# 2 模糊線性迴歸的介紹

## 2.1 模糊數及其運算

現今科學所欲研究物件之結構複雜性日益增加，人類的知識語言因本身的主觀意識、不同時間、環境的變遷，使得科學家無法清楚研究物件的真實本質，因此透過建立適當的數學模式，對於一些較無確定性的資料給予一個集合，稱之爲模糊數。

利用隸屬度函數可以描述模糊數的性質，他是模糊理論的最基本概念，透過隸屬度函數，我們才能對模糊數進行量化，也才有可能利用精準的數學方法，去分析和處理模糊性的資訊。通常隸屬度函數的定義方式可以分成數值及函數兩種：數值定義方式又稱爲離散化(discretization)隸屬度函數，他是直接給定有限模糊集合內每個元素的隸屬度。函數定義的方式又可稱爲連續化(continuous)隸屬度函數，函數定義的表現，可以是無限模糊集合的元素及其隸屬度之間的關係，也可以是有限模糊集合的元素及其隸屬度之間的關係，但在處理實際問題上僅討論有限的情況。以下是具有相同類型隸屬度函數之模糊數的運算，其定義爲：

定義 **2.1** *設 $A = (m_a, \alpha_a, \beta_a), B = (m_b, \alpha_b, \beta_b), \lambda \in \mathbb{R}^+$，其中 $m_a, m_b$ 為中心，$\alpha_a, \alpha_b$ 與 $\beta_a, \beta_b$ 分別為左右分展度，且 $A$ 與 $B$ 有相同類型的隸屬度函數，則我們有以下的運算：*

1. $A + B = (m_a + m_b, \alpha_a + \alpha_b, \beta_a + \beta_b)$
2. $\lambda A = (\lambda m_a, \lambda\alpha_a, \lambda\beta_a)$

即兩個模糊數相加的結果，爲中心與中心相加，左分展度與右分展度相加；若一個大於零的數乘上一個模糊數，則分別爲中心和左右分展度分別乘上此正數。但是在此，我們無法定義出一個模糊數乘上一個小於零的實數，也因

此無法定義出模糊數之間的減法。我們需要一套尋找模糊數間距離的方法，在2.3節中將介紹常見模糊數間距離的定義。

## 2.2 一般模糊線性迴歸模型

考慮下列的模糊線性迴歸模型，爲了方便起見稱其爲模型I：

$$\text{模型I} \quad y_i = A_0 + A_1 x_{i1} + A_2 x_{i2} + ... + A_p x_{ip}, \quad i = 1, 2, ..., n \tag{2.1}$$

其中自變數值 $x_{ij}$ 爲實數，應變數值 $y_i = [c_i - s_i, c_i + s_i]$ 爲模糊數，且 $c_i$ 爲資料中心，$s_i$ 爲資料分展度，$A_m = [a_m - r_m, a_m + r_m]$ 是模糊迴歸參數，具有和 $y_i$ 相同的隸屬度函數。在沒有模糊數距離概念下，我們如何估計 $A_m$ 呢？首先將 $y_{Li} = c_i - s_i$,$y_{Ri} = c_i + s_i$ 視爲樣本的左右端點。對左端點 $\{(y_{Li}, x_{i1}, x_{i2}, ..., x_{ip})|i = 1, 2, ..., n\}$ 資料(右端點 $\{(y_{Ri}, x_{i1}, x_{i2}, ..., x_{ip})|i = 1, 2, ..., n\}$ 資料)配置線性模型 $y = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p$，分別得到：

$$\hat{y}_{Li} = \hat{L}_0 + \hat{L}_1 x_{i1} + \hat{L}_2 x_{i2} + ... + \hat{L}_p x_{ip}, i = 1, 2, ...n$$

$$\hat{y}_{Ri} = \hat{R}_0 + \hat{R}_1 x_{i1} + \hat{R}_2 x_{i2} + ... + \hat{R}_p x_{ip}, i = 1, 2, ...n$$

則 $\hat{A}_m = [\hat{a}_m - \hat{r}_m, \hat{a}_m + \hat{r}_m]$ 其中 $\hat{a}_m = \frac{\hat{L}_m + \hat{R}_m}{2}, \hat{r}_m = \frac{|\hat{R}_m - \hat{L}_m|}{2}$。如此方法所得到的結果，相當直觀而且好處理，不過其缺點是太注重於左右端點值，完全忽略了隸屬度函數描述資料的功能，這樣的估計方式有失"模糊"的概念。若在使用最小平方法時，能加入模糊數間距離的概念，或許可得到模糊迴歸參數 $A_m$ 更適當的估計。

## 2.3 簡單距離公式

爲了得到(2.1)式模型I中，模糊迴歸參數的最小平方估計，我們必須考慮模糊數間的距離。

定義**2.2** *(簡單距離公式)* 設 $A=(c_a,s_a)$,$B=(c_b,s_b)$ 是對稱形式的模糊數，其中 $c_a,c_b$ 為中心，$s_a,s_b$ 為分展度，則 $A$ 與 $B$ 間的距離 $D$ 為：

$$D = \sqrt{(c_a-c_b)^2+(s_a-s_b)^2} \tag{2.2}$$

若 $A=(c_a,p_a,q_a)$，$B=(c_b,p_b,q_b)$ 是不對稱形式的模糊數，其中 $c_a,c_b$ 為中心，$p_a,p_b$ 為左分展度，$q_a,q_b$ 為右分展度，則 $A$ 與 $B$ 之間的距離 $D$ 為：

$$D = \sqrt{(c_a-c_b)^2\pi_c+(p_a-p_b)^2\pi_p+(q_a-q_b)^2\pi_q} \tag{2.3}$$

其中 $\pi_c$、$\pi_p$ 與 $\pi_q$ 分別為任意正的比重。

在上述的距離公式中，若模糊數為對稱形式，則距離為中心減去中心的平方，加上分展度減去分展度平方，再開根號。若模糊數為不對稱形式，則距離為中心減去中心的平方，加上左右分展度差的平方，並分別在中心與左右分展度處乘上一個正的比重後再開根號。在不對稱的距離公式中，需加上對模糊距離的主觀判斷，給予中心與分展度一個適當的比重。由於上述模糊數間的距離公式，只考慮到模糊數的中心與分展度，稱其為簡單距離公式，以便和3.1節中Yang和Ko的距離公式做一區隔。

在有了模糊數的距離定義後，可考慮用最小平方法對(2.1)式的傳統模糊線性迴歸進行參數估計。假設 $y_i=(c_i,s_i)$ 與 $A_m=(a_m,r_m)$ 具有相同的隸屬度函數，且經過適當的平移後可使得所有 $x_{ij}>0$。(2.1)式可表示為：

$$(c_i,s_i) = (a_o,r_o)+(a_1,r_1)x_{i1}+(a_2,r_2)x_{i2}+...+(a_p,r_p)x_{ip}$$

在(2.2)式的簡單距離定義下，$a_i,r_i$ 的最小平方估計，即為使 $D^2$ 最小之 $a_i,r_i$ 值：

$$D^2 \equiv \sum_{i=1}^{n}(c_i-(a_o+a_1x_{i1}+...+a_px_{ip}))^2+(s_i-(r_o+r_1x_{i1}+...+r_px_{ip}))^2$$

令 $\|v\|$ 表示向量 $v$ 的長度，上述式子若以向量及矩陣表達，在處理上會更方便，經改寫後可得：

$$D^2 = \|Xa-C\|^2+\|Xr-S\|^2$$

其中$X=(x_{ij})$是$n\times(p+1)$的自變數值矩陣且$a=(a_o,a_1,...a_p)'$,$r=(r_0,r_1,...r_p)'$,$C=(c_1,c_2,...c_n)'$ ,$S=(s_1,s_2,...s_n)'$。令$\frac{\partial D^2}{\partial a}=0$與$\frac{\partial D^2}{\partial r}=0$，可得$a$與$r$的解如下：

$$\begin{aligned}\hat{a} &= (X'X)^{-1}X'C \\ \hat{r} &= (X'X)^{-1}X'S \qquad (2.4)\end{aligned}$$

簡單的說，這種處理方式相當於對中心和分展度分別做迴歸，所獲得的估計結果自然和隸屬度函數無關，但實例分析上又呈現出這種估計結果較其他方法爲佳。

## 2.4 對稱和不對稱雙重模糊線性迴歸模型

在模型I的架構下，若使用最小平方法與簡單距離公式，所獲得最小平方估計的結果是中心與分展度分別做迴歸，該結果顯示中心與分展度無任何關聯，但D'Urso和Gastaldi(2000)認爲模糊數之中心與其分展度，常存在一些相關性，因此提出以下雙重模糊線性迴歸模型(簡稱模型II)，並將應變數$y_i=[c_i-s_i,c_i+s_i]$考慮爲對稱三角形模糊數。

$$\text{模型II} \quad C = C^*+\varepsilon_c \qquad C^*=Xa \qquad (2.5)$$

$$S = S^*+\varepsilon_s \qquad S^*=C^*b+\mathbf{1}d \qquad (2.6)$$

其中$C=(c_1,c_2,...,c_n)'$,$S=(s_1,s_2,...,s_n)'$, $a$是$(p+1)\times 1$的參數向量，$\mathbf{1}$是每一分量皆爲1的$n\times 1$向量，$b,d$是第二迴歸模型參數。模型II的結構是先對中心的部份建立起核心迴歸模型(core regression model)，再利用核心迴歸模型，對分展度的部份建立起與中心有關的分展度迴歸模型(spread regression model)。

D'Urso和Gastaldi利用最小平方法以及(2.2)的簡單距離公式，求參數$a,b,d$使誤差平方和

$$\begin{aligned}D^2 &\equiv \|C-C^*\|^2+\|S-S^*\|^2 \\ &= C'C-2C'Xa+a'X'Xa(1+b^2)+S'S \\ &\quad -2S'Xab-2S'\mathbf{1}d+2a'X'\mathbf{1}bd+nd^2\end{aligned}$$

最小。令$\frac{\partial D^2}{\partial a}=0$,$\frac{\partial D^2}{\partial b}=0$與$\frac{\partial D^2}{\partial d}=0$可得下列的方程式：

$$
\begin{aligned}
\frac{\partial D^2}{\partial a} &= -X'C+X'Xa(1+b^2)-X'Sb+X'\mathbf{1}bd=0 \\
\frac{\partial D^2}{\partial b} &= a'X'Xab-S'Xa+a'X'\mathbf{1}d=0 \\
\frac{\partial D^2}{\partial d} &= -S'\mathbf{1}+a'X'\mathbf{1}b+nd=0
\end{aligned}
\tag{2.7}
$$

並由上述的方程式，求得$a$,$b$與$d$的遞迴關係式如下：

$$
\begin{aligned}
a &= \frac{1}{1+b^2}((X'X)^{-1}X'(C+rb-\mathbf{1}bd)) \\
b &= (a'X'Xa)^{-1}(r'Xa-a'X'\mathbf{1}d) \\
d &= \frac{1}{n}(r'\mathbf{1}-a'X'\mathbf{1}b)
\end{aligned}
\tag{2.8}
$$

事實上，我們可由(2.7)式解出$a$,$b$,$d$的通解，其過程如下：
由(2.7)的第1個方程式知$a=\frac{1}{1+b^2}((X'X)^{-1}X'(C+Sb-\mathbf{1}bd))$，將其代入(2.7)的第2及第3個方程式，可得：

$$
\|\hat{C}\|^2b+C'\hat{S}b^2-n\bar{C}b^2d-S'\hat{C}-\|\hat{S}\|^2b+2n\bar{S}bd+n\bar{C}d-nbd^2 = 0 \tag{2.9}
$$

$$
b\bar{C}-\bar{S}+d = 0 \tag{2.10}
$$

其中$X(X'X)^{-1}X'C=\hat{C}$,$X(X'X)^{-1}X'S=\hat{S}$,$\frac{1}{n}\mathbf{1}'C=\bar{C}$, $\frac{1}{n}\mathbf{1}'S=\bar{S}$。由(2.10)可得$d=\bar{S}-b\bar{C}$，再將其代回(2.9)式中，則可以簡化成一個$b$的二次方程式：

$$
M_1b^2+M_2b+M_3 = 0
$$

其中$M_1=C'\hat{S}-n\bar{C}\bar{S}$,$M_2=\|\hat{C}\|^2-\|\hat{S}\|^2+n\bar{S}^2-n\bar{C}^2$, $M_3=n\bar{C}\bar{S}-S'\hat{C}$。解上述二次方程式可得：

$$
\hat{b} = \frac{-M_2\pm\sqrt{M_2^2-4M_1M_3}}{2M_1}
$$

所對應的$a,d$為：

$$
\begin{aligned}
\hat{d} &= \bar{S}-\hat{b}\bar{C} \\
\hat{a} &= \frac{1}{1+\hat{b}^2}((X'X)^{-1}X'(c+r\hat{b}-\mathbf{1}\hat{b}\hat{d}))
\end{aligned}
$$

亦有兩組解，將兩組$\hat{a}$,$\hat{b}$,$\hat{d}$之值代回$D^2$中，找出使誤差平方和$D^2$較小的即爲最小平方估計。由$a, b, d$的估計式可看出，無論應變數部份是何種形態的模糊數，所獲得的參數估計結果都相同，因此在簡單距離公式下所得到的最小平方估計，無法適度反映模糊數的形態。

在模糊應變數$y_i = [c_i - q_i, c_i + p_i]$爲非對稱三角形的情況下，D'Urso(2001)提出以下的模糊線性迴歸模型，爲了方便起見稱其爲模型III：

$$\begin{aligned} \text{模型III} \qquad C &= C^* + \varepsilon \qquad C^* = Xa \\ p &= p^* + \lambda \qquad p^* = C^*b + \mathbf{1}d \\ q &= q^* + \rho \qquad q^* = C^*g + \mathbf{1}h \end{aligned} \tag{2.11}$$

式中$C = (c_1, c_2, ..., c_n)', p = (p_1, p_2, ..., p_n)', q = (q_1, q_2, ..., q_n)'$，$a$是$(p+1) \times 1$的參數向量，$\mathbf{1}$是每一分量皆爲1的$n \times 1$向量，$b, d, g, h$是第二迴歸模型參數。模型III可稱爲不對稱的雙重模糊線性迴歸模型。

D'Urso在(2.3)的簡單距離公式下，利用最小平方法，導出$a, b, d, g, h$的最小平方估計滿足下面的遞迴關係式：

$$\begin{aligned} a &= \frac{1}{\pi_c + b^2\pi_p + g^2\pi_q}[(X'X)^{-1}X'(c\pi_c + (p - \mathbf{1}d)b\pi_p) + (q - \mathbf{1}h)g\pi_q] \\ b &= (a'X'Xa)^{-1}(a'X'p - a'X'\mathbf{1}d) \\ d &= \frac{1}{n}(p'\mathbf{1} - a'X'\mathbf{1}b) \\ g &= (a'X'Xa)^{-1}(a'X'q - a'X'\mathbf{1}h) \\ h &= \frac{1}{n}(q'\mathbf{1} - a'X'q - a'X'\mathbf{1}h) \end{aligned} \tag{2.12}$$

由(2.12)的遞迴關係式，可看出模型III的參數估計和$\pi_c, \pi_p, \pi_q$權重的選取有密切關係，這不是一套客觀的處理方式，同時由(2.12)也可看出參數估計的結果和隸屬度函數無關，因此在下一章中，我們將設法獲得客觀並與隸屬度函數有關聯的最小平方估計。