บทที่ 2
ตัวสถิติและผลงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณของตัวแบบการถดถอยเชิงเส้นโดยวิธีกำลังสองน้อยสุด วิธีการถดถอยแบบริดจ์ และวิธีการประมาณตัวประมาณริดจ์ รวมถึงผลงานวิจัยที่เกี่ยวข้องในการวิจัยครั้งนี้ ซึ่งรายละเอียดต่างๆ มีดังต่อไปนี้

2.1 ตัวแบบทั่วไป

ตัวแบบที่แสดงความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามในการวิเคราะห์ความถดถอยพหุคูณในรูปเชิงเส้นทั่วไป อยู่ในรูปของ

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \qquad \text{.....................(2.1.1)}$$

เมื่อ $\underset{\sim}{y}$ เป็น เวกเตอร์ของตัวแปรตามขนาด n x 1

X เป็น เมทริกซ์ของตัวแปรอิสระขนาด n x (p+1)

$\underset{\sim}{\beta}$ เป็น เวกเตอร์ของสัมประสิทธิ์การถดถอยขนาด (p+1) x 1

และ $\underset{\sim}{\varepsilon}$ เป็น เวกเตอร์ของความคลาดเคลื่อนที่เกิดขึ้นขนาด n x 1 ซึ่งมีการแจกแจงปกติ

โดยที่ $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$, $E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}') = \sigma^2 \mathbf{I}_n$

ในการวิเคราะห์การถดถอยแบบริดจ์นั้นมักพิจารณาตัวแบบที่อยู่ในรูปมาตรฐานเพราะฉะนั้น ในการศึกษาครั้งนี้เมทริกซ์ตัวแปรอิสระ (X) จะมีขนาด n x p และเวกเตอร์ของสัมประสิทธิ์การถดถอยมีขนาด p x 1 จากสมการตัวแบบ (2.1.1) ตัวประมาณสัมประสิทธิ์การถดถอยที่ได้จากวิธีกำลังสองน้อยสุด อยู่ในรูปของ

$$\hat{\underset{\sim}{\beta}} = (X'X)^{-1} X' \underset{\sim}{y} \qquad \text{.....................(2.1.2)}$$

ซึ่งตัวประมาณ $\hat{\underset{\sim}{\beta}}$ ที่ได้นี้จะมีคุณสมบัติเป็นตัวประมาณไม่เอนเอียงและมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุดในบรรดาตัวประมาณที่ไม่เอนเอียงเชิงเส้นทั้งหลาย กล่าวคือ

$$E(\hat{\underset{\sim}{\beta}}) = E\left[(X'X)^{-1}X'\underset{\sim}{y}\right]$$
$$= E\left[(X'X)^{-1}X'(X\underset{\sim}{\beta}+\underset{\sim}{\varepsilon})\right]$$
$$= E(\underset{\sim}{\beta}) + (X'X)^{-1}X'E(\underset{\sim}{\varepsilon})$$
$$= \underset{\sim}{\beta}$$

และค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสอง คือ

$$MSE(\hat{\underset{\sim}{\beta}}) = Cov(\hat{\underset{\sim}{\beta}}) + bias^2(\hat{\underset{\sim}{\beta}})$$
$$= Cov\left(X'X)^{-1}X'\underset{\sim}{y}\right)$$
$$= \sigma^2(X'X)^{-1}$$

เมื่อ $Cov(\hat{\underset{\sim}{\beta}})$ เป็น เมทริกซ์ความแปรปรวนร่วมของตัวประมาณสัมประสิทธิ์การถดถอย

เราจะเห็นได้ว่าค่าเฉลี่ยของความคลาดเคลื่อนยกกำลังสองเป็นฟังก์ชันของ $(X'X)^{-1}$ ดังนั้น ในการประมาณค่าสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยสุดนี้จะต้องมีข้อสมมุติว่า ตัวแปรอิสระจะต้องไม่มีพหุสัมพันธ์กัน มิฉะนั้นนั้นตัวประมาณที่ได้อาจจะไม่ใช่ตัวประมาณที่มีความคลาดเคลื่อนยกกำลังสองต่ำสุด

ถ้ากำหนดให้ L_1 คือ ความแตกต่างระหว่าง $\hat{\underset{\sim}{\beta}}$ และ $\underset{\sim}{\beta}$ ซึ่ง $L_1^2 = (\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})'(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})$

ค่าเฉลี่ยกำลังสองของความแตกต่างระหว่าง $\hat{\underset{\sim}{\beta}}$ และ $\underset{\sim}{\beta}$ คือ

$$E(L_1^2) = E\left[(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})'(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})\right]$$
$$= E(\hat{\underset{\sim}{\beta}}'\underset{\sim}{\beta}) - \underset{\sim}{\beta}'\underset{\sim}{\beta}$$
$$\because \quad E(\hat{\underset{\sim}{\beta}}'\underset{\sim}{\beta}) = \underset{\sim}{\beta}'\underset{\sim}{\beta} + \sigma^2 trace(X'X)^{-1}$$
$$\therefore \quad E(L_1^2) = \sigma^2 trace(X'X)^{-1} \qquad \text{.....................(2.1.3)}$$

และเพราะว่า $\underset{\sim}{\varepsilon}$ มีการแจกแจงปกติ

$$\therefore Var(L_1^2) = 2\sigma^4 trace(X'X)^{-2} \qquad \text{.....................(2.1.4)}$$

จากสมการที่ (2.1.3) และ (2.1.4) จะเห็นได้ว่าทั้งสองสมการต่างก็อยู่ในรูปของผลบวกในแนวเส้นทแยงมุมของเมทริกซ์ $X'X$ $(trace(X'X))$ ดังนั้นเราสามารถเขียนสมการทั้งคู่ในรูปของค่าเฉพาะ (eigen value) ซึ่งจะได้ว่า

$$E(L_1^2) = \sigma^2\sum_{i=1}^{p}(\frac{1}{\lambda_i}) \qquad \text{.....................(2.1.5)}$$

$$Var(L_1^2) \quad = \quad 2\sigma^4 \sum_{i=1}^{p} (\frac{1}{\lambda_i})^2 \qquad \text{.....................(2.1.6)}$$

เมื่อ $\sum_{i=1}^{p} \lambda_i \;=\; trace(X'X)$

$\lambda_{\max} \;=\; \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p = \lambda_{\min} > 0$

p เป็น จำนวนตัวแปรอิสระ

และ λ_i เป็น ค่าเฉพาะของเมทริกซ์ $(X'X)$; $i = 1,2,\ldots, p$

ในกรณีที่เกิดปัญหาตัวแปรอิสระมีพหุสัมพันธ์กันค่าเฉพาะของเมทริกซ์ $X'X$ บางค่าจะมีค่าน้อยมากทำให้ค่า $|X'X|$ มีค่าน้อยลงไปด้วยเพราะ $|X'X|$ มีค่าเท่ากับผลคูณของค่าเฉพาะส่งผลให้ค่าเฉลี่ยของความแตกต่างระหว่าง $\hat{\underset{\sim}{\beta}}$ และ $\underset{\sim}{\beta}$ มีค่ามาก รวมทั้งค่าเฉลี่ยของความคลาดเคลื่อนยกกำลังสองของ $\hat{\underset{\sim}{\beta}}$ ที่ประมาณด้วยวิธีกำลังสองน้อยสุดจะมีค่าสูงเกินไปด้วย

2.2 การประมาณค่าสัมประสิทธิ์โดยวิธีการถดถอยแบบริดจ์ [1]

ในปี ค.ศ. 1970 โฮเอริน์ (Hoerl) และเคนนาร์ด (Kennard) ได้เสนอวิธีแก้ปัญหาตัวแปรอิสระมีพหุสัมพันธ์ซึ่งเรียกว่า **การถดถอยแบบริดจ์ (Ridge regression)** วิธีการนี้จะให้ค่าเฉลี่ยของความคลาดเคลื่อนยกกำลังสองต่ำกว่าวิธีกำลังสองน้อยสุดเพราะหลักการของวิธีนี้คือ การบวกค่าคงที่ k ที่มีค่ามากกว่าศูนย์กับสมาชิกทุกตัวบนเส้นทแยงมุม(diagonal) ของเมทริกซ์ $X'X$ เพื่อจะทำให้ค่าเฉลี่ยของความแตกต่างระหว่าง $\hat{\underset{\sim}{\beta}}$ และ $\underset{\sim}{\beta}$ ลดลง ตัวประมาณสัมประสิทธิ์การถดถอยแบบริดจ์ อยู่ในรูปของ

$$\hat{\underset{\sim}{\beta}}_R \quad = \quad (X'X + kI)^{-1} X' \underset{\sim}{y} \quad ; \quad k \geq 0 \qquad \text{.....................(2.2.1)}$$

เพื่อความสะดวกในการเปรียบเทียบคุณสมบัติต่างๆ ของการประมาณค่าพารามิเตอร์ $\underset{\sim}{\beta}$ จะกำหนดสัญลักษณ์ต่างๆ ดังต่อไปนี้

ถ้ากำหนดให้ $W \;=\; (X'X + kI)^{-1}$ และแทนในสมการ (2.2.1) จะได้ว่า $\hat{\underset{\sim}{\beta}}_R \;=\; WX'\underset{\sim}{y}$ และ เราสามารถเขียน $\hat{\underset{\sim}{\beta}}_R$ ให้อยู่ในรูปของ $\hat{\underset{\sim}{\beta}}$ ได้โดยจากสมการปกติ $X'X\hat{\underset{\sim}{\beta}} \;=\; X'\underset{\sim}{y}$ เพราะฉะนั้น

---

[1] จิรายุส พู่มนตรี. การเปรียบเทียบตัวประมาณริดจ์สำหรับการวิเคราะห์การถดถอยแบบริดจ์. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย,2534.

ตัวประมาณ $\hat{\underset{\sim}{\beta}}_R$ จะเขียนอยู่ในรูปของ $\hat{\underset{\sim}{\beta}}_R = \left[I + k(X'X)^{-1}\right]^{-1}\hat{\underset{\sim}{\beta}} = Z\hat{\underset{\sim}{\beta}}$ เมื่อ

$Z = \left[I + k(X'X)^{-1}\right]^{-1}$

คุณสมบัติที่จำเป็นของ Z, W และ $\hat{\underset{\sim}{\beta}}_R$

1. ให้ $\xi_i(W)$ และ $\xi_i(Z)$ เป็น ค่าเฉพาะของ W และ Z ตามลำดับ ซึ่ง

$$\xi_i(W) = \frac{1}{\lambda_i + k} \quad ; \quad i = 1,2,\ldots,p \qquad \ldots\ldots\ldots\ldots\ldots\ldots(2.2.2)$$

และ $$\xi_i(Z) = \frac{\lambda_i}{\lambda_i + k} \quad ; \quad i = 1,2,\ldots,p \qquad \ldots\ldots\ldots\ldots\ldots\ldots(2.2.3)$$

ซึ่งค่าเหล่านี้คำนวณได้จากสมการเฉพาะ (characteristic equation)

$$|W - \xi I| = 0 \quad \text{และ} \quad |Z - \xi I| = 0$$

2. เราสามารถเขียน ค่า Z ในรูปของ W ได้ดังนี้

$$Z = I - k(X'X + kI)^{-1} = I - kW \qquad \ldots\ldots\ldots\ldots\ldots\ldots(2.2.4)$$

3. ค่าประมาณ $\hat{\underset{\sim}{\beta}}_R$ จะมีค่าน้อยกว่า $\hat{\underset{\sim}{\beta}}$ เมื่อ $k > 0$ นั่นคือ

$$\hat{\underset{\sim}{\beta}}_R{}' \hat{\underset{\sim}{\beta}}_R < \hat{\underset{\sim}{\beta}}'\hat{\underset{\sim}{\beta}}$$

จาก $\hat{\underset{\sim}{\beta}}_R = Z\hat{\underset{\sim}{\beta}}$ โดยที่เมทริกซ์ $X'X$ และ Z เป็นเมทริกซ์ที่มีคุณสมบัติสมมาตรและเป็นบวกแน่นอน (symmetric positive definite matrix) [2]

$$\hat{\underset{\sim}{\beta}}_R{}' \hat{\underset{\sim}{\beta}}_R = (Z\hat{\underset{\sim}{\beta}})'(Z\hat{\underset{\sim}{\beta}})$$

$$= \sum_{i=1}^{p} \xi_i^2(Z)\hat{\underset{\sim}{\beta}}'\hat{\underset{\sim}{\beta}} \le \xi_{i(\max)}^2(Z)\hat{\underset{\sim}{\beta}}'\hat{\underset{\sim}{\beta}}$$

เมื่อ $\xi_{\max}(Z) = \left(\frac{\lambda_i}{\lambda_i + k}\right)$

และ λ_1 เป็น ค่าเฉพาะที่มีค่ามากที่สุดของเมทริกซ์ $X'X$

จากสมการที่ (2.2.3) และ (2.2.4) ณ จุดที่ $k = 0$ ค่า $Z = I$ และเมื่อ $k \to \infty$ ค่า Z จะมีค่าเข้าใกล้ศูนย์

ดังนั้น $$\hat{\underset{\sim}{\beta}}_R{}' \hat{\underset{\sim}{\beta}}_R < \hat{\underset{\sim}{\beta}}'\hat{\underset{\sim}{\beta}}$$

และเนื่องจากผลบวกของความคลาดเคลื่อนยกกำลังสองของตัวประมาณ $\hat{\underset{\sim}{\beta}}_R$ อยู่ในรูปของ

---

[2] เมทริกซ์ $A_{(pxp)}$ จะมีคุณสมบัติสมมาตรและเป็นบวกแน่นอน ถ้า $\underset{\sim}{\alpha}' A \underset{\sim}{\alpha} > 0$, $\forall \underset{\sim}{\alpha} \in \Re^p$, $\forall \underset{\sim}{\alpha} \ne 0$

$$\Phi^*(\underset{\sim}{\hat{\beta}}_R) = (\underset{\sim}{y} - X\underset{\sim}{\hat{\beta}}_R)'(\underset{\sim}{y} - X\underset{\sim}{\hat{\beta}}_R)$$

$$= \underset{\sim}{y}'\underset{\sim}{y} - \underset{\sim}{\hat{\beta}}'_R X'\underset{\sim}{y} - k\underset{\sim}{\hat{\beta}}'_R \underset{\sim}{\hat{\beta}}_R \qquad \text{.....................(2.2.5)}$$

ผลบวกของความคลาดเคลื่อนยกกำลังสองของตัวประมาณ $\underset{\sim}{\hat{\beta}}$ อยู่ในรูปของ

$$\Phi(\underset{\sim}{\hat{\beta}}) = \underset{\sim}{y}'\underset{\sim}{y} - \underset{\sim}{\hat{\beta}}'X'\underset{\sim}{y} \qquad \text{.....................(2.2.6)}$$

จากสมการที่ (2.2.5) และ (2.2.6) เราจะเห็นได้ว่าตัวประมาณ $\underset{\sim}{\hat{\beta}}_R$ มีความคลาดเคลื่อนยกกำลังสองต่ำกว่าตัวประมาณ $\underset{\sim}{\hat{\beta}}$ ทั้งนี้เพราะผลบวกความคลาดเคลื่อนกำลังสองของตัวประมาณ $\underset{\sim}{\hat{\beta}}_R$ ขึ้นอยู่กับระยะทางกำลังสองของตัวประมาณ $\underset{\sim}{\hat{\beta}}_R$ ด้วย

2.3 ค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสองของตัวประมาณริดจ์

เราสามารถแสดงรูปแบบทั่วไป (general form) ของตัวแบบในสมการที่ (2.1.1) และ (2.1.2) ให้อยู่ในรูปของค่าเฉพาะได้เพื่อความสะดวกในการพิสูจน์และการประมาณค่า โดยการลดรูปเมทริกซ์ $X'X$ ให้เป็นเมทริกซ์แนวทแยง (diagonal metrix) กล่าวคือ

กำหนดให้ $\mathbf{P}$ เป็นเมทริกซ์เชิงตั้งฉาก และ $X^* = XP$

จะได้ว่า $X^{*\prime}X^* = P'X'XP = \Lambda = diag(\lambda_i)$

และ $\underset{\sim}{\beta}^* = P'\underset{\sim}{\beta}$

เพราะฉะนั้นตัวประมาณกำลังสองน้อยที่สุดของ $\underset{\sim}{\beta}$ คือ

$$\underset{\sim}{\hat{\beta}}^* = \left(X^{*\prime}X^*\right)^{-1}X^{*\prime}\underset{\sim}{y}$$

และตัวประมาณสัมประสิทธิ์การถดถอย ริดจ์ของ $\underset{\sim}{\beta}$ คือ

$$\underset{\sim}{\hat{\beta}}^*_R = (\Lambda + K)^{-1}X^{*\prime}\underset{\sim}{y}$$

$$= (\Lambda + K)^{-1}X^{*\prime}X^*\underset{\sim}{\hat{\beta}}^*$$

$$= (\Lambda + K)^{-1}\Lambda\underset{\sim}{\hat{\beta}}^* \qquad \text{.....................(2.3.1)}$$

เมื่อ $K = diag(k_1, k_2, \ldots, k_n)$

ดังนั้น $$\hat{\beta}^*_{R(i)} = \left(\frac{\lambda_i}{\lambda_i + k_i}\right)\hat{\beta}^*_i$$

และเราสามารถหาค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสองของ $\underset{\sim}{\hat{\beta}}^*_R$ คือ

$$MSE(\underset{\sim}{\hat{\beta}}^*_R) = E\left[L_1^2(k)\right]$$

$$= E(\hat{\underset{\sim}{\beta}}^*_R - \underset{\sim}{\beta}^*)'(\hat{\underset{\sim}{\beta}}^*_R - \underset{\sim}{\beta}^*)$$

$$\because \quad \hat{\underset{\sim}{\beta}}^*_R = (\Lambda + K)^{-1}\Lambda\hat{\underset{\sim}{\beta}}^* = A\hat{\underset{\sim}{\beta}}^*$$

จะได้ว่า $$E\left[L_1^2(k)\right] = E(A\hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^*)'(A\hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^*) \qquad \text{............(2.3.2)}$$

เมื่อ $$A\hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^* = A\hat{\underset{\sim}{\beta}}^* - A\underset{\sim}{\beta}^* + A\underset{\sim}{\beta}^* - \underset{\sim}{\beta}^*$$

$$= A(\hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^*) + (A - I)\underset{\sim}{\beta}^* \qquad \text{............(2.3.3)}$$

และ $$(A\hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^*)' = (\hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^*)'A' + \underset{\sim}{\beta}^{*\prime}(A - I)' \qquad \text{............(2.3.4)}$$

เราแทนค่า สมการ (2.3.3) และ (2.3.4) ลงในสมการ (2.3.2) จะได้ว่า

$$E\left[L_1^2(k)\right] = E\left[(\hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^*)'A' + \underset{\sim}{\beta}^{*\prime}(A-I)'\right]\left[A(\hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^*) + (A-I)\underset{\sim}{\beta}^*\right]$$

$$= E\left[(\hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^*)'A'A(\hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^*) + (\hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^*)'A'(A-I)\underset{\sim}{\beta}^*\right.$$

$$\left. + \underset{\sim}{\beta}^{*\prime}(A-I)'A(\hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^*) + \underset{\sim}{\beta}^{*\prime}(A-I)'(A-I)\underset{\sim}{\beta}^*\right]$$

$$\because \quad E(\hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^*) = 0$$

ดังนั้น

$$E\left[L_1^2(k)\right] = E\left[(\hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^*)'A'A(\hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^*)\right] + \underset{\sim}{\beta}^{*\prime}(A-I)'(A-I)\underset{\sim}{\beta}^*$$

$$= \textit{Varience} + \textit{bias}^2$$

$$= \gamma_1 k + \gamma_2 k$$

$$\because \quad \hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^* = \left[\left(X^{*\prime}X^*\right)^{-1}X^{*\prime}\underset{\sim}{y} - \underset{\sim}{\beta}^*\right]$$

$$= \left[\left(X^{*\prime}X^*\right)^{-1}X^{*\prime}(X^*\underset{\sim}{\beta}^* + \underset{\sim}{\varepsilon}) - \underset{\sim}{\beta}^*\right]$$

$$= \left(X^{*\prime}X^*\right)^{-1}X^{*\prime}\underset{\sim}{\varepsilon}$$

จะได้ว่า

$$\gamma_1 k = E\left[(\hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^*)'A'A(\hat{\underset{\sim}{\beta}}^* - \underset{\sim}{\beta}^*)\right]$$

$$= E\left[\underset{\sim}{\varepsilon}'X^*\left(X^{*\prime}X^*\right)^{-1}A'A\left(X^{*\prime}X^*\right)^{-1}X^{*\prime}\underset{\sim}{\varepsilon}\right]$$

$$= \sigma^2\, \textit{trace}\left[(X^{*\prime}X^*)^{-1}A'A\right]$$

$$= \sigma^2\, \textit{trace}\left[\Lambda^{-1}((\Lambda+K)^{-1}\Lambda)'(\Lambda+K)^{-1}\Lambda\right]$$

$$= \sigma^2\sum_{i=1}^{p}\frac{\lambda_i}{(\lambda_i + k_i)^2} \qquad \text{...................(2.3.5)}$$

และ $\gamma_2 k = \underset{\sim}{\beta}^{*\prime}(A-I)'(A-I)\underset{\sim}{\beta}^{*}$

$$= \sum_{i=1}^{p} \frac{k_i^2 \beta_i^{*2}}{(\lambda_i + k_i)^2} \qquad \text{....................(2.3.6)}$$

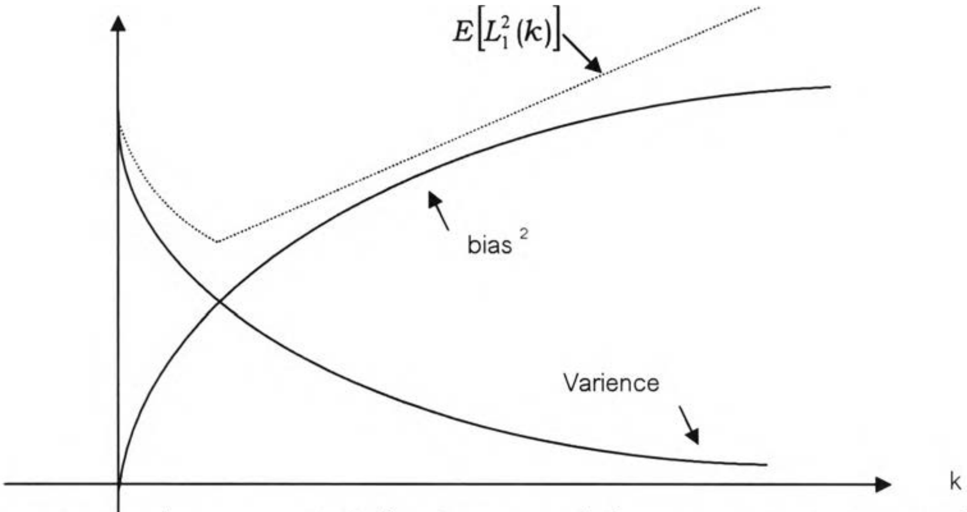
จากสมการที่ (2.3.5) และ (2.3.6) จะได้ว่า

$$E\left[L_1^2(k)\right] = \sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i + k_i)^2} + \sum_{i=1}^{p} \frac{k_i^2 \beta_i^{*2}}{(\lambda_i + k_i)^2} \qquad \text{..............(2.3.7)}$$

$\therefore$ เราสามารถเขียน $E\left[L_1^2(k)\right]$ ได้ในอีกรูปแบบหนึ่งคือ

$$E\left[L_1^2(k)\right] = E(\underset{\sim}{\hat{\beta}}_R^* - \underset{\sim}{\hat{\beta}}^*)'(\underset{\sim}{\hat{\beta}}_R^* - \underset{\sim}{\hat{\beta}}^*) + \sigma^2 \, trace\left[(A-K)^{-1}\Lambda^{-1}(A-K)^{-1}\right]$$

$$= \sum_{i=1}^{p} \frac{k_i^2}{(\lambda_i + k_i)^2}\left(\frac{\sigma^2}{\lambda_i} + \beta_i^{*2}\right) + \sigma^2 \sum_{i=1}^{p} \frac{(\lambda_i - k_i)}{\lambda_i(\lambda_i + k_i)} \qquad \text{......(2.3.8)}$$



รูปที่ 1 แสดงกราฟแสดงความสัมพันธ์ระหว่าง ความแปรปรวน (Varience) , ความเอนเอียงยกกำลังสอง (bias ^{2}) และค่าคงที่ k

จากกราฟจะเห็นได้ว่าฟังก์ชันของความแปรปรวน $(\gamma_1 k)$ มีลักษณะเป็นฟังก์ชันที่ต่อเนื่อง ถ้าเราทำการหาอนุพันธ์อันดับหนึ่งของฟังก์ชัน $\gamma_1 k$ เทียบกับ k นั่นคือ $\frac{\partial(\gamma_1 k)}{\partial k} = -2\sigma^2 \sum_{i=1}^{p} \left(\frac{\lambda_i}{\lambda_i + k_i}\right)^3$ จะได้ฟังก์ชันที่มีค่าเป็นลบ ดังนั้น $\lim_{k \to 0} \frac{\partial(\gamma_1 k)}{\partial k}$ จะมีลักษณะเป็นฟังก์ชันที่ลดลงทางเดียว (monotonically decreasing) ในรูปของค่า k และจะลู่เข้าสู่ $-\infty$ เมื่อ $k \to 0$ และ $\lambda_p \to 0$

สำหรับฟังก์ชันของความเอนเอียงยกกำลังสอง $(\gamma_2 k)$ ซึ่งมีลักษณะเป็นฟังก์ชันต่อเนื่องเช่นเดียวกัน ถ้าเราทำการหาอนุพันธ์อันดับหนึ่งของฟังก์ชัน $\gamma_2 k$ เทียบกับ k นั่นคือ $\frac{\partial(\gamma_2 k)}{\partial k} = -2k_i \sum_{i=1}^{p} \frac{\lambda_i \beta_i^{*2}}{(\lambda_i + k_i)^3}$ จะได้ฟังก์ชันที่มีค่าเป็นบวก ดังนั้น $\lim_{k \to 0} \frac{\partial(\gamma_2 k)}{\partial k}$ จะมีลักษณะเป็นฟังก์ชันที่เพิ่มขึ้นทางเดียว (monotonically increasing) และจะมีค่าเข้าใกล้ศูนย์เมื่อ $k \to 0$ จากรูปที่ 1 เราจะพบว่า สามารถเลือกค่า k บางค่าที่สามารถทำให้ค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสองของตัวประมาณที่ได้จากการถดถอยริดจ์มีค่าน้อยกว่าตัวประมาณที่ได้จากวิธีกำลังสองน้อยที่สุดได้

2.4 การประมาณค่าคงที่ k ในการถดถอยแบบริดจ์

ปัญหาในการวิเคราะห์การถดถอยแบบริดจ์นั้นคือ เราไม่สามารถกำหนดค่าคงที่ k ที่แน่นอนได้ ได้มีผู้นำเสนอวิธีการประมาณค่าคงที่ k ไว้หลายวิธี แต่ยังไม่มีวิธีใดที่ให้ผลสรุปได้ชัดเจน ในการศึกษาครั้งนี้ ได้ทำการศึกษาเปรียบเทียบวิธีการประมาณค่า k ดังต่อไปนี้

2.4.1 วิธี HK

โฮเอริน์ (Hoerl) และ เคนนาร์ด(Kennard) ได้เสนอวิธีนี้ขึ้นมา โดยให้ข้อสังเกตว่า การเลือกค่า k ที่เหมาะสมนั้น ค่า k ที่ได้จะต้องทำให้ค่าเฉลี่ยของความแตกต่างระหว่าง $\hat{\beta}$ กับ β มีค่าต่ำสุด นั่นคือการทำให้ $E\left[L_1^2(k)\right]$ มีค่าต่ำสุด โดยการหาอนุพันธ์เทียบกับค่า k_i และกำหนดให้มีค่าเท่ากับศูนย์ นั่นคือ

$$\frac{dE\left[L_1^2(k)\right]}{dk_i} = 0$$

จะได้ว่า

$$\frac{d}{d(k_i)}\left(\frac{\sigma^2\lambda_i + \beta_i^{*2}k_i^2}{(\lambda_i + k_i)^2}\right) = 0$$

กล่าวคือ

$$\frac{-2(\lambda_i + k_i)\lambda_i\sigma^2 + 2(\lambda_i + k_i)^2\beta_i^{*2}k_i - 2(\lambda_i + k_i)\beta_i^{*2}k_i^2}{(\lambda_i + k_i)^4} = 0$$

$$-\lambda_i\sigma^2 + \lambda_i\beta_i^{*2}k_i = 0$$

$$k_i = \frac{\sigma^2}{\beta_i^{*2}}$$

จะเห็นได้ว่าค่า k_i เป็นฟังก์ชันที่ขึ้นกับค่า σ^2 และ β_i^* แต่เนื่องจาก $\beta^* = P'\beta$ ดังนั้นจึงอาจจะมีกรณีที่ $\beta_i^* = 0$ มีค่าเท่ากับศูนย์และจะทำให้ไม่สามารถหาค่า k ได้ ดังนั้นเพื่อความ สะดวกจึงกำหนดค่า k เพียงค่าเดียวที่สามารถปรับเมตริกซ์ $X'X$ ที่มีผลทำให้ค่าเฉลี่ยความ

คลาดเคลื่อนยกกำลังสองมีค่าต่ำสุด ซึ่งค่า k ที่เหมาะสมนั้นควรมีค่าไม่มากนัก เพราะค่า k มีผลกับความความเอนเอียงยกกำลังสองและค่าความแปรปรวนของตัวประมาณ ดังนั้นจึงควรกำหนดค่า β_i^* เป็นค่าที่มากที่สุด และเนื่องจากเราไม่ทราบค่าพารามิเตอร์ ดังนั้นเราจะแทนค่าพารามิเตอร์ด้วยค่าประมาณที่มาจากกลุ่มตัวอย่าง

$$\hat{k} = \frac{\hat{\sigma}^2}{(\max(\hat{\beta}_1^*, \hat{\beta}_2^*, \ldots, \hat{\beta}_p^*))^2}$$

เมื่อ $\hat{\sigma}^2 = \dfrac{y'y - \hat{\underset{\sim}{\beta}}^{*\prime} X'y}{n-p}$

$\hat{\underset{\sim}{\beta}}^* = P'\hat{\underset{\sim}{\beta}}$

n คือ จำนวนตัวอย่าง

และ p คือ จำนวนตัวแปรอิสระ

2.4.2 วิธี SEQ [3]

วิธีการค้นหาข้อมูลแบบลำดับนี้ เป็นวิธีที่ใช้ค้นหาข้อมูลได้ทั้งข้อมูลที่มีลักษณะต่อเนื่องและไม่ต่อเนื่อง อีกทั้งข้อมูลที่นำมาใช้จะเรียงลำดับหรือไม่เรียงลำดับก็ได้ จากผลการวิจัยของธันยากร ต้นชลขันธ์ พบว่า การใช้วิธีการค้นหาข้อมูลแบบลำดับในการการหาค่า k ในการถดถอยแบบริดจ์นั้นให้ผลดี แต่อาจต้องใช้เวลานานในการค้นหาข้อมูล แต่วิธีนี้ยังคงเหมาะสมเพราะค่าของ k นั้นมีค่าไม่ห่างจากศูนย์มากนัก เมื่อพิจารณาค่าเฉลี่ยของความคลาดเคลื่อนยกกำลังสองของตัวประมาณสัมประสิทธิ์การถดถอยแล้วกราฟที่ได้มีจุดต่ำสุดเพียงจุดเดียวไม่มีจุดวกกลับ สัญลักษณ์ที่ใช้ในวิธีการนี้คือ

กำหนด

$MSE(\hat{\underset{\sim}{\beta}}_R(k))$ คือ ค่าเฉลี่ยของความคลาดเคลื่อนยกกำลังสอง ณ จุด k

d คือ ค่าคงที่ที่กำหนดขึ้นเพื่อเป็นช่วงห่างของค่า k

k_{opt} คือ ค่า k ที่ทำให้ $MSE(\hat{\underset{\sim}{\beta}}_R(k))$ มีค่าต่ำสุด

k_{sta} คือ จุดเริ่มต้นของ k ในการค้นหาข้อมูล

k_{d01} คือ จุดที่มีการเว้นช่วงห่างของค่า k = 0.01 (d = 0.01)

k_{b01} คือ จุดที่อยู่ก่อนค่า k ที่เหมาะสมและมีการเว้นช่วงห่างของค่า k = 0.01 (d = 0.01)

---

[3] ธันยากร ต้นชลขันธ์ , การเปรียบเทียบตัวประมาณ ริดจ์สำหรับการวิเคราะห์การถดถอยแบบริดจ์ , วิทยานิพนธ์ปริญญามหาบัณฑิต , ภาควิชาสถิติ , จุฬาลงกรณ์มหาวิทยาลัย , 2539

$\mathbf{k}_{a01}$ คือ จุดที่อยู่หลังค่า k ที่เหมาะสมและมีการเว้นช่วงห่างของค่า k = 0.01 (d = 0.01)

$\mathbf{k}_{d001}$ คือ จุดที่มีการเว้นช่วงห่างของค่า k = 0.001 (d = 0.001)

$\mathbf{k}_{end}$ คือ ค่าที่ใช้ตรวจสอบยุติการหาค่า k ที่เหมาะสมเมื่อค้นหาข้อมูลทางด้านขวาของค่า k ที่เหมาะสม

$\mathbf{k}_{fit}$ คือ ค่า k ที่เหมาะสมจริง

หลักการของวิธีการค้นหาข้อมูลแบบลำดับ

<u>ขั้นตอนที่ 1 ค่า d = 0.01</u>

1.1) กำหนดให้ค่า d = 0.01 และค่า $\mathbf{k}_{sta} = 0.0$

1.2) คำนวณค่า $\mathbf{k}_{d01} = \mathbf{k}_{sta} + \mathbf{d}$

1.3) คำนวณค่า $MSE(\hat{\underset{\sim}{\beta}}_R(k_{sta}))$ และ $MSE(\hat{\underset{\sim}{\beta}}_R(k_{d01}))$

1.4) เปรียบเทียบค่า $MSE(\hat{\underset{\sim}{\beta}}_R(k_{sta}))$ กับ $MSE(\hat{\underset{\sim}{\beta}}_R(k_{d01}))$

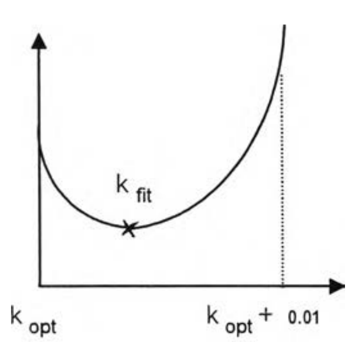
ก. ถ้า $\mathrm{MSE}(\hat{\underset{\sim}{\beta}}_R(\mathbf{k}_{sta})) < MSE(\hat{\underset{\sim}{\beta}}_R(k_{d01}))$

กำหนดให้ $k_{opt} = k_{sta}$ ยุติการประมวลผล

ข. ถ้า $MSE(\hat{\underset{\sim}{\beta}}_R(k_{sta})) \geq MSE(\hat{\underset{\sim}{\beta}}_R(k_{d01}))$

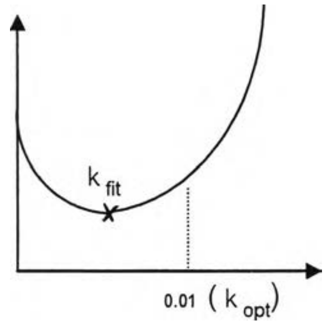
กำหนดให้ $\mathbf{k}_{sta} = \mathbf{k}_{d01}$ และทำการวิเคราะห์ตั้งแต่ข้อ 1.2) ใหม่

เมื่อพิจารณาถึงลักษณะของกราฟ $MSE(\hat{\underset{\sim}{\beta}}_R)$ จะเห็นได้ว่าเป็นโค้งหงายไม่มีจุดวกกลับ ฉะนั้นเมื่อมีการกำหนดค่า d อาจทำให้เกิดลักษณะที่เป็นไปได้ในการได้มาซึ่ง k_{opt} ดังนี้คือ



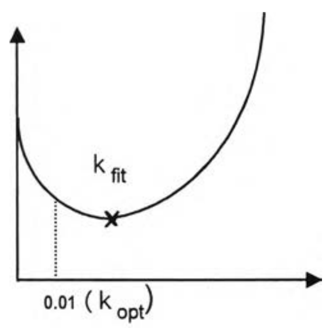
<u>ลักษณะที่ 1</u>

ลักษณะเช่นนี้จะเกิดขึ้นต่อเมื่อ เรากำหนด ช่วง d ห่างเกินไป จุด k_{opt} ที่ได้อาจจะเป็นจุดเริ่มแรกที่กำหนด คือ $k_{opt} = 0.0$ ซึ่งจะแก้ปัญหาโดยจะมีการค้นหาใหม่ทางด้านขวาของ k_{opt} คือตั้งแต่ $k_{opt} = 0.0$ จนถึง $k_{opt} = 0.0 + 0.01$ โดยใช้ช่วงห่างเป็น d = 0.001



ลักษณะที่ 2

ลักษณะเช่นนี้ เราจะได้ค่า k ที่ตกอยู่หลังค่า k เหมาะสมจริง (k_{fit}) จะแก้ปัญหาโดยการค้นหาค่า k ใหม่ ทางด้านซ้ายของ k_{opt} คือ ตั้งแต่ $k_{opt} - 0.01$ จนถึง k_{opt} โดยใช้ ช่วงห่างเป็น d = 0.001



ลักษณะที่ 3

ลักษณะเช่นนี้เราจะได้ค่า k ที่ตกอยู่ก่อนหน้าค่า k ที่เหมาะสมจริง (k_{fit}) จะแก้ปัญหาโดยการค้นหาค่า k ใหม่ทางด้านขวาของ k_{opt} คือ ตั้งแต่ k_{opt} จนถึง $k_{opt} + 0.01$ โดยใช้ช่วงห่างเป็น d = 0.001

ขั้นตอนที่ 2 ค่า d = 0.001

ในขั้นตอนนี้จะทำการตรวจสอบก่อนว่าค่า k_{opt} ที่ได้มีค่าเท่ากับศูนย์หรือไม่ เพื่อตรวจสอบความเป็นไปได้ของลักษณะที่ 1 แต่ถ้าค่า k_{opt} ที่ได้มีค่าไม่เท่ากับศูนย์จะทำการค้นหาค่า k ทางด้านซ้ายของค่า k_{opt} ก่อนถ้าไม่พบจะทำการค้นหาทางด้านขวาของ k_{opt} ใหม่ ซึ่งมีลำดับการค้นหาดังนี้

2.1) ตรวจสอบค่า $k_{opt} = 0.0$ หรือไม่

ก. ถ้า $k_{opt} = 0.0$ ให้ทำการวิเคราะห์ตามข้อ 2.3 (ค้นหาทางด้านขวาของค่า k_{opt})

ข. ถ้า $k_{opt} \neq 0.0$ ให้ทำการวิเคราะห์ตามข้อ 2.2 (ค้นหาทางด้านซ้ายของค่า k_{opt})

2.2) กำหนดให้ $k_{b01} = k_{opt} - 0.01$

2.2.1 กำหนดให้ $k_{d001} = k_{b01}$ + d

2.2.2 เปรียบเทียบค่า k_{d001} กับ k_{opt}

ก. ถ้า $k_{d001} \geq k_{opt}$ ให้ทำการวิเคราะห์ตามข้อ 2.3

ข. ถ้า $k_{d001} < k_{opt}$ จะทำการคำนวณค่า $MSE(\hat{\beta}_R(k_{b01}))$ และ

$MSE(\hat{\beta}_R(k_{d001}))$ และทำการเปรียบเทียบค่าดังกล่าว

- ถ้า $MSE(\hat{\beta}_R(k_{d001})) > MSE(\hat{\beta}_R(k_{b01}))$ กำหนดให้

$k_{opt} = k_{b01}$ และยุติการประมวลผล

- ถ้า $MSE(\hat{\beta}_R(k_{d001})) \leq MSE(\hat{\beta}_R(k_{b01}))$ กำหนดให้

$k_{b01} = k_{d001}$ แล้วทำการวิเคราะห์ตามข้อ 2.2.1 อีกครั้งจนกว่าจะเข้ากรณีที่ $k_{d001} \geq k_{opt}$ หรือ กรณีที่ $MSE(\hat{\beta}_R(k_{d001})) > MSE(\hat{\beta}_R(k_{b01}))$

2.3) กำหนดให้ $k_{end} = k_{opt} + 0.01$

2.3.1 กำหนดให้ $k_{d001} = k_{opt}$

2.3.2 กำหนดให้ $k_{a01} = k_{d001} + \mathrm{d}$ เปรียบเทียบค่า k_{a01} กับ k_{end}

ก. ถ้า $k_{a01} \geq k_{end}$ ยุติการประมวลผล

ข. ถ้า $k_{a01} < k_{end}$ จะทำการคำนวณค่า $MSE(\hat{\beta}_R(k_{a01}))$ และ

$MSE(\hat{\beta}_R(k_{d001}))$ และทำการเปรียบเทียบค่าดังกล่าว

- ถ้า $MSE(\hat{\beta}_R(k_{a01})) > MSE(\hat{\beta}_R(k_{d001}))$ กำหนดให้

$k_{opt} = k_{d001}$ และยุติการประมวลผล

- ถ้า $MSE(\hat{\beta}_R(k_{a01})) \leq MSE(\hat{\beta}_R(k_{d001}))$ กำหนดให้

$k_{d001} = k_{a01}$ แล้วทำการวิเคราะห์ตามข้อ 2.3.2 อีกครั้งจนกว่าจะเข้ากรณีที่ $k_{a01} \geq k_{end}$ หรือ กรณีที่ $MSE(\hat{\beta}_R(k_{a01})) > MSE(\hat{\beta}_R(k_{d001}))$

จากการวิจัยของธันยากร ต้นชลขันธ์ พบว่า เมื่อขยายช่วงของ d จาก 0.001 เป็น 0.0001 ค่า $MSE(\hat{\beta}_R(k))$ ณ จุดที่ d = 0.001 กับ d = 0.0001 มีค่าไม่แตกต่างกัน ดังนั้นเราจะไม่ทำการขยายช่วงของ d เพิ่มมากขึ้น

2.4.3 วิธีเบส์

ในปี ค.ศ. 1996 ทรอสกี้ (Troskie) และชาร์ลตัน (Chalton) ได้เสนอวิธีเบส์ในการหาค่า k โดย เปรียบเทียบกับวิธีของ Hoerl & Kennard (HK) ,Hemmerle & Carey (HC) , Noor & Metha (NM) พบว่า วิธีในการหาค่า k วิธีเบส์ให้ผลดีที่สุด ซึ่งวิธีเบส์แสดงได้ดังนี้

เนื่องจาก $\underset{\sim}{\beta}^{*} = P'\underset{\sim}{\beta}$ ซึ่งเมตริกซ์ P เป็นเมทริกซ์ที่ทราบค่า เราจึงสามารถศึกษาคุณสมบัติ $\underset{\sim}{\beta}^{*}$ จาก $\underset{\sim}{\beta}$ ได้ ให้ $\underset{\sim}{y} = (y_1, y_2, \ldots, y_n)'$ เป็นเวกเตอร์ของค่าสังเกต n ค่า โดยมีการแจกแจงความน่าจะเป็นร่วม ($p(\underset{\sim}{y} \mid \underset{\sim}{\beta}, \sigma)$) (joint probability distribution) ซึ่งขึ้นอยู่กับพารามิเตอร์ $\underset{\sim}{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$ และ σ ซึ่ง $(\underset{\sim}{\beta}, \sigma)'$ มีการแจกแจงความน่าจะเป็นร่วม $p(\underset{\sim}{\beta}, \sigma)$ ดังนั้น [4]

$$p(\underset{\sim}{y} \mid \underset{\sim}{\beta}, \sigma) \cdot p(\underset{\sim}{\beta}, \sigma) = p(\underset{\sim}{\beta}, \sigma \mid \underset{\sim}{y}) \cdot p(\underset{\sim}{y})$$

เราจะได้ว่าการแจกแจงที่มีเงื่อนไข (joint conditional distribution) ของ $\underset{\sim}{\beta}$ และ σ เมื่อกำหนด $\underset{\sim}{y}$ อยู่ในรูปของ

$$p(\underset{\sim}{\beta}, \sigma \mid \underset{\sim}{y}) = \frac{p(\underset{\sim}{y} \mid \underset{\sim}{\beta}, \sigma) \cdot p(\underset{\sim}{\beta}, \sigma)}{p(\underset{\sim}{y})} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(2.4.3)$$

ดังนั้น จากสมการ 2.4.3 จะได้ว่า

$$p(\underset{\sim}{\beta}, \sigma \mid \underset{\sim}{y}) \;\alpha\; p(\underset{\sim}{y} \mid \underset{\sim}{\beta}, \sigma) \cdot p(\underset{\sim}{\beta}, \sigma)$$

เมื่อ $p(\underset{\sim}{y} \mid \underset{\sim}{\beta}, \sigma)$ เป็น การแจกแจงความน่าจะเป็นร่วม (joint probability distribution)

$p(\underset{\sim}{\beta}, \sigma)$ เป็น การแจกแจงก่อนร่วม (joint prior distribution) สำหรับ $(\underset{\sim}{\beta}, \sigma)'$

และ $p(\underset{\sim}{\beta}, \sigma \mid \underset{\sim}{y})$ เป็น การแจกแจงภายหลังร่วม (joint posterior distribution) สำหรับ $(\underset{\sim}{\beta}, \sigma)'$

กำหนดให้

1. การแจกแจงความน่าจะเป็นร่วม (joint probability distribution ; $p(\underset{\sim}{y} \mid \underset{\sim}{\beta}, \sigma)$) ของค่าสังเกต $\underset{\sim}{y} = (y_1, y_2, \ldots, y_n)'$ คือ

$$p(\underset{\sim}{y} \mid \underset{\sim}{\beta}, \sigma) = \left[\frac{1}{2\pi\sigma^2}\right]^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\underset{\sim}{y} - X\underset{\sim}{\beta})'(\underset{\sim}{y} - X\underset{\sim}{\beta})\right\}$$

$$\alpha \quad \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}(\underset{\sim}{y} - X\underset{\sim}{\beta})'(\underset{\sim}{y} - X\underset{\sim}{\beta})\right\} \qquad \ldots\ldots\ldots(2.4.3.1)$$

---

[4] วีรพา ฐานะปรัชญ์ , การวิเคราะห์เชิงเบส์สำหรับตัวแบบการถดถอยเชิงเส้นเชิงเดียว , วิทยานิพนธ์ปริญญามหาบัณฑิต , ภาควิชาสถิติ , จุฬาลงกรณ์มหาวิทยาลัย , 2542 .

2. การแจกแจงก่อนร่วม (joint prior distribution ; $p(\underset{\sim}{\beta},\sigma)$) สำหรับ $(\underset{\sim}{\beta},\sigma)'$ คือ

$$p(\underset{\sim}{\beta},\sigma) \quad = \quad p(\underset{\sim}{\beta}\,|\,\sigma)\cdot p(\sigma)$$

เมื่อ การแจกแจงที่มีเงื่อนไขสำหรับ $\underset{\sim}{\beta}$ เมื่อกำหนด σ $(p(\underset{\sim}{\beta}|\sigma))$ เป็นการแจกแจงปกติหลายตัวแปร $(N(\underset{\sim}{\bar{\beta}},\sigma^2 A^{-1}))$ กล่าวคือ

$$\begin{aligned} p(\underset{\sim}{\beta}|\sigma) \quad &= \quad \left\{\frac{1}{2\pi\sigma}\right\}^{\frac{1}{2}p} |A|^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\underset{\sim}{\beta}-\underset{\sim}{\bar{\beta}})'A(\underset{\sim}{\beta}-\underset{\sim}{\bar{\beta}})\right\} \\ &\alpha \quad \sigma^{-p} \exp\left\{-\frac{1}{2\sigma^2}(\underset{\sim}{\beta}-\underset{\sim}{\bar{\beta}})'A(\underset{\sim}{\beta}-\underset{\sim}{\bar{\beta}})\right\} \qquad \text{......(2.4.3.2)} \end{aligned}$$

เมื่อ $\underset{\sim}{\bar{\beta}}$ คือ ค่าเฉลี่ยของ $\underset{\sim}{\beta}$

และ การแจกแจงส่วนริมสำหรับ σ $(p(\sigma))$ เป็นการแจกแจงไค-กำลังสองผกผัน ที่มีระดับขั้นความเสรี (degree of freedom) เท่ากับ υ (inverted χ^2_υ) กล่าวคือ

$$\begin{aligned} p(\sigma) \quad &= \quad \frac{2}{\Gamma\left(\frac{\upsilon}{2}\right)\sigma^{(\upsilon+1)}}\left[\frac{\upsilon}{2}s_0^2\right]^{\frac{1}{2}\upsilon} \exp\left\{-\frac{\upsilon s_0^2}{2\sigma^2}\right\} \\ &\alpha \quad \sigma^{-(\upsilon+1)} \exp\left\{-\frac{\upsilon s_0^2}{2\sigma^2}\right\} \qquad \text{..........(2.4.3.3)} \end{aligned}$$

ซึ่งค่าพารามิเตอร์ $\underset{\sim}{\bar{\beta}}$, A , υ และ s_0^2 ขึ้นอยู่กับข้อมูลก่อน (prior information) เพราะฉะนั้นเราจะได้ว่า

$$E(\underset{\sim}{\beta}\,|\,\sigma) \;=\; E(\underset{\sim}{\beta}) \;=\; \underset{\sim}{\bar{\beta}}$$

$$COV(\underset{\sim}{\beta}\,|\,\sigma) \;=\; \sigma^2 A^{-1}$$

$$E(\sigma) \;=\; \frac{\Gamma[(\upsilon-1)/2]}{\Gamma(\upsilon/2)}\left(\frac{\upsilon}{2}\right)^{1/2} s_0$$

$$E(\sigma^2) \;=\; \frac{\upsilon s_0^2}{\upsilon-2} \;=\; s_0^{*2}$$

จากสมการที่ (2.4.3.2) และ (2.4.3.3) เราจะได้การแจกแจงก่อนร่วมสำหรับ $(\underset{\sim}{\beta},\sigma)'$ อยู่ในรูปของ

$$\begin{aligned} p(\underset{\sim}{\beta},\sigma) \quad &= \quad p(\underset{\sim}{\beta}\,|\,\sigma)\cdot p(\sigma) \\ &\alpha \quad \sigma^{-(\upsilon+1)-p} \exp\left\{-\frac{1}{2\sigma^2}(\upsilon s_0^2+(\underset{\sim}{\beta}-\underset{\sim}{\bar{\beta}})'A(\underset{\sim}{\beta}-\underset{\sim}{\bar{\beta}}))\right\} \end{aligned}$$

...............(2.4.3.4)

เราจะเรียกการแจกแจงก่อนร่วมนี้ว่า การแจกแจงก่อนคู่สังยุคโดยธรรมชาติ (natural conjugate - prior distribution)

เพราะฉะนั้นจะได้ว่าการแจกแจงภายหลังร่วมสำหรับ $(\underset{\sim}{\beta},\sigma)'$ $(p(\underset{\sim}{\beta},\sigma)\mid \underset{\sim}{y}))$ คือ

$$p(\underset{\sim}{\beta},\sigma\mid \underset{\sim}{y}) = p(\underset{\sim}{y}\mid \underset{\sim}{\beta},\sigma)\cdot p(\underset{\sim}{\beta},\sigma)$$

$$\alpha \quad \sigma^{-(\upsilon+1)-p-n}\exp\left[-\frac{1}{2\sigma^2}\left\{\upsilon s_0^2+(\underset{\sim}{y}-X\underset{\sim}{\beta})'(\underset{\sim}{y}-X\underset{\sim}{\beta})+ (\underset{\sim}{\beta}-\overline{\underset{\sim}{\beta}})'(\underset{\sim}{\beta}-\overline{\underset{\sim}{\beta}})\right\}\right] \qquad \text{......(2.4.3.5)}$$

เราจะสามารถแสดงให้เห็นได้ว่า สมการที่ (2.4.3.5) เป็นรูปแบบฟังก์ชันปกติ-แกมมา โดยเราสามารถจัดรูปแบบสมการที่ (2.4.3.5) ได้ใหม่ดังต่อไปนี้

$$\begin{aligned}\upsilon s_0^2+(\underset{\sim}{y}-X\underset{\sim}{\beta})'(\underset{\sim}{y}-X\underset{\sim}{\beta})+(\underset{\sim}{\beta}-\underset{\sim}{\beta}) &= \upsilon s_0^2+\underset{\sim}{y}'\underset{\sim}{y}+2\underset{\sim}{\beta}'X'\underset{\sim}{y}+\underset{\sim}{\beta}'X'X\underset{\sim}{\beta}+\\ &\quad \underset{\sim}{\beta}'A\underset{\sim}{\beta}-2\underset{\sim}{\beta}'A\overline{\underset{\sim}{\beta}}+\overline{\underset{\sim}{\beta}}'A\overline{\underset{\sim}{\beta}}\\ &= \upsilon s_0^2+\underset{\sim}{y}'\underset{\sim}{y}+\underset{\sim}{\beta}'(A+X'X)\underset{\sim}{\beta}+2\underset{\sim}{\beta}'(A+X'X)\overline{\overline{\underset{\sim}{\beta}}}+\\ &\quad \overline{\overline{\underset{\sim}{\beta}}}(A+X'X)\overline{\overline{\underset{\sim}{\beta}}}-\overline{\overline{\underset{\sim}{\beta}}}(A+X'X)\overline{\overline{\underset{\sim}{\beta}}}+\overline{\underset{\sim}{\beta}}'A\overline{\underset{\sim}{\beta}}\\ &= \overline{\upsilon}\overline{s}^2+(\underset{\sim}{\beta}-\overline{\underset{\sim}{\beta}})'(A+X'X)(\underset{\sim}{\beta}-\overline{\underset{\sim}{\beta}})\end{aligned}$$

..................(2.4.3.6)

เมื่อ $\overline{\overline{\underset{\sim}{\beta}}} = (A+X'X)^{-1}(A\overline{\underset{\sim}{\beta}}+X'\underset{\sim}{y})$

$= (A+X'X)^{-1}(A\overline{\underset{\sim}{\beta}}+X'X\hat{\underset{\sim}{\beta}})$ ....(2.4.3.7)

$\overline{\upsilon}\overline{s}^2 = \upsilon s_0^2+\underset{\sim}{y}'\underset{\sim}{y}+\overline{\underset{\sim}{\beta}}'A\overline{\underset{\sim}{\beta}}-\overline{\overline{\underset{\sim}{\beta}}}'(A+X'X)\overline{\overline{\underset{\sim}{\beta}}}$ ....(2.4.3.8)

และ $\overline{\upsilon} = n+\upsilon$

เราแทนค่าสมการที่ (2.4.3.6) ในสมการที่ (2.4.3.5) จะได้ว่า

$$\begin{aligned}p(\underset{\sim}{\beta},\sigma\mid \underset{\sim}{y}) \quad &\alpha \quad \sigma^{-p}\exp\left[-\frac{1}{2\sigma^2}(\underset{\sim}{\beta}-\overline{\overline{\underset{\sim}{\beta}}})'(A+X'X)(\underset{\sim}{\beta}-\overline{\overline{\underset{\sim}{\beta}}})\right]\\ &\cdot \quad \sigma^{-(\overline{\upsilon}+1)}\exp\left(-\frac{\overline{\upsilon}\overline{s}^2}{\sigma^2}\right) \qquad \text{...... ....(2.4.3.9)}\\ &\alpha \quad p(\underset{\sim}{\beta}\mid \sigma,\underset{\sim}{y})\cdot p(\sigma\mid \underset{\sim}{y})\end{aligned}$$

รูปแบบในสมการ (2.4.3.9) เป็นรูปแบบของการแจกแจงปกติ-แกมมา เพราะว่า

1. การแจกแจงภายหลังที่มีเงื่อนไขสำหรับ $\underset{\sim}{\beta}$ เมื่อกำหนด σ $(p(\underset{\sim}{\beta}\mid \sigma,\underset{\sim}{y}))$ เป็นการแจกแจงปกติหลายตัวแปร ที่มีเวกเตอร์ค่าเฉลี่ย $\overline{\overline{\underset{\sim}{\beta}}}$ และเมทริกซ์ความแปรปรวนร่วม คือ $\sigma^2(A+X'X)^{-1}$

2. การแจกแจงภายหลังขอบ (marginal posterior) สำหรับ σ $(p(\sigma \mid \underset{\sim}{y}))$ เป็นการแจกแจงแกมมาผกผันที่มีพารามิเตอร์ $\bar{\upsilon}$ และ $\bar{s}^2$

เพราะฉะนั้นจากสมการ (2.4.3.9) เราสามารถหาการแจกแจงภายหลังขอบ $p(\underset{\sim}{\beta} \mid \underset{\sim}{y})$ ได้ โดย

$$p(\underset{\sim}{\beta} \mid \underset{\sim}{y}) = \int_0^\infty p(\underset{\sim}{\beta}, \sigma \mid \underset{\sim}{y}) d\sigma$$

$$\alpha \int_0^\infty \sigma^{-\bar{\upsilon}-p-1} \exp\left(\frac{-h}{2\sigma^2}\right) d\sigma \qquad \text{..............(2.4.3.10)}$$

เมื่อ $h = \bar{\upsilon}\bar{s}_0^2 + (\underset{\sim}{\beta} - \underset{\sim}{\bar{\bar{\beta}}})'(A + X'X)(\underset{\sim}{\beta} - \underset{\sim}{\bar{\bar{\beta}}})$

เราสามารถอินทิกรัลสมการ (2.4.3.10) โดยใช้คุณสมบัติของฟังก์ชันแกมมา ซึ่งเราจะได้ว่า

$$p(\underset{\sim}{\beta} \mid \underset{\sim}{y}) \quad \alpha \quad h^{-(\bar{\upsilon}+p)/2}$$

$$\alpha \left[\bar{\upsilon}\bar{s}_0^2 + (\underset{\sim}{\beta} - \underset{\sim}{\bar{\bar{\beta}}})'(A + X'X)(\underset{\sim}{\beta} - \underset{\sim}{\bar{\bar{\beta}}}) \right]^{-(\bar{\upsilon}+p)/2}$$

..............(2.4.3.11)

การแจกแจงภายหลังขอบสำหรับ $\underset{\sim}{\beta}$ ที่ได้เป็นการแจกแจงทีหลายตัวแปร (Multivariate - t distribution) ที่มี

1. เวกเตอร์ค่าเฉลี่ย คือ $\underset{\sim}{\bar{\bar{\beta}}} = (A + X'X)^{-1}(A\underset{\sim}{\bar{\beta}} + X'X\underset{\sim}{\hat{\beta}})$
2. เมทริกซ์ความแปรปรวนร่วม คือ $COV(\underset{\sim}{\beta}) = \bar{s}^2(A + X'X)^{-1}$
3. ค่าคาดหวังของ σ^2 คือ $E(\sigma^2) = \left(\frac{\bar{\upsilon}}{\bar{\upsilon}-2}\right)\bar{s}^2$

เนื่องจากวิธีเบส์เป็นวิธีที่เราจะทราบการแจกแจงก่อนของพารามิเตอร์ $\underset{\sim}{\beta}$ ดังนั้นค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสองในสมการที่ (2.3.5) จะเป็นแบบที่มีเงื่อนไข นั่นคือ

$$E\left(L_1^2(k) \,\middle|\, \underset{\sim}{\beta}^*, \sigma^2\right) = \sum_{i=1}^{p} \frac{k_i^2}{(\lambda_i + k_i)^2}\left(\frac{\sigma^2}{\lambda_i} + \beta_i^{*2}\right) + \sigma^2 \sum_{i=1}^{p} \frac{(\lambda_i - k_i)}{\lambda_i(\lambda_i + k_i)}$$

................(2.4.3.12)

และเนื่องจาก

$$E\left[\left(\frac{\sigma^2}{\lambda_i} + \beta_i^{*2}\right) \,\middle|\, \sigma^2\right] = \frac{\sigma^2}{\lambda_i} + E\left(\beta_i^{*2} \,\middle|\, \sigma^2\right)$$

$$E\left[\left(\frac{\sigma^2}{\lambda_i}+\beta_i^{*2}\right)\,\middle|\,\sigma^2\right] = \frac{\sigma^2}{\lambda_i}+E\left[\left(\beta_i^*-\overline{\beta}_i^* + \overline{\beta}_i^*\right)^2\,\middle|\,\sigma^2\right]$$

$$= \frac{\sigma^2}{\lambda_i}+E\left[\left(\left(\beta_i^*-\overline{\beta}_i^*\right)^2+\overline{\beta}_i^{*2}+2\overline{\beta}_i^*\left(\beta_i^*-\overline{\beta}_i^*\right)\right)\,\middle|\,\sigma^2\right]$$

$$= \frac{\sigma^2}{\lambda_i}+E\left[\left(\beta_i^*-\overline{\beta}_i^*\right)^2\,\middle|\,\sigma^2\right]+\overline{\beta}_i^{*2}$$

และ $E\left(\beta_i^*-\overline{\beta}_i^*\right) = 0$

ดังนั้น

$$E\left(\left(\frac{\sigma^2}{\lambda_i}+\beta_i^{*2}\right)\,\middle|\,\sigma^2\right) = \frac{\sigma^2}{\lambda_i}+\frac{\sigma^2}{a_i}+\overline{\beta}_i^{*2}$$

$$E(\sigma^2) = \left(\frac{\upsilon}{\upsilon-2}\right)s_0^2 = s_0^{*2}$$

เพราะฉะนั้นเมื่อเราแทนค่าต่าง ๆ ลงในสมการที่ (2.4.3.12) จะได้ว่าค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสองแบบไม่มีเงื่อนไข อยู่ในรูปของ

$$E\left(L_1^2(k)\right) = \sum_{i=1}^{p}\frac{k_i^2}{(\lambda_i+k_i)^2}\left(\frac{s_0^{*2}}{\lambda_i}+\frac{s_0^{*2}}{a_i}+\overline{\beta}_i^{*2}\right)+s_0^{*2}\sum_{i=1}^{p}\frac{(\lambda_i-k_i)}{\lambda_i(\lambda_i+k_i)} \qquad \text{............(2.4.3.13)}$$

เราจะทำการหาอนุพันธ์สมการ (2.4.3.13) เทียบกับ k_i และกำหนดให้เท่ากับศูนย์เพื่อให้ได้ค่า k_i ที่ทำให้ $\mathrm{E}\left[L_1^2(k)\right]$ มีค่าน้อยที่สุด กล่าวคือ $\frac{\partial E(L_1^2(k))}{\partial k_i} = 0$

จะได้ว่า

$$\frac{\partial E(L_1^2(k))}{\partial k_i} = \left[\frac{2k_i\lambda_i}{(\lambda_i+k_i)^3}\right]\left(\frac{s_0^{*2}}{\lambda_i}+\frac{s_0^{*2}}{a_i}+\overline{\beta}_i^{*2}\right)-\frac{2s_0^{*2}}{(\lambda_i+k_i)^2} = 0$$

$$\left[\frac{2k_i\lambda_i}{(\lambda_i+k_i)^3}\right]\left(\frac{s_0^{*2}}{\lambda_i}+\frac{s_0^{*2}}{a_i}+\overline{\beta}_i^{*2}\right)-\frac{2s_0^{*2}(\lambda_i+k_i)}{(\lambda_i+k_i)^3} = 0$$

$$k_is_0^{*2}+\left(\frac{k_i\lambda_is_0^{*2}}{a_i}\right)+k_i\lambda_i\overline{\beta}_i^{*2}-\lambda_is_0^{*2}-k_is_0^{*2} = 0$$

$$\left(\frac{k_i\lambda_is_0^{*2}}{a_i}\right)+k_i\lambda_i\overline{\beta}_i^{*2}-\lambda_is_0^{*2} = 0$$

$$k_i\lambda_i\left(\frac{s_0^{*2}+a_i+\overline{\beta}_i^{*2}}{a_i}\right)-\lambda_is_0^{*2} = 0$$

$$k_i = \frac{a_i\,s_0^{*2}}{s_0^2+a_i\,\overline{\beta}_i^{*2}} \qquad \text{...... (2.4.3.15)}$$

เราสามารถจัดรูปแบบในสมการ (2.4.3.15) ได้ใหม่คือ

$$k_i = \frac{a_i}{1 + \dfrac{a_i \bar{\beta}_i^{*2}}{s_0^{*2}}}$$

เมื่อ a_i เป็น สมาชิกตัวที่ i ของเมทริกซ์แนวทแยงมุมของเมทริกซ์ $(A + X'X)^{-1}$

ในการศึกษาครั้งนี้ เราจะศึกษาตัวประมาณเบส์ภายใต้ฟังก์ชันความสูญเสียเป็นแบบความผิดพลาดยกกำลังสอง ซึ่งเราจะได้ว่า ตัวประมาณเบส์ สำหรับ $\underset{\sim}{\beta}$ คือ ค่าเฉลี่ยของการแจกแจงภายหลัง ดังนั้นตัวประมาณเบส์สำหรับ $\underset{\sim}{\beta}$ ในวิธีนี้คือ

$$\underset{\sim}{\hat{\beta}}_B = \underset{\sim}{\bar{\bar{\beta}}} = (A + X'X)^{-1} (A\underset{\sim}{\bar{\beta}} + X'X\underset{\sim}{\hat{\beta}})$$

$$\hat{\sigma}_B^2 = \left(\frac{\bar{\upsilon}}{\bar{\upsilon} - 2}\right)\bar{s}_0^2$$

และ $$\underset{\sim}{\hat{\beta}}_B^* = P'\underset{\sim}{\hat{\beta}}_B$$

เมื่อเราใช้การแจกแจงก่อนคู่สังยุคโดยธรรมชาติ ซึ่งเป็นการแจกแจงก่อนที่ให้ข้อมูล (information prior) เกี่ยวกับตัวแปรสุ่มอย่างแน่ชัด จึงเหมือนกับว่าเรามีข้อมูลตัวอย่างตามทฤษฎี (hypothetical sample) ถ้านำข้อมูลตัวอย่างนี้ไปรวมกับข้อมูลที่มีอยู่ในปัจจุบันและข้อมูลที่ใช้การแจกแจงก่อนที่ไม่ให้ข้อมูล (noninformation prior) ซึ่งเป็นการแจกแจงที่ให้ข้อมูลเกี่ยวกับตัวแปรสุ่มไม่แน่ชัดหรือคลุมเครือ(vague) แล้ว ในทางทฤษฎีจะแทนค่า

$$A = X'X$$

$$X'\underset{\sim}{y} = A\underset{\sim}{\bar{\beta}}$$

$$\hat{\sigma}^2 = s_0^{*2}$$

เพราะฉะนั้นในเทอมของ k_i จะได้ว่า

$$k_i = \frac{a_i}{1 + \dfrac{a_i \hat{\beta}_{i(B)}^*}{\hat{\sigma}_B^2}} \qquad \text{..................(2.4.3.16)}$$

โดยทั่วไปแล้วเราจะไม่มีการแจกแจงก่อนคู่สังยุคโดยธรรมชาติที่เหมาะสม และข้อมูลที่ดีพอ เราจึงต้องใช้ค่าของตัวประมาณที่มาจากกลุ่มตัวอย่าง ซึ่งเราจะได้ว่า

$$k_i = \frac{\lambda_i}{1 + \dfrac{\lambda_i \hat{\beta}_i^{*2}}{\hat{\sigma}^2}}$$

เมื่อ $\hat{\sigma}^2 = \dfrac{\underset{\sim}{y}'\underset{\sim}{y} - \hat{\underset{\sim}{\beta}} X'\underset{\sim}{y}}{n-p}$

$\hat{\underset{\sim}{\beta}}^{*} = P'\hat{\underset{\sim}{\beta}}$

และ $a_i = \lambda_i$

โดยที่ λ_i เป็น ค่าเจาะจงของเมทริกซ์ $X^{*\prime} X^{*}$

$\hat{\underset{\sim}{\beta}}$ เป็น ตัวประมาณสัมประสิทธิ์การถดถอยจากวิธีกำลังสองน้อยสุด

และ P เป็น เมทริกซ์ของเวกเตอร์เฉพาะ