}|\mathcal{I}_{obs}}(c_{new}|\mathcal{I}_{obs}) = \Psi(x_{new})\hat{\beta} \quad (3.8)$$

$$E_{\mathcal{I}_{new}|\mathcal{I}_{obs}}(z_{new}|\mathcal{I}_{obs}) = \Psi(x_{new})\hat{\gamma}.$$

至於矩陣 t 分配的大樣本漸進分配可由定理 3 可知, 當 $n \rightarrow \infty$ 時, $\mathcal{I}_{new}|\mathcal{I}_{obs}$ 的大樣本分配為

$$N_2(\text{vec}(\Psi_m(x_{new})\hat{\theta}), Q),$$

$$Q = (1 - \Psi_m(x_{new})(\Psi_m(X)' \Psi_m(X) + \Psi_m(x_{new})' \Psi_m(x_{new}))^{-1} \Psi_m(x_{new})')^{-1} \\ \otimes (\mathcal{I}_{obs} - \Psi_m(X)\hat{\theta})'(\mathcal{I}_{obs} - \Psi_m(X)\hat{\theta}).$$

由此也可推得當 $n \rightarrow \infty$ 時, $e^{z_{new}} | \mathcal{I}_{obs}$ 的大樣本分配為

$$\text{LogNormal}(\Psi_m(x_{new})\hat{\gamma}, Q_{2,2})$$

$$Q = (1 - \Psi_m(x_{new})(\Psi_m(X)' \Psi_m(X) + \Psi_m(x_{new})' \Psi_m(x_{new}))^{-1} \Psi_m(x_{new})')^{-1} \\ \otimes (\mathcal{I}_{obs} - \Psi_m(X)\hat{\theta})'(\mathcal{I}_{obs} - \Psi_m(X)\hat{\theta}).$$

故

$$E(e^{z_{new}} | \mathcal{I}_{obs}) \approx e^{\Psi_m(x_{new})\hat{\gamma} + \frac{1}{2}Q_{2,2}} \quad (3.9)$$

$$\text{Var}(e^{z_{new}} | \mathcal{I}_{obs}) \approx e^{2\Psi_m(x_{new})\hat{\gamma} + Q_{2,2}}(e^{Q_{2,2}} - 1), \quad (3.10)$$

$$Q = (1 - \Psi_m(x_{new})(\Psi_m(X)' \Psi_m(X) + \Psi_m(x_{new})' \Psi_m(x_{new}))^{-1} \Psi_m(x_{new})')^{-1} \\ \otimes (\mathcal{I}_{obs} - \Psi_{m_z}(X)\hat{\theta})'(\mathcal{I}_{obs} - \Psi_{m_z}(X)\hat{\theta}).$$



第 4 章

套用區間距離函數

第 4 章推導的結果對任意的區間距離函數都適用，本節將兩種區間距離函數分別套用到之前的結果上。

4.1 第一種區間距離函數

兩區間差異的大小取決於它們端點的距離，也就是區間 1 下界和區間 2 下界，以及區間 1 上界和區間 2 上界的遠近。因此衡量兩區間的差異的一個做法是計算端點的差值。若令 $\mathcal{I}_1 = [u_1, v_1]$, $\mathcal{I}_2 = [u_2, v_2]$ 為兩區間，則由上述的想法我們可以定出一個函數

$$d^*(\mathcal{I}_1, \mathcal{I}_2) = |u_1 - u_2| + |v_1 - v_2|.$$

很明顯 d^* 滿足 3.1 節區間距離函數的 2 個條件，故 d^* 為一區間距離函數，為兩絕對值函數的和。但絕對值函數存在不可微分的點，因此做一些深入的處理時會比較麻煩。故我們修改一下 d^* ，定出另一函數為

$$d(\mathcal{I}_1, \mathcal{I}_2) = (u_1 - u_2)^2 + (v_1 - v_2)^2,$$

距離 d 滿足 3.1 節區間距離函數的 2 個條件也是顯而易見的，此時 d 就不像 d^* 存在不可微分的點。由於我們採用模型一區間反應變數的形式，故 u_i, v_i 透過 (2.2) 式之轉換，

再代入 d 中可得

$$\begin{aligned} d(\mathcal{I}_1, \mathcal{I}_2) &= (u_1 - u_2)^2 + (v_1 - v_2)^2 \\ &= (c_1 - e^{z_1}) - (c_2 - e^{z_2})^2 + ((c_1 + e^{z_1}) - (c_2 + e^{z_2}))^2 \\ &= 2[(c_1 - c_2)^2 + (e^{z_1} - e^{z_2})^2]. \end{aligned}$$

將 d 代入 (3.1) 式的區間損失函數 $L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})$ 中可得

$$\begin{aligned} L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j}) &= 2[(\hat{c}_{m,x_j} - c_{rep,j})^2 + (e^{\hat{z}_{m,x_j}} - e^{z_{rep,j}})^2] \\ &\quad + 2k[(\hat{c}_{m,x_j} - c_{x_j})^2 + (e^{\hat{z}_{m,x_j}} - e^{z_{x_j}})^2], \end{aligned}$$

故貝氏預測風險函數為

$$\begin{aligned} &E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})] \\ &= 2E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[(\hat{c}_{m,x_j} - c_{rep,j})^2 + (e^{\hat{z}_{m,x_j}} - e^{z_{rep,j}})^2] \\ &\quad + 2k[(\hat{c}_{m,x_j} - c_{x_j})^2 + (e^{\hat{z}_{m,x_j}} - e^{z_{x_j}})^2] \\ &= 2E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[(\hat{c}_{m,x_j} - c_{rep,j})^2] + 2k[(\hat{c}_{m,x_j} - c_{x_j})^2] \\ &\quad + 2E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[(e^{\hat{z}_{m,x_j}} - e^{z_{rep,j}})^2] \\ &\quad + 2k[(exp(\hat{z}_{m,x_j}) - e^{z_{x_j}})^2]. \end{aligned}$$

因此整個實驗的貝氏預測風險函數 f 為

$$\begin{aligned} &f(\hat{\mathcal{I}}_{m,1}, \dots, \hat{\mathcal{I}}_{m,n}) \\ &= \sum_{j=1}^n E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})] \\ &= 2 \sum_{j=1}^n \{E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[(\hat{c}_{m,x_j} - c_{rep,j})^2] + k[(\hat{c}_{m,x_j} - c_{x_j})^2]\} \\ &\quad + 2 \sum_{j=1}^n \{E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[(e^{\hat{z}_{m,x_j}} - e^{z_{rep,j}})^2] + k[(e^{\hat{z}_{m,x_j}} - e^{z_{x_j}})^2]\} \\ &= f_c(\hat{c}_{m,x_1}, \dots, \hat{c}_{m,x_n}) + f_z(\hat{z}_{m,x_1}, \dots, \hat{z}_{m,x_n}). \end{aligned} \tag{4.1}$$

其中

$$f_c(\hat{c}_{m,x_1}, \dots, \hat{c}_{m,x_n}) = 2 \sum_{j=1}^n \{E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[(\hat{c}_{m,x_j} - c_{rep,j})^2] + k[(\hat{c}_{m,x_j} - c_{x_j})^2]\},$$

$$f_z(\hat{z}_{m,x_1}, \dots, \hat{z}_{m,x_n}) = 2 \sum_{j=1}^n \{E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[(e^{\hat{z}_{m,x_j}} - e^{z_{rep,j}})^2] + k[(e^{\hat{z}_{m,x_j}} - e^{z_{x_j}})^2]\}.$$

由 (4.1) 式可知找出 f 的最小值與分別找出 f_c, f_z 的最小值再相加是一樣的。故我們可對 f_c 與 f_z 分別處理。

4.1.1 區間預測

$f_c(\hat{c}_{m,x_1}, \dots, \hat{c}_{m,x_n})$ 表示模型 m 下, c_x 的貝氏預測風險函數, 現我們要對它求最小值。爲了符號簡便, 將 $E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}$ 寫成 E , $p_l = \hat{c}_{m,x_l}$,

$$f_c(p_1, \dots, p_n) = 2 \sum_{l=1}^n E[(c_{rep,l} - p_l)^2] + k(c_{x_l} - p_l)^2, \quad (4.2)$$

此時相當於是找 $(p_1^*, p_2^*, \dots, p_n^*)$ 使得 $f_c(p_1^*, p_2^*, \dots, p_n^*)$ 達到最小。將 f_c 對 p_l, p_k 微分可得

$$\begin{aligned} \frac{\partial f_c}{\partial p_l} &= 2(p_l - E(c_{rep,l}|\mathcal{I}_{obs})) + 2k(p_l - c_{x_j}) \\ \frac{\partial^2 f_c}{\partial p_l \partial p_k} &= 0 \\ \frac{\partial^2 f_c}{\partial p_l^2} &= 2(k+1), \end{aligned}$$

令 $\frac{\partial f_c}{\partial p_l} = 0$ 得到

$$p_l^* = \frac{E(c_{rep,l}|\mathcal{I}_{obs}) + kc_{x_l}}{k+1}.$$

且

$$\begin{pmatrix} \frac{\partial^2 f_c}{\partial p_1 \partial p_1} & \frac{\partial^2 f_c}{\partial p_1 \partial p_2} & \dots & \frac{\partial^2 f_c}{\partial p_1 \partial p_n} \\ \frac{\partial^2 f_c}{\partial p_2 \partial p_1} & \frac{\partial^2 f_c}{\partial p_2 \partial p_2} & \dots & \frac{\partial^2 f_c}{\partial p_2 \partial p_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f_c}{\partial p_n \partial p_1} & \frac{\partial^2 f_c}{\partial p_n \partial p_2} & \dots & \frac{\partial^2 f_c}{\partial p_n \partial p_n} \end{pmatrix} = \begin{pmatrix} 2(k+1) & 0 & \dots & 0 \\ 0 & 2(k+1) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 2(k+1) \end{pmatrix}$$

爲一正定矩陣，故當 $p_l^* = \frac{E(c_{rep,l}|\mathcal{I}_{obs}) + kc_{x_l}}{k+1}$ 時， $f(p_1^*, \dots, p_n^*)$ 有最小值，將之代回 (4.2)

式得

$$f_c(p_1^*, \dots, p_n^*) = \sum_{l=1}^n [Var(c_{rep,l}|\mathcal{I}_{obs}) + \frac{k}{k+1}(E(c_{rep,l}|\mathcal{I}_{obs}) - c_{x_l})^2].$$

而對 f_z 求最小值的部分作法類似，爲了符號簡便， $E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}$ 寫成 E ，令 $q_l = \hat{z}_{m,l}$ ，

$$f_z(q_1, \dots, q_n) = 2 \sum_{l=1}^n [E[(e^{z_{rep,l}} - e^{q_l})^2] + k(e^{z_{x_l}} - e^{q_l})^2], \quad (4.3)$$

此時相當於是找 $(q_1^*, q_2^*, \dots, q_n^*)$ 使得 $f_z(q_1^*, q_2^*, \dots, q_n^*)$ 達到最小。

$$\begin{aligned} \frac{\partial f_z}{\partial q_l} &= 2(2(k+1)e^{2q_l} - (ke^{2q_l} + E(e^{z_{rep,l}}|\mathcal{I}_{obs})e^{q_l})) \\ \frac{\partial^2 f_z}{\partial q_l \partial q_k} &= 0 \\ \frac{\partial^2 f_z}{\partial q_l^2} &= 2e^{q_l}[2(k+1)e^{q_l} - (ke^{z_{x_l}} + E(e^{z_{rep,l}}|\mathcal{I}_{obs}))] \end{aligned}$$

令 $\frac{\partial f_z}{\partial q_l} = 0$ 得到

$$q_l^* = \ln\left(\frac{E(e^{z_{rep,l}}|\mathcal{I}_{obs}) + ke^{z_{x_l}}}{k+1}\right).$$

將 q_l^* 代入 $\frac{\partial^2 f_z}{\partial q_l^2}$ 得到

$$2\left(\frac{E(e^{z_{rep,l}}|\mathcal{I}_{obs}) + ke^{z_{x_l}}}{k+1}\right)(ke^{z_{x_l}} + E(e^{z_{rep,l}}|\mathcal{I}_{obs})) > 0,$$

因此

$$\begin{pmatrix} \frac{\partial^2 f_z}{\partial q_1 \partial q_1} & \frac{\partial^2 f_z}{\partial q_1 \partial q_2} & \cdots & \frac{\partial^2 f_z}{\partial q_1 \partial q_n} \\ \frac{\partial^2 f_z}{\partial q_2 \partial q_1} & \frac{\partial^2 f_z}{\partial q_2 \partial q_2} & \cdots & \frac{\partial^2 f_z}{\partial q_2 \partial q_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f_z}{\partial q_n \partial q_1} & \frac{\partial^2 f_z}{\partial q_n \partial q_2} & \cdots & \frac{\partial^2 f_z}{\partial q_n \partial q_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 f_z}{\partial q_1^2} & 0 & \cdots & 0 \\ 0 & \frac{\partial^2 f_z}{\partial q_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\partial^2 f_z}{\partial q_n^2} \end{pmatrix}$$

爲一正定矩陣，

故 $q_l^* = \ln\left(\frac{E(e^{z_{rep,l}}|\mathcal{I}_{obs}) + ke^{z_{x_l}}}{k+1}\right)$ 時， $f_z(q_1^*, \dots, q_n^*)$ 有最小值，將之代回 (4.3) 式得

$$f_z(q_1^*, \dots, q_n^*) = \sum_{l=1}^n [Var(e^{z_{rep,l}}|\mathcal{I}_{obs}) + \frac{k}{k+1}(E(e^{z_{rep,l}}|\mathcal{I}_{obs}) - e^{z_{x_l}})^2].$$

由 (3.4) 式可知,

$$E(c_{rep}|\mathcal{I}_{obs}) = \Psi_m(X)\hat{\beta},$$

由 (3.6) 式可知,

$$E(e^{z_{rep,l}}|\mathcal{I}_{obs}) = \exp(\Psi_{m_z}(X)\hat{\gamma}_l + \frac{1}{2}Q_{l+n,l+n})$$

$$Q = (I_n + \Psi_{m_z}(X)(\Psi_{m_z}(X)'\Psi_{m_z}(X))^{-1}\Psi_{m_z}(X)') \otimes (\mathcal{I}_{obs} - \Psi_{m_z}(X)\hat{\theta})'(\mathcal{I}_{obs} - \Psi_{m_z}(X)\hat{\theta}).$$

故當子模型是 $m = (m_c, m_z)$, 對因子設定 x_l 的預測區間為:

$$p_l^* = \hat{c}_{m_c, x_l} = \frac{[\Psi_{m_c}(X)\hat{\beta}]_l + kc_{x_l}}{k+1}$$

$$q_l^* = \hat{z}_{m_z, x_l} = \ln\left(\frac{\exp([\Psi_{m_z}(X)\hat{\gamma}]_l + \frac{1}{2}Q_{l+n,l+n}) + ke^{z_{x_l}}}{k+1}\right)$$

$$Q = (I_n + \Psi_{m_z}(X)(\Psi_{m_z}(X)'\Psi_{m_z}(X))^{-1}\Psi_{m_z}(X)') \otimes (\mathcal{I}_{obs} - \Psi_{m_z}(X)\hat{\theta})'(\mathcal{I}_{obs} - \Psi_{m_z}(X)\hat{\theta}).$$

當要預測一在新的因子設定 x_{new} 下的區間 \mathcal{I}_{new} 時, 由 (3.3) 式可知當子模型是 $m =$

(m_c, m_z) , 其預測區間為

$$\hat{\mathcal{I}}_{m, new}^* = \arg\{\min_{\hat{\mathcal{I}}_{m, new}} E_{\mathcal{I}_{new}|\mathcal{I}_{obs}}[L(\hat{\mathcal{I}}_{m, new}, \mathcal{I}_{new}; \mathcal{I}_{obs, j})]\}.$$

經由類似之前的推導, 可得

$$\hat{c}_{m_c, x_{new}} = E_{\mathcal{I}_{new}|\mathcal{I}_{obs}}(c_{new}|\mathcal{I}_{obs}),$$

$$\hat{z}_{m_z, x_{new}} = \ln(E_{\mathcal{I}_{new}|\mathcal{I}_{obs}}(e^{z_{new}}|\mathcal{I}_{obs})).$$

再由 (3.8) 與 (3.9) 式可得

$$\hat{c}_{m_c, x_{new}} = \Psi_{m_c}(x_{new})\hat{\beta}$$

$$\hat{z}_{m_z, x_{new}} = \ln(\exp(\Psi_{m_z}(x_{new})\hat{\gamma} + \frac{1}{2}Q_{2,2})).$$

4.1.2 模型選取準則

給定子模型 m 下, $R_k(m) = f_c(p_1^*, \dots, p_n^*) + f_z(q_1^*, \dots, q_n^*)$, 其中

$$f_c(p_1^*, \dots, p_n^*) = \sum_{l=1}^n [Var(c_{rep,l}|\mathcal{I}_{obs}) + \frac{k}{k+1}(E(c_{rep,l}|\mathcal{I}_{obs}) - c_{x_l})^2]$$

$$f_z(q_1^*, \dots, q_n^*) = \sum_{l=1}^n [Var(e^{z_{rep,l}}|\mathcal{I}_{obs}) + \frac{k}{k+1}(E(e^{z_{rep,l}}|\mathcal{I}_{obs}) - e^{z_{x_l}})^2].$$

我們已經有了的 $E(c_{rep,l}|\mathcal{I}_{obs})$ 與 $E(e^{z_{rep,l}}|\mathcal{I}_{obs})$ 形式, 因此接著我們來推導 $\sum_{l=1}^n Var(c_{rep,l}|\mathcal{I}_{obs})$ 與 $\sum_{l=1}^n Var(e^{z_{rep,l}}|\mathcal{I}_{obs})$ 。由 (3.5) 式可知

$$\begin{aligned}
& \sum_{l=1}^n Var(c_{rep,l}|\mathcal{I}_{obs}) + \sum_{l=1}^n Var(z_{rep,l}|\mathcal{I}_{obs}) \\
&= tr[Cov(\mathcal{I}_{rep}|\mathcal{I}_{obs})] \\
&= tr\left[\frac{(I_n + \Psi_m(X)(\Psi_m(X)'\Psi_m(X))^{-1}\Psi_m(X)') \otimes (\mathcal{I}_{obs} - \Psi_m(X)\hat{\theta})'(\mathcal{I}_{obs} - \Psi_m(X)\hat{\theta})}{n - p - 3}\right]
\end{aligned} \tag{4.4}$$

且注意到

$$\begin{aligned}
& (\mathcal{I}_{obs} - \Psi_m(X)\hat{\theta})'(\mathcal{I}_{obs} - \Psi_m(X)\hat{\theta}) \\
&= \begin{pmatrix} (c_{obs} - \Psi_m(X)\hat{\beta})'(c_{obs} - \Psi_m(X)\hat{\beta}) & (c_{obs} - \Psi_m(X)\hat{\beta})'(z_{obs} - \Psi_m(X)\hat{\gamma}) \\ (z_{obs} - \Psi_m(X)\hat{\gamma})'(c_{obs} - \Psi_m(X)\hat{\beta}) & (z_{obs} - \Psi_m(X)\hat{\gamma})'(z_{obs} - \Psi_m(X)\hat{\gamma}) \end{pmatrix} \\
&= \begin{pmatrix} SSE_{c,m} & (c_{obs} - \Psi_m(X)\hat{\beta})'(z_{obs} - \Psi_m(X)\hat{\gamma}) \\ (z_{obs} - \Psi_m(X)\hat{\gamma})'(c_{obs} - \Psi_m(X)\hat{\beta}) & SSE_{z,m} \end{pmatrix}.
\end{aligned}$$

若取 $P = \Psi_m(X)(\Psi_m(X)'\Psi_m(X))^{-1}\Psi_m(X)'$, 令 p_{ij} 為其第 i 列第 j 行的元素, 則

$$I_n + P = \begin{pmatrix} 1 + p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & 1 + p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & 1 + p_{nn} \end{pmatrix},$$

$$\begin{aligned}
& (I_n + P) \otimes (\mathcal{I}_{obs} - \Psi_m(X)\hat{\theta})'(\mathcal{I}_{obs} - \Psi_m(X)\hat{\theta}) \\
&= \begin{pmatrix} (1 + p_{11})SSE_{c,m} & & & \\ & (1 + p_{11})SSE_{z,m} & & \\ & & \ddots & \\ & & & (1 + p_{nn})SSE_{z,m} \end{pmatrix}_{2n \times 2n}.
\end{aligned}$$

故 (4.4) 式可進一步寫成

$$\begin{aligned}
& \sum_{l=1}^n \text{Var}(c_{rep,l} | \mathcal{I}_{obs}) + \sum_{l=1}^n \text{Var}(z_{rep,l} | \mathcal{I}_{obs}) \\
&= \text{tr} \left[\frac{(I_n + \Psi_m(X)(\Psi_m(X)' \Psi_m(X))^{-1} \Psi_m(X)') \otimes (\mathcal{I}_{obs} - \Psi_m(X)\hat{\theta})'(\mathcal{I}_{obs} - \Psi_m(X)\hat{\theta})}{n-p-3} \right] \\
&= \frac{1}{n-p-3} \left[(n + \sum_{l=1}^n p_{lu}) SSE_{c,m} + (n + \sum_{l=1}^n p_{lu}) SSE_{z,m} \right].
\end{aligned}$$

因此

$$\begin{aligned}
& \sum_{l=1}^n \text{Var}(c_{rep,l} | \mathcal{I}_{obs}) \\
&= \frac{1}{n-p-3} \left\{ \left[(n + \sum_{l=1}^{\lfloor \frac{n}{2} \rfloor} p_{lu}) SSE_{c,m} + (n + \sum_{l=1}^{\lfloor \frac{n}{2} \rfloor} p_{lu}) SSE_{z,m} \right] + \left(\lfloor \frac{n+1}{2} \rfloor - \lfloor \frac{n}{2} \rfloor \right) SSE_{c,m} \right\}.
\end{aligned} \tag{4.5}$$

其中 $\lfloor \frac{n}{2} \rfloor$ 為不大於 $\frac{n}{2}$ 的最大整數。而 $\sum_{l=1}^n \text{Var}(e^{z_{rep,l}} | \mathcal{I}_{obs})$ 直接由 (3.7) 式可得

$$\sum_{l=1}^n \text{Var}(e^{z_{rep,l}} | \mathcal{I}_{obs}) = e^{2\Psi_m(X)\hat{\gamma}_l + Q_{ll}} (e^{Q_{l+n,l+n}} - 1). \tag{4.6}$$

將 (4.5) 式與 (4.6) 式代入 f_c, f_z 得到

$$\begin{aligned}
& f_c(p_1^*, \dots, p_n^*) \\
&= \frac{1}{n-p-3} \left\{ \left[(n + \sum_{l=1}^{\lfloor \frac{n}{2} \rfloor} p_{lu}) (SSE_{c,m_c} + SSE_{z,m_c}) \right] + \left(\lfloor \frac{n+1}{2} \rfloor - \lfloor \frac{n}{2} \rfloor \right) SSE_{c,m_c} \right\} \\
&+ \frac{k}{k+1} \sum_{l=1}^n (\Psi_{m_c}(X)\hat{\beta} - c_{x_l})^2.
\end{aligned}$$

$$f_z(q_1^*, \dots, q_n^*) = \sum_{l=1}^n \left[e^{2[\Psi_m(X)\hat{\gamma}]_l + Q_{l+n,l+n}} (e^{Q_{l+n,l+n}} - 1) + \frac{k}{k+1} (e^{\Psi_m(X)\hat{\gamma}_l + \frac{1}{2}Q_{l+n,l+n}} - e^{z_{x_l}})^2 \right],$$

$$Q = (I_n + \Psi_{m_z}(X)(\Psi_{m_z}(X)' \Psi_{m_z}(X))^{-1} \Psi_{m_z}(X)') \otimes (\mathcal{I}_{obs} - \Psi_{m_z}(X)\hat{\theta})'(\mathcal{I}_{obs} - \Psi_{m_z}(X)\hat{\theta}).$$

由於 $R_k(m) = f_c + f_z$ ，因此第一種區間距離函數下的模型選取準則為找一 m^* 使得 $m^* = \arg\{\min_m R_k(m)\}$ 。

4.2 第二種區間距離函數

第二種區間距離函數除了考慮兩區間端點的差異之外，也多考慮了兩區間重疊長度的資訊。我們將觀測區間和預測區間重疊的長度去除以預測區間的長度，這個比例代表預測區間中預測正確的部分，然而正確的比例越高，該預測區間產生的損失就應該越小。不僅如此，我們也希望對於擁有相同距離一大小的兩區間，超出觀測區間的損失要比落在觀測區間內的損失要大。此想法是源自於第1章中提及的焊接實驗之實務上的需求，對此焊接實驗而言，若電流的預測區間是落在圖1.1中『良好的焊接點』之範圍內，則其遠比落到此範圍外更佳。故除了考慮觀測區間與預測區間端點的差異外，我們還希望能盡量讓預測區間落在觀測區間內。這個目標可透過定義適當的區間距離函數來達成。

基於以上這兩個想法，我們將第一種區間距離函數的值再多加入一個區間重疊比例的懲罰項。令 $\hat{\mathcal{I}}$ =預測區間， \mathcal{I}_{obs} =觀測區間，符號 $|\cdot|$ 表示區間的長度。由於預測區間中預測正確的比例為

$$\frac{|\hat{\mathcal{I}} \cap \mathcal{I}_{obs}|}{|\hat{\mathcal{I}}|},$$

錯誤的比例為

$$1 - \frac{|\hat{\mathcal{I}} \cap \mathcal{I}_{obs}|}{|\hat{\mathcal{I}}|}.$$

由於預測正確的部分不應被納入懲罰的考量，故我們將第二種區間距離函數 d_2 定為

$$\begin{aligned} d_2(\hat{\mathcal{I}}, \mathcal{I}_{obs}) &= d_1(\hat{\mathcal{I}}, \mathcal{I}_{obs}) + d_1(\hat{\mathcal{I}}, \mathcal{I}_{obs}) \times \left(1 - \frac{|\hat{\mathcal{I}} \cap \mathcal{I}_{obs}|}{|\hat{\mathcal{I}}|}\right) + 0 \times \frac{|\hat{\mathcal{I}} \cap \mathcal{I}_{obs}|}{|\hat{\mathcal{I}}|} \\ &= d_1(\hat{\mathcal{I}}, \mathcal{I}_{obs}) \times \left(2 - \frac{|\hat{\mathcal{I}} \cap \mathcal{I}_{obs}|}{|\hat{\mathcal{I}}|}\right). \end{aligned}$$

其中 $d_1(\hat{\mathcal{I}}, \mathcal{I}_{obs})$ 為 $\hat{\mathcal{I}}$ 和 \mathcal{I}_{obs} 第一種區間距離函數值。因此我們可以寫出任意兩區間 $\mathcal{I}_1, \mathcal{I}_2$ 的第二種區間距離函數如下

$$\begin{aligned} d(\mathcal{I}_1, \mathcal{I}_2) &= d_1(\mathcal{I}_1, \mathcal{I}_2) + d_1(\mathcal{I}_1, \mathcal{I}_2) \times \left(1 - \frac{|\mathcal{I}_1 \cap \mathcal{I}_2|}{|\mathcal{I}_1|}\right) + 0 \times \frac{|\mathcal{I}_1 \cap \mathcal{I}_2|}{|\mathcal{I}_1|} \\ &= d_1(\mathcal{I}_1, \mathcal{I}_2) \times \left(2 - \frac{|\mathcal{I}_1 \cap \mathcal{I}_2|}{|\mathcal{I}_1|}\right). \end{aligned}$$

很明顯 d 滿足4.1節區間距離函數的2個條件。

接著我們由圖 4.1 來說明上述的距離函數可以達到超出觀測區間的損失要比落在觀測區間內的損失要大的目的。圖 4.1 的數字 0,1,2,3 分別代表觀測區間與 3 個預測區

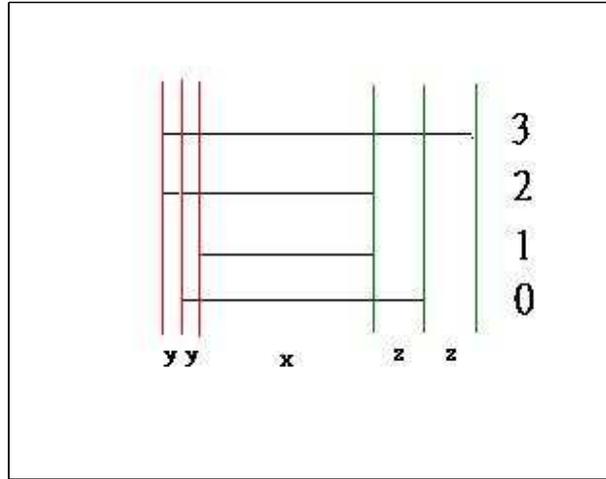


圖 4.1: 第二種區間距離函數

間，由圖可知預測區間一的兩端點落在觀測區間內，預測區間三的兩端點則皆落在觀測區間外且包住整個觀測區間。而預測區間二則有一端點落於觀測區間內，一端點落於觀測區間外。值得注意的是，在上一節的距離一之下，這 3 個預測區間與該觀測區間的差距都相同，都是 $(y^2 + z^2)$ 。但在距離二下，

$$(y^2 + z^2)\left(2 - \frac{x - y - z}{x - y - z}\right) = (y^2 + z^2), \quad (4.7)$$

第二個預測區間與觀測區間的距離為

$$(y^2 + z^2)\left(2 - \frac{x - z}{x + y - z}\right), \quad (4.8)$$

第三個預測區間與觀測區間的距離為

$$(y^2 + z^2)\left(2 - \frac{x}{x + y + z}\right), \quad (4.9)$$

很容易可以驗證 (4.7) 式最小, (4.8) 次之, (4.9) 是最大。故此距離可以達到我們要求。

令 $\mathcal{I}_1 = [u_1, v_1], \mathcal{I}_2 = [u_2, v_2]$, 並令

$$d_1(\mathcal{I}_1, \mathcal{I}_2) = [(u_1 - u_2)^2 + (v_1 - v_2)^2],$$

則由以上的討論知 d 可寫為

$$d(\mathcal{I}_1, \mathcal{I}_2) = \begin{cases} 2d_1(\mathcal{I}_1, \mathcal{I}_2), & u_2 > v_1; \\ (2 - \frac{v_1 - u_2}{v_1 - u_1})d_1(\mathcal{I}_1, \mathcal{I}_2), & u_1 < u_2 < v_1 < v_2; \\ (2 - \frac{v_2 - u_2}{v_1 - u_1})d_1(\mathcal{I}_1, \mathcal{I}_2), & u_1 < u_2 < v_2 < v_1; \\ (2 - \frac{v_1 - u_1}{v_1 - u_1})d_1(\mathcal{I}_1, \mathcal{I}_2), & u_2 < u_1 < v_1 < v_2; \\ (2 - \frac{v_2 - u_1}{v_1 - u_1})d_1(\mathcal{I}_1, \mathcal{I}_2), & u_2 < u_1 < v_2 < v_1; \\ 2d_1(\mathcal{I}_1, \mathcal{I}_2), & u_1 > v_2. \end{cases}$$

由 (2.2) 式之轉換有

$$[u_1, v_1] = [c_1 - e^{z_1}, c_1 + e^{z_1}],$$

$$[u_2, v_2] = [c_2 - e^{z_2}, c_2 + e^{z_2}].$$

令

$$d_1(\mathcal{I}_1, \mathcal{I}_2) = 2[(c_1 - c_2)^2 + (e^{z_1} - e^{z_2})^2],$$

故 d 可寫成

$$d(\mathcal{I}_1, \mathcal{I}_2) =$$

$$\begin{cases} 2d_1(\mathcal{I}_1, \mathcal{I}_2) & , c_2 - e^{z_2} > c_1 + e^{z_1}; \\ (2 - \frac{(c_1 + e^{z_1}) - (c_2 - e^{z_2})}{2e^{z_1}})d_1(\mathcal{I}_1, \mathcal{I}_2) & , c_1 - e^{z_1} < c_2 - e^{z_2} < c_1 + e^{z_1} < c_2 + e^{z_2}; \\ (2 - \frac{(c_2 + e^{z_2}) - (c_2 - e^{z_2})}{2e^{z_1}})d_1(\mathcal{I}_1, \mathcal{I}_2) & , c_1 - e^{z_1} < c_2 - e^{z_2} < c_2 + e^{z_2} < c_1 + e^{z_1}; \\ (2 - \frac{(c_1 + e^{z_1}) - (c_1 - e^{z_1})}{2e^{z_1}})d_1(\mathcal{I}_1, \mathcal{I}_2) & , c_2 - e^{z_2} < c_1 - e^{z_1} < c_1 + e^{z_1} < c_2 + e^{z_2}; \\ (2 - \frac{(c_2 + e^{z_2}) - (c_1 - e^{z_1})}{2e^{z_1}})d_1(\mathcal{I}_1, \mathcal{I}_2) & , c_2 - e^{z_2} < c_1 - e^{z_1} < c_2 + e^{z_2} < c_1 + e^{z_1}; \\ 2d_1(\mathcal{I}_1, \mathcal{I}_2) & , c_1 - e^{z_1} > c_2 + e^{z_2}. \end{cases}$$

將 d 代入 (3.1) 式的貝氏預測風險函數 f

$$\begin{aligned} f(\hat{\mathcal{I}}_{m,1}, \dots, \hat{\mathcal{I}}_{m,n}) &= \sum_{j=1}^n E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}} [L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})] \\ &= \sum_{j=1}^n E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}} [d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j})] + kd(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{obs,j}), \end{aligned}$$

其中計算 $E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}} [d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j})]$ 時，為了符號簡單，令

$$\hat{\mathcal{I}}_{m,j} = \begin{pmatrix} p_{m,j} & q_{m,j} \end{pmatrix}, \mathcal{I}_{rep,j} = \begin{pmatrix} w_j & y_j \end{pmatrix}, d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}) = 2[(w_j - p_{m,j})^2 + (y_j - q_{m,j})^2]$$

，可以看出 d 將平面分成互斥的6個區域，如圖 4.2。

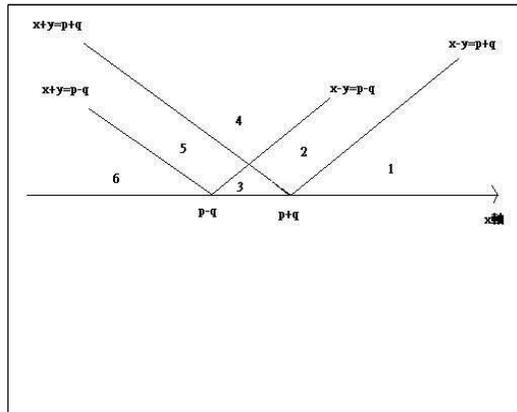


圖 4.2: 第二種區間距離函數將上半平面的切割

$$\begin{aligned}
& E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j})] \\
&= \int_0^\infty \int_{p_{m,j}+q_{m,j}+y_j}^\infty 2d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j})p(w_j, y_j|\mathcal{I}_{obs})dw_jdy_j \\
&+ \int_0^{q_{m,j}} \int_{p_{m,j}+q_{m,j}-y_j}^{p_{m,j}+q_{m,j}+y_j} \left[2 - \frac{(p_{m,j} + q_{m,j}) - (w_j - y_j)}{2q_{m,j}}\right] d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j})p(w_j, y_j|\mathcal{I}_{obs})dw_jdy_j \\
&+ \int_{q_{m,j}}^\infty \int_{p_{m,j}-q_{m,j}+y_j}^{p_{m,j}+q_{m,j}+y_j} \left[2 - \frac{(p_{m,j} + q_{m,j}) - (w_j - y_j)}{2q_{m,j}}\right] d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j})p(w_j, y_j|\mathcal{I}_{obs})dw_jdy_j \\
&+ \int_0^{q_{m,j}} \int_{p_{m,j}-q_{m,j}+y_j}^{p_{m,j}+q_{m,j}+y_j} \left[2 - \frac{(w_j + y_j) - (w_j - y_j)}{2q_{m,j}}\right] d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j})p(w_j, y_j|\mathcal{I}_{obs})dw_jdy_j \\
&+ \int_{q_{m,j}}^\infty \int_{p_{m,j}+q_{m,j}-y_j}^{p_{m,j}-q_{m,j}+y_j} \left[2 - \frac{(p_{m,j} + q_{m,j}) - (p_{m,j} - q_{m,j})}{2q_{m,j}}\right] d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j})p(w_j, y_j|\mathcal{I}_{obs})dw_jdy_j \\
&+ \int_0^{q_{m,j}} \int_{p_{m,j}-q_{m,j}-y_j}^{p_{m,j}-q_{m,j}+y_j} \left[2 - \frac{(w_j + y_j) - (p_{m,j} - q_{m,j})}{2q_{m,j}}\right] d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j})p(w_j, y_j|\mathcal{I}_{obs})dw_jdy_j \\
&+ \int_{q_{m,j}}^\infty \int_{p_{m,j}-q_{m,j}-y_j}^{p_{m,j}+q_{m,j}+y_j} \left[2 - \frac{(w_j + y_j) - (p_{m,j} - q_{m,j})}{2q_{m,j}}\right] d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j})p(w_j, y_j|\mathcal{I}_{obs})dw_jdy_j \\
&+ \int_0^\infty \int_{-\infty}^{p_{m,j}-q_{m,j}-y_j} 2d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j})p(w_j, y_j|\mathcal{I}_{obs})dw_jdy_j. \tag{4.10}
\end{aligned}$$

$p(w_j, y_j|\mathcal{I}_{obs})$ 為 $\mathcal{I}_{rep,j}|\mathcal{I}_{obs}$ 的貝氏預測密度函數。

4.2.1 區間預測與模型選取準則

將 (4.10) 式代入 (3.2) 式可得對因子設定 x_1, \dots, x_n 的區間預測式。由於最小化貝氏預測風險函數的計算很複雜，故這部分由電腦程式來處理。用到的技巧是蒙地卡羅積分 (Monte Carlo integration)。作法如下：

- 1：生成服從 $p(w_j, y_j|\mathcal{I}_{obs})$ 的隨機變數 500 筆。
- 2：看每筆亂數落在那 6 個區域的哪一個，代入 $E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j})]$ 求值，將這 500 個值平均，就得到一個給定 $p_{m,j}, q_{m,j}$ 下的貝氏預測風險函數值。
- 3：找出使得 $E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})]$ 最小的 $(p_{m,j}^*, q_{m,j}^*)$ ，即為 x_j 的區間預測值。

$(p_{m,j}^*, q_{m,j}^*)$ 為給定子模型 m 下的最佳區間預測值，將之代入 $E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})]$ 後記為 $E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[L(\hat{\mathcal{I}}_{m,j}^*, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})]$ 。此值可看成模型 m 下，預測區間造成的最小平均損失。預測一個未做過實驗的區間 \mathcal{I}_{new} 方法類似：

- 1 : 生成服從 $p(\mathcal{I}_{new}|\mathcal{I}_{obs})$ 的隨機變數500筆。
- 2 : 看每筆亂數落在那6個區域的哪一個, 代入 $E_{\mathcal{I}_{new}|\mathcal{I}_{obs}}[d(\hat{\mathcal{I}}_{m,new}, \mathcal{I}_{new})]$ 求值, 將這500個值平均, 就得到一個給定 $p_{m,new}, q_{m,new}$ 下的貝氏預測風險函數值。
- 3 : 找出使得 $E_{\mathcal{I}_{new}|\mathcal{I}_{obs}}[L(\hat{\mathcal{I}}_{m,new}, \mathcal{I}_{new}; \mathcal{I}_{obs,j})]$ 最小的 $(p_{m,new}^*, q_{m,new}^*)$, 即為 x_{new} 的區間預測值。

模型選取的部分, 由於我們希望所選出的模型能夠擁有最小的『最小平均損失』, 故若記

$$R_k(m) = \sum_{j=1}^n E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[L(\hat{\mathcal{I}}_{m,j}^*, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})].$$

則選模準則為:

$$\text{找一 } m^* \text{ 使得 } m^* = \arg\{\min_m R_k(m)\}.$$

使用蒙地卡羅積分 (Monte Carlo integration) 的作法如下:

- 1 : 生成服從 $p(w_j, y_j|\mathcal{I}_{obs})$ 的隨機變數500筆。
- 2 : 看每筆亂數落在那6個區域的哪一個, 代入 $E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j})]$ 求值, 將這500個值平均, 就得到一個給定 p_j, q_j 下的貝氏預測風險函數值。
- 3 : 找出使得 $E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[L(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})]$ 最小的 $(p_{m,j}^*, q_{m,j}^*)$, 將之代回記為 $E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[L(\hat{\mathcal{I}}_{m,j}^*, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})]$ 。
- 4 : 找一 m^* 使得 $m^* = \arg\{\min_m \sum_{j=1}^n E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}}[L(\hat{\mathcal{I}}_{m,j}^*, \mathcal{I}_{rep,j}; \mathcal{I}_{obs,j})]\}$ 。

第 5 章

實例分析

在Li et al. (2001)的論文中談到了一個焊接實驗,此實驗的目的在找出合適的電流範圍以產生良好的焊接點。該實驗因子有7個,如表 5.1。在焊接實驗中電流強度與

表 5.1: 實驗因子描述

因子	定義
Poor fitup(Ft)	兩片金屬板是否有放置好
Axial misal(Ax)	上下電極是否對齊
Angular misal(An)	金屬板與電極是否垂直
Electrode size(Ed)	電極大小
Force(F)	電極施壓強度
Weld time(T)	焊接時間
Current(I)	電流強度

焊接時間為兩個主要的實驗因子。由於高電流長時間會導致焊接點產生噴濺的現象,使得強度降低。而低電流短時間又無法提供足夠的能量產生焊接點,故一個合適的實驗便應該避開這兩種實驗區間,而將實驗點分佈於能產生品質良好焊接點的區域內。但合適的實驗區間一開始是未知的,所以該篇論文提出一個兩階段實驗法。第一階段的實驗是在找出合適的電流強度範圍,找法是先固定其他6個因子的水準組合,然後讓電流強度由大到小,當焊接點從有噴濺變成無噴濺時,將該電流強度記錄下來作為上界。當焊接點從有焊接點變成無焊接點時,將該電流強度記錄下來作為下界。我們對這6個因子每一水準組合,都如此的將電流強度區間紀錄下來,並把它視為區間反應變數,進而使用本文提出的方法來建模及選出最具預測力的模型。

在本實驗中, 區間反應變數是電流強度的上界與下界, 我們可透過 (2.2) 式將之轉換為電流區間中心值和 \ln (電流區間半長值)。本實驗的6個因子共有18組因子設定, 我們用 x_1, \dots, x_{18} 代表其水準組合, 其設計矩陣與觀測值如列於表 5.2。其中 u_x, v_x 為因子設定 x 時的電流強度下界與上界, 將它們透過 (2.2) 式之轉換得到 c_x, z_x 。

為了檢驗所選出之模型預測的能力, 我們只用後17組水準組合來分析, 而將第1個

表 5.2: 設計矩陣與觀測值

因子設定	Ft	Ax	An	Ed	F	T	u_x	v_x	c_x	z_x
x_1	0	0	0	6.35	3.3	12	10.56	12.44	11.50	-0.03
x_2	5	0	10	6.35	4.2	8	11.25	13.25	12.25	0
x_3	0	0	10	6.35	4.2	16	10.4	12.5	11.45	0.02
x_4	0	0	10	7.87	2.4	8	18.26	22.41	20.34	0.32
x_5	5	0	10	7.87	2.4	12	16.00	17.10	16.55	-0.26
x_6	0	0	0	7.87	4.2	16	16.40	19.00	17.7	0.11
x_7	0	0	0	4.50	2.4	8	8.51	10.16	9.34	-0.08
x_8	0	0	10	4.50	3.3	12	8.47	9.83	9.15	-0.17
x_9	5	0	10	4.50	3.3	16	7.85	9.75	8.8	-0.02
x_{10}	5	1.5	0	6.35	4.2	12	11.42	13.48	12.45	0.01
x_{11}	5	1.5	10	6.35	3.3	8	11.40	12.60	12.00	-0.22
x_{12}	0	1.5	0	6.35	2.4	8	11.50	13.50	12.50	0
x_{13}	5	1.5	10	4.50	2.4	16	7.10	8.50	7.80	-0.15
x_{14}	0	1.5	0	4.50	4.2	16	8.85	11.25	10.05	0.08
x_{15}	5	1.5	0	4.50	3.3	8	9.50	11.10	10.30	-0.09
x_{16}	0	1.5	0	7.87	3.3	12	15.67	17.87	16.77	0.04
x_{17}	5	1.5	10	7.87	4.2	12	16.16	17.61	16.88	-0.14
x_{18}	5	1.5	0	7.87	2.4	16	14.25	15.75	15.00	-0.12

水準組合的結果保留下來以供與預測區間相比較。為了簡化計算, 我們刪除掉一些因子, 而只挑了2個重要的因子 Electrode size 和 Force 來分析。以下將 Electrode size 用 A, Force用 B 來代表。由於因子 A,B 都是3水準, 各有線性效應與二次效應, 再加上兩者的交互作用 (包含 A 的線性效應對 B 的線性效應、A 的線性效應對 B 的二次效應、A 的二次效應對 B 的線性效應、A 的二次效應對 B 的二次效應), 因此模型矩陣中總共有8個效應。在模型一的假設下,8個效應可生成 $255(= 2^8 - 1)$ 個

子模型，下面我們分別使用兩個準則來選取最佳的子模型。第一個是在迴歸分析中常見的 *PRESS*，第二個則為本文提出的方法。選出模型後，再對原本的 17 個與第 1 個因子設定作區間預測，觀察它們之間的關係。

5.1 選模準則：*PRESS*

在迴歸分析中，當模型選取的目的是為了『預測』的時候，常用的準則是 *PRESS*。其計算公式為：

$$PRESS = \sum_i \hat{\epsilon}_{(i)},$$

其中 $\hat{\epsilon}_{(i)}$ 為去除掉第 i 筆數據後由迴歸模型得到的殘差項。由於我們目前處理數據的反應值是一個區間，因此我們分別對區間中心與 $\ln(\text{區間半長})$ 建立迴歸模型，再選出各自 *PRESS* 值最小的子模型。接著再對前 17 個因子設定下的區間作預測。對於 c_x ，選出的效應為 $\psi_{A,l}(x), \psi_{A,q}(x)$ ，分別代表因子 A 的線性效應與二次效應。而對於 z_x ，選出的效應為 $\psi_{B,q}(x)$ ，表示因子 B 的二次效應。將模型選取結果與 *PRESS* 值整理成表 5.3。由於我們是對 c_x 與 z_x 各別選取模型，故它們會有各自的 *PRESS* 值。而本文

表 5.3: *PRESS* 選出的模型

反應變數	選取效應	<i>PRESS</i> 值
c_x 的模型	$\psi_{A,l}(x), \psi_{A,q}(x)$	23.497
z_x 的模型	$\psi_{A,l}(x), \psi_{B,q}(x), \psi_{A,q}(x)\psi_{B,q}(x)$	0.354

提出的方法由於將 c_x 與 z_x 視為區間反應變數，因此只會有一個對應的值 (*MinRisk*，會在下節說明)。利用 *PRESS* 所選取的模型之區間預測式為 c_x 與 z_x 各自的最小平方估計式：

$$\begin{aligned} \hat{c}_x &= 12.8586 + 3.9838\psi_{A,l}(x) + 0.3643\psi_{A,q}(x) \\ \hat{z}_x &= -0.0488 + 0.0319\psi_{A,l}(x) + 0.0253\psi_{B,q}(x) - 0.0233\psi_{A,q}(x)\psi_{B,q}(x) \end{aligned} \quad (5.1)$$

利用 (5.1) 式預測出這 18 個因子設定的區間後，將它與觀察值比較如圖 5.1。圖中符號 『*』 代表利用 (5.1) 得到的預測區間，方形的符號代表原始的觀測數據形成

的區間, 以下簡稱觀測區間。圖 5.1 中區間的順序由下往上分別是 x_1 到 x_{18} 。我們發現在 x_4, x_{13}, x_{18} 的設定下, 預測區間與觀測區間的差異較大, 這顯示了可能是預測的不好, 也有可能該觀測區間是離群值。而 x_1 是一新的水準組合, 我們可以看到(5.1) 式得到的預測區間與 x_1 的觀測區間差異不大。

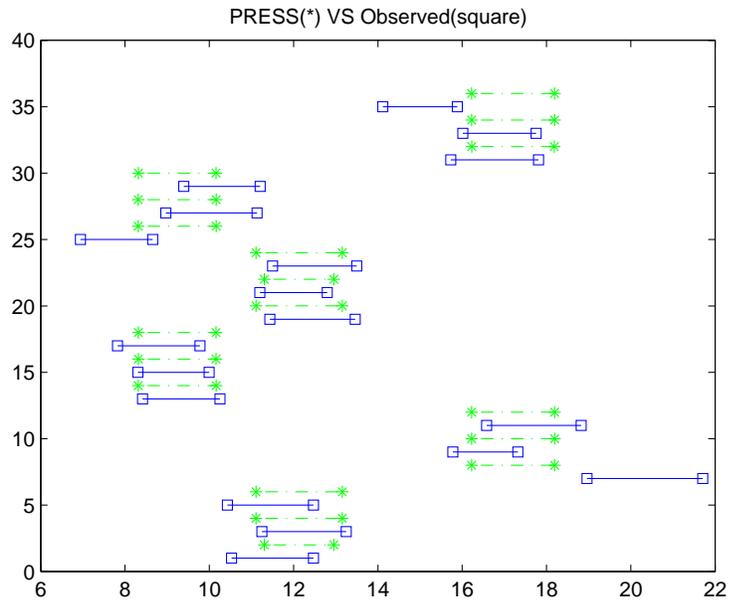


圖 5.1: PRESS與觀測區間之比較

5.2 選模準則: 最小貝氏預測風險函數

我們要找一個 c_x 子模型與 z_x 子模型的組合, 看哪一種組合會有最小的最小貝氏預測風險值。正如在第四章所提及, 第一種區間距離函數由於是平方損失函數, 故可以分解成 c_x 跟 z_x 各別的风险函數的和, 所以可知這與各別對 c_x, z_x 來作模型選取的意思是等價的。但第二種區間距離函數的形式就不像第一種區間距離函數可做如此分解, 故要針對每種 c_x 和 z_x 的子模型組合來看哪一種的最小貝氏預測風險值最小。底下我們亦想觀察模型擬合權重 k 的大小對區間預測的影響, 故對每個損失函數都取 $k = 0, 1, 100$ 這3個值。

5.2.1 第一種區間距離函數

首先取 $k = 0$ ，代表不考慮模型擬合，則模型選取的結果如表 5.4。 $MinRisk$ 為固

表 5.4: 距離一, $k=0$ 選出的模型 M_0^*

反應變數	選取效應	$MinRisk$ 值
c_x 的模型	$\psi_{A,l}(x), \psi_{A,q}(x)$	89.402
z_x 的模型	$\psi_{A,l}(x), \psi_{B,q}(x), \psi_{A,q}(x)\psi_{B,q}(x)$	

定 c_x 和 z_x 的模型組合下，最小的貝氏預測風險函數值，因此不像上節有兩個 $PRESS$ 值。

模型選取的結果與表 5.3 的模型是相同的。若記表 5.4 中選出的模型為 M_0^* ，則對 x_2, \dots, x_{18} 的區間預測式為：

$$(\hat{\mathcal{I}}_{m^*,2}^*, \hat{\mathcal{I}}_{m^*,3}^*, \dots, \hat{\mathcal{I}}_{m^*,18}^*) = \arg\{ \min_{\hat{\mathcal{I}}_{m^*,2}^*, \dots, \hat{\mathcal{I}}_{m^*,18}^*} \sum_{j=2}^{18} E_{\mathcal{I}_{rep,j} | \mathcal{I}_{obs}} [d(\hat{\mathcal{I}}_{m,j}^*, \mathcal{I}_{rep,j})] \}, \quad (5.2)$$

而對 x_1 的區間預測式為：

$$\hat{\mathcal{I}}_{m^*,1}^* = \arg\{ \min_{\hat{\mathcal{I}}_{m^*,1}^*} E_{\mathcal{I}_1 | \mathcal{I}_{obs}} [L(\hat{\mathcal{I}}_{m^*,1}^*, \mathcal{I}_1; \mathcal{I}_{obs,j})] \} \quad (5.3)$$

利用 (5.2) 式預測出後 17 個因子設定的區間後，將它與觀測區間比較如圖 5.2。圖中符號『。』代表利用 (5.2) 得到的預測區間，方形的符號代表觀測區間，圖 5.2 中區間的順序由下往上分別是 x_1 到 x_{18} 。我們發現在 x_4, x_{13}, x_{18} 下，預測區間與觀測區間的差異較大，有可能是預測的不好，也有可能該觀測區間為離群值。使用 (5.3) 式得到 x_1 的預測區間，可發現 x_1 的數據雖然沒有納入模型選取的步驟中，但該預測區間與觀測區間也相當接近。圖 5.3 展示了 (5.1) 式與 (5.2), (5.3) 式之預測區間的比較。由圖 5.3 可發現這兩種方法得到的預測區間相當近似，而且最小平方估計式得到的預測區間會稍微短一點。

接著取 $k = 1$ ，其表示模型擬合與模型預測同等重要。模型選取的結果如表 5.5。與 $k = 0$ 得到的模型相同，但 $MinRisk$ 的值較大。若記該選出的模型為 M_1^* ，則區間

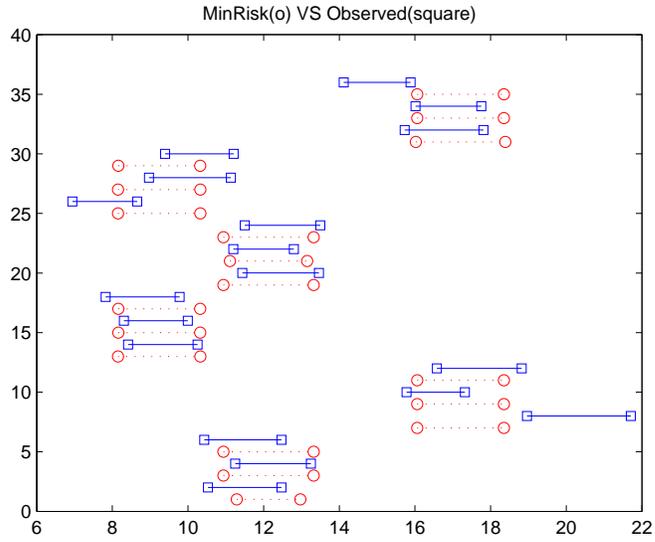


圖 5.2: 距離一, $k=0$, 與觀測區間之比較

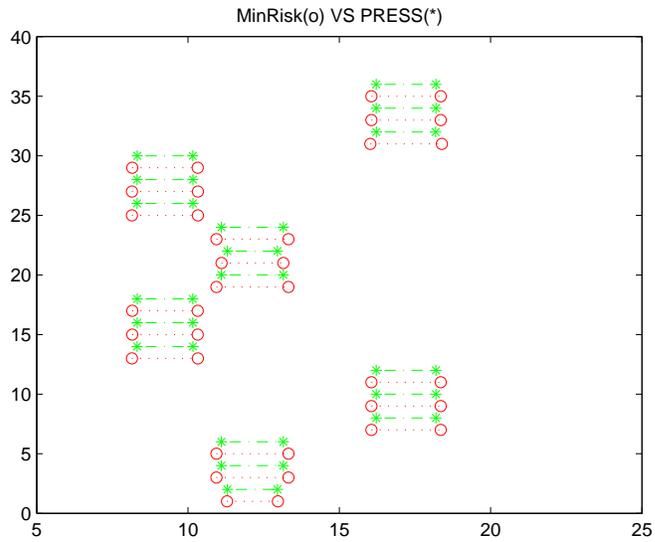


圖 5.3: 距離一, $k=0$, 與 $PRESS$ 之比較

表 5.5: 距離一, k=1 選出的模型 M_1^*

反應變數	選取效應	MinRisk值
c_x 的模型	$\psi_{A,l}(x), \psi_{A,q}(x)$	110.557
z_x 的模型	$\psi_{A,l}(x), \psi_{B,q}(x), \psi_{A,q}(x)\psi_{B,q}(x)$	

預測式為:

$$(\hat{\mathcal{I}}_{m^*,2}^*, \hat{\mathcal{I}}_{m^*,3}^*, \dots, \hat{\mathcal{I}}_{m^*,18}^*) = \arg\{ \min_{\hat{\mathcal{I}}_{m^*,2}^*, \dots, \hat{\mathcal{I}}_{m^*,18}^*} \sum_{j=2}^{18} (E_{\mathcal{I}_{rep,j} | \mathcal{I}_{obs}} [d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j})] + d((\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{obs,j})) \}, \quad (5.4)$$

利用 (5.4) 式預測出後 17 個因子設定的區間後，將它與觀測區間比較如圖 5.4。圖中符號『。』代表利用 (5.4) 得到的預測區間，方形的符號代表觀測區間。圖 5.4 中區間的順序由下往上分別是 x_1 到 x_{18} 。我們發現預測區間與觀測區間的差異有減小的趨勢，這是因為對模型擬合的要求提高導致。 x_1 由於我們將之視為未做實驗的設定，故預測式和 (5.3) 相同，得到的預測區間也和 $k = 0$ 時一樣。圖 5.5 展示了 (5.1) 式與

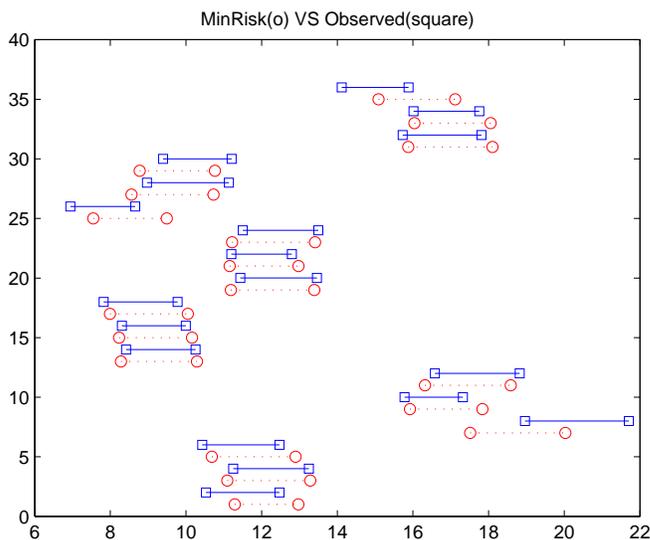


圖 5.4: 距離一, k=1, 與觀測區間之比較

(5.3), (5.4) 式之預測區間的比較。由圖可發現這兩種方法得到的預測區間除了 x_4 以

外都相當近似, 而且 (5.1) 式得到的預測區間會稍微短一點。至於在 x_4 時, 預測區間明顯往觀測區間移動, 因此兩者差異較 $k = 0$ 要明顯。

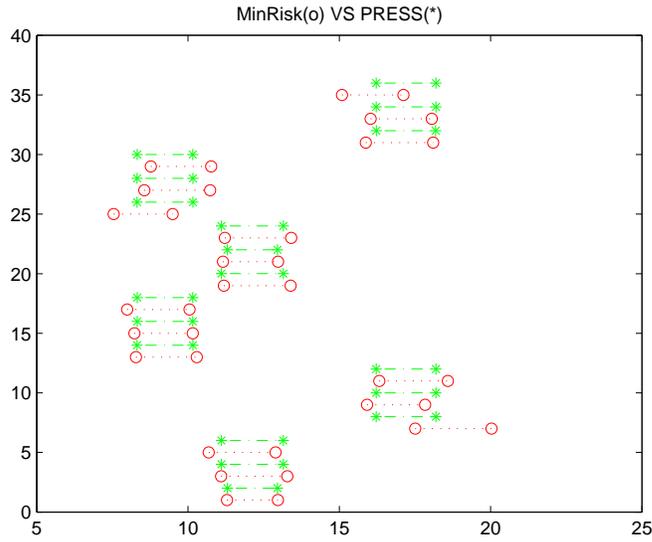


圖 5.5: 距離一, $k=1$, 與 $PRESS$ 之比較

接著取 $k = 100$, 其表示重視模型擬合的程度遠超過模型預測, 由前面的經驗, 我們預期此時後 17 個設定的預測區間應該要和觀測區間相當接近。模型選取的結果如表 5.6。與 $k = 0$ 或 1 時得到的模型相同, 但 $MinRisk$ 的值更大。若記該選出的模型

表 5.6: 距離一, $k=100$ 選出的模型 M_{100}^*

反應變數	選取效應	$MinRisk$ 值
c_x 的模型	$\psi_{A,l}(x), \psi_{A,q}(x)$	131.291
z_x 的模型	$\psi_{A,l}(x), \psi_{B,q}(x), \psi_{A,q}(x)\psi_{B,q}(x)$	

為 M_{100}^* , 則區間預測式為:

$$(\hat{\mathcal{I}}_{m^*,2}^*, \hat{\mathcal{I}}_{m^*,3}^*, \dots, \hat{\mathcal{I}}_{m^*,18}^*) = \arg\left\{ \min_{\hat{\mathcal{I}}_{m^*,2}^*, \dots, \hat{\mathcal{I}}_{m^*,18}^*} \sum_{j=2}^{18} (E_{\mathcal{I}_{rep,j} | \mathcal{I}_{obs}} [d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j})] + 100d((\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{obs,j})) \right\}, \quad (5.5)$$

利用 (5.5) 式預測出後 17 個因子設定區間的預測值後, 將它與觀測區間比較如圖 5.6。圖中符號『.』代表利用 (5.5) 得到的預測區間, 方形的符號代表觀測區間。圖 5.6 中

區間的順序由下往上分別是 x_1 到 x_{18} 。我們發現預測區間與觀測區間的幾乎重疊，與我們的預期一樣。而 x_1 仍用(5.3)來做預測，故結果和之前一樣。圖5.7展示了(5.1)式

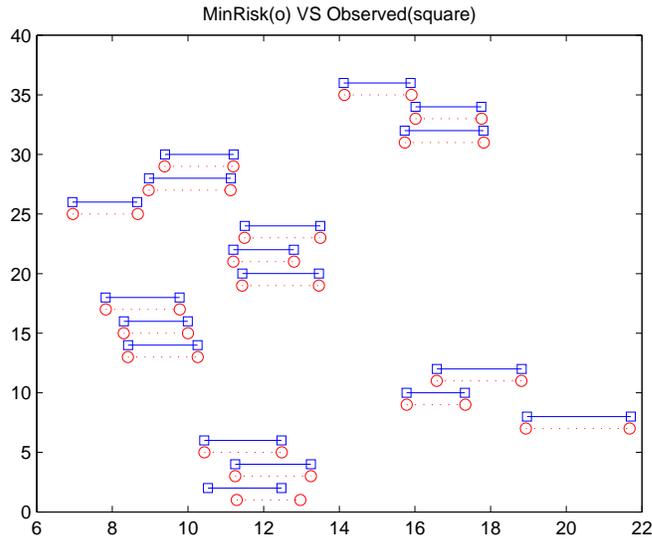


圖 5.6: 距離一, $k=100$, 與觀測區間之比較

與(5.3), (5.5)式之預測區間的比較, 可發現此圖與圖5.1幾乎一樣。

由以上的觀察, 我們發現 k 扮演的角色就是在預測與擬合之間做取捨。 $k = 0$ 時, 利用(5.1)式得到的預測區間和(5.2)式差不多, 但當 k 變大到100時, (5.5)式得到的預測區間幾乎和原本觀察到的區間一樣。這其實可以從下面的性質得知:

$$\lim_{k \rightarrow \infty} \frac{E(c_{rep,l} | \mathcal{I}_{obs}) + kc_{x_l}}{k+1} = c_{x_l}$$

$$\lim_{k \rightarrow \infty} \ln\left(\frac{E(e^{z_{rep,l}} | \mathcal{I}_{obs}) + ke^{z_{x_l}}}{k+1}\right) = z_{x_l}.$$

因此當 k 夠大的時候預測區間必定會和觀察區間相當接近。然而這就類似迴歸分析中直接拿觀測值來當做預測值一樣, 並不一定可靠。

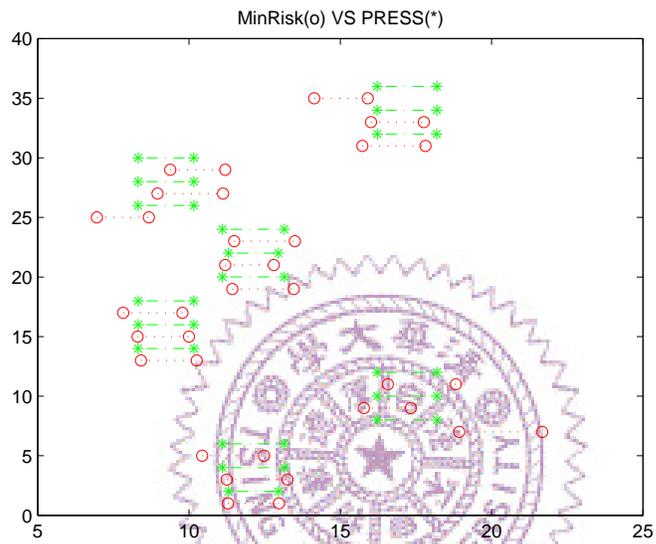


圖 5.7: 距離一, $k=100$, 與 $PRESS$ 之比較

5.2.2 第二種區間距離函數

首先取 $k = 0$ ，代表不考慮模型擬合，則模型選取的結果如表 5.7。與結果與用距

表 5.7: 距離二, $k=0$ 選出的模型 M_0^*

反應變數	選取效應	<i>MinRisk</i> 值
c_x 的模型	$\psi_{A,l}(x), \psi_{A,q}(x)$	14.759
z_x 的模型	全部的效應	

離一, $k = 0$ 時不同，其中 c_x 的部分保留 A 的線性與二次效應，而 z_x 的部分則選入所有
的效應。若記表 5.7 中選出的模型為 M_0^* ，則對 x_2, \dots, x_{18} 的區間預測式為：

$$(\hat{\mathcal{I}}_{m^*,2}^*, \hat{\mathcal{I}}_{m^*,3}^*, \dots, \hat{\mathcal{I}}_{m^*,18}^*) = \arg\{\min_{\hat{\mathcal{I}}_{m^*,2}^*, \dots, \hat{\mathcal{I}}_{m^*,18}^*} \sum_{j=2}^{18} E_{\mathcal{I}_{rep,j}} | \mathcal{I}_{obs} [d(\hat{\mathcal{I}}_{m,j}^*, \mathcal{I}_{rep,j})]\}, \quad (5.6)$$

而對 x_1 的區間預測式為：

$$\hat{\mathcal{I}}_{m^*,1}^* = \arg\{\min_{\hat{\mathcal{I}}_{m^*,1}^*} E_{\mathcal{I}_1} | \mathcal{I}_{obs} [L(\hat{\mathcal{I}}_{m^*,1}^*, \mathcal{I}_1; \mathcal{I}_{obs,j})]\} \quad (5.7)$$

利用 (5.6) 式預測出後 17 個因子設定的區間後，將它與觀測區間比較如圖 5.8。圖中符號『。』代表利用 (5.6) 得到的預測區間，方形的符號代表觀測區間，圖 5.8 中區間的順序由下往上分別是 x_1 到 x_{18} 。我們發現在 x_4, x_{13}, x_{18} 下，預測區間與觀測區間的差異較大，有可能是離群值。使用 (5.7) 式得到 x_1 的預測區間，可發現 x_1 的數據雖然沒有納入模型選取的步驟中，但預測區間與觀測區間重疊也相當接近。圖 5.9 展示了 (5.1) 式與 (5.6), (5.7) 式之預測區間的比較。由圖 5.9 可發現這兩種方法得到的預測區間相當近似，而且最小平方估計式得到的預測區間會稍微短一點。

接著取 $k = 1$ ，其表示模型擬合與模型預測同等重要。模型選取的結果如表 5.8。與 $k = 0$ 得到的模型相同，但 *MinRisk* 的值較大。若記該選出的模型為 M_1^* ，則區間預測式為：

$$(\hat{\mathcal{I}}_{m^*,2}^*, \hat{\mathcal{I}}_{m^*,3}^*, \dots, \hat{\mathcal{I}}_{m^*,18}^*) = \arg\{\min_{\hat{\mathcal{I}}_{m^*,2}^*, \dots, \hat{\mathcal{I}}_{m^*,18}^*} \sum_{j=2}^{18} (E_{\mathcal{I}_{rep,j}} | \mathcal{I}_{obs} [d(\hat{\mathcal{I}}_{m,j}^*, \mathcal{I}_{rep,j})] + d((\hat{\mathcal{I}}_{m,j}^*, \mathcal{I}_{obs,j})))\}, \quad (5.8)$$

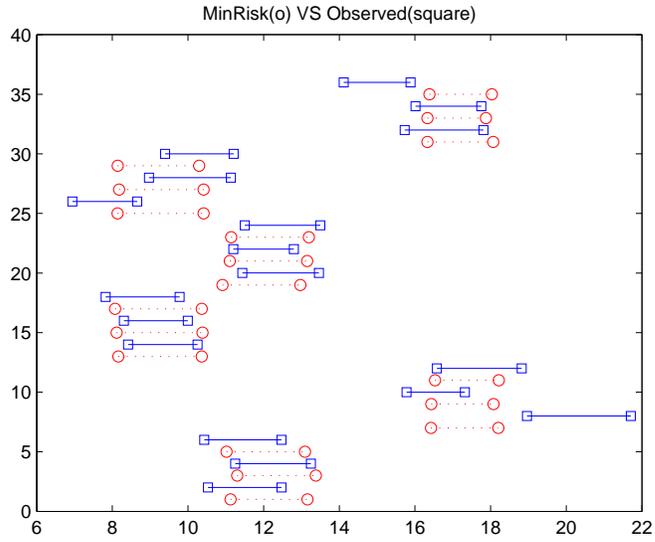


圖 5.8: 距離二, $k=0$, 與觀測區間之比較

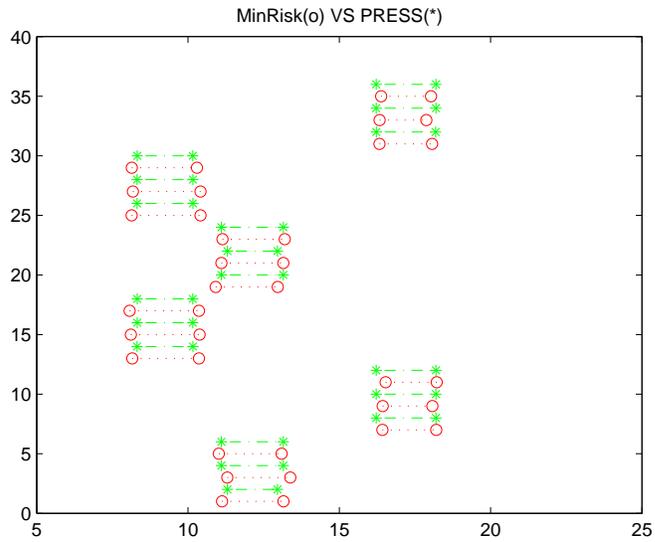


圖 5.9: 距離二, $k=0$, 與 $PRESS$ 之比較

表 5.8: 距離二, $k=1$ 選出的模型 M_1^*

反應變數	選取效應	$MinRisk$ 值
c_x 的模型	$\psi_{A,l}(x), \psi_{A,q}(x)$	19.526
z_x 的模型	全部的效應	

利用 (5.8) 式預測出後 17 個因子設定的區間後，將它與觀測區間比較如圖 5.10。圖中符號『。』代表利用 (5.8) 得到的預測區間，方形的符號代表觀測區間。圖 5.10 中區間的順序由下往上分別是 x_1 到 x_{18} 。我們發現預測區間與觀測區間的差異有減小的趨勢，尤其是 x_4 設定最明顯，這是因為對模型擬合的要求提高導致。 x_1 由於我們將之視為未做實驗的設定，故預測式和 (5.7) 相同，得到的預測區間也和 $k = 0$ 時一樣。圖 5.11

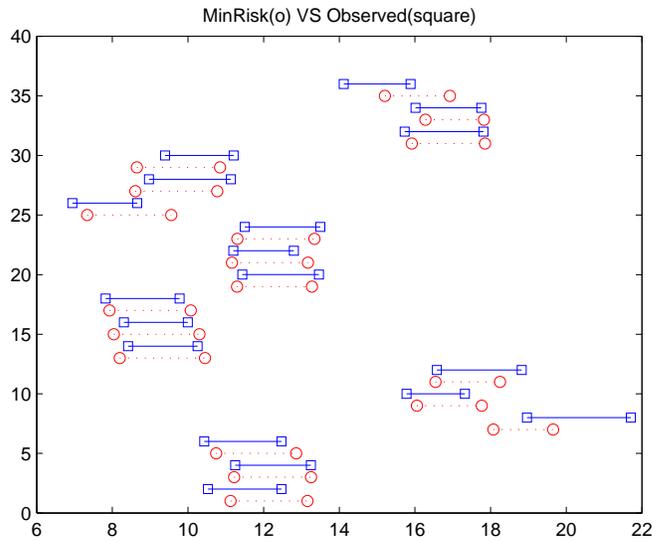


圖 5.10: 距離二, $k=1$ ，與觀測區間之比較

展示了 (5.1) 式與 (5.7), (5.8) 式之預測區間的比較。由圖可發現這兩種方法得到的預測區間除了 x_4 以外都相當近似，而且 (5.1) 式得到的預測區間會稍微短一點。至於在 x_4 時，預測區間明顯往觀測區間移動，因此兩者差異較 $k = 0$ 要明顯。

接著取 $k = 100$ ，其表示重視模型擬合的程度遠超過模型預測，由前面的經驗，我們預期此時後 17 個設定的預測區間應該要和觀測區間相當接近。模型選取的結果如表 5.9。與 $k = 0$ 或 1 時得到的模型相同，但 $MinRisk$ 的值更大。若記該選出的模型

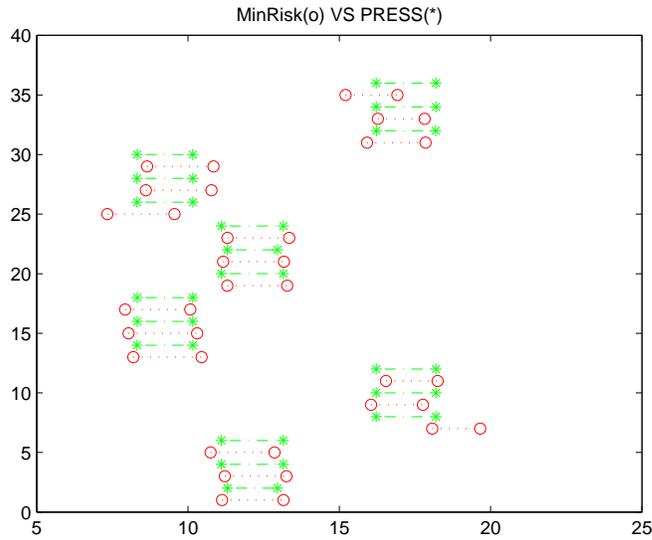


圖 5.11: 距離二, $k=1$, 與 $PRESS$ 之比較

表 5.9: 距離二, $k=100$ 選出的模型 M_{100}^*

反應變數	選取效應	$MinRisk$ 值
c_x 的模型	$\psi_{A,l}(x), \psi_{A,q}(x)$	13.805
z_x 的模型	全部的效應	

為 M_{100}^* , 則區間預測式為:

$$(\hat{\mathcal{I}}_{m^*,2}^*, \hat{\mathcal{I}}_{m^*,3}^*, \dots, \hat{\mathcal{I}}_{m^*,18}^*) = \arg\left\{ \min_{\hat{\mathcal{I}}_{m^*,2}^*, \dots, \hat{\mathcal{I}}_{m^*,18}^*} \sum_{j=2}^{18} (E_{\mathcal{I}_{rep,j}|\mathcal{I}_{obs}} [d(\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{rep,j})] + 100d((\hat{\mathcal{I}}_{m,j}, \mathcal{I}_{obs,j})) \right\}, \quad (5.9)$$

利用 (5.9) 式預測出後 17 個因子設定區間的預測值後, 將它與觀測區間比較如圖 5.12。圖中符號『。』代表利用 (5.9) 式得到的預測區間, 方形的符號代表觀測區間。圖 5.12 中區間的順序由下往上分別是 x_1 到 x_{18} 。我們發現預測區間與觀測區間的幾乎重疊, 與我們的預期一樣。而 x_1 仍用 (5.7) 式來做預測, 故結果和之前一樣。圖 5.13 展示了 (5.1) 式與 (5.7), (5.9) 式之預測區間的比較, 可發現此圖與圖 5.1 幾乎一樣。

由以上的觀察, 我們發現 k 扮演的角色就是在預測與擬合之間做取舍。 $k = 0$ 時, 利用 (5.1) 式得到的預測區間和 (5.6) 式差不多, 但當 k 變大到 100 時, (5.9) 式得到

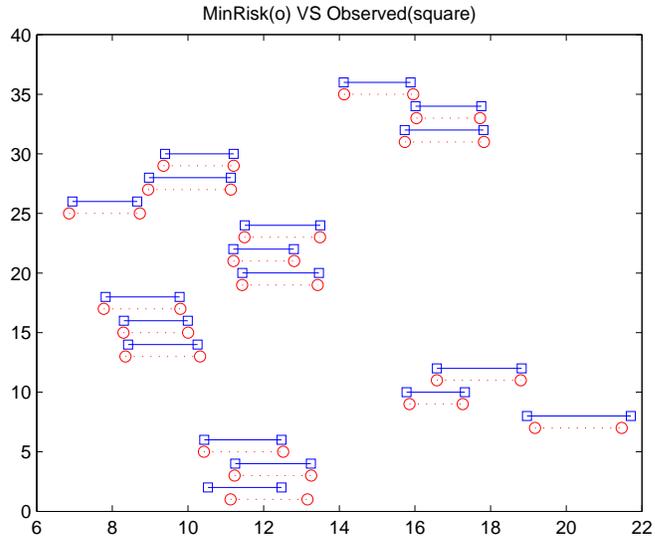


圖 5.12: 距離二, $k=100$, 與觀測區間之比較

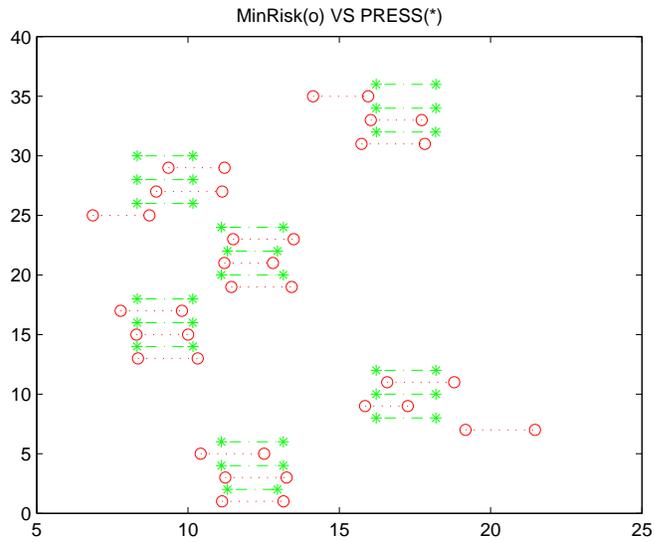


圖 5.13: 距離二, $k=100$, 與 $PRESS$ 之比較

的預測區間幾乎和原本觀察到的區間一樣。

5.2.3 兩種區間距離函數的比較

從圖 5.14 到圖 5.16, 我們發現不管 k 為多少, 第一種區間距離函數所預測出來的區間都比第二種的要長一些, 這是因為第二種區間距離函數對預測區間超出與不及觀測區間的部分不再是給予一樣的懲罰, 而會納入它們重疊部分的資訊來做調整, 使得超出觀測區間的懲罰較重。

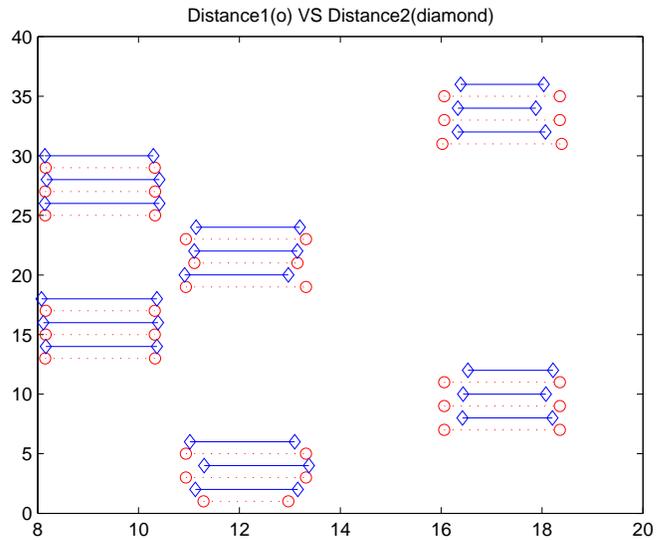


圖 5.14: $k=0$, 兩種距離的比較

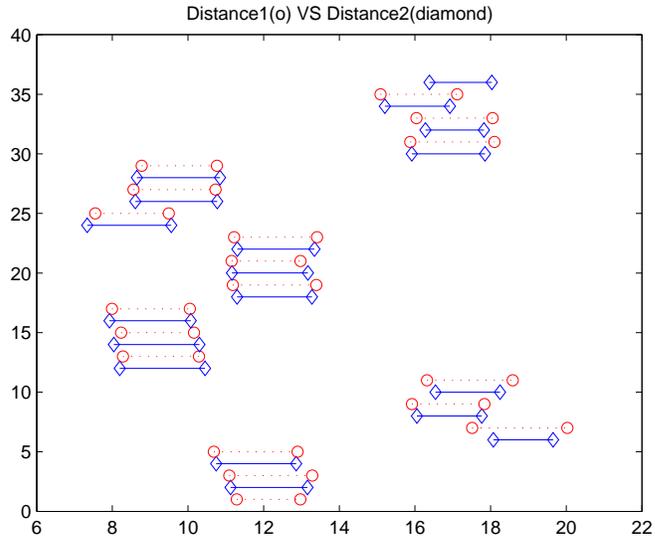


圖 5.15: $k=1$, 兩種距離的比較

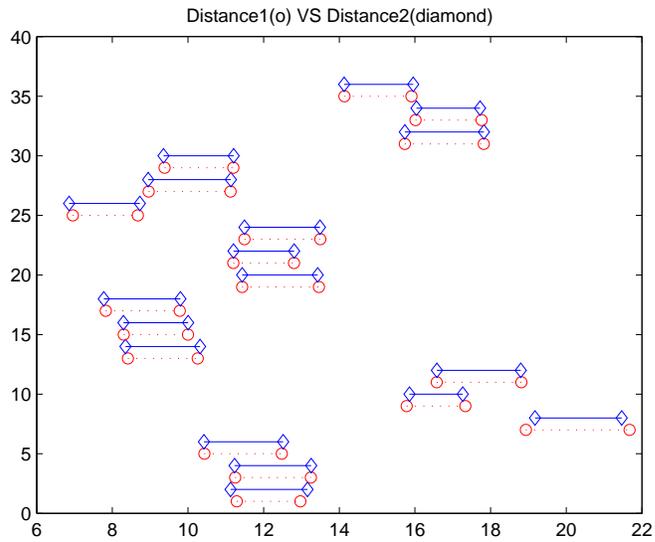


圖 5.16: $k=100$, 兩種距離的比較

第 6 章

結論

本文在決策理論的架構下發展了針對區間反應變數的預測式與模型選取準則。我們定義了兩區間的距離與損失函數，接著在一般化的條件下推導出區間預測式，再利用該式進而得到模型選取的準則。接著我們將區間距離函數指定形式，第一種類似單一反應變數中的平方誤差損失，雖然其中有些分配不易推導，但可用大樣本的漸進分佈來處理。第二種區間距離函數則是在第一種距離中再加入了兩區間重疊長度比例的考量，形式複雜許多，推導不易故用電腦程式來模擬。在第 5 章我們實際用了 Li et al. (2001) 的實驗數據來檢驗這些區間預測式與模型選取準則的表現，而我們發現使用本文的方法在模型選取時會選出一個效應較多的模型，尤其在第二種區間距離函數下更是明顯。而區間預測式隨著 k 值增大時，也符合預期的接近了觀測區間。由於第二種區間距離函數不像第一種只考慮兩區間端點的差異，而多考慮了兩區間重疊的長度，故在 Li et al. (2001) 的實驗數據下，我們看到使用第二種區間距離函數得到的區間預測值幾乎都比第一種的要短，這也是符合我們所預期的。

由於本文在數據背後的假設只考慮了模型一，在後續的研究上可以試著用其它的模型。然而區間距離函數的形式也很多樣化，本文並未說明這兩種區間距離函數有什麼好的性質，而純粹以直觀的角度建立了兩種區間距離函數，故找出一個具有好的性質的區間距離函數也是一個可以繼續努力的目標。

參考文獻

- Gelfand, A.E. and Ghosh, S.K. (1998), “Model choice:a minimum poerior predictive loss approach”, *Biometrika*, 85, 1–11.
- Gupta, A.K. and Nagar, D.K. (2000), *Matrix variate distribution*, Chapman Hall.
- Li, W., Cheng, S.W., Hu, S.J., and Shriver, J. (2001), “Statistical investigation on resistance spot welding quality using a two-stage,sliding-level experiment”, *Journal of Manufacturing Science and Engineering*, 123, 513–520.
- Ng, V.M. (2000), “A note on predictive inference for multivariate elliptically contoured distributions”, *Communication in Statistics-Theory and Methods*, 29, 477–483.
- Wu, C.F. and Hamada, Michael (2000), *Experiments:Planning,Analysis,and Parameter Design Optimization*, Wiley Series.

